

T.C.
ANKARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
BİLGİ VE BELGE YÖNETİMİ ANABİLİM DALI

**VERİ GÖRSELLEŞTİRME: TR DİZİN'DE İNDEKSLENEN ANKARA
ÜNİVERSİTESİ ADRESLİ MAKALELERİN ÜST VERİ KAYITLARI
ÜZERİNE BİR ARAŞTIRMA**

Doktora Tezi

Recep IŞIK

Ankara, 2024

T.C.
ANKARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
BİLGİ VE BELGE YÖNETİMİ ANABİLİM DALI

**VERİ GÖRSELLEŞTİRME: TR DİZİN'DE İNDEKSLENEN ANKARA
ÜNİVERSİTESİ ADRESLİ MAKALELERİN ÜST VERİ KAYITLARI
ÜZERİNE BİR ARAŞTIRMA**

Doktora Tezi

Recep IŞIK

Tez Danışmanı:

Prof. Dr. Tülay OĞUZ

Ankara, 2024

T.C.
ANKARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
BİLGİ VE BELGE YÖNETİMİ ANABİLİM DALI

VERİ GÖRSELLEŞTİRME: TR DİZİN'DE İNDEKSLENEN ANKARA
ÜNİVERSİTESİ ADRESLİ MAKALELERİN ÜST VERİ KAYITLARI
ÜZERİNE BİR ARAŞTIRMA

DOKTORA TEZİ

Tez Danışmanı:

Prof. Dr. Tülay OĞUZ

TEZ JÜRİSİ ÜYELERİ

Adı ve Soyadı

Prof. Dr. Tülay OĞUZ (Danışman)	
Prof. Dr. Özlem GÖKKURT DEMİRTEL	
Prof. Dr. Sacit ARSLANTEKİN	
Doç. Dr. Mustafa BAYTER	
Dr. Öğr. Üyesi Neslihan ER-KOÇOĞLU	

Tez Savunması Tarihi

04.06.2024

T.C.
ANKARA ÜNİVERSİTESİ
Sosyal Bilimler Enstitüsü Müdürlüğü'ne

Prof. Dr. Tülay OĞUZ danışmanlığında hazırladığım “Veri Görselleştirme: TR Dizin'de İndekslenen Ankara Üniversitesi Adresli Makalelerin Üst Veri Kayıtları Üzerine Bir Araştırma (Ankara-2024)” adlı doktora tezindeki bütün bilgilerin akademik kurallara ve etik davranış ilkelerine uygun olarak toplanıp sunulduğunu, başka kaynaklardan aldığım bilgileri metinde ve kaynakçada eksiksiz olarak gösterdiğimi, çalışma sürecinde bilimsel araştırma ve etik kurallarına uygun olarak davrandığımı ve aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul edeceğimi beyan ederim.

03.07.2024

Recep IŞIK

ÖNSÖZ

Günümüzün veri odaklı dünyasında, büyük ve karmaşık veri kümelerinden içgörülerini etkili bir şekilde iletme durumu, çeşitli sektörlerde kritik bir gereksinim haline gelmiştir. Geniş ve dağınık veri kümeleri içinde gizlenmiş mesajları iletmek için yeni temsil araçlarının ortaya çıkışı, bu artan gereksinimi karşılamak için önemli bir çözüm olarak "veri görselleştirmenin" yükselişini sağlamıştır. Veri görselleştirme uygulamaları medya ve iletişim, eğlence, yayıncılık, sağlık, eğitim, ulaşım gibi birçok sektörde ve pek çok farklı alanda veri analizi çalışmalarının ayrılmaz bir parçası haline gelmiştir.

Bu çalışma, ülkemizde kütüphanecilik ve bilgi bilimi alanına özgü olarak bir veri görselleştirme sürecini tüm adımlarıyla ele alan ve bu bağlamda bir uygulama geliştiren alanımızdaki ilk çalışma olması bakımından özgün bir değere sahiptir. Çalışmanın hem kütüphanecilik ve bilgi bilimi alanında hem de diğer alanlarda veri analizi ve veri görselleştirme üzerine çalışma yapacak araştırmacılara rehberlik edeceği ve yarar sağlayacağı öngörülmektedir. Ayrıca Ankara Üniversitesi'nin TR Dizin'de indekslenen yayınlarını içeren veri setleri ile gerçekleştirilen bu çalışma, bir ulusal yayın veri tabanının görselleştirme açısından sunduğu olanakların avantajlı ve dezavantajlı yönlerinin anlaşılmasına da katkı sunmaktadır. Bu yönüyle ulusal veri tabanının geliştirilmesinde yararlanılabilecek bulgular ortaya koymaktadır.

Öncelikle akademik çalışmam konusunda beni teşvik eden, tez yazım sürecinde bilgisi ve deneyimleri ile titiz rehberliği, tavsiye ve yönlendirmelerinden dolayı tez danışmanım Sayın Prof. Dr. Tülay OĞUZ'a çok teşekkür eder, şükranlarımı sunarım.

Tez izleme komitesinde bulunmayı kabul ederek değerli katkılar sunan hocalarım Prof. Dr. Özlem GÖKKURT DEMİRTEL ve Doç. Dr. Haydar YALÇIN'a verdikleri değerli desteklerden dolayı çok teşekkür ediyorum.

Yönlendirmeleri ile araştırmaya önemli katkılar sağlayan değerli hocalarım Prof. Dr. Sacit ARSLANTEKİN, Doç. Dr. Mustafa BAYTER ve Dr. Öğr. Üyesi Neslihan ER-KOÇOĞLU'na değerlendirmeleri, fikirleri ve destekleri için çok teşekkür ediyorum.

Çalışmam süresince verdikleri manevi destek ve motive edici yaklaşımlarından dolayı, Kütüphaneler ve Yayınlar Genel Müdürlüğündeki mesai arkadaşlarım Şube Müdürü Selma DEMİRÇİN'e, Sevinç CANBOLAT'a ve Şenol KARADENİZ'e çok teşekkür ediyorum.

Bu yolculukta yanımda olan, maddi ve manevi katkılarını esirgemeyen sevgili eşim Demet IŞIK'a ve çocuklarım Fırat IŞIK'a ve Baran IŞIK'a sabırla, anlayışla ve hoşgörüyle verdikleri destekten dolayı minnetlerimi sunuyorum. İyi ki varsınız canlarım.

Son olarak çalışmalarından ilham aldığım, yararlandığım, bu çalışmanın sonundaki kaynakçada belirterek kendilerine atıf yaptığım dünyanın dört bir yanından tüm yazarlara ve araştırmacılara şükranlarımı sunarım.

Recep IŞIK, Ankara/2024



İÇİNDEKİLER

ÖNSÖZ.....	i
İÇİNDEKİLER.....	iii
ŞEKİL LİSTESİ.....	vii
TABLO LİSTESİ	xii
KOD LİSTESİ	xiv
KISALTMALAR.....	xx
1.BÖLÜM: GİRİŞ	1
1. 1 Araştırmanın Konusu ve Önemi	1
1.2 Araştırmanın Amacı.....	7
1.3 Araştırmanın Problemi, Soruları, Alt Soruları ve Hipotezi	8
1.4 Araştırmanın Sınırlılıkları.....	10
1.5 Araştırmanın Kapsamı, Evreni ve Örneklemi.....	10
1.6 Araştırmanın Yöntemi ve Veri Toplama Teknikleri.....	10
1.7 Araştırmanın Özgün Değeri	14
1.8 Araştırmanın Düzeni.....	14
1.9 Kaynaklar.....	15
2. BÖLÜM: KAVRAMSAL ÇERÇEVE.....	17
2.1 Veri Görselleştirmenin Tarihsel Gelişimi.....	19
2.2 İlgili Literatür	32
2.3 Veri Görselleştirme Yöntem ve Araçları.....	47
2.3.1 Veri Görselleştirme Yöntemleri.....	48
2.3.1.1 Veri Görselleştirmede Kodlama.....	53
2.3.1.2 Veri Görselleştirme Türleri	57
2.3.2. Görselleştirme araçları	67
3. BÖLÜM: YÖNTEM	74

3.1 Veri Toplama	77
3.1.1 Veri Kaynağı Kategorisi, Veri Alanları ve Veri Türlerinin Belirlenmesi	78
3.1.2 Veri Toplama Aşamasında Kullanılan Araçlar / Yöntemler.....	82
3.1.3 Web Kazıma ile Veri Seti Oluşturma	85
3.2 Veri Ön işleme (Data preprocessing)	103
3.2.1 Veri Seti Ön İzleme	104
3.2.2 Veri Temizleme	108
3.2.2.1 Ayıklama İşlemi	109
3.2.2.2 Veri Türlerinin Dönüştürülmesi	110
3.2.2.3 Veri Alanlarında Boş Değerlerin Tespit Edilmesi	114
3.2.2.4 Mükerrer Kayıtların Tespit Edilmesi	116
4.BÖLÜM: BULGULAR.....	117
4.1 Keşifsel Veri Analizi ve Görselleştirmeler.....	117
4.1.1 Soru 1: Makalelerin temel konu kategorilerine göre dağılımı nedir?	120
4.1.2 Soru 2: Makalelerin dillere göre dağılımı nedir?	123
4.1.2.1 Soru 2.1 Türkçe dışındaki dillerde yayınlanmış makaleler hangi konu kategorisinde yer almaktadır?	124
4.1.3 Soru 3: Makalelerin türlere göre atıf alma durumları nedir?	125
4.1.3.1 Soru 3.1 Türlerin konu ve yıla göre dağılımı nedir?	126
4.1.3.2 Soru 3.2 En çok atıf alan makale türleri hangileridir?	129
4.1.4 Soru 4: Makale sayıları ile atıf sayılarının yıllara göre dağılımı nedir?	133
4.1.4.1 Soru 4.1 Makalelerin yıllara göre görüntülenme ve indirilme sayıları nedir?	134
4.1.5 Soru 5: Ulaşılabilirlik açısından makalelerin erişim türüne (açık/kapalı) göre dağılımı nedir?	136
4.1.5.1 Soru 5.1 Hangi temel konulu makaleler erişime kapalıdır; bunlar yıllar içinde artış gösterme eğiliminde midir?	137
4.1.6 Soru 6: Yazar iş birliği (tek yazarlı/ortak yazarlı) eğilimi ne düzeydedir?	138

4.1.6.1 Soru 6.1 Yazar sayısına göre yayın sayısı dağılımı nedir?	141
4.1.6.2 Soru 6.2 Makale başına düşen yazar sayısına göre atıf alma durumu nedir?. 145	
4.1.6.3 Soru 6.3 Tek yazarlı makaleler ile çok yazarlı makalelerin atıf alma oranları ne düzeydedir?	146
4.1.6.4 Soru 6.4 Kurumsal iş birliği ne düzeydedir?	147
4.1.6.5 Soru 6.5 Kurumsal iş birliğinin yıla göre dağılımı nedir?.....	151
4.1.6.6 Soru 6.6 Kamu iş birliği olan makalelerde ilk yazara göre yazar kurum dağılımı nedir?.....	153
4.1.7. Soru 7 Makalelerin konusal olarak dağılımı nedir?	156
4.1.7.1 Türkçe makalelerin konusal olarak yıl ve genel konu kategori bazındaki dağılımı nedir?.....	162
4.1.7.2 İngilizce makalelerin konusal olarak yıl ve genel konu kategorisi bazındaki dağılımı nedir?.....	164
4.1.7.3 Türkçe anahtar kelimelerin yıllara göre dağılımı nedir?	165
4.1.7.4 İngilizce anahtar kelimelerin yıllara göre dağılımı nedir?	171
4.1.8 Soru 8: Makalelerin yayınlandığı dergi sayılarının yıllara göre dağılımı nedir? ...	172
4.1.8.1 Soru 8.1: Dergi sayısı artış oranı ile yayın sayısı artış oranı arasında bir korelasyon var mıdır?.....	174
4.1.8.2 Soru 8.2: Yıllara göre dergi başına düşen makale sayıları nedir?	175
4.1.8.3 Soru 8.3: AÜ yazarları hangi dergileri tercih etmektedir?	177
4.1.8.4 Soru 8.4: AÜ makalelerinin yayınlandığı dergilerin yayın hayatına başlama tarihine göre dağılımı nedir?	180
4.1.8.5 Soru 8.5: AÜ makalelerinin yayınlandığı dergilerin yıllık yayın periyodları, makale sayıları, aldığı atıf sayıları ve kendine atıf sayıları nedir?	182
5. BÖLÜM: SONUÇ VE ÖNERİLER.....	185
KAYNAKÇA	197
EKLER.....	212
EK-1: Terimler Listesi	212

ÖZET.....	215
SUMMARY.....	217



ŞEKİL LİSTESİ

Şekil 1: Veri görselleştirmenin görsel tanımı	18
Şekil 2: 1482'de basıldığı şekliyle Batlamyus'un dünya haritası	21
Şekil 3: Yedi göksel cismin uzay ve zaman üzerindeki değişen konumu	22
Şekil 4: Michael Florent van Langren [1600-1675] tarafından hazırlanan, Toledo'dan Roma'ya olan mesafeyi boylam olarak belirleyen grafik	23
Şekil 5: Örnek grafik türleri (çizgi, çubuk, pasta, histogram, eşyükselti, dağılım ve zaman serisi)	24
Şekil 6: 1801 tarihli Playfair'in "Ülkelerin nüfus, arazi genişliği, vergi yükü ve siyasi durumu" temalı grafiği	24
Şekil 7: Alexander von Humboldt'un 1805 tarihli Naturgemalde adlı eseri	25
Şekil 8: John Snow'un Londra'daki kolera salgını haritası.....	26
Şekil 9: İngiliz askerlerinin ölüm sebebini gösteren gül diyagramı-Florence Nightingale-1858	27
Şekil 10: Messier 87 (M87) gökadasındaki bir kara deliğin ilk görüntüsü	28
Şekil 11: Yaz saati uygulamasının etkileri	29
Şekil 12: Veri görselleştirme tarihinde kilometre taşları zaman çizelgesi.....	30
Şekil 13.a: WoS Research Area alanında “data visualization” terimi ile yapılan tarama sonucuna göre yayınların dağılımı (en çok yayın yapılan konu kategorileri).....	33
Şekil 13.b: WoS Research Area alanında “data visualization” terimi ile yapılan tarama sonucuna göre yayınların dağılımı (en az yayın yapılan konu kategorileri)	34
Şekil 14: “Data visualization” konusunda yapılan yayınların yıllara göre yayın dağılımı.....	35
Şekil 15: Bilimsel veri görselleştirme çerçevesi.....	40
Şekil 16: College of Charleston ders saatleri ve kitap harcama oranları	44
Şekil 17: Bir öğretme ve öğrenme metodolojisini özetleyen süreç diyagramı	45
Şekil 18: Veri görselleştirme modeli	48
Şekil 19: Hayvanların yiyecek araması ile insanların bilgi arama davranışları.....	51
Şekil 20: Yan yana tasarımı şeklinde iki zaman serisinin karşılaştırılması	52

Şekil 21: Üst üste bindirme tasarımı şeklinde iki zaman serisinin karşılaştırılması.....	52
Şekil 22: Açık temsil tasarımı şeklinde iki zaman serisinin karşılaştırılması.....	53
Şekil 23: İşaretlerin (marks) kullanıldığı örnek görselleştirmeler	54
Şekil 24: “Data visualisation: a handbook for data driven design” kitabının ilk bölümünde kullanılan kelimelerin sıklığı.....	57
Şekil 25: Sosyal medya platformlarının 2021-2024 arası dünyadaki kullanım oranları	58
Şekil 26: Dünya çapındaki internet tarayıcı pazar payı	59
Şekil 27: Petrol kulelerinin dünya çapındaki dağılımları	59
Şekil 28: COVID-19 pandemisi göçü	60
Şekil 29: ABD- King County'deki müstakil ev satışlarının düştüğünü gösteren histogram örneği.....	61
Şekil 30: Almanya'nın Avrupa ve Kuzey Amerika'ya karşı küresel rekabet edebilirlik puanlarının karşılaştırması	62
Şekil 31: Avrupa nüfusunun arttığı ve azaldığı yerler	63
Şekil 32: ABD’de kadınların bölümlere göre üniversite mezuniyet oranları	64
Şekil 33: 1995-2014 yılları arası İngiltere ve Galler'deki ortalama günlük doğum oranları ...	65
Şekil 34: 20 Kasım 2021 tarihli kripto para hacimleri ve fiyat durumları.....	66
Şekil 35: Veri görselleştirme süreci.....	74
Şekil 36: Python programlama dili “pandas” kod kütüphanesi veri türleri	80
Şekil 37: Pandas veri yapısı örneği.....	81
Şekil 38: TR Dizin Basit Arama Çubuğu	82
Şekil 39: TR Dizin Dışarı Aktarma Arayüzü.....	82
Şekil 40: Web kazıma için kullanılan araçlar	83
Şekil 41: Jupyter Notebook’da “print” fonksiyonunun nasıl çalıştığını gösteren örnek ekran görüntüsü	85
Şekil 42: TR Dizin robots.txt dosyası	86
Şekil 43: TR Dizin makale özet sayfası üst veri alanları	87

Şekil 44: TR Dizin makale detay sayfası üst veri alanları	88
Şekil 45: TR Dizin dergi detay bilgi alanı	88
Şekil 46: Anahtar kelime ve konu alanı verilerinin toplandığı DergiPark arayüzü	97
Şekil 47: Eksik verisi olan değişkenler	115
Şekil 48: Yıllara göre makale konu kategorisi dağılımı	121
Şekil 49: Yıllara göre yayın sayısı dağılımı	122
Şekil 50: Makalelerin dillere göre dağılımı	123
Şekil 51: Dil alt veri setinin “Treemap” olarak görselleştirilmesi	125
Şekil 52: Yayınların makale türüne göre dağılımı	126
Şekil 53: Makale türlerinin yıllara göre dağılımı	127
Şekil 54: Konu kategorisine göre kümelenmiş makale türü dağılımı	129
Şekil 55: Boxplot grafik türünün özellikleri	130
Şekil 56: Makale türü kategorilerinin aldıkları atıf sayılarına göre dağılımı	131
Şekil 57: Makale türü kategorilerinin aldıkları atıf sayılarına göre dağılımı	132
Şekil 58: Yıllık toplam yayın sayısı ile yıllık toplam atıf sayılarının karşılaştırılması	134
Şekil 59: Makale yılına göre makale görüntülenme ve indirilme sayıları dağılımı	135
Şekil 60: Yayınların erişim türüne göre dağılımı	136
Şekil 61: Erişim türü bibliyografik veri olan kayıtların temel konu kategorisine göre dağılımı	138
Şekil 62: Yayın başına yazar sayısının “describe()” metodu ile incelenmesi	141
Şekil 63: Tek yazarlı ve ortak yazarlı yayınların oranı	142
Şekil 64: Yazar sayısına göre makale sayısının “barh plot” grafiği ile incelenmesi	143
Şekil 65: Yazar sayısına göre makale sayısının “bar plot” grafiği ile incelenmesi (makale türü “araştırma makalesi” olan makaleler)	144
Şekil 66: Makale başına düşen düşen yazar sayısı ile atıf sayılarının yayın yılına göre incelenmesi	145
Şekil 67: Makale başına düşen yazar sayısının “box plot” grafiği ile incelenmesi	146

Şekil 68: Makale başına düşen atıf sayısının “box plot” grafiği ile incelenmesi	146
Şekil 69: Tek ve ortak yazarlı yayınların atıf oranları	147
Şekil 70: En çok iş birliği yapılan kurumlar (20 kurum)	149
Şekil 71: Ortak yazarlı makalelerde iş birliği oranı	151
Şekil 72: Kamu iş birliği durumu	152
Şekil 73: Kamu iş birliği durumu (Yayın yılına göre).....	152
Şekil 74: Diğer kurumlarla iş birliği yapılan makalelerdeki ilk kurum durumu	155
Şekil 75: Makalelerin DOI durumu	156
Şekil 76: DOI linkinin çalışma durumu	157
Şekil 77: DOI hedef URL adresi belirleme	157
Şekil 78: DOI adresi DergiPark olan makaleler	158
Şekil 79: Yayınların yıl, dil, konu kategorisi ve makale konusuna göre dağılımı.....	160
Şekil 80: Konuların dillere göre dağılımı	161
Şekil 81: Türkçe yayınların “konu_df” veri setindeki “yıl”, “kkategori” ve “konu” değişkenlerine göre dağılımı (treemap).....	163
Şekil 82: İngilizce yayınların “konu_df” veri setindeki “yıl”, “kkategori” ve “konu” değişkenlerine göre dağılımı (treemap).....	164
Şekil 83: Anahtar kelime sayısının yıllara göre dağılımı	167
Şekil 84: Anahtar kelime sayısının dillere göre dağılımı	168
Şekil 85: Anahtar kelime sayısının konu kategorilerine göre dağılımı	168
Şekil 86: Türkçe anahtar kelimelerin yıllara göre kelime bulutu şeklinde dağılımı.....	170
Şekil 87: İngilizce anahtar kelimelerin yıllara göre kelime bulutu şeklinde dağılımı	171
Şekil 88: AÜ makalelerinin yıllara göre yayınlandığı dergi sayısı dağılımı	173
Şekil 89: Makale yılına göre toplam makale ve dergi sayısı dağılımı (1) (çizgi).....	175
Şekil 90: Makale yılına göre toplam makale ve dergi sayısı dağılımı (2) (çubuk).....	175
Şekil 91: Yıllara göre dergi başına düşen makale sayısı	176
Şekil 92: Tekil dergi sayısının " makale_konu_Kategorisi"ne göre dağılımı.....	177

Şekil 93: Yazarların tercih ettiği Sosyal konu kategorisindeki dergiler	178
Şekil 94: Yazarların tercih ettiği Fen konu kategorisindeki dergiler	179
Şekil 95: Dergi sayılarının yayın hayatına başlama tarihine göre dağılımı	180
Şekil 96: Dergilerin yayına başlama tarihine göre makale sayıları	181
Şekil 97: "der_yillikYayin_sayisi, der_makaleSayisi, der_aldigi_atifSayisi, der_kendine_atifSayisi" değişkenlerinin boxplot ile incelenmesi	183
Şekil 98: Dergi yıllık yayınlanma sayısına göre makale sayısı dağılımı	184
Şekil 99: Veri görselleştirme modeli	194



TABLO LİSTESİ

Tablo 1: İşaret kodlama sınıflandırması	54
Tablo 2: Nitelik kodlama sınıflandırması	55
Tablo 3: Veri görselleştirme için kullanılan çeşitli görselleştirme türleri	66
Tablo 4: Veri görselleştirme araçları	69
Tablo 5: Etkileşimli ağ görselleştirme araçlarının değerlendirilmesi	73
Tablo 6: TR Dizin Makale üst veri alanları ve veri seti değişken adları	86
Tablo 7: TR Dizin Dergi veri alanları ve veri seti değişken adları	87
Tablo 8: Veri alanları (değişken) özellikleri	105
Tablo 9: Veri setinin ilk beş satırının görünümü	107
Tablo 10: Veri setinin son beş satırı görünümü	107
Tablo 11: Veri setinden ayıklanan değişken listesi	109
Tablo 12: Değişken veri türleri görünümü	111
Tablo 13: Veri türü dönüştürülecek değişkenler	112
Tablo 14: Dönüştürme işlemi sonrası değişken veri türleri görünümü	113
Tablo 15: Değişkenlere ait kayıp değer sayıları görünümü	114
Tablo 16: Veri setine yöneltilen sorular	118
Tablo 17: Makale yılı ve makale konu kategorisinden oluşan alt veri seti	121
Tablo 18: “dil_df” alt veri seti son beş kayıt görüntüsü	124
Tablo 19: “tur_df” alt veri seti görünümü	127
Tablo 20: Güncel “tur_df” alt veri setinin ilk beş kaydı	128
Tablo 21: “makale_atif_sayisi” alt veri seti	133
Tablo 22: Makale yılına göre toplam makale görüntüleme ve indirme sayıları	135
Tablo 23: “bibliyografik_df” veri seti ilk beş kayıt	137
Tablo 24: “yazar_Adi” değişkenindeki ilk beş kaydın görünümü	139
Tablo 25: “yazar_df” alt veri setine makale “ilk_Kurum” değişkeninin eklenmesi	155

Tablo 26: “konu_df” alt veri setinin ilk beş satırı	159
Tablo 27: “keywords_df” alt veri setinin ilk beş satırı.....	167
Tablo 28: “dergiSayisi_df” alt veri seti görünümü	173
Tablo 29: “dergiSayisi_df” alt veri seti görünümü	176



KOD LİSTESİ

Kod 1: Veri türü dönüştürme.....	81
Kod 2: Google Drive için izinlerin tanımlanması	89
Kod 3: Web kazıma için kütüphanelerin aktifleştirilmesi	89
Kod 4: TR Dizin sayfası için oluşturulan sorgu fonksiyonu	90
Kod 5: Makale özet alanı için liste değişkenlerinin tanımlanması	91
Kod 6: Makale özet alanı verilerinin toplanması.....	91
Kod 7: Makale özet alanı liste değişkenlerinin DataFrame’e aktarılması.....	91
Kod 8: Yazar verilerinin toplanması için liste değişken tanımlanması	92
Kod 9: Yazar adlarının toplanması ve “yazarAdı” liste değişkenine aktarılması	92
Kod 10: “yazarAdı” liste değişkeninin DataFrame’e değişken olarak eklenmesi.....	92
Kod 11: “yazar_Kurum” liste değişkeninin tanımlanması	92
Kod 12: Yazar kurum verilerinin toplanması ve “yazar_Kurum” liste değişkenine aktarılması	93
Kod 13: “yazar_Kurum” geçici liste değişkeninin DataFrame’e değişken olarak aktarılması.....	93
Kod 14: Makale detay alanı için liste değişkenlerin oluşturulması.....	93
Kod 15: Makale detay alanı verilerinin toplanması ve liste değişkenlerine aktarılması	94
Kod 16: Makale detay alanı liste değişkenlerinin DataFrame’e değişken olarak eklenmesi ..	94
Kod 17: Görüntülenme ve indirme sayısı verileri için liste değişkenlerinin oluşturulması	94
Kod 18: Görüntülenme ve indirme verilerinin toplanması ve liste değişkenlerine aktarılması	95
Kod 19: Görüntülenme ve indirme liste değişkenlerinin DataFrame’e aktarılması	95
Kod 20: DOI verisi için liste değişkeni oluşturulması	95
Kod 21: DOI adreslerinin toplanması ve liste değişkenine aktarılması	96
Kod 22: “makaleDoi” geçici liste değişkeninin DataFrame’e değişken olarak eklenmesi	96
Kod 23: Anahtar kelimeler için liste değişken tanımlanması.....	97
Kod 24: Validators paketinin yüklenmesi ve aktifleştirilmesi	97

Kod 25: DergiPark'taki anahtar kelimelerin toplanması ve “keywords” liste değişkenine aktarılması	98
Kod 26: Anahtar kelimeler verilerini taşıyan liste değişkenin DataFrame'e eklenmesi	98
Kod 27: DergiPark makale konu alanları için liste değişkenlerin tanımlanması	98
Kod 28: DergiPark makale konu alanlarının ve URL adreslerinin toplanması	99
Kod 29: “topics ve page_url” liste değişkenlerinin DataFrame'e değişken olarak eklenmesi	99
Kod 30: TR Dizin dergi sayfası için liste değişkenlerinin tanımlanması	100
Kod 31: TR Dizin dergi sayfası verilerinin toplanması (1)	101
Kod 32: TR Dizin dergi sayfası verilerinin toplanması (2)	101
Kod 33: TR Dizin dergi sayfası verilerinin toplanması (3)	102
Kod 34: TR Dizin dergi sayfası verilerinin toplanması (4)	102
Kod 35: Liste değişkenlerinin DataFrame'e değişken olarak aktarılması	103
Kod 36: TR Dizin'den toplanan verilerin dosya olarak Google Drive'a kaydedilmesi	103
Kod 37: Google Drive için erişim izinlerinin tanımlanması	104
Kod 38: Veri seti ön analiz işlemleri için kod kütüphanelerinin aktifleştirilmesi	104
Kod 39: Ön analiz işlemleri yapılacak veri setinin yüklenmesi ve DataFrame'e dönüştürülmesi	104
Kod 40: Veri seti satır ve sütun sayısının görüntülenmesi	105
Kod 41: Veri seti özet bilgilerinin görüntülenmesi	105
Kod 42: Değişkenlerin veri türüne göre gruplanması	106
Kod 43: Veri setinde yer alan değişkenlerin (sütunların) görüntülenmesi	106
Kod 44: Veri setinde yer alan ilk beş satırın görüntülenmesi	106
Kod 45: Veri setinde yer alan son beş satırın görüntülenmesi	107
Kod 46: Veri setinde yer alan mükerrer kayıt sayısının görüntülenmesi	108
Kod 47: Analizlerde kullanılmayacak değişkenlerin ayıklanması	109
Kod 48: Veri seti değişkenleri (sütun adları)	110
Kod 49: Yazar kurum verilerinin başlık formatına dönüştürülmesi	110

Kod 50: Değişkenlerin veri türlerinin toplu olarak görüntülenmesi.....	111
Kod 51: “atif_Sayisi” değişkenindeki boş kayıtlara varsayılan değer atama işlemi	112
Kod 52: Değişkenlere ait veri türlerinin dönüştürülmesi	113
Kod 53: Değişkenlere ait boş kayıt sayısının görüntülenmesi	114
Kod 54: Değişkenlerin eksik kayıt sayılarının görselleştirilmesi	115
Kod 55: Mükerrer satırların tespit edilmesi ve ayıklanması.....	116
Kod 56: Makale temel konu kategorisi ögeleri	120
Kod 57: Makale yılı ve makale konu kategorisinden oluşan “yil_df” alt veri setinin oluşturulması	120
Kod 58: “yil_df” alt veri setinin görselleştirilmesi.....	121
Kod 59: Yıllara göre yayın sayısı dağılımı.....	122
Kod 60: Yayınların dilleri göre dağılımı	123
Kod 61: “dil_df” alt veri setinin oluşturulması	124
Kod 62: “Dil” alt veri setinin “Treemap” olarak görselleştirilmesi	124
Kod 63: Makale türüne göre dağılım.....	125
Kod 64: “tur_df” alt veri setinin oluşturulması	126
Kod 65: “tur_df” alt veri setinin görselleştirilmesi	127
Kod 66: “tur_df” alt veri setinin güncellenmesi	128
Kod 67: Temel konu kategorilerine göre makale türlerinin dağılımı	128
Kod 68: Makale türlerinin atıf sayılarına göre dağılımı	131
Kod 69: Makale türlerinin atıf sayılarına göre dağılımı	132
Kod 70: “makale_atif_sayisi” alt veri setinin oluşturulması	133
Kod 71: Yıllık toplam makale sayısı ile yıllık toplam atıf sayılarının karşılaştırılması.....	133
Kod 72: “view_downloads_df” alt veri setinin oluşturulması.....	134
Kod 73: Makale yılına göre makale görüntülenme ve indirme sayıları dağılımı	135
Kod 74: Yayınların erişim türüne göre dağılımı	136

Kod 75: Eriřim türü bibliyografik veri olan kayıtların alt veri seti	137
Kod 76: Sadece bibliyografik verilerine erişilen makalelerin yıllara göre konu dağılımı	137
Kod 77: “yazar_df” alt veri setinin oluşturulması	139
Kod 78: Yazar sayısı sütununun “describe” metodu ile incelenmesi	140
Kod 79: Yazar sayısı sütununun “describe” metodu ile incelenmesi	140
Kod 80: “yazar_df” alt veri setine “durum” değişkeninin eklenmesi	141
Kod 81: Tek yazarlı ve ortak yazarlı yayınların oranı	142
Kod 82: Yazar sayısına göre yayın sayısı dağılımı	143
Kod 83: Yazar sayısına göre makale sayısı dağılımı (araştırma makaleleri)	144
Kod 84: Makale başına düşen yazar sayısı ile atıf sayılarının yayın yılına göre incelenmesi	145
Kod 85: Makale başına düşen yazar sayısının “box plot” grafiğı ile incelenmesi	146
Kod 86: Makale başına düşen atıf sayılarının “box plot” grafiğı ile incelenmesi	146
Kod 87: “atıflar_df” alt veri setinin oluşturulması	147
Kod 88: “atıflar_df” alt veri setinin görselleştirilmesi	147
Kod 89: Makaleler için kurum adlarının tekilleştirilmesi	148
Kod 90: “kurum_df” alt veri setinin oluşturulması	148
Kod 91: En çok iş birliğı yapılan kurumlar (20 kurum)	148
Kod 92: Diğer kurumlar ile iş birliğı sayısı	149
Kod 93: İş birliğı yapılan tekil kurum sayısı	149
Kod 94: Kurum içi diğer akademisyenlerle iş birliğı sayısı	149
Kod 95: Kurum iş birliğı durumunun tespit edilmesi	150
Kod 96: Ortak yazarlı makalelerdeki kurum iş birliğinin görselleştirilmesi	150
Kod 97: İş birliğı durumunun veri setine sütun olarak eklenmesi	152
Kod 98: Kamu iş birliğı durumu	152
Kod 99: Kamu iş birliğı durumu (Yayın yılına göre)	152

Kod 100: “araştırma, derleme ve olgu sunumu” türünde olan makalelerin filtrelenmesi	153
Kod 101: “yazar_df” veri setindeki ortak yazarlı makalelerin filtrelenmesi.....	154
Kod 102: AÜ yazarlarının diğer kurumlarla yaptıkları iş birliği sayısı.....	154
Kod 103: İlk kurum değişkeninin oluşturulması	154
Kod 104: Diğer kurumlarla iş birliği ile oluşturulan makalelerdeki ilk kurum durumu	155
Kod 105: Makalelerin DOI durumu	156
Kod 106: DOI linkinin çalışma durumu	157
Kod 107: DOI hedef URL adresi belirleme.....	157
Kod 108: DOI adresi DergiPark olan makaleler	158
Kod 109: “konu_df” alt veri setinin oluşturulması.....	158
Kod 110: “konu_df” alt veri setinin yıl, dil, konu kategorisi ve makale konusuna göre dağılımı.....	159
Kod 111: Konuların dillere göre dağılımı	161
Kod 112: Türkçe makale konuları sayısı.....	161
Kod 113: İngilizce makale konuları sayısı	161
Kod 114: Türkçe ve İngilizce olmayan makale konu sayısı.....	162
Kod 115: Türkçe yayınların “konu_df” veri setindeki “yıl”, “kkategori” ve “konu” değişkenlerine göre dağılımı (treemap).....	162
Kod 116: İngilizce yayınların “konu_df” veri setindeki “yıl”, “kkategori” ve “konu” değişkenlerine göre dağılımı (treemap).....	164
Kod 117: Anahtar kelime alanı dolu olan makale sayısı (TR Dizin ve DergiPark)	165
Kod 118: Anahtar kelime alanı dolu olan kayıtlar (DOI adresi DergiPark olanlar).....	165
Kod 119: Anahtar kelime alanı dolu olan kayıtlar (TR Dizin üst veri sayfası).....	166
Kod 120: “keywords_df” anahtar kelime alt veri setinin oluşturulması	166
Kod 121: Anahtar kelime sayısının yıllara göre dağılımı.....	167
Kod 122: Anahtar kelime sayısının dillere göre dağılımı.....	168
Kod 123: Anahtar kelime sayısının konu kategorilerine göre dağılımı.....	168

Kod 124: Türkçe anahtar kelimelerin yıllara göre kelime bulutu şeklinde dağılımı	169
Kod 125: İngilizce anahtar kelimelerin yıllara göre kelime bulutu şeklinde dağılımı	171
Kod 126: 2017-2021 arası makalelerin yayımlandığı tekil dergi sayısı	172
Kod 127: “dergiSayisi_df” alt veri setinin oluşturulması.....	172
Kod 128: AÜ makalelerinin yıllara göre yayınlandığı dergi sayısı dağılımı	173
Kod 129: “dergiSayisi_df” veri setine “makale_Sayisi” değişkeninin eklenmesi	174
Kod 130: Makale yılına göre toplam makale ve dergi sayısı dağılımı (1)	174
Kod 131: Makale yılına göre toplam makale ve dergi sayısı dağılımı (2)	174
Kod 132: “sayi” değişkeninin “dergiSayisi_df” alt veri setine eklenmesi	175
Kod 133: Yıllara göre dergi başına düşen makale sayısı.....	176
Kod 134: Dergi adlarının başlık formatına dönüştürülmesi	177
Kod 135: Tekil dergi sayısının " makale_konu_Kategorisi"ne göre dağılımı.....	177
Kod 136: Yazarların tercih ettiği Sosyal konu kategorisindeki dergiler	178
Kod 137: Yazarların tercih ettiği Fen konu kategorisindeki dergiler.....	179
Kod 138: “dergi_df” alt veri setinin oluşturulması	180
Kod 139: Dergi sayılarının yayın hayatına başlama tarihine göre dağılımı	180
Kod 140: Dergilerin yayına başlama tarihine göre makale sayıları	181
Kod 141: "der_yillikYayin_sayisi, der_makaleSayisi, der_aldigi_atifSayisi, der_kendine_atifSayisi" değişkenlerinin boxplot ile incelenmesi	182
Kod 142: Dergi yıllık yayınlanma sayısına göre makale sayısı	183

KISALTMALAR

APA	American Psychological Association
API	Application Programming Interface
AÜ	Ankara Üniversitesi
bkz.	Bakınız
CSV	Comma separated values-Virgülle ayrılmış değerler
EDA	Exploratory Data Analysis-Keşifsel Veri Aanalizi
HTML	Hyper Text Markup Language
HTTP	Hyper Text Transfer Protocol-Hiper-Metin Transfer Protokolü
ISO	International Organization for Standardization
M.S.	Milattan Sonra
s.	Sayfa
ss.	Sayfa aralığı
t.y.	Tarih yok
ULAKBİM	Ulusal Akademik Ağ ve Bilgi Merkezi
URL	Uniform Resource Locator-Tekdüzen Kaynak Bulucu
ÜAK	Üniversitelerarası Kurul
XML	Extensible Markup Language
YÖK	Yüksek Öğretim Kurulu

1.BÖLÜM: GİRİŞ

Çalışmanın bu bölümünde araştırmanın konusu ve önemi, araştırmanın amaçları, soruları ve hipotezlerine yer verilmiş ve araştırmanın kapsamı, düzeni ve yöntemi açıklanmış, araştırmanın özgün değeri ortaya konmaya çalışılmıştır.

1. 1 Araştırmanın Konusu ve Önemi

Teknolojinin sürekli olarak gelişimi sayesinde dünyadaki veri hacmi ve çeşitliliği, insanlık tarihinde daha önce hiç görülmediği hızda artmıştır (Aktan, 2018, s.1). Uydu, kamera, giyilebilir teknolojiler, nesnelerin interneti, dijitalleşme, çevrimiçi ticaret vb. gibi kaynaklardan akan verilerle, internet üzerindeki veri miktarı, hızı ve çeşitliliği her geçen gün artmakta ve bu verilerin bütünleşik bir şekilde işlenmesine gereksinim duyulmaktadır. Dolayısıyla günümüzde karşı karşıya olduğumuz zorluğu “tüm bu verilerle nasıl başa çıkacağız?” (Phetteplace, 2012) şeklinde ifade edebiliriz. Bilgi sistemlerinin ham maddesi olan verilerin; örneğin kuruluşlarda karar vericiler ve bilim dünyasında ise araştırmacılar tarafından kendi amaçları doğrultusunda erişilebilir ve kullanılabilir olması son derece önemli hale gelmiştir. Uygun veriye doğru zamanda ulaşabilmek kadar, uygun veriye ulaşıldıktan sonra toplanan yapısal ve yarı yapısal verilerin bir bütün olarak ilişkilendirilerek işlenmesi ve stratejik kararlar verme aşamasında kullanılabilmesi için veri analizi son derece önem kazanmıştır (Aktan, 2018, s.10). Elde edilen veri kümeleri içerisinde neyin önemli/yararlı olabileceğini ya da olmayacağını anlamak, giderek daha zor hale gelmektedir. Analiz yapma ve analiz sonuçlarını sunma tekniklerindeki gelişmeler, veri kümelerinin barındırdığı örüntülerin açığa çıkmasını sağlamakta, verilerin daha kolay anlaşılmasına ve yorumlanmasına aracılık etmektedir. Veri görselleştirme, veri kümeleri içerisinde neyin önemli olup neyin olmadığını anlamak için fırsatlar sunmaktadır.

Artık büyük ve dağınık veri kümelerinin mesajını iletmek için yeni temsil araçlarına ihtiyaç duyulmakta ve “veri görselleştirme”, günümüzde bu ihtiyacı karşılamak için güncel bir konu olarak gündemdeki yerini almış bulunmaktadır. İletişim, medya ve eğlence sektörlerinden enerji sektörüne, sağlık, eğitim, ulaşım hizmetlerinden siyasi kampanyalara kadar birçok alanda veri analiz çalışmaları kapsamında veri görselleştirme teknikleri kullanılmaktadır. Görselleştirme teknikleri ve araçları, kullanıcıların veri ile ilgili genel bir

bakış ve içgörü¹ edinmelerini ve görülmesi zor olan veri veya üst verilerdeki örüntüleri görmelerini sağlar (Shneiderman, 2002; Sterman ve Borda, 2017, s.5).

Veri görselleştirme, veri veya bilginin açık ve etkili bir şekilde grafiksel veya resimsel olarak temsil edilmesidir (Brigham, 2016, ss.215-216). Veri görselleştirme verilerin; grafikler, haritalar, etiket bulutları, simülasyonlar şeklinde temsil edilmesi için kullanılan yazılım uygulamaları olarak tanımlanabilir (Reitz, 2014b). Bir başka tanımada ise "bilişi güçlendirmek için verilerin bilgisayar destekli, etkileşimli görsel temsillerinin kullanılması" (Card, Mackinlay ve Shneiderman, 1999'dan aktaran Sterman ve Borda, 2017, s. 5) şeklinde ifade edilmektedir.

Veri görselleştirme, insanların bir alandaki verileri daha etkili bir şekilde anlamalarına ve yorumlamalarına yardımcı olmak için verileri grafiksel olarak temsil etmeyi sağlayan, veri analizinin önemli bir bileşenidir. Bilginin görselleştirme yoluyla temsili, karmaşık bilgilerin daha kolay algılanmasına, anlaşılmasına ve sezgisel olarak yakalanmasına yardımcı olur. Görsel temsiller, veri içindeki ham sayısal biçimde ayırt edilmesi zor olabilecek örüntülerin, eğilimlerin ve ilişkilerin kavranmasına olanak tanır. Görselleştirme, uzman olmayanlar dâhil, çeşitli hedef kitlelere içgörülerin açık ve kısa bir şekilde iletilmesini kolaylaştırır. Sunumlar, raporlar ve akademik yayınlar için veri görselleştirme güçlü bir araçtır. Ayrıca bilgiyi daha anlaşılır ve ilgi çekici hale getirir.

Veri görselleştirme, rasyonel kararlar almak için gerekli olan örüntülerin, eğilimlerin ve aykırı değerlerin belirlenmesine yardımcı olur. Araştırmacıların ve analistlerin verilerdeki anormallikleri veya düzensizlikleri tespit etmesine yardımcı olarak daha fazla araştırma yapılmasının önünü açabilir. Karar vericiler, verilerin görsel temsillerini hızlı bir şekilde yorumlayarak daha rasyonel kararlar alabilirler. Veri görselleştirme, verilerin kapsamlı bir görünümünü sunarak karar vericilerin farklı seçimlerin sonuçlarını anlamalarını kolaylaştırır.

Veri görselleştirme yeni bir konu değildir. Tarih boyunca resimler aracılığıyla anlam aktarma, yıldızların ve diğer gök cisimlerinin konumlarının belirlenmesi, konum tespiti ve keşiflere yardımcı olması amacıyla kullanılmış ve oluşturulan birçok görselleştirme ürünü ve yöntemi günümüze kadar ulaşmıştır. Veri görselleştirmenin en eski biçimleri; mağara resimleri ile kil, taş veya papirüs üzerinde sayıların, sembollerin veya yönlerin temsil edildiği basit

¹ Türk Dil Kurumu (2020) "içgörü"yü "kendi duygularını, kendi kendini anlayabilme yeteneği" olarak açıklamıştır. Başka bir ifade ile içgörü: sezgisel anlayış yoluyla bir şeyin gerçek doğasını anlama durumu, zihinsel görüş veya ayırt etme, temeldeki gerçeği görme yeteneğidir (Insight, 2020).

çizimler şeklinde karşımıza çıkmaktadır. Örneğin: Mısır'daki Torino Papirüs Haritası (Turin Papyrus Map) coğrafi verileri ve madencilik verilerini göstermektedir (Hillery, 2020; Think Insights, 2017). Julius Caesar ve Augustus tarafından kullanılan Roma haritası; siyasi sınırları, yolları, şehirleri ve doğal kaynakları (Dilke, 1987; Think Insights, 2017), Su Song'un Çin gök atlası (MS 1092) ise yıl boyunca güneşin, ayın, gezegenlerin ve yıldızların konumlarını (Bucur, 2022, s.16; Think Insights, 2017) göstermektedir. Bu bağlamda birçok eski uygarlığın, navigasyon, astronomi, tarım ve savaş gibi çeşitli amaçlar için verileri görselleştirdiği bilinmektedir.

16. yüzyılda veri türlerini temsil etmek için renk ve sembollerin kullanıldığı ilk tematik haritalar görülmeye başlamıştır. Rönesans ve Aydınlanma dönemleri, veri görselleştirmeye yönelik doğruluğu ve kesinliği artıran yeni yöntem ve araçların geliştirilmesine tanık olmuştur. Örneğin; Abraham Ortelius 1570 yılında Avrupa'nın siyasi sınırlarını ve nüfus istatistiklerini gösteren haritaların bulunduğu ilk modern atlası yayımlamıştır (D'Orgeix, 2020). Christoph Scheiner ise, 1618'de güneşin aktivitelerini tahmin etmede kullanılan güneş lekelerinin görüntülerini oluşturmuştur (Think Insights, 2017). Michael Florent van Langren (1663) istatistiksel verileri daha doğal bir şekilde göstermek için kontur haritalarını² kullanmıştır (Friendly, 2008, s.3; Tufte, 1997, s.15). William Playfair (1748), sayısal verileri görüntülemek için histogramlar ve nokta grafikleri gibi istatistiksel grafikleri tanıtmıştır. John Snow (1853), su kaynaklı hastalık salgınının kaynağını belirlemek için Londra'daki kolera salgınlarının haritasını oluşturmuştur (DundasBI, 2019; Hillery, 2020).

Sanayi Devrimi, teknolojiye ve toplumda görselleştirme için mevcut verilerin hacmini ve çeşitliliğini artıran hızlı değişiklikleri beraberinde getirmiştir. Örneğin, Charles Minard (1879), renk kodlu semboller kullanarak savaş sırasında Rus askerlerinin hareketlerini gösteren bir harita oluşturmuştur (Friendly, 2000, s.1; Wood ve Bolch, 2018, s.11). Florence Nightingale (1858), İngiliz askerlerinin ölüm sebeplerini yaş gruplarına göre farklı renklerle kullanarak göstermek üzere bir gül grafiği oluşturmuştur (Friendly, 2006, s.14; Friendly, Chen, Härdle, Unwin, 2008; Ward, Grinstein ve Keim, 2015; Kirk, 2019; Erkurt, 2020). Luigi Perozzo (1896), farklı perspektiflere sahip iki görüntüyü birleştiren stereogramları kullanarak nesnelerin 3 boyutlu temsillerini oluşturmuştur (Laurels, 2009; Think Insights, 2017). Galton (1884), sıcaklık değişimlerinin renklerle gösterildiği meteorolojik haritalar oluşturmuş, Charles

² Kontur gabari; yapılacak bir binanın belediyece öngörülen azami yüksekliğine denir. Mevcut bir binanın parsel üzerindeki zemin oturumunu, ayrıca binanın parsel ile olan mesafesi ve bina kesitleri ile bina yüksekliğini gösteren belgedir (Kontur gabari nedir, 2016).

Booth (1889) ise renkleri, bölgelere göre gelir düzeyleri için kullanarak yoksulluk haritaları oluşturmuştur.

20. ve 21. yüzyıllarda, veri görselleştirmede etkileşim, animasyon, hikâye anlatımı ve erişilebilirlik gibi yeni eğilimlerin ortaya çıktığı görülmüştür. Örneğin Herman Chernoff (1952), bilgisayarda oluşturulan yüz görüntülerinin yaş veya cinsiyet gibi parametrelerinde değişiklik yapmak suretiyle farklı yüz ifadelerinin elde edilebileceğini gösteren, Chernoff yüzleri tekniğini yaratmıştır (Chernoff, 1973, s.2; Raciborski, 2009, s.374). Richard A. Becker ve William S. Cleveland (1975), kullanıcıların herhangi bir temel kod veya yapıyı değiştirmeden ekrandaki parametreleri değiştirmesine olanak tanıyan etkileşimli görselleştirmeleri tanıtmıştır (Think Insights, 2017).

Teknolojik gelişmelere paralel olarak yöntem ve yazılımlarda ilerleme kaydedilmesiyle veri görselleştirme yeni bir boyut kazanmıştır. Verileri keşfetme ve anlamlandırma aracı olarak 1970'lerde John Tukey tarafından, “Exploratory Data Analysis-EDA” olarak bilinen “keşifsel veri analizi” nin tanıtılmasıyla veri görselleştirmenin gücü daha da pekişmiştir (Few ve Edge, 2007). Keşifsel veri analizi, verileri araştırmak ve analiz etmek için güçlü bir görsel yaklaşımdır (Tukey, 1977; H. M. Chen, 2017). Örneğin; Jason Dykes (2007), kullanıcıların etkileşimli haritalar kullanarak coğrafi özellikler arasındaki mekânsal ilişkileri keşfetmelerine olanak tanıyan bir kartografik veri görselleştiricisi yaratmıştır.

Büyük veri dünyasında, büyük miktarda veriyi analiz etmek ve veriye dayalı kararlar vermek için veri görselleştirme yöntem ve yazılımları artık vazgeçilmez hale gelmiştir. Veri görselleştirme yöntemleri, görsel temsilleri oluşturmak bakımından önem kazanmaktadır. Veri görselleştirme türleri ise, bu temsillerin çizelge, grafik, harita vb. biçimlerde sunumunu ifade etmektedir. Veri görselleştirme, başta iş ve bilim dünyası olmak üzere birçok alanda sıklıkla uygulanır olmuştur.

Veri görselleştirme için Tableau, Google Charts, PowerBI gibi hazır yazılımlar kullanılabileceği gibi, D3.js, Python veya R gibi programlama dilleri kapsamındaki açık kaynak kod kütüphaneleri de kullanılabilmektedir.

Veri analizi sürecinde uygun görselleştirme türlerini kullanmak; hem veri noktaları arasındaki önemli ilişkileri veya örüntüleri anlamayı ve sunmayı, hem de yeni fikirler için öngörüler elde etmeyi sağlar. Bunun için en uygun görselleştirme türünü seçmek önemlidir. Günümüz bilgisayar teknolojisi ve veri görselleştirme programları büyük oranda veri görselleştirme türlerine olanak vermektedir. Etiket bulutu (word cloud), çubuk grafik (bar

chart), çizgi grafik (line chart), pasta grafik (pie chart), dağılım grafiği (scatter plot), histogram, radar grafik (radar chart), haritalar, piktogramlar, ısı haritaları ve ağaç haritası (treemap) veri görselleştirme türlerine örnek olarak gösterilebilir.

Uygun görselleştirme seçimi, veri türüne ve iletilmek istenen mesaja bağlı olarak değişkenlik gösterebilir. Örneğin; sürekliliği olan veriler için çizgi grafik, histogram, dağılım grafiği; kategorik veriler için çubuk grafik, pasta grafik; hiyerarşik veriler için ağaç haritası ve güneş ışığı grafiği gibi türler kullanılabilir. Ayrıca jeo-uzaysal (geospatial) veriler için; koroplet haritalar (choropleth maps) ve kabarcık haritalar (bubble haritalar), çok boyutlu veriler için ise ısı haritası (heatmap) ve paralel koordinatlar (parallel coordinates) görselleştirme türleri kullanılabilir.

Görselleştirme seçimi, yararlanıcı grubun belirli özelliklerine ve ihtiyaçlarına bağlı olarak da değişiklik gösterebilir. Örneğin yararlanıcı grup özet veriler tercih ediyorsa çubuk grafikler veya pasta grafikler gibi daha basit görselleştirmeler sunmak uygundur. Öte yandan hedef kitlenin ayrıntılı içgörülere ihtiyaç duyan veri analistleri veya uzmanlardan oluşması durumunda ısı haritaları veya dağılım grafikleri gibi daha karmaşık görselleştirmeler uygun olacaktır. Genel olarak veri görselleştirmeleri; karar vericiler için içgörü elde etmek, bulguların ve önerilerin paydaşlara daha net bir şekilde iletilmesini kolaylaştırarak etkili bir iletişim kurmak ve eylemlerin bilinçli bir şekilde yönlendirilmesi için güçlü araçlar olarak hizmet eder.

Veri görselleştirmede, veriler arasındaki ilişkilerin doğru bir şekilde temsil edilmesinin ve bilgilerin etkili bir şekilde iletilmesinin önemi göz ardı edilemez. Ancak görselleştirmelerde kullanılan verilerin tamlığı ve aykırı değerlerin belirlenmesi, mükerrer kayıtların ayıklanması, eksik verilere değer atama vb. veri temizleme işlemlerinin yapılmış olmasından emin olmak da aynı derecede önemlidir. Bunun nedeni, eksik veya ön işleme sürecinden geçmemiş verilerin yanıltıcı görselleştirmelere yol açabilmesi ve bunun da hatalı yorumlara ve sonuçlara yol açabilmesidir. Bu nedenle veri görselleştirmenin başarısı yalnızca verileri erişilebilir ve anlaşılır bir şekilde sunma becerisine değil, aynı zamanda kullanılan verilerin kalitesine de bağlıdır.

Çalışmanın konusu bağlamında veri görselleştirme, bir üniversitenin araştırma eğilimlerinin saptanmasında ve araştırma politikalarının oluşturulmasında kurum yöneticilerinin karar süreçlerine katkı sağlayabilir. Örneğin; akademik yayınlarının üst veri alanlarına ait veri setlerine ulaşmak veya bunları farklı sistemlerden toplayarak veri setleri oluşturmak hem kurum yöneticilerine hem de araştırmacılara birçok fayda sağlayabilir.

Bilimsel yayınlar ve araştırmacılar üzerine yapılan atıf çalışmaları ile etki değerlendirmelerinin yapılması sürecinde veri görselleştirme sıklıkla başvurulmuş bir yöntemdir. Bu yöntemin kurumlar tarafından yayınların üst verilerinin analizine yönelik de kullanılması, karar destek mekanizmaları için girdi sağlayabilir. Bunun yanı sıra araştırmacılar için de veri kümelerindeki özelliklerin, örüntülerin ve ilişkilerin keşfedilmesine ve anlaşılmasına, böylece yeni hipotezlerin ve içgörülerin üretilmesine yardımcı olabilir.

Bu çalışmada TR Dizin’de indekslenen Ankara Üniversitesi adresli, beş yıllık (2017-2021)³ döneme ait akademik yayınların üst verileri üzerine bir görselleştirme süreci uygulanmıştır. Böylelikle üst verilerin görselleştirme aracılığı ile üniversitenin araştırma eğilimleri hakkında bilgi verebileceği, karar vericiler açısından stratejik bir unsur olarak değerlendirilebileceği ortaya konmaya çalışılmıştır. Bunun yanı sıra veri görselleştirmenin başarısı için sürecin nasıl gerçekleştirileceği, süreçte ne gibi sorunlarla karşılaşılacağı, bu sorunlara ne gibi çözümler üretilebileceğine dair bir yöntem sunmak ve uygulama ortaya konmaya çalışılmıştır. Bu doğrultuda çalışmada araştırmacı tarafından geliştirilen uygulama, araştırmanın alana özgün katkısı niteliğindedir.

TR Dizin, ülkemizdeki akademik yayınların bir arada sunulduğu ulusal akademik bir veri tabanı olması nedeniyle bu çalışma için uygun bir veri kaynağıdır. Bu çalışma kapsamında, kurumsal bilgiler ve elde edilen bulgular doğrultusunda, kurumsal hizmetler ve karar destek süreçlerinde kullanılmak üzere bir veri görselleştirme uygulaması tasarlanmış, uygulamanın tüm adımları detaylı bir şekilde sunulmuştur. Oluşturulan veri görselleştirme uygulaması, özelde Ankara Üniversitesi ulusal yayınlarının analizinde ve değerlendirilmesinde kullanılabilirken genelde benzer amaçlar için yeni veri setlerinin görselleştirilmesinde de kullanılabilir nitelikte örnek bir uygulamadır. Bu çalışmada ortaya konulan sorular doğrultusunda gerçekleştirilen veri görselleştirme uygulaması ile açığa çıkarılan bilgilerin; kurumun stratejik planı, performans değerlendirmeleri ve politika belgelerinin oluşturulması sürecinde katkı sunacağı ve karar verme aşamalarını destekleyeceği düşünülmektedir. TR Dizin’de ulusal yayınların yer alması dolayısıyla diğer üniversitelerin de akademik yayınlarını analiz etme ve değerlendirme gereksinimlerini karşıladıkları bir veri tabanı olduğu düşünüldüğünde çalışmada sunulan uygulama, Türkiye’deki diğer üniversiteler için de bir model olma özelliği taşımaktadır.

³ Hazırlanan veri seti 2022 yılında oluşturulduğu için 2021 yılına kadar olan makaleler kapsama alınmıştır.

Bir üniversitenin akademik yayınlarının analiz edilmesi, o üniversitenin bilimsel yayınları temelinde araştırma eğilimlerinin saptanması ve araştırma stratejilerinin değerlendirilmesinin önemli bir bileşenidir. Bilimsel yayınların üst verilerinin görselleştirme teknikleri aracılığı ile analiz edilip sunulması yoluyla kurumlar; araştırma çıktılarının etkileri hakkında ve güçlü oldukları araştırma alanları ve/veya gelişme eğilimi gösteren alanlar hakkında değerli bilgiler edinebilirler. Bu süreç yalnızca akran kurumlarla kıyaslama yapılmasına yardımcı olmakla kalmaz, aynı zamanda kurum içinde stratejik karar alma ve kaynak tahsisi konularında da önemli bilgiler sağlayabilir.

Yapılan literatür incelemesinde, kütüphanecilerin veri görselleştirme sürecinde önemli bir rol alabilecekleri konusunun sıkça vurgulandığı tespit edilmiştir. Bilginin organizasyonu, bilgi erişim ve bilgi yönetimi konularındaki uzmanlıklarıyla kütüphaneciler, akademik yayınlarla ilgili üst verilerin oluşturulması ve düzenlenmesi ile üst verilerin görselleştirilmesi süreçlerini gerçekleştirecek donanımına sahiptirler.

Veri görselleştirme araçlarının kullanımı konusundaki yetkinlik; ham verileri eyleme geçirilebilir içgörülere dönüştürmeyi ve yayın verileri içindeki eğilimleri ve korelasyonları tespit edebilmeyi sağlar. Bunun için veri görselleştirme yöntem ve yazılımlarına ilişkin eğitim ve yönlendirme konusu önemli hale gelmektedir. Kütüphane ve yönetim iş birliği ile birlikte bu uygulamaların benimsenmesi ve gerekli planlamanın yapılması kurumlara; araştırma çalışmalarında stratejik karar alma ve inovasyon için veriye dayalı içgörülerden faydalanma olanağı sağlayacaktır.

1.2 Araştırmanın Amacı

Bu çalışmanın temel amacı; veri görselleştirmeye ilişkin kavramsal çerçeve sunmak, veri görselleştirme araçlarını tanımlamak; bir veri görselleştirme sürecinin nasıl uygulanacağını göstermek ve karşılaşılabilecek sorunları ortaya koyarak çözüm önerilerinde bulunmaktır. Bu doğrultuda tezin bir diğer amacı TR Dizin’de yer alan Ankara Üniversitesi adresli, beş yıllık (2017-2021) döneme ait akademik yayınların üst verileri üzerinden elde edilen veri setlerine görselleştirme teknikleri uygulayarak araştırma eğilimlerinin keşfedilmesini sağlamak; görselleştirmenin sunduğu bilginin karar vericiler için stratejik önemine dikkat çekmek ve politika geliştirmede nasıl kullanılacağı konusunda önerilerde bulunmaktır.

Bu bağlamda; çalışmanın amaçları şöyle sıralanabilir:

- Veri görselleştirmeye ilişkin bir kavramsal çerçeve oluşturmak,
- Veri görselleştirme yöntem ve araçlarını belirlemek,
- Veri görselleştirme yöntem ve araçlarının üst veri alanları üzerine kullanılabilirliğini ortaya koymak ve veri görselleştirme sürecinde karşılaşılabilecek sorunları saptayarak çözüm önerilerinde bulunmak,
- Ankara Üniversitesi adresli bilimsel yayınların üst verilerine dayanarak bir veri görselleştirme uygulaması geliştirmek,
- Veri görselleştirme yoluyla karar süreçlerine girdi oluşturabilecek ne gibi bilgiler edinilebileceğini göstermek,
- Görselleştirme araçlarının etkin kullanımı konusunda farkındalık yaratmak, personel ve kullanıcı eğitimi açısından önerilerde bulunmak.

Ankara Üniversitesi'nin "2024-2028 Stratejik Planı"nda; kurumsal yapının güçlendirilmesi amacıyla veri yönetimine, araştırma faaliyetlerinin izlenmesine ve iyileştirilmesine yönelik hedeflere yer verilmiştir. Bu bağlamda veri görselleştirmenin değerlendirme yapma, içgörü edinme, yeni stratejiler belirleme gibi hususlarda karar vericilere önemli katkılar sunacağı açıktır. Bu çalışmada araştırmacının geliştirdiği ve Ankara Üniversitesi özelinde yayınların üst verilerinin kullanıldığı görselleştirme uygulamasının, araştırmacıların araştırma eğilimlerinin saptanmasında, üniversitenin araştırma stratejisinin belirlenmesinde kullanılabileceğini ortaya koymak amaçlanmıştır; çalışmanın bu doğrultuda kurumsal olarak üniversitelere ve TR Dizin bünyesinde ulusal değerlendirmelere örnek teşkil edeceği ve katkı sağlayacağı düşünülmektedir.

1.3 Araştırmanın Problemi, Soruları, Alt Soruları ve Hipotezi

Veri görselleştirme; eğilimlerin ne yönde olduğu, nerelerde boşluklar bulunduğu ve nerelerde öbeklenme/kümelenme olduğunun daha net bir şekilde algılanmasını sağlayarak büyük miktarda verinin hızlı bir şekilde anlaşılmasına olanak tanır (Serman ve Borda, 2017, s.2). Etkileşimli veri görselleştirme ile dijital ortamdaki bilimsel çalışmalar hakkında daha fazla bilginin keşfedilmesi mümkündür.

Araştırmanın veri kaynağı olan TR Dizin, ulusal bilimsel dergileri kapsayan ve denetimli olarak ULAKBİM çatısı altında yürütülen bir veri tabanıdır. TR Dizin veri tabanı,

ulusal bilimsel yayınlara erişim sağlamanın ötesinde zengin bir veri havuzu olarak da değerlendirilebilir.

Veri görselleştirme tekniklerinden yararlanılarak veriler; karar vericiler, araştırmacılar veya diğer kullanıcılara etkili bir şekilde iletilebilir. Veri görselleştirme yoluyla, veri tabanına gömülü karmaşık eğilimler ve ilişkiler açığa çıkarılabilir, bu da kullanıcıların altta yatan bilgileri daha sezgisel bir şekilde keşfetmesine ve anlamasına olanak tanıyabilir. Ayrıca veri tabanındaki yayınlar üzerinde veri görselleştirmeler yapılarak yeni araştırma eğilimlerinin ortaya çıkartılması, politikaların etkisinin değerlendirilmesi veya yaygınlaştırılması sağlanabilir. Bu sayede veri tabanının potansiyelinden daha etkili bir şekilde fayda sağlanmış olacaktır. Bu kapsamda araştırmanın problemi: “Bilimsel araştırmaların çıktıları olan yayınların üst verileri üzerine uygulanacak görselleştirme teknikleri, atıfa ve yayın sayısına dayalı değerlendirmelerin yanı sıra araştırmacılara/yetkililere; araştırma eğilimlerinin saptanması –disiplinlerarasılık, bilimsel iş birliği, araştırma konuları- ve araştırma stratejilerinin belirlenmesi konularında bilgi sağlayabilir mi?” dir.

Bu doğrultuda çalışmada TR Dizin kapsamında yer alan Ankara Üniversitesi adresli bilimsel yayınların üst verileri ele alınmış ve bir görselleştirme uygulaması geliştirilerek üst verilerin araştırma eğilimlerinin saptanması ve araştırma stratejilerinin belirlenmesi konularında keşifsel veri analizi için uygun olup olmadığı ve araştırmacılara ne gibi bilgiler sağlayacağı araştırılmıştır.

Araştırma kapsamında aşağıdaki araştırma sorularına yanıt aranmıştır:

- TR Dizin, akademik yayınlara ait olan ve keşifsel veri analizi yapmaya uygun, hangi üst veri alanlarını barındırmaktadır? Analiz ve görselleştirmeler için var olan üst veri alanları yeterli midir?
- Üst veri alanlarındaki veriler, veri görselleştirme için uygun mudur?
- Hangi tür veriler için hangi görselleştirmeler uygundur?
- Veri görselleştirme uygulaması karar süreçlerinde kullanılabilecek yeterli ve doğru bilgi sağlayabilmekte midir?

Belirlenen bu araştırma problemleri doğrultusunda araştırmanın hipotezi;

“Bilimsel akademik yayınlara ait üst verilerden oluşan veri setlerine, veri görselleştirme destekli keşifsel veri analizi teknikleri uygulayarak veri alanlarının görselleştirmeye elverişli olup olmadığı ve içgörü sağlayıp sağlayamayacağı tespit edilebilir.” şeklindedir.

1.4 Araştırmanın Sınırlılıkları

- Araştırmada TR Dizin’de yer alan Ankara Üniversitesi adresli makalelerin üst verilerinden oluşan veri setleri kullanılmıştır.
- Yıl aralığı olarak Ankara Üniversitesi adresli 2017-2021 yılları arasındaki makaleler çalışma kapsamına alınmıştır.
- Disiplinlerarası iş birliğinin belirlenmesi ve anahtar kelime, makale konusu analizleri için TR Dizin üst veri alanlarında verilerin eksik olması nedeniyle ilgili üst veri alanları DergiPark’tan derlenmek zorunda kalınmıştır.
- Anahtar sözcük veya konu analizlerinde bir otorite dizin kullanılmaması analizleri güçleştirmiştir.

1.5 Araştırmanın Kapsamı, Evreni ve Örneklemi

Bu araştırmanın evrenini 2017-2021 yılları arasında TR Dizin’de yer alan Ankara Üniversitesi adresli bilimsel yayınların tamamı oluşturmaktadır. Araştırma kapsamında uygulama için TR Dizin bir veri kaynağı olarak alınmıştır.

1.6 Araştırmanın Yöntemi ve Veri Toplama Teknikleri

Bu çalışma kapsamında öncelikle alanyazın taraması yapılmıştır. Alanyazın taraması kavramsal çerçeveyi oluşturmak, sorunu sınırlamak ve yeni yaklaşımlar bulabilmek için kullanılmaktadır (Doğanay, Ataizi, Şimşek, Balaban Salı ve Akbulut, 2018, ss.56-57). Alanyazın taraması; yazarların sorunu tanımlaması ve netleştirmesi, okuyucuyu araştırmanın durumu hakkında bilgilendirmek için önceki araştırmaları özetlemesi, literatürdeki ilişkileri, çelişkileri, boşlukları ve tutarsızlıkları belirlemesi ve sorunun çözümünde bir sonraki adım veya adımları önermesi açısından öğretici bilgiler sunmaktadır (American Psychological Association, 2013, s.10). Kavramsal çerçeve ile “araştırılmak üzere seçilen sorunun kuramsal temelleri, disiplinlerarası boyutları, alanda geçerli olan tanımlamalar/sınıflamalar, incelenen

konunun tarihsel gelişimi ve konunun genel görünümü oluşturulmaya çalışılır” (Doğanay ve diğerleri 2018, s.56). Sorunu sınırlamak ile anlatılmak istenen; araştırılan konunun ya da sorunun tanımlamasını ve işedönük biçimde sınırlandırılmasını yapabilmek için o güne kadar alanda yapılmış olan çalışmaları incelemeye gereksinim olduğudur (Doğanay ve diğerleri 2018, s.57). “Alanyazın taraması sayesinde çalışılan alan ya da araştırılan sorunla ilgili olarak başka çalışmalarda kullanılan yaklaşımlar incelenerek, yapılacak yeni araştırmalar için alternatif yaklaşımlar geliştirilebilir” (Doğanay ve diğerleri 2018, s.57).

Alanyazın taramasında veri görselleştirme kavramı başta olmak üzere veri görselleştirme ile ilgili kavramlar, veri görselleştirme yöntemleri, türleri ve araçları hem uluslararası hem de ulusal kaynaklardan detaylı olarak incelenmiştir.

Bilimsel yöntem, nicel ve nitel paradigmaların kullanılması ile ayrılmaktadır (Doğanay ve diğerleri 2018, s.84). Bu çalışma, nicel araştırma türüne bir örnektir. “Nicel (kantitatif) araştırmalar, bir örneklem üzerinde yapılan ölçümleri ana kitleye genellemek amacıyla yapılan; ne kadar, ne miktarda gibi soruları cevaplamaya yönelik araştırmalardır.” (İslamoğlu ve Alınışık, 2016, s.47). Yöntemsel açıdan nicel araştırma tümdengelimci bir yaklaşımı benimser ve bu nedenle kuramsal temelin kurulması ya da alanyazın taraması yapılması önemlidir (Doğanay ve diğerleri 2018, s.85). Bu çalışmada nicel araştırma modellerinden genel tarama modeli tercih edilmiştir. “Genel tarama modelleri, örnekleme yoluyla evren hakkında kestirimlerde bulunma ve genellemeler yapma amacını gütmektedir. Bu modeller özellikle evrene ilişkin eğilimlerin belirlenmesinde yararlı olduğu için olabildiğince geniş bir örneklemden veri toplanır.” (Doğanay ve diğerleri 2018, s.92).

Araştırmaya dair veriler TR Dizin’de yer alan 2017-2021 yılları arasındaki Ankara Üniversitesi adresli akademik yayınların üst verilerinden oluşmaktadır. Akademik yayınlara ilişkin üst veriler çekilerek çeşitli veri setleri oluşturulmuş ve veri görselleştirme uygulamasında bu veri setlerinden yararlanılmıştır.

Görselleştirme uygulamasının geliştirilmesinde Phyton programlama dilinin pandas, numpy, beautifulsoup, re, request, matplotlib, seaborn ve plotly kod kütüphaneleri kullanılmıştır. Veri seti depolama için Google Drive ortamı, kod yazmak için ise Google Colab platformu üzerinden Jupyter Notebook aracı kullanılmıştır. Görsellerin elde edilmesinde veri setine sorular sorulmuştur (bkz. Tablo 16). Her bir soru için uygun görselleştirme aracı

belirlenmiş ve keşifsel veri analizi aracılığı ile verilerin doğru bilgi veren görseller üretip üretmediği test edilmiştir.

1.6.1 Araştırmada Kullanılan Veri Setleri ile Veri Toplama, Analiz ve Değerlendirme Teknikleri

Araştırma kapsamında kullanılan veri setleri hakkında ayrıntılı bilgiler çalışmanın üçüncü bölümünde yer almaktadır. Bu bölümün iki alt başlığında veri türlerinin belirlenmesi ve veri toplamada kullanılan yöntem ve araçlara değinilmiştir. Veri setlerinde analiz ve değerlendirme yapabilmek için veri setine aşağıdaki sorular sorulmuştur. Böylece her bir soru yanıtlanırken çeşitli görselleştirmeler kullanılmıştır.

Makale Demografik Bilgileri Bağlamındaki Sorular	
Soru 1	Makalelerin temel konu kategorilerine ⁴ göre dağılımı nedir?
Soru 2	Makalelerin dillere göre dağılımı nedir? 2.1. Türkçe dışındaki dillerde yayınlanmış makaleler hangi konu kategorisinde yer almaktadır?
Soru 3	Makalelerin türlere göre atıf alma durumları nedir? 3.1 Türlerin konu ve yıla göre dağılımı nedir? 3.2 En çok atıf alan makale türleri hangileridir?
Soru 4	Makale sayıları ile atıf sayılarının yıllara göre dağılımı nedir? 4.1 Makalelerin yıllara göre görüntülenme ve indirilme sayıları nedir?
Soru 5	Ulaşılabilirlik açısından makalelerin erişim türüne göre dağılımı nedir? 5.1 Hangi temel konulu makaleler erişime kapalıdır; bunlar yıllar içinde artış gösterme eğiliminde midir?
Soru 6	Yazar iş birliği (tek yazarlı/ortak yazarlı) eğilimi ne düzeydedir? 6.1 Yazar sayısına göre yayın sayısı dağılımı nedir? 6.2 Makale başına düşen yazar sayısına göre atıf alma durumu nedir?

⁴ TR Dizin’de makalelere ait bibliyografik kimliklerin gruplandırıldıkları temel kategoriler.

	6.3 Tek yazarlı makaleler ile çok yazarlı makalelerin atıf oranları ne düzeydedir? 6.4 Kurumsal iş birliği ne düzeydedir? 6.5 Kurumsal iş birliğinin yıla göre dağılımı nedir? 6.6 Kamu iş birliği olan makalelerde ilk yazara göre yazar kurum dağılımı nedir?
Soru 7	Makalelerin konusal dağılımı nedir? 7.1 Türkçe makalelerin konusal olarak yıl ve genel konu kategorisi bazındaki dağılımı nedir? 7.2 İngilizce makalelerin konusal olarak yıl ve genel konu kategorisi bazındaki dağılımı nedir? 7.3 Türkçe anahtar kelimelerin yıllara göre dağılımı nedir? 7.4 İngilizce anahtar kelimelerin yıllara göre dağılımı nedir?
Soru 8	Makalelerin yayınlandığı dergi sayılarının yıllara göre dağılımı nedir? 8.1 Dergi sayısı artış oranı ile yayın sayısı artış oranı arasında bir korelasyon var mıdır? 8.2 Yıllara göre dergi başına düşen makale sayıları nedir? 8.3 AÜ yazarları hangi dergileri tercih etmektedir? 8.4 AÜ makalelerinin yayınlandığı dergilerin yayın hayatına başlama tarihine göre dağılımı nedir? 8.5 AÜ makalelerinin yayınlandığı dergilerin yıllık yayın periyodları, makale sayıları, aldığı atıf sayıları ve kendine atıf sayıları nedir?

1.6.2 Kullanılan araç ve yazılımlar

Veri toplama sürecinde kullanılan araçlar ve yazılımlar; Python Programlama Dilinde bulunan pandas, numpy, beautifulsoup, re, request kod kütüphaneleridir. Bu kod kütüphaneleri aracılığıyla toplanan veriler CSV formatında Google Drive'a kaydedilmiştir. Keşifsel veri analizi ve görselleştirmeler için ise Python Programlama Dilinde bulunan Pandas ve numpy veri analiz kod kütüphaneleri ile matplotlib, seaborn, plotly veri görselleştirme kod

kütüphaneleri kullanılmıştır. Kodlar Google Colab platformunda bulunan Jupyter Notebook'ta oluşturulmuş ve aynı platformda uygulanmıştır.

1.7 Araştırmanın Özgün Değeri

Son yıllarda dünyada büyük veri ve veri yönetimi önem kazanmakta, veri bilimi ve yapay zekâ araştırmaları hızla artmaktadır. Veri görselleştirme konusu da gündemde olan konuların başında gelmektedir. Bu çalışmanın kütüphanecilik ve bilgi bilimi alanında özgün katkısı; keşifsel veri analizi ve görselleştirmenin bilginin sunumunda/iletilmesinde güçlü bir yöntem ve araç olduğunu göstermesi, bu anlamda üst verilerin potansiyel değerini ortaya koyması ve örnek bir görselleştirme uygulaması geliştirmesidir. Çalışmanın hem kütüphanecilik ve bilgi bilim alanına hem de diğer alanlarda veri analizi ve veri görselleştirme üzerine çalışma yapan araştırmacılara yarar sağlayacağı öngörülmektedir. Ayrıca bu çalışma TR Dizin'de indekslenen Ankara Üniversitesi yayınlarının üst verilerine dayalı veri setleri üzerinden yapıldığından ulusal nitelikli bu veri tabanının sunduğu olanakları da ortaya koymaktadır.

1.8 Araştırmanın Düzeni

Araştırma 5 bölümden oluşmaktadır:

1. Bölüm: Bu bölümde öncelikle araştırmanın konusu ve önemine değinilmiştir. Bunu izleyen alt bölümde araştırmanın amaçları ele alınmıştır. Araştırmanın problemi, soruları ve hipoteze değinildikten sonra araştırmanın kapsamı, evreni ve örnekleme hakkında bilgilere yer verilmiştir. Daha sonra araştırmanın yöntemi ile analiz ve değerlendirme teknikleri alt başlığı yer almaktadır. Son olarak bu bölümde araştırmanın özgün değeri tartışılmış, çalışmanın düzenine ve çalışmada kullanılan kaynaklara ilişkin bilgilere yer verilmiştir.

2. Bölüm: Bu bölüm çalışmanın kavramsal çerçevesine ilişkindir. “Veri Görselleştirme Yöntem ve Araçları” başlıklı bu bölümde öncelikle veri görselleştirmenin tanımı ve tarihçesi ile veri görselleştirmenin amaç ve işlevlerine değinilmiştir. Daha sonra veri görselleştirmede kullanılan yöntemler ve araçlar ile veri görselleştirme kodları ve veri görselleştirme türleri detaylı bir şekilde ele alınmıştır.

3. Bölüm: Çalışmanın bu bölümünde uygulanan yöntem ele alınmıştır. Bu bölümde veri görselleştirme süreci hakkında bilgilere yer verilmektedir. Veri toplama aşaması on adım olarak düzenlenmiştir. Tüm adımlar ayrıntılı bir biçimde ve bir modele (Andrews, Segonds-Pichon, Virk ve Montgomery, 2023) dayalı olarak ele alınmaktadır. Veri ön işleme konusu ele alındıktan sonra keşifsel veri analizi işlemi ayrıntılı şekilde sunulmuştur.

4. Bölüm: Bulguların ele alındığı çalışmanın bu bölümünde veri setine sorular sorularak yapılan analizlere yer verilmiş ve görselleştirmeler ile bu sorulara yanıtlar aranmaya çalışılmıştır.

5. Bölüm: Sonuç ve öneriler ile veri görselleştirme modelinin yer aldığı bölümdür.

1.9 Kaynaklar

Araştırma kapsamında yapılan alanyazın taramasında kullanılan kaynaklar aşağıdaki gibi olup, taramalar 2020 yılından bu yana belli aralıklarla güncellenmiştir:

- DergiPark
- Emerald
- Google Books (<http://books.google.com>)
- Google Scholar (<http://scholar.google.com>)
- Proquest (1986-)
- ScienceDirect (1980-)
- Scopus
- TR Dizin
- UMI ProQuest Digital Dissertations
- Web of Science (WoS)
- YÖK Ulusal Tez Merkezi

Alanyazın taraması yapılırken kullanılan Türkçe ve İngilizce anahtar terimler aşağıdaki gibidir;

- Veri görselleştirme (data visualization),
- Veri görselleştirmenin tarihi (history of data visualization),

- Veri görselleştirme yöntemleri (data visualization methods),
- Veri görselleştirme araçları (data visualization tools),
- Veri seti (data-set)
- Veri seti oluşturma (creating data-set),
- Web kazıma (web scraping),
- Python-Pandas,
- BeautifulSoup,
- matplotlib,
- seaborn.

Veri görselleştirme araçlarına yönelik olarak kullanılan kaynaklar için ise aşağıdaki çevrimiçi sayfalar referans alınmıştır:

- Python: <https://www.python.org>
- BeautifulSoup4 - Web Scraping with Python: <https://pypi.org/project/beautifulsoup4/>
- NumPy: <https://numpy.org>
- Pandas - Python Data Analysis Library: <https://pandas.pydata.org>
- Matplotlib - Visualization with Python: <https://matplotlib.org>
- Seaborn - Statistical Data Visualization: <https://seaborn.pydata.org>
- Plotly - Python Graphing Library: <https://plotly.com/python/>

Araştırma raporunun yazımında ve tez düzeninde Ankara Üniversitesi Sosyal Bilimler Enstitüsü (2020) tarafından oluşturulan “Tez Yazım Yönergesi”nde yer alan talimatlar uygulanmıştır. Metin içi gönderme ve kaynakçanın oluşturulmasında ise APA (American Psychological Association) atıf stilinin 7. edisyonu referans alınmıştır. Ayrıca referans yönetim aracı olarak Mendeley’in çevrimiçi uygulaması kullanılmıştır.

Kaynaklardan elde edilen bilgilerin ve kişisel notların depolanmasında; not alma, notları düzenleme, arşivleme özellikleri olan “Notion.so” adlı çevrimiçi uygulama kullanılmıştır. Veri setinin depolanması için bir bulut uygulaması olan Google Drive ve uygulama kodlarının yazılması için ise yine bulut ortamında bulunan Google Colab platformunda yer alan Jupiter Notebook IDE kullanılmıştır.

2. BÖLÜM: KAVRAMSAL ÇERÇEVE

"Bir resim bin kelimeye bedeldir" (Barnard, 1921).

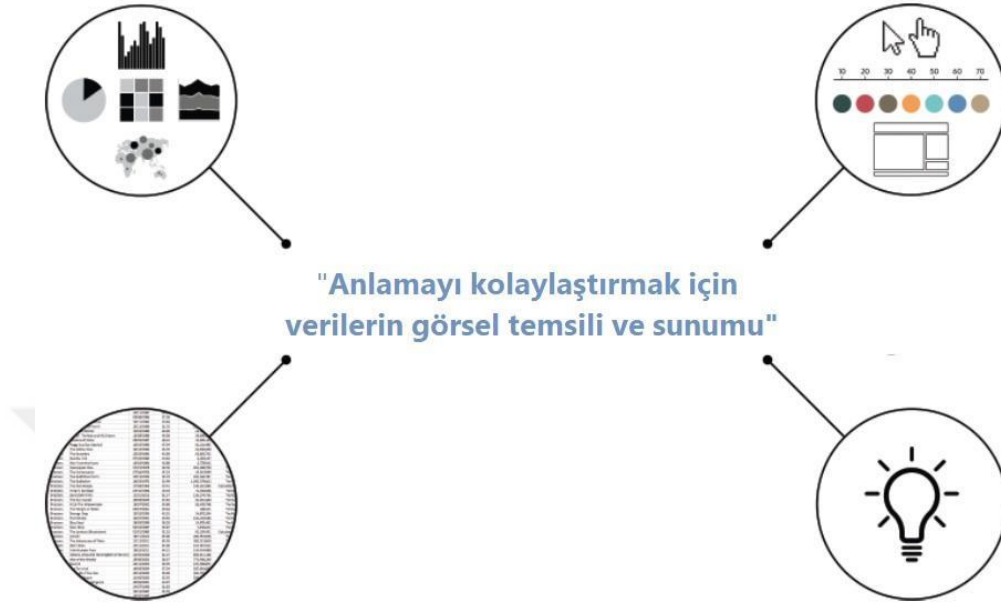
Görselleştirme (visualization), resimler, diyagramlar, çizimler veya animasyonlar gibi görsel formları kullanarak bir bilgiyi temsil etmek ve iletmek çabasıdır. Merriam-Webster (t.y) çevrimiçi sözlüğe göre görselleştirme; görsel olarak yorumlama veya gözle görünür bir forma sokma eylemi veya süreci veya zihinsel görsel imgelerin oluşumu olarak tanımlanmıştır. İmgeleme⁵ yoluyla görselleştirme, bilinen en eski görseller olan mağara resimlerinden bu yana hem soyut fikirleri hem de somut olayları iletmeyen etkili bir yolu olmuştur.

Görselleştirme, insanoğlunun tarih boyunca iletişim kurmak, olayları betimlemek, konum belirlemek vb. amaçlar için sıkça başvurduğu bir yöntem olmuştur. İnsanlar; yeryüzündeki toplumsal olayların ve fenomenlerin (doğa olayları) anlaşılması, duygu ve düşüncelerin aktarılması ve iletişim amacıyla (mağara resimleri, semboller, hiyeroglifler vb yollarla) bilginin temsili (information representation) kolaylaştıran görseller kullanmışlardır. Yıldızların ve diğer gök cisimlerinin konumlarının belirlenmesi, haritacılıkta konum tespiti (navigasyon) ve keşiflere yardımcı olacak haritaların yapımı gibi girişimler için toplanan veriler ve bu verilerdeki hikâyenin sunumu ve anlaşılmasının sağlanması amacı ise, görselleştirmenin daha yoğun bir şekilde kullanılmasının önünü açmıştır.

Veri görselleştirme; günümüzde çeşitli amaçlar doğrultusunda derlenen verinin, açık ve etkili bir şekilde grafiksel olarak temsil edilmesidir. Reitz'e (2014b) göre veri görselleştirme verileri; grafikler, haritalar, etiket bulutları, simülasyonlar veya içeriği daha kolay anlaşılır kılmak amacıyla grafik şeklinde temsil etmek için yazılım uygulamalarının kullanılmasıdır. Bilgisayar bilimciler veri görselleştirmeyi bir biliş güçlendirme aracı olarak görmekte ve "verilerin bilgisayar destekli, etkileşimli görsel temsillerinin kullanılması" (Card, Mackinlay ve Shneiderman, 1999; Sterman ve Borda, 2017) olarak açıklamaktadır. Börner, Chen ve Boyack (2003) veri görselleştirme kavramını veri nesnelerinin ve bunların ilişkilerinin görsel görünümünün tasarımı şeklinde tanımlamaktadır.

⁵ İmgeleme: Kelimeleri temsilen zihinde görsellerin yaratılmasıdır (Öznur ve Bayram, 2021, s.526).

Kirk (2019, s.15) ise Şekil 1’de görsel olarak yaptığı açıklamada veri görselleştirme adımlarını; veri toplama, uygun görselleştirme yöntemleri ile keşif ve analiz etme, uygun tasarım ile sunuş ve içgörü elde etme olarak sıralamaktadır.



Şekil 1: Veri görselleştirmenin görsel tanımı (Kirk, 2019, s.15)

Buradaki adımlar verilerin temsil biçimlerinin gelişimini yansıtmaktadır. Tarihsel süreçte yeni görsel temsil düzeyleri üretilmiştir. Örneğin; tablo, sayısal verilerin ilişkili temsili verirken bir üst düzeye taşınıp grafik temsil biçimi bulunmuş, daha sonra renklerin de kullanılmasıyla pasta, çubuk vb. grafik türleri oluşturulmuştur. Kullanılan teknikler verinin sunum biçimini de belirleyerek ilerlemektedir.

İnceleme ve öngörü elde etmek amacıyla verilerin görselleştirilmesi; veri görselleştirme (data visualization) veya bilgi görselleştirme (information visualization) gibi kavramlarla ifade edilmektedir. Veri görselleştirme ve bilgi görselleştirme birbiriyle ilişkili kavramlar olsa da literatürde bu kavramlar için farklı yaklaşımlar bulunmaktadır. Falkowitz’in (2019) veri görselleştirme ile bilgi görselleştirmeyi ele aldığı çalışmasında, her iki terim için, kullanım amaçları bağlamında ayrıştırıcı tanımlamalar yaptığı görülmektedir. Veri görselleştirmenin, bir veri seti ile bir analist veya görselleştirme tasarımcısı arasında aracılık eden bir iletişim formu olduğunu, tasarımcının veri setinin belirgin özelliklerini tanımlamaya ve göstermeye çalıştığını ifade etmektedir. Bilgi görselleştirmenin ise, bir tasarımcı ile karar vericilerden oluşan bir izleyici kitlesi arasındaki iletişim olarak düşünülebileceğini; tasarımcının görselleştirme yoluyla karar vericileri etkileme niyetinin olduğunu söylemektedir.

Amadebai (t.y.), veri görselleştirmenin, bir veri kümesinin dikkate değer önemli noktalarını tanımlamak için kullanıldığını, veri görselleştirmenin keşfetmeye, araştırmaya veya gerektiğinde veriyi değiştirmeye izin verdiğini, ayrıca veri görselleştirme işleminin bilgi görselleştirme işleminden önce geldiğini, bilgi görselleştirmenin veri görselleştirmenin bir sonucu olduğuna dikkat çekmektedir.

Veri görselleştirme ve bilgi görselleştirme için yapılan diğer çeşitli tanımlamalarda ise (Friendly, 2006; Kirk, 2019; Keim, 2002; Mazza, 2009; C. Chen, 2010; Brigham, 2016; Rezitz, 2020; Card, Mackinlay, and Shneiderman, 1999; Serman ve Borda, 2017; Börner ve diğerleri, 2003) genellikle benzer ifadelerin kullanıldığı, aynı amaçların belirtildiği, aynı yöntem ve araçların tavsiye edildiği, kesin bir ayırımın yapılmadığı tespit edilmiştir.

Kirk'in veri görselleştirme tanımı için oluşturduğu Şekil 1 ile Andrews ve diğerlerinin (2023) Şekil 35'te sundukları ve görselleştirme sürecini betimleyen model, Falkowitz'in (2019) aksine (veri ve bilgi safhaları olarak ayırdığı) görselleştirmeyi bir bütünlük içinde sunmaktadır. Bu çalışmada; Andrews ve diğerlerinin (2023) ortaya koyduğu bütüncül model benimsenmiştir; bu modelin, veri seti-tasarımcı-izleyici iletişimini yansıtmaya bakımından daha elverişli olduğu düşünülmektedir. Bu doğrultuda çalışmada "veri görselleştirme" teriminin kullanılması tercih edilmiştir. Çalışmanın bu bölümünde öncelikle veri görselleştirme kavramı açıklanmış, veri görselleştirmenin amaç ve işlevleri verilmiştir. Bundan sonraki alt bölümlerde veri görselleştirme tarihsel olarak ele alınacak ve literatür değerlendirilecektir. Son olarak veri görselleştirme yöntemleri ve yazılımları detaylı olarak incelenecektir.

2.1 Veri Görselleştirmenin Tarihsel Gelişimi

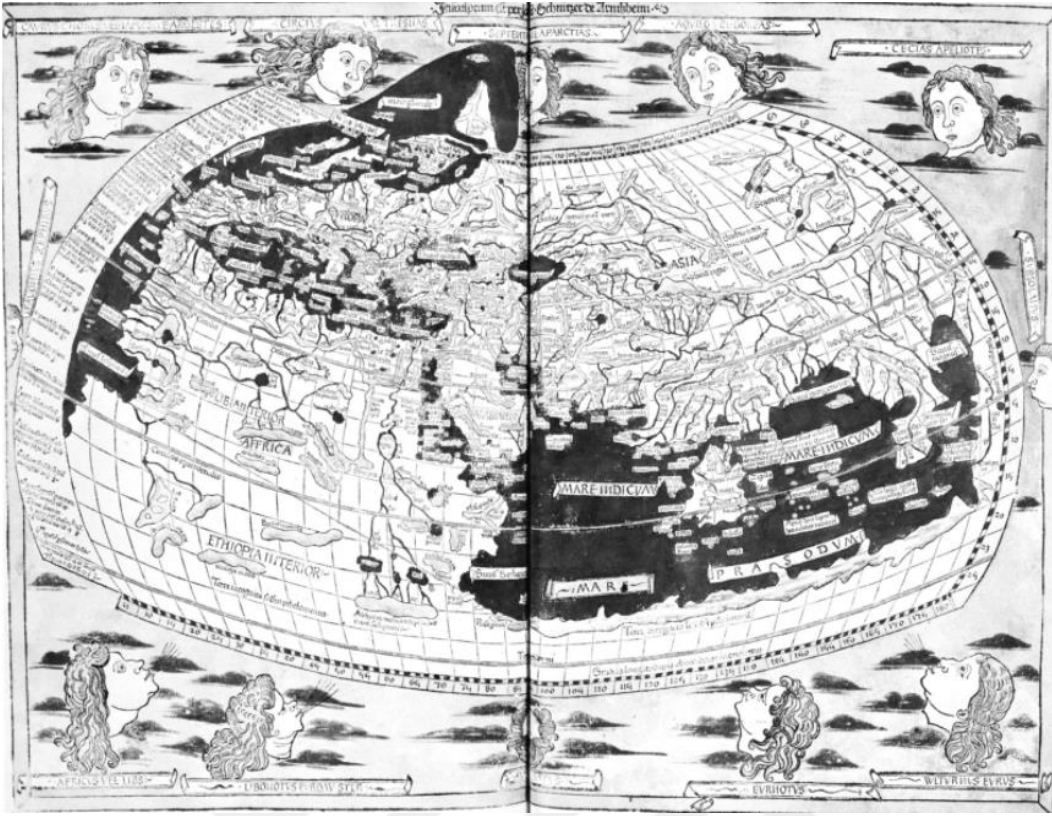
Veri görselleştirme yeni bir konu değildir. Tarih boyunca resimler aracılığıyla anlam aktarma, yıldızların ve diğer gök cisimlerinin konumlarının belirlenmesi ve uzamda konum tespiti gibi amaçlar doğrultusunda birçok görselleştirme günümüze kadar ulaşmıştır. İnsanların bu sabit/durağan görüntüler aracılığıyla veya resimleme yoluyla toplumsal olayları (avlanma, savaş, tören vb.) betimleyerek bilgiyi temsil etme ve iletme çabaları, yaklaşık otuz bin yıl öncesine kadar uzanmaktadır.

Veri görselleştirmenin ilk örnekleri ile daha sonraki görselleştirmeler arasındaki farklar, kullanılan tekniklerin giderek karmaşık hale gelmesinden kaynaklanmaktadır. Kayalar veya kil tabletler üzerine kaydedilmiş resim, çizim gibi görselleştirmelerin, (dönemin insanların yaşamları hakkında bize bilgi sağlayan) betimleme, kayıt tutma ve iletişim

amacıyla kullanıldığı ve görsel temsilin ilkel biçimleri olduğu söylenebilir (TRUECHART, 2021; Dragicevic ve Jansen, 2012). Buna karşılık zaman içinde görselleştirmeler, hesaplamalı veri içeren ve teknik açıdan daha gelişmiş, çekici ve bilgilendirici hale gelmiştir. Veri görselleştirmede bilgiyi temsil etme biçimi ve okuyanların bu bilgiyi anlamaya çalışması, iki taraflı bir çabayı gerektirmekte olup, bu durum veri görselleştirme okuryazarlığı ve farkındalık yaratma ihtiyacını da beraberinde getirmiştir.

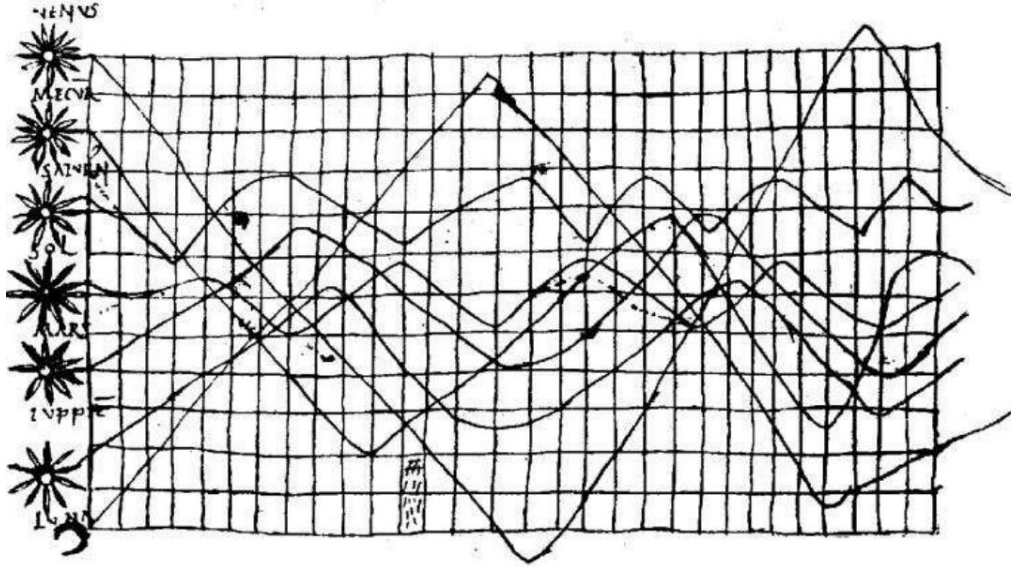
17. yüzyıla kadar görselleştirme, esas olarak arazileri işaretlemek ve şehirleri, yolları ve doğal kaynakları gösteren haritalar oluşturmak amacıyla başvurulan bir yöntem olmuştur. Haritacılık alanından örnek vermek gerekirse; M.S. 2. yüzyıla kadar Mısır'da şehirlerin dünyevi ve göksel (ilahi) konumlarının tespit edilmesi için enlem ve boylam belirlemeye benzer bir yaklaşım kullanılmıştır (Few ve Edge, 2007, s.2; Friendly, 2008, s.3). M.S. 100-170 yılları arasında yaşayan İskenderiyeli Yunan matematikçi, coğrafyacı ve astronom Klaudyos Batlamyus tarafından ortaya konan, dünya küresinin enlem ve boylamlara dayalı olarak haritalanması yöntemi, 14. yüzyıla kadar referans alınan standart bir yöntem olarak kullanılmıştır (Friendly, 2006, s.3). Batlamyus'un 'Totius Orbis Habitabilis Brevis Descriptio' (tüm yaşanabilir dünyanın kısa bir açıklaması) başlıklı dünya haritası, insan şeklinde tasvir edilen on iki rüzgârı ve zodyak⁶ kuşağı işaretlerini göstermektedir.

⁶ Zodyak: Astronomide zodyak, ekliptiğin her iki yanında 9° uzanan, Ay'ın ve ana gezegenlerin yörüngelerini kapsayan bir kuşağı ifade eder. Ekliptik üzerinde merkezlenen gökyüzü koordinat merkezinin bir özelliğidir, ekinoks noktasının doğusunda derece cinsinden ölçülen gök boylamının ölçülen değeridir (Zodyak, 2024).



Şekil 2: 1482'de basıldığı şekliyle Batlamyus'un dünya haritası (Kaynak: <https://cdn.britannica.com/16/2316-050-3E41AB58/map-Ptolemy-world-Ulm-Ger.jpg>)

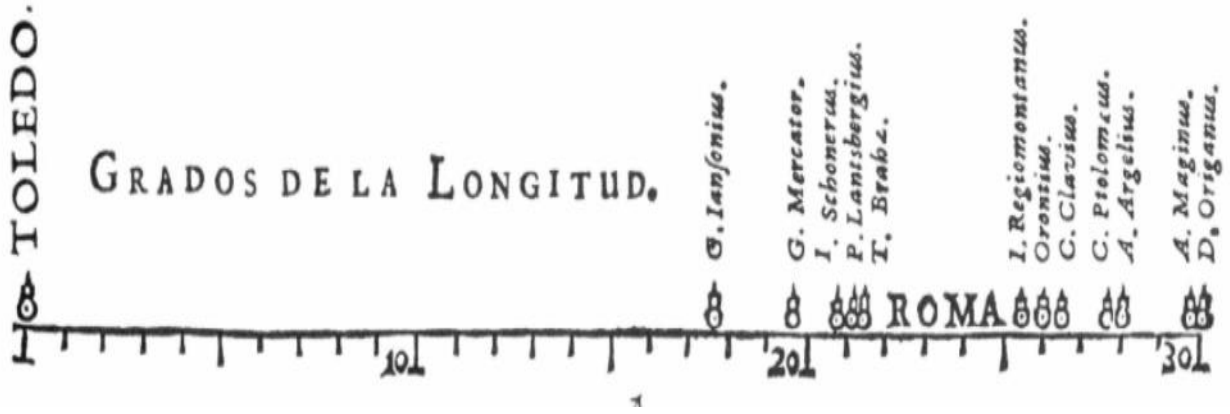
Nicel bilgilerin en eski grafik tasvirleri arasında yer alan, “önde gelen yedi göksel cismin uzay ve zaman üzerindeki değişen konumu” temalı anonim bir 10. yüzyıl çoklu zaman serisi grafiği Şekil 3'de verilmiştir.



Şekil 3: Yedi göksel cismin uzay ve zaman üzerindeki değişen konumu (Funkhouser'dan (1936) aktaran Tufte, 2001, s.28)

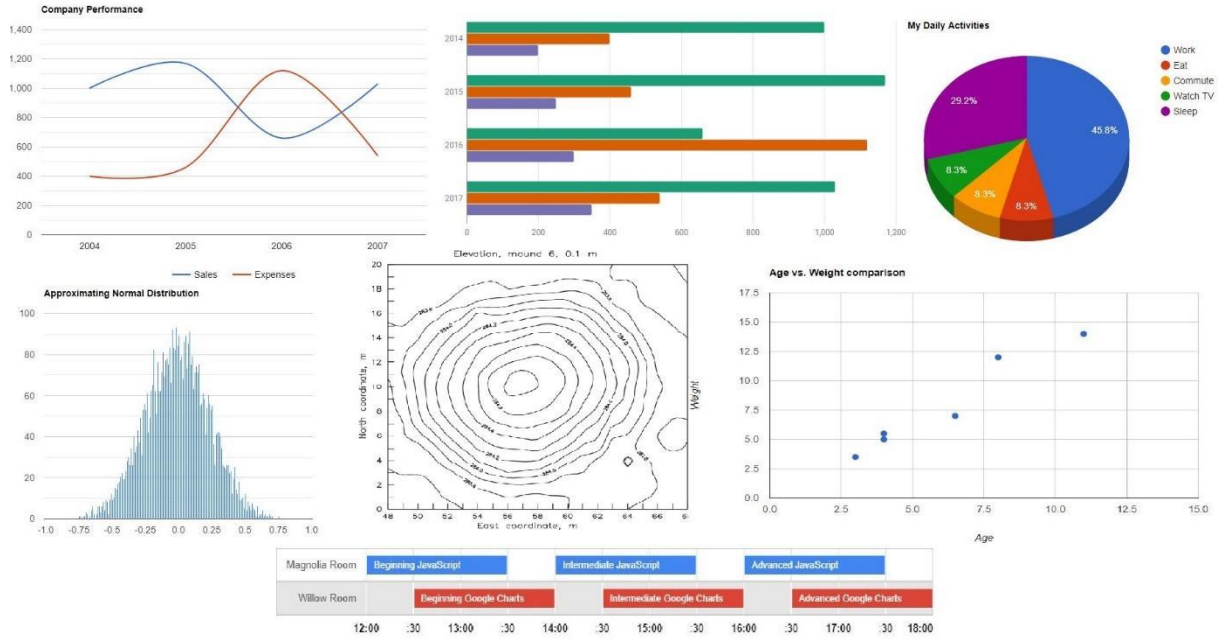
Veri görselleştirmenin ilk örneklerinin astronomi ve topoğrafya bilgisini temsil etmek üzere geliştirildiği, insanlığın gökyüzü ve dünya hakkında bilgi edinme çabasını yansıttığı görülmektedir. Bu durum, veri görselleştirmenin tarihsel evrimine de ışık tutmaktadır. Astronomi ve topoğrafya gibi alanlarda tablolar ve haritalar kullanılarak verilerin görsel olarak temsil edilmesi, insanların karmaşık bilgileri anlamalarına ve keşfetmelerine yardımcı olmuştur. Veri görselleştirmenin aslında insanlığın doğa ve evren hakkındaki merakını tatmin etme, bilgiyi anlama ve paylaşma arzusunu desteklediğini göstermektedir. Bu bağlamda, veri görselleştirme, tarihsel kökenleri itibariyle insanlığın bilgiye olan doğal eğilimini ve keşfetme isteğini yansıtan önemli bir araç olmuştur.

H. M. Chen (2017), 17. yüzyılda analitik geometrinin yükselişinin ve zaman, mesafe ve uzayı tahmin etmek için kullanılan ölçümlerin ve teorilerin doğuşunun, fenomenlere yönelik çözümlerin üretilme şeklini önemli ölçüde geliştirdiğini ve böylece görsel düşünmenin yolunu açtığını belirtmektedir. Verilerin istatistiksel analiz ve grafiksel olarak sunumunda görselleştirme tekniklerinin kullanımı da anılan yüzyılda (1644 yılında) olmuştur. İspanya sarayında görev yapan ve Flaman kökenli bir astronom olan Michael Florent van Langren [1600-1675], Toledo'dan Roma'ya olan mesafeyi boylam olarak göstermeyi amaçlayan bir grafik (Friendly, 2006, s.3; Tufte, 1997, s.15) geliştirmiştir. Söz konusu grafik istatistiksel verilerin kullanıldığı ilk görsel temsil olarak bilinmektedir (bkz. Şekil 4).



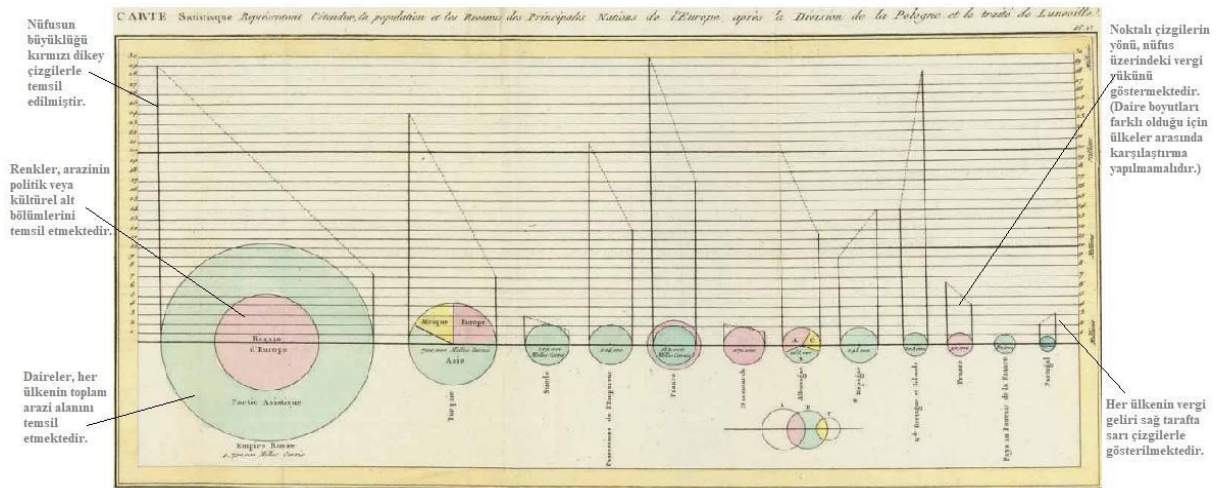
Şekil 4: Michael Florent van Langren [1600-1675] tarafından hazırlanan, Toledo'dan Roma'ya olan mesafeyi boylam olarak belirleyen grafik (Tufte, 1997, s.15)

18. ve 19. yüzyıllar, görselleştirme alanında tematik haritalamanın başladığı dönem olarak bilinmektedir. Histogramlar, zaman serisi grafikleri, kontur (eşyüksekti) grafikleri, dağılım grafikleri gibi birçok istatistiksel grafik türü bu dönemde icat edilmiştir. Bu dönem aynı zamanda bugün kullandığımız en popüler grafiklerin (çizgi/line, çubuk/bar ve pasta/pie grafikler) çoğunun mucidi olarak kabul edilen İskoç mühendis William Playfair'in de dönemidir (Few ve Edge, 2007, s.3). Playfair, bilgiyi iletmenin en hızlı ve en sezgisel yolunun "göze hitap etmek" olduğunu savunmuştur (Classics of Data Visualization, 2021). Playfair, aynı zamanda İngiltere adına çalıştığı için, görselleştirmeleri genellikle politik ve ekonomik niteliktedir.



Şekil 5: Örnek grafik türleri (çizgi, çubuk, pasta, histogram, eşyükselti, dağılım ve zaman serisi) (Google Charts, t.y.)

1801 yılında Playfair'in "Statistical Breviary" adlı eserinden alınan "ülkelerin nüfus, arazi genişliği, vergi yükü ve siyasi durumu" temalı grafik Şekil 6'da verilmiştir.

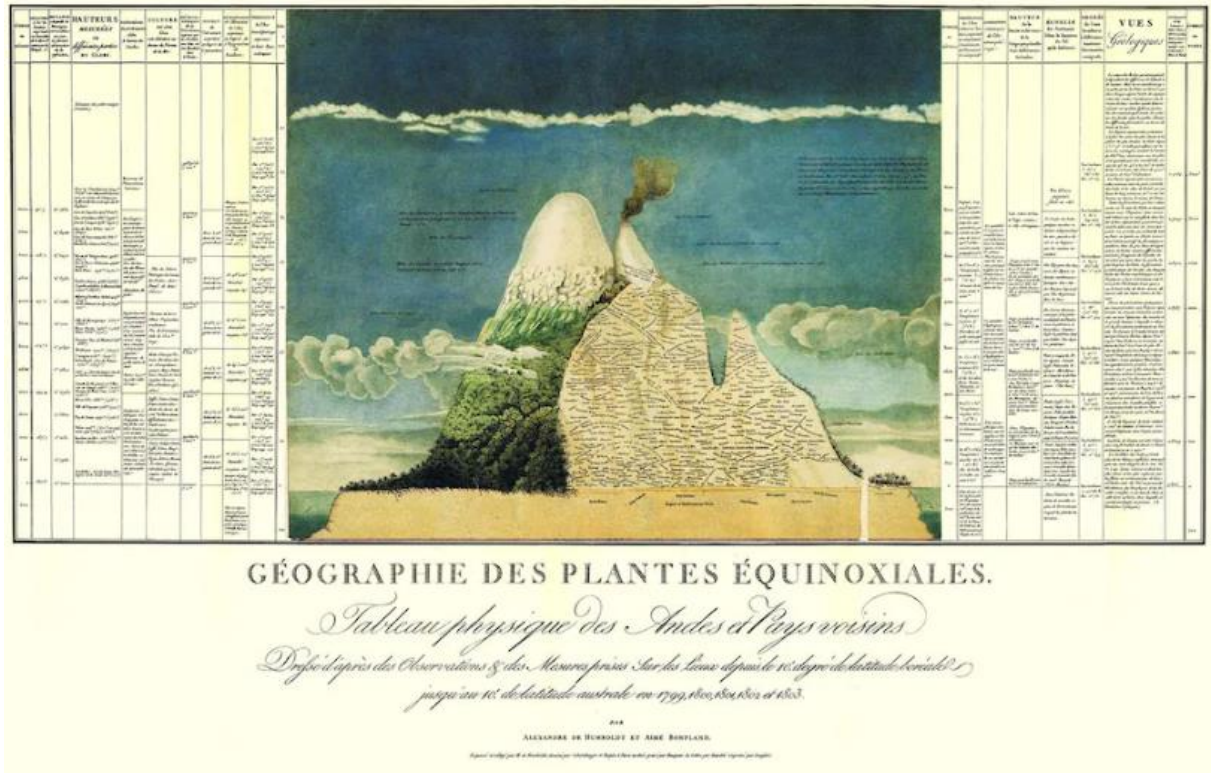


Şekil 6: 1801 tarihli Playfair'in "Ülkelerin nüfus, arazi genişliği, vergi yükü ve siyasi durumu" temalı grafiği (Classics of Data Visualization, 2021)

1800'lü yıllarda, görselleştirmede diyagram kullanımının yaygınlaştığı matematiksel ispatların ve işlevlerin temsilde yeni yöntemlerin kullanılmaya başladığı görülmektedir (Friendly, 2006; H. M. Chen, 2017). Hesaplamalara yardımcı olan nomogramlar geliştirildikçe görselleştirmede de bir evrim yaşanmıştır. Özellikle bu dönemde, karmaşık verileri ve olayları

anlamak için nomogramlar⁷ kullanılmıştır. Nomogramların yanı sıra, çizgi grafikler, histogramlar⁸, dağılım grafikleri gibi formların kullanımı yaygınlaşmış ve bunlar özellikle sosyal alandaki yeni araştırmalarda görselleştirme türleri olarak uygulanmıştır (H. M. Chen, 2017). Görsel düşünme (visual thinking) ile beraber, birçok toplumsal alanda sayısal verilerin kullanımını içeren görselleştirilmiş istatistiki temsiller, bu dönemde hükümetler tarafından politika oluşturma ve faaliyet planlaması gibi çalışmalar için kullanılmıştır.

Almanya Berlin doğumlu (1769) iktisatçı ve coğrafyacı Alexander von Humboldt'un, Latin Amerika'nın fiziki coğrafyası ve doğal alanları hakkında bilimsel verileri kaydetmek için yaptığı Latin Amerika seyahatinde, Ekvador'da Chimborazo adlı bir yanardağın eteklerinde karşılaştığı bitkileri, bulundukları rakıma göre yerleştirdiği Naturgemalde (Unity of Nature) adlı eseri Şekil 7'de sunulmuştur. Humboldt resmin yanındaki iki sütunu, karşılaştığı bitkiler hakkında kapsamlı ayrıntılar vermek için kullanmıştır.

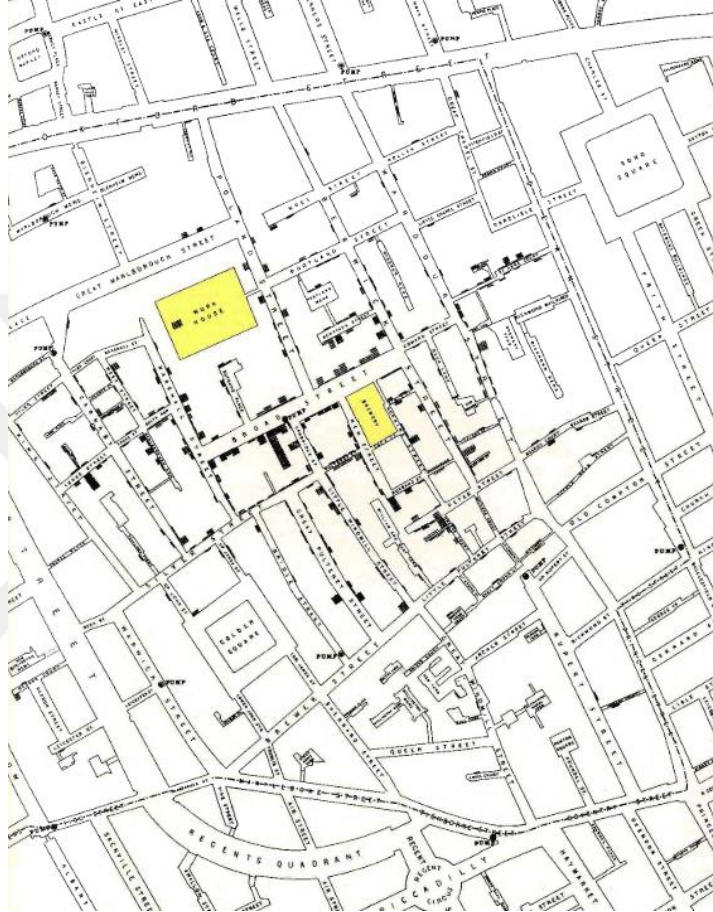


Şekil 7: Alexander von Humboldt'un 1805 tarihli Naturgemalde adlı eseri (Classics of Data Visualization, 2021)

⁷ Nomogram: Bilinen değerleri iki çizgi üzerinde birleştirmek için bir düz kenar kullanarak bilinmeyen bir değer başka bir çizgiyle kesişme noktasında okunabileceği şekilde ölçeklendirilmek üzere işaretlenmiş ve düzenlenmiş birkaç çizgiden oluşan bir grafik temsidir (Merriam-Webster, t.y.).

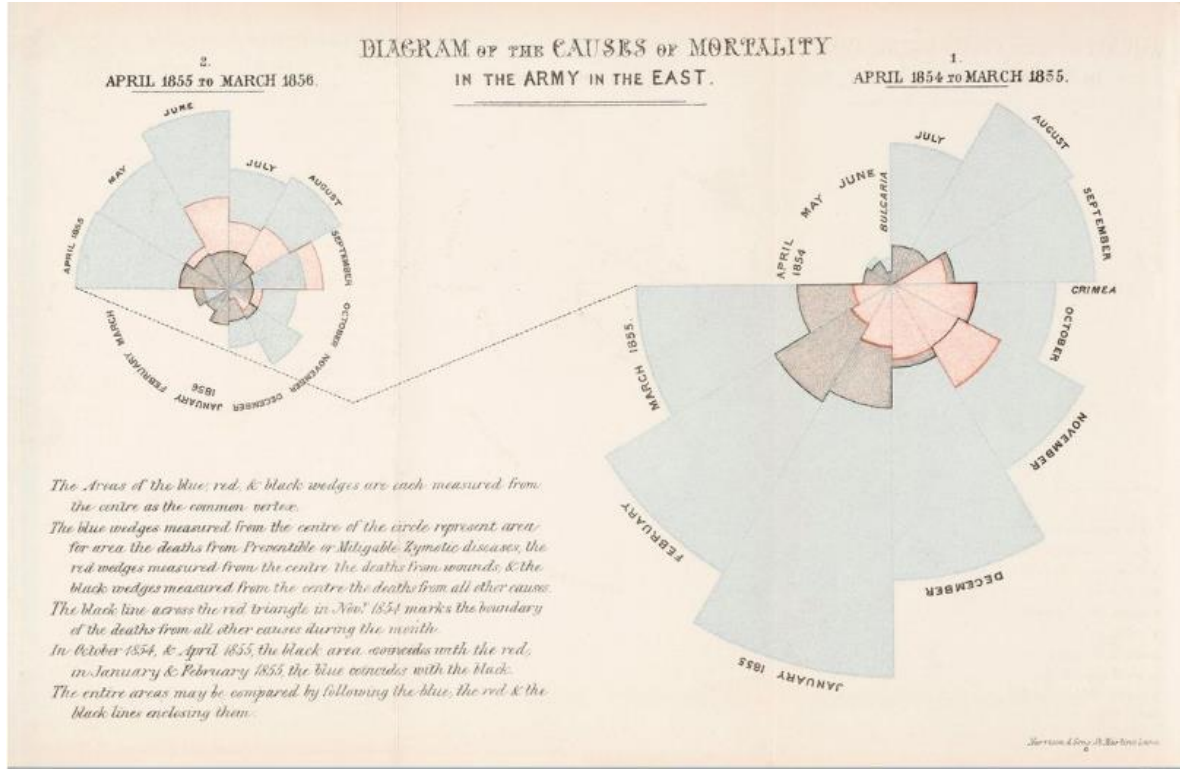
⁸ Histogram: Genişlikleri sınıf aralıklarını temsil eden ve alanları karşılık gelen frekanslarla orantılı olan dikdörtgenler aracılığıyla bir frekans dağılımının temsidir (Merriam-Webster, t.y.).

19. yüzyılda sayısal verilere dayalı olarak oluşturulan önemli veri görselleştirme çalışmalarına örnek olarak; John Snow'un 1854'te Londra'daki kolera salgınları haritası (bkz. Şekil 8), Charles Minard'ın Napolyon'un 1812'deki Rus seferini gösteren 1869 tarihli tablosu ve Florence Nightingale'in, 1854-1856 yıllarındaki Kırım Savaşı'nda İngiliz askerlerinin ölüm sebeplerini anlattığı "gül diyagramı" verilebilir (bkz. Şekil 9).



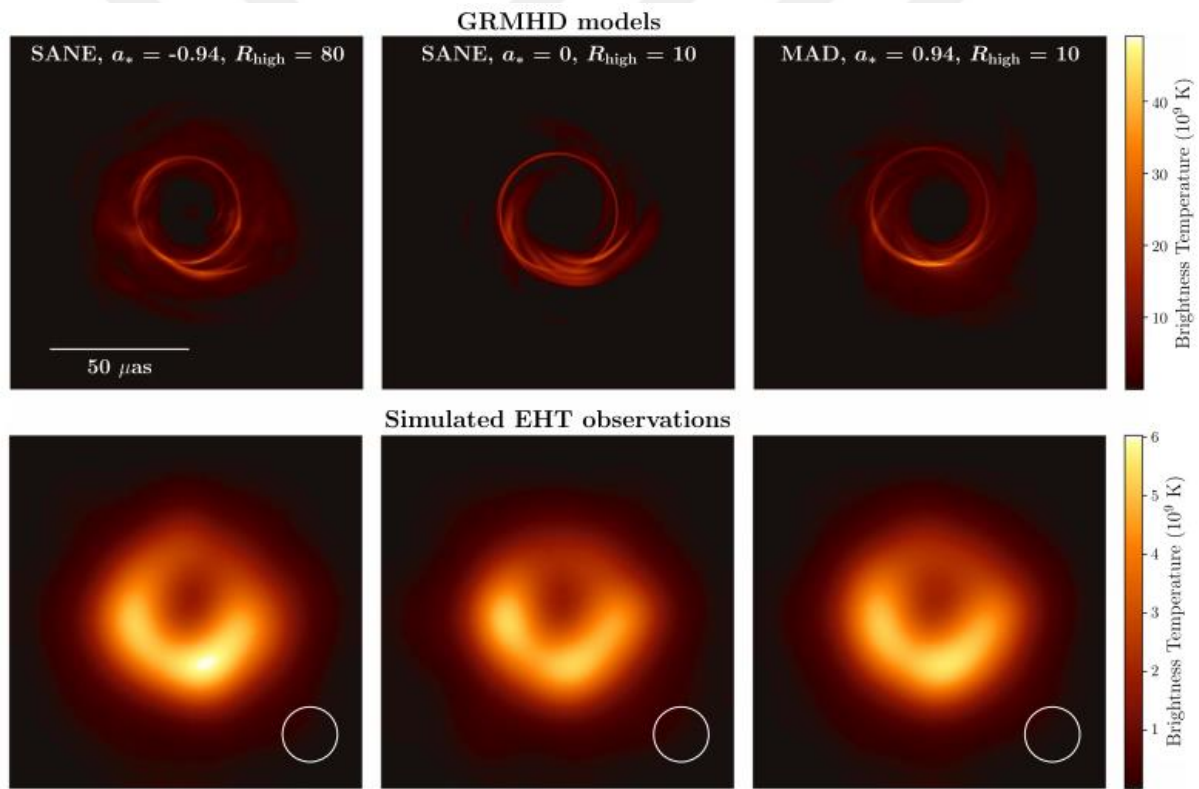
Şekil 8: John Snow'un Londra'daki kolera salgını haritası (Tufte, 1997, s. 31)

Modern hemşirelik hizmetlerinin kurucusu olarak kabul edilen ve aynı zamanda istatistiksel grafiklerin de öncülerinden olan Florence Nightingale [1820–1910], 1857 yılında "coxcombs" (gül diyagramları) olarak adlandırılan bir görselleştirme gerçekleştirmiştir. Nightingale çalışmasında, 1854-1856 yılları arasında Kırım Savaşı'nda ölen İngiliz askerlerinin ölüm sebeplerini betimlemiştir (Friendly, 2006, s.14; Chen, Hårdle, Unwin, 2008; Ward ve diğerleri, 2015; Kirk, 2019; Erkurt, 2020). Şekil 9'da görüleceği üzere Nightingale, İngiliz askerlerinin ölüm sebeplerini "gül diyagramı" yoluyla anlatmış; askerlerin ölüm nedenlerini farklı renklerle temsil etmiştir. Hastalıktan ölenleri mavi, savaş yaralarından ölenleri kırmızı ve diğer nedenlerden dolayı ölenleri siyah renk ile göstermiştir.



adımdır. Bilgisayar kullanımındaki artış ve görselleştirme yöntem ve araçlarındaki ilerleme, keşifsel veri analizi için daha büyük hacimli verilerin görüntülenebilmesini ve dinamik etkileşimini sağlamıştır.

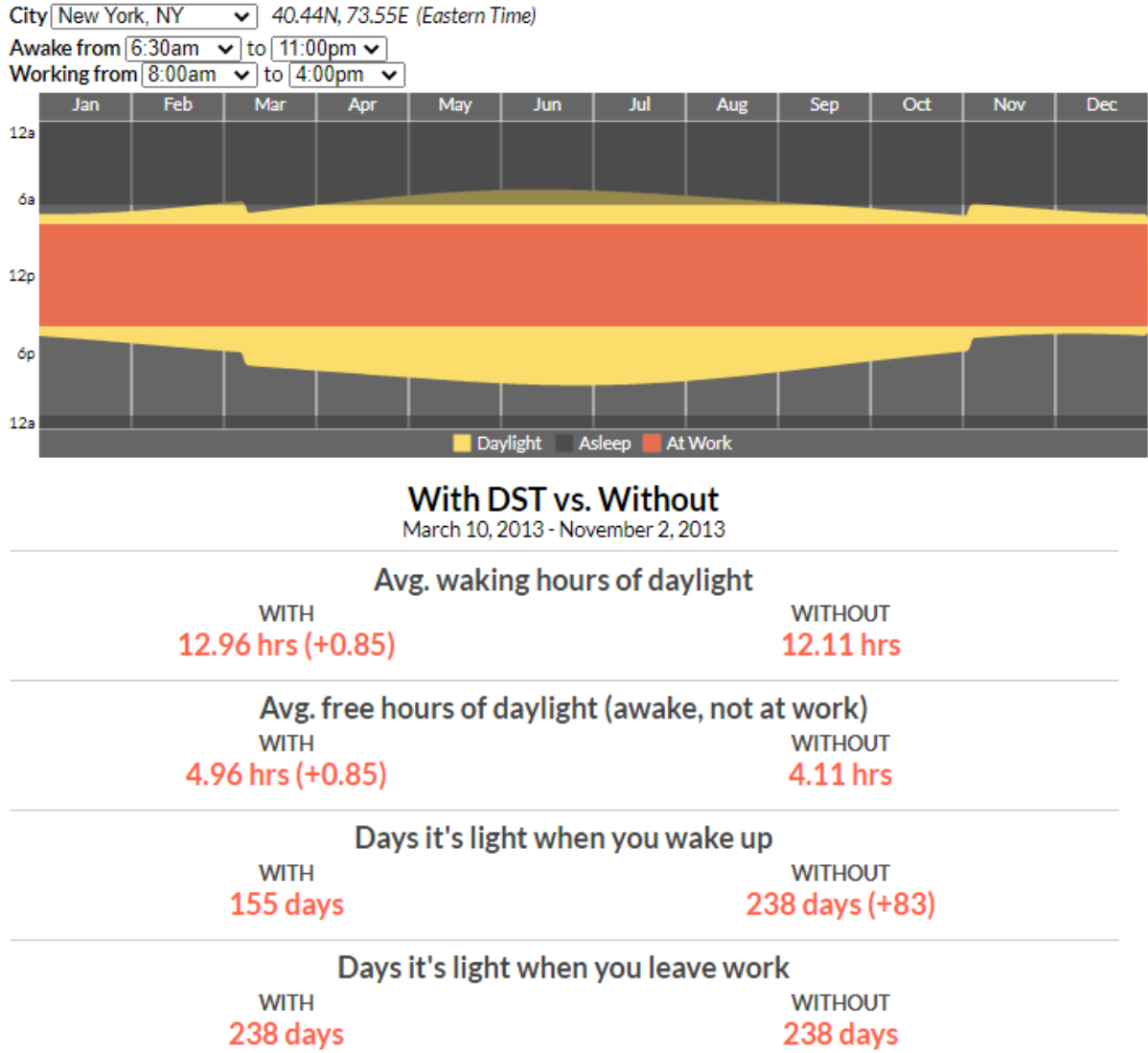
2017 yılında The Event Horizon Telescope Collaboration tarafından, dünyayı muazzam bir radyo teleskobuna dönüştürmek için dünya çapında bir teleskop ağı (altı adet teleskop) oluşturulmuştur. Bu teleskop ağı ile 55 milyon ışık yılı uzaklıktaki Messier 87 (M87) isimli gökadanın merkezinde bulunan süper kütleli bir kara delik gözlemlenmiştir. Elde edilen veriler ile ilk defa bir kara deliğin gerçek görüntüsü elde edilmiştir. Yaklaşık olarak 5 petabyte (5 milyon gigabyte) veri, birbirinden bağımsız veri analiz ve veri görselleştirme uzmanlarından oluşan gruplar tarafından analiz edilerek Şekil 10'daki görsel elde edilmiştir ve Nisan 2019'da dünya basınında tanıtılmıştır (Horizon Telescope Collaboration, T. E., 2019).



Şekil 10: Messier 87 (M87) gökadasındaki bir kara deliğin ilk görüntüsü (Horizon Telescope Collaboration, T. E., 2019)

Veri görselleştirme, analiz ve sunum için kullanılan bilişsel bir süreçtir, verilerin daha iyi anlaşılmasını sağlar ve sunduğu anlayışa göre hareket etme fırsatı sunar. Veri görselleştirme, ham verilerin barındırdığı eğilimlerin tespit edilmesine yardımcı olur. Veri görselleştirme sürecinde üretilen görseller statik olabileceği gibi birçok görselleştirme aracı, etkileşim olanağı da sunmaktadır. Bu etkileşimlilik, yalnızca tek ve tartışılmaz bir gerçeği değil, aynı zamanda

son kullanıcıya verileri keşfetme ve muhtemelen beklenmedik sonuçlara ulaşma fırsatı da verir. Veri görselleştirmede etkileşimli araçlara örnek olarak Noah Veltman'ın web sitesinde (Veltman, t.y.) yer alan projelerden biri olan ve Şekil 11'de sunulan "Effects of Daylight Saving Time (Yaz Saati Uygulamasının Etkileri)" adlı grafik verilebilir.

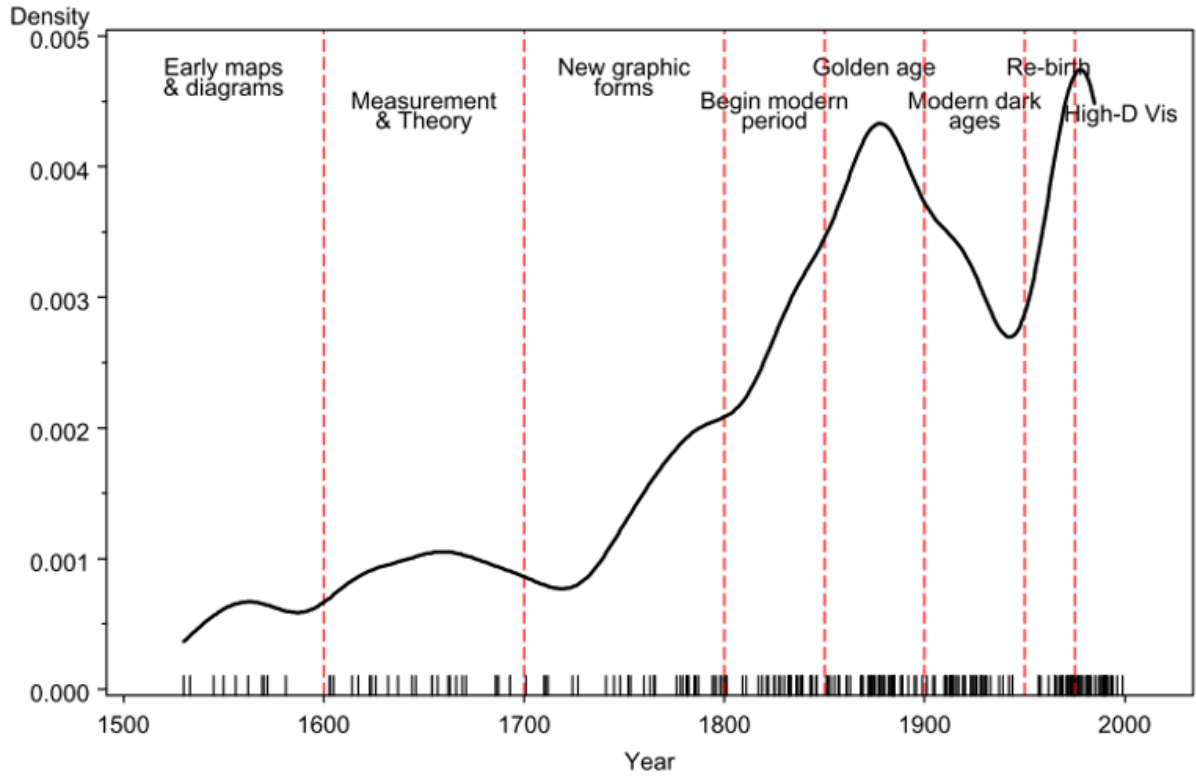


Şekil 11: Yaz saati uygulamasının etkileri (Veltman, t.y.)

Şekil 11'deki görselde kullanıcıya belirli parametreleri değiştirme olanağı sunulmaktadır. Kullanıcı; şehir, uyanık kalınan saat aralığı ve çalışma saat aralığı gibi parametre değerlerini değiştirebilmektedir; böylelikle kullanıcı değişimi gözlemleyebilmekte ve karşılaştırma yapabilmektedir.

York Üniversitesinden (Kanada) Profesör Michael Friendly tarafından oluşturulan "Görselleştirmenin Kilometre Taşları Projesi" (Milestones Project) (<http://www.datavis.ca/>); veri görselleştirme, istatistiksel grafikler ve tematik haritacılık tarihlerindeki önemli olayların

kapsamlı ve görsel bir özetini sunmaktadır. Friendly (2006), orta çağdan modern zamanlara kadar veri görselleştirmenin entelektüel tarihine genel bir bakış sunarak, bazı önemli ilerlemeleri örnekler ile göstermiştir. Bu projenin temel amacı modern veri görselleştirmeye yol veren çok çeşitli alanlardaki önemli gelişmeleri tek bir yerde toplamak, kataloglamak ve belgelemektir. Friendly (2006, s.3) çalışmasında orta çağdan günümüze veri görselleştirmenin kilometre taşlarını bir görselleştirme türü olan zaman serisi (time-line) ile vermiştir (bkz. Şekil 12).



Şekil 12: Veri görselleştirme tarihinde kilometre taşları zaman çizelgesi (Friendly, 2006, s.3)

Tarih boyunca çeşitli amaçlarla başvurulmuş bir yöntem olan veri görselleştirmedeki çoğu gelişmenin son iki yüzyılda, ağırlıklı olarak da son yarım yüzyılda meydana geldiği görülmektedir. Alberto Cairo (2012, s.16) görselleştirme ile ilgili düşüncelerini şu şekilde dile getirmektedir:

“Yıllar boyunca beni büyüleyen bir gizemi seçmemi isteseydiniz, bu beynimizin farklı soyutlama derecelerine sahip görsel temsiller yaratma ve anlama yeteneği olurdu: verileri, kavramları, bağlantıları ve coğrafi konumları kodlayan grafikler...”

İnsan zihni, görsel biçimde aktarılan bilgiyi yazılı veya sözlü bilgiye kıyasla çok daha hızlı, verimli ve kalıcı bir şekilde algılayabilmektedir. Shneiderman'a (2003, s.2) göre, görsellerin hızlı algılanabilmesi sayesinde insanlar; görüntüleri hızla tarayabilir, tanıyabilir ve

anımsayabilir ve boyut, renk, şekil, hareket veya örüntülerdeki değişiklikleri algılayabilirler. Beynin diğer veri formlarına oranla, görselleştirilmiş formları işlemesi daha kolaydır (H. M. Chen, 2017, s.7). Görseller, analizi çok daha kolay ve daha hızlı hale getirmekte ve önemli olanı bir bakışta görme olanağı sunmaktadır. Ayrıca insanlar görsellere metinden çok daha hızlı tepki göstermektedirler.

Veri görselleştirmenin amacı, verilerin anlaşılmasına yardımcı olmaktır. H. M. Chen (2017), görselleştirmenin amacının sadece verileri resmetmek değil, içgörü kazanmak olduğunu söylemektedir. Ayrıca içgörü kazanmayı sağlamlaştıran etkili bir iletişim ve sunum aracı olduğuna da vurgu yapmaktadır. Kirk (2019) ise veri görselleştirmenin temel amacının açıklama ve keşfetme olduğunu belirtmektedir.

Veri görselleştirme;

- karmaşık verileri basitleştirir,
- verileri bağlamsallaştırır, argümanları sunabilmek için açıklama fırsatı sunar (Kirk, 2019),
- büyük verileri görseller kullanarak analiz etmeye ve keşfetmeye yardımcı olur,
- dikkat edilmesi veya iyileştirilmesi gereken alanların belirlenmesini sağlar,
- veriler arasındaki ilişkiyi tanımlar,
- yeni örüntüleri araştırır ve verilerdeki gizli örüntüleri ortaya çıkararak öngörüler elde etmeyi sağlar.

Veri görselleştirme, başkalarına yeni içgörüler iletmek, onlarla iş birliği yapmak ve etkileşimde bulunmak için büyük miktarda veriyi yorumlamanın açık ve verimli bir yöntemidir. Verilerin görselleştirilmesi; kullanıcıların dikkatini çekmek, verilerin anlaşılması için gereken süreyi azaltmak, karşılaştırma yapılmasını sağlamak, resim üstünlüğü etkisi (Picture Superiority Effect¹⁰) ile ana mesajın daha akılda kalıcı hale getirilmesini sağlamak gibi kolaylıklar sunar. Ayrıca bilgiyi başka dilleri konuşan kullanıcılar için daha anlaşılabilir hale getirir (Krum, 2014, s.288). Veri odaklı görselleştirmeler tasarlamak, karmaşık bilgilerin anlaşılmasını kolaylaştırarak kullanıcılar için ilgili konuyu çekici hale getirir.

Doğası gereği görüntülerin kelimelerden daha kolay anlaşılır olduğu görsel bir dünyada yaşadığımız için, veri görselleştirme, büyük veri ve veri analiz projeleri söz konusu olduğunda özellikle önemlidir.

¹⁰ Picture Superiority Effect: Resim üstünlüğü etkisi, resimlerin ve resimlerin hatırlanma olasılığının kelimelerden daha fazla olduğu fenomenini ifade eder (Picture Superiority Effect, 2024).

Günümüzde, daha ustaca gerçek zamanlı (veya neredeyse gerçek zamanlı) veri görselleştirmeleri oluşturmak için yeni ve farklı seçenekler sağlayan yeni teknolojiler ve uygulamalar keşfedilmekte, geliştirilmekte ve uygulanmaktadır. Teknolojik gelişmeler veri görselleştirmeye ilişkin bazı zorlukları da beraberinde getirmektedir. Miller (2017, s.9), doğru ve etkili görselleştirme yöntem ve türlerinin nasıl seçileceğine dair bir anlayış kazanmanın önemli olduğunu belirtmekte ve bu anlayışın kazanılabilmesi için veri görselleştirme öncesinde aşağıdaki soruların cevaplanması gerektiğini ifade etmektedir:

- Görselleştirilecek verilerin boyutu ve hacmi nedir?
- Verilerin önem derecesi ve bağlamı nedir?
- Hedef kitle kimlerden oluşmaktadır ya da bu bilgiyi kimler tüketecektir?
- Verilmek istenen mesajlar/aktarılmak istenen bilgi, hedef kitleye en iyi şekilde ne tür bir görsel ile iletebilir?

Veri görselleştirmeye başlamadan önce yukarıdaki soruların yanıtlanması, görsel temsillerin yalnızca görsel olarak çekici olmasını değil, aynı zamanda bilgilendirici, alakalı ve hedef kitlenin ihtiyaçlarını karşılayacak şekilde uyarlanmış olmasını da sağlayacaktır. Böylece; karar verme süreçlerinin yönlendirilmesi ve büyük veri kümelerinden eyleme geçirilebilir bilgiler elde edilmesi için veri görselleştirmenin etki gücü arttırılmış olacaktır.

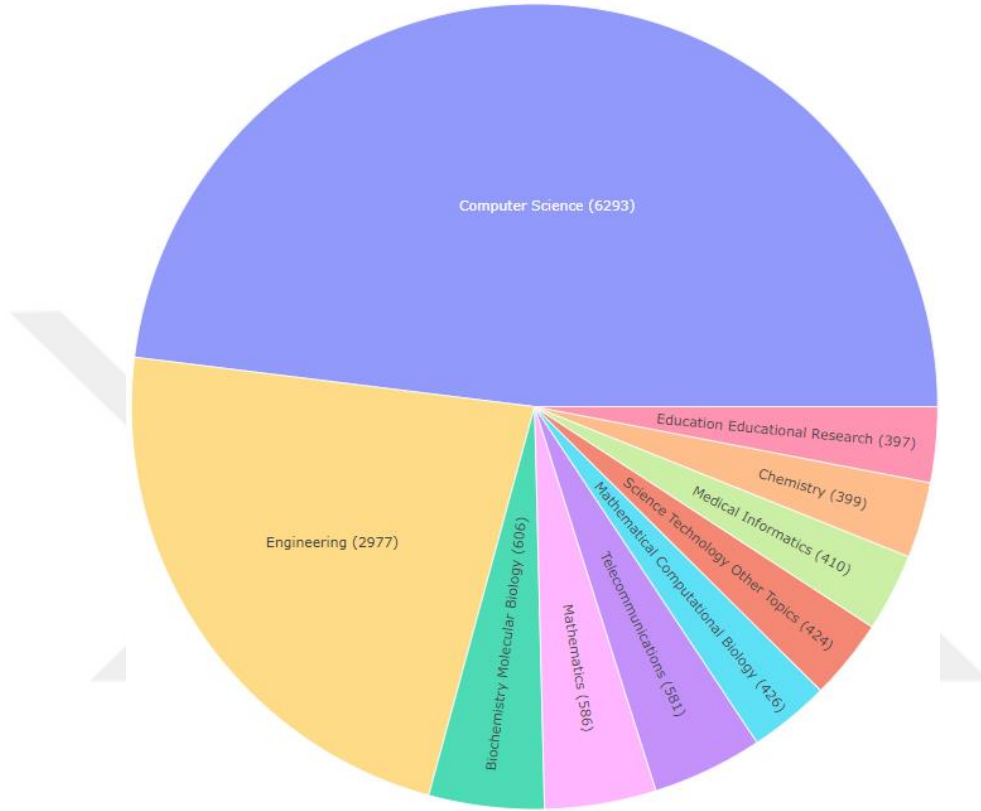
Günümüzün veri odaklı dünyasında, büyük veri analitiğinin önemi sektör sınırlarını aşarak her alanda ilgi odağı haline gelmiştir. Geniş ve çeşitli veri kümeleriyle karakterize edilen büyük veriler, değerli bilgiler elde etmek, karar verme süreçlerini geliştirmek ve rekabet avantajı kazanmak isteyen kuruluşlar için bir mihenk taşı haline gelmiştir. Ancak büyük verilerin hacmi ve karmaşıklığı, anlamlı yorumların verimli bir şekilde elde edilmesinde zorluklara yol açmaktadır. Veri görselleştirmenin, veri analitiğinin önemli bir bileşeni olarak ortaya çıktığı yer burasıdır.

Büyük veri kümelerine veri görselleştirmelerin uygulanması, ham veri kümelerinin görsel olarak ilgi çekici temsillere dönüştürülmesine olanak tanır ve kullanıcılara karar verme süreçlerine yardımcı olan bilgilendirici görseller oluşturma olanağı sağlar.

2.2 İlgili Literatür

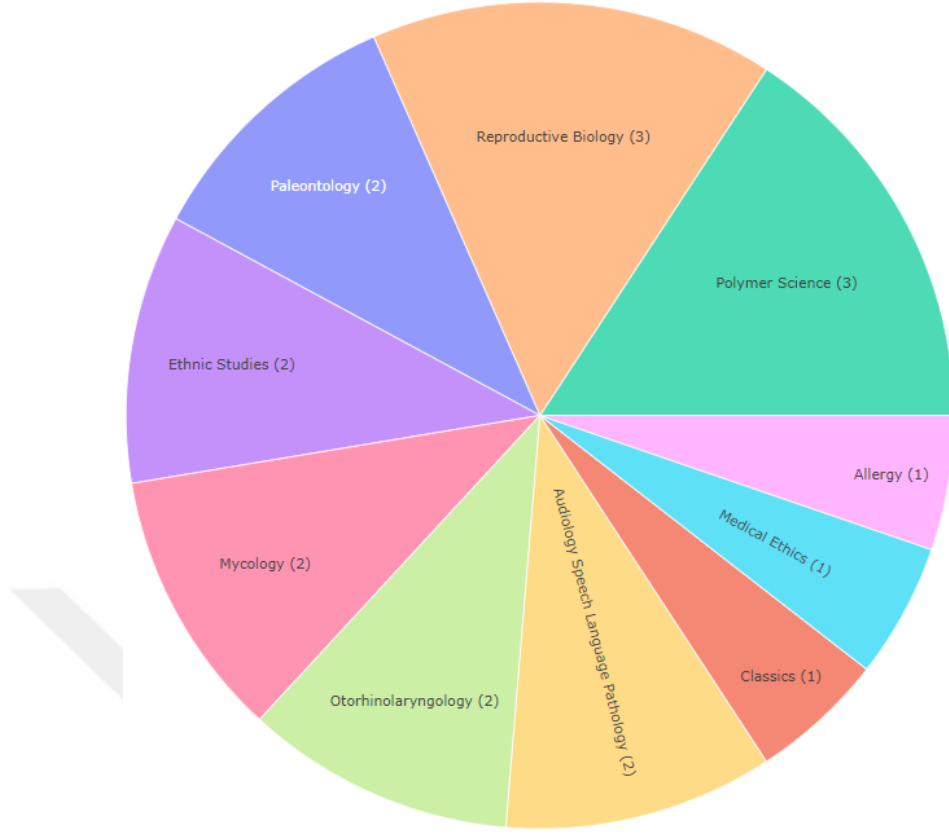
Son yıllarda veri görselleştirme, birçok araştırmacının ilgisini çekmiş, birçok disiplinde üzerinde araştırma yapılan bir konu haline gelmiştir. Web of Science veri tabanında "data visualization (veri görselleştirme)" terimiyle yapılan aramada, bu konu ile ilgili geniş bir yayın

kümesi elde edilmiştir. Veri görselleştirme konusunun hangi disiplinlerde ele alındığının anlaşılması için araştırma alanı (research area) ve yapılan yayın sayısı verileri toplanarak Şekil 13.a ve Şekil 13.b'deki görseller oluşturulmuştur. Toplam 151 araştırma alanında 21.245 yayın saptanmıştır. Şekil 13.a'da en çok yayın yapılan 10 araştırma alanı ve Şekil 13.b'de ise en az yayın yapılan 10 araştırma alanı görüntülenmiştir.



Şekil 13.a: WoS Research Area alanında “data visualization” terimi ile yapılan tarama sonucuna göre yayınların dağılımı (en çok yayın yapılan konu kategorileri)

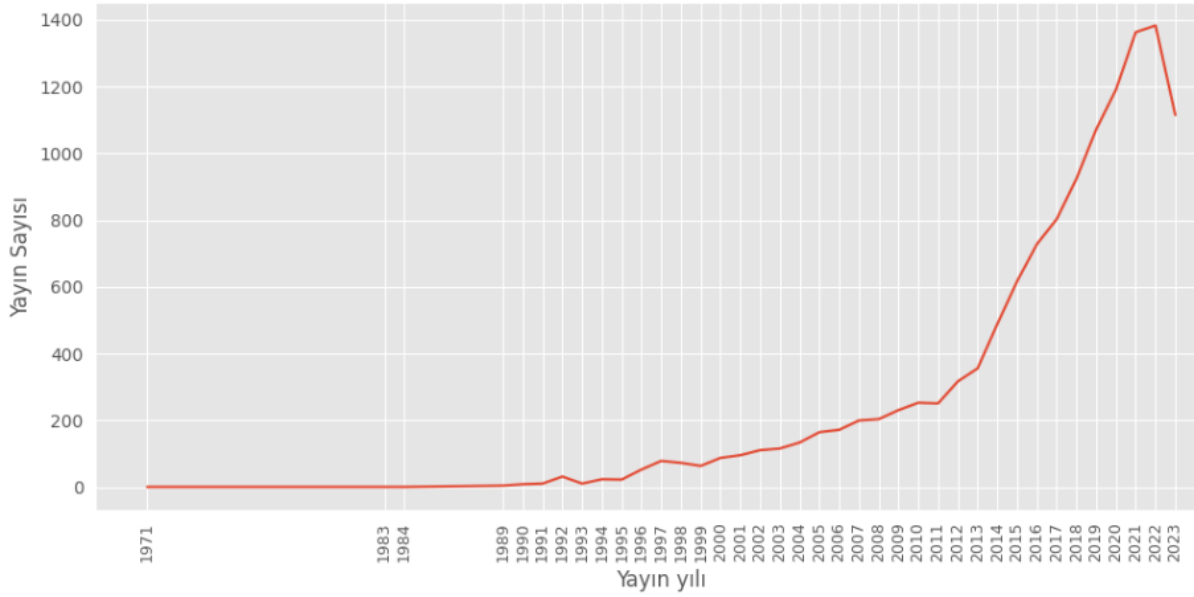
Şekil 13.a incelendiğinde, uluslararası alanda veri görselleştirme konusunda en çok bilimsel yayının bilgisayar bilimleri (computer science) ve mühendislik (engineering) araştırma alanlarında üretildiği görülmektedir.



Şekil 13.b: WoS Research Area alanında “data visualization” terimi ile yapılan tarama sonucuna göre yayınların dağılımı (en az yayın yapılan konu kategorileri)

Şekil 13.b incelendiğinde ise veri görselleştirme konusunda en az bilimsel yayının alerji-allergy, tıp etiği-medical ethics ve klasikler-classics araştırma alanlarında üretildiği görülmektedir.

Sonuçlar yıllara göre (Şekil 14) gruplandığında ise “veri görselleştirme-data visualization” konusu ile ilgili yapılan yayın sayısında düzenli bir artışın olduğu gözlemlenmiştir. Şekil 14’te görüleceği üzere veri görselleştirme konusunda uluslararası alanda yıllara göre yapılan yayın sayısında sürekli bir artış bulunmaktadır.



Şekil 14: “Data visualization” konusunda yapılan yayınların yıllara göre yayın dağılımı

TR Dizin, DergiPark ve Google Akademik’te yer alan, veri görselleştirme konusunda ülkemizde yapılmış yayınlar incelendiğinde, konunun birçok farklı alanda ele alındığı tespit edilmiştir. Başta bilgisayar bilimleri olmak üzere veri görselleştirme uygulamalarının edebiyat, sanat, emlak vb. alanlarda da kullanıldığı tespit edilmiştir.

Artut (2017) araştırmasında, internet kullanıcılarının “Distopya-Ütopya” (www.dystopia-utopia.com) adlı internet sitesindeki, distopik filmlerle ilgili görüşlerini içeren veriler üzerine bir görselleştirme uygulaması gerçekleştirmiş ve “Bilim Kurgu Sineması” nda toplumsal açıdan önem içeren söylemleri sorgulamıştır. Çalışmada görselleştirme türlerinden radar grafik (radar chart) ve hiyerarşik kümeleme (hierarchical clustering) görselleştirme türleri kullanılmıştır.

Altın (2018), karma yöntem kullanarak farklı sektörlerden üç grubun görüşlerine başvurarak yaptığı çalışmada, “verinin günümüz dünyasında en çok değer biçilen kavramlardan biri olduğunu, ancak yoğun enformasyon bombardımanında önemli olanın sağlanan verilerin herkes için hızlı ve kolay anlaşılır olmasına bağlı olduğunu” (Altın, 2018, s.575) belirtmektedir. Çalışma, “görselleştirme ve infografik disiplinlerinin tasarım eğitimi içerisindeki yerini ortaya koyma” (Altın, 2018, s.575) amacı ile gerçekleştirilmiştir. Yazar, infografik ve veri görselleştirme alanlarına tasarım eğitimi içerisinde yer verilmesiyle, öğrencilerin kapsamlı bilgi kümelerini analiz ederek bilgiyi estetik ve etkili bir biçimde aktarabileceğini, hiyerarşik yapı oluşturma becerisi edinebileceğini, böylece analiz edilebilen tasarımlar oluşturabileceğini söylemektedir (Altın, 2018, s.585).

Çelik ve Akdamar (2018) "Büyük veri ve veri görselleştirme" adlı çalışmalarında, veri görselleştirmenin bilgisayar bilimlerinde oldukça yeni ve umut verici bir araştırma alanı olduğunu vurgulamışlardır. Çalışmada öncelikle büyük veri tanıtılmış, daha sonra yaygın olarak kullanılan görselleştirme türlerine dair açıklayıcı bilgiler, görsel örnekleriyle verilerek veri görselleştirmenin önemi vurgulanmıştır. "Büyük verinin interaktif görselleştirilmesi: Tableau üzerine öğrenci deneyimleri" adlı çalışmada Eken (2020), Kocaeli Üniversitesi Bilgisayar Mühendisliği Bölümü'nde verilen, Büyük Veri Analizine Giriş dersi kapsamında gerçekleştirilen bir örnek durum çalışmasından söz etmektedir. Buna göre öğrenciler Tableau programı kullanılarak, interaktif olarak büyük verinin görselleştirilmesi konusunda bilgilendirilmiş, bir uygulama yapmaları sağlanmış ve sonrasında öğrencilere bir anket uygulanmıştır. Anket sonuçlarına göre öğrencilerin açık kaynak teknolojilerini kullanabilme, herkese açık veri kümesi bulabilme ve Tableau kullanarak veri analizi ve görselleştirme yapabilme yetkinliklerinin arttığı gözlenmiştir. Çalışmada çeşitli açık veri setlerinden yararlanılmış ve görselleştirme türleri olarak bar chart (çubuk), pie chart (pasta), line chart (çizgi), treemap (ağaç harita), bubble (baloncuk), scatter plot (dağılım grafiği), haritalar ve çeşitli infografikler kullanılmıştır.

Çalışkan ve Kahraman (2020) "Viyana Toplum ve Ekonomi Müzesi'nin sergileme tasarımında verinin görselleştirilmesi" adlı çalışmalarında; bir istatistik veri görselleştirme yöntemi olan ISOTYPE¹¹ (International System of Typographic Picture Education), sergileme tasarımı kapsamında incelenmiştir. Çalışmada, "küreselleşme ve teknolojik gelişmelere bağlı olarak sergi ve müzelerde bilgi aktarım yöntemlerinin gelişimi ele alınmış, herhangi bir nesneye ihtiyaç duyulmaksızın bilginin her kesime eşit oranda aktarımının sağlanmasının önemine" vurgu yapılmıştır (s.3).

Atalay ve Neslihanoglu (2021) tarafından, hava kirliliğinde etkin bir madde olan partikül madde (PM10) miktarının Türkiye'deki şehirlerde etkisinin değerlendirilmesi, R programlama dilindeki görselleştirme paketleri (ggplot2, ganimate) kullanılarak harita üzerinde görselleştirilmiştir. İlgili görselleştirme çalışmasında, veri seti olarak, Çevre, Şehircilik ve İklim Değişikliği Bakanlığı Hava Kalitesi Bülteninin 2019 yılı sayısında yer alan 2014-2018 yılları arasındaki PM10 verileri kullanılmıştır. Görselleştirmeler için R programlama dilinin "ggplot2" kod kütüphanesinin kullanıldığı bir diğer çalışmada ise Cansız (2019); çok

¹¹ ISOTYPE (International System of Typographic Picture Education): Otto Neurath ve Gerd Arntz iş birliği ile "kelimesiz küresel bir dil" yaratmak için tasarlanan ISOTYPE bir istatistiksel veri görselleştirme yöntemidir (Çalışkan ve Kahraman, 2020).

değişkenli, fazla gözleme sahip veri setlerinde görselleştirme için uygulanan parçalı dairesel (Circle Segments) veri görselleştirme türünün iki boyutlu düzlemlerde noktasal olarak nasıl kullanılabileceğini irdelemiştir.

Erkut ve Yıldırım (2021), "Büyük Veri Görselleştirme: Emlak Sektörüne İlişkin Bir Uygulama" adlı çalışmalarında, öncelikle İstanbul ili konut ilanlarından oluşan bir veri seti oluşturmuş ve yazarlar konut tercihinde yararlanılabilecek kullanıcı dostu infografikler üretmişlerdir. Çalışmada; konut ilan istatistikleri karşılaştırmalı olarak verilerek kullanıcılara konut tercihinde seçim yapma kolaylığı sunmak amaçlanmıştır. Veri görselleştirme için Tableau Reader ve R programlama dili görselleştirme aracı olarak kullanılmıştır.

Doğanay, Orman ve Dener (2022), literatürde yer alan çalışmalara uygulanan model testlerinin, yeni ve güncel veri setleri üzerinde yapılmasının siber güvenlik için alınacak önlemleri de güncelleyeceği yönündeki görüşlerini sundukları çalışmalarında (s.1573), büyük veri görselleştirmelerine dair literatürü değerlendirmiş ve bir siber güvenlik büyük veri seti olan BETH¹² üzerinden görselleştirme uygulaması yapmışlardır. Çalışmada “BETH veri seti, Principle Component Analysis (PCA)¹³ kullanılarak görselleştirilmiş ve eski tarihli bir veri seti olan KDD¹⁴ veri setinin görselleştirilmesiyle karşılaştırılmıştır” (ss.1573-1574). Çalışmada görselleştirme türü olarak çeşitli dağılım grafikleri kullanılmıştır.

Zengin ve Cemelelioğlu (2022), "Tanzimat'tan günümüze en çok eser vermiş 10 edebiyatçı kadına yönelik veri görselleştirme projesi" adlı çalışmalarında, veri görselleştirmenin, görsel-algisal yöntemler kullanılarak soyut ve büyük verilerin anlaşılabilmesini, veriler arasındaki ilişkilerin gözlemlenebilmesini ve bilgilerin güçlendirilmesini sağladığını ifade etmişlerdir. Çalışmada, Tanzimat'tan bugüne en çok eser vermiş 10 kadın yazar seçilmiş ve veri seti; bu 10 kadın yazarın eser türleri, eser sayıları ve eser tarihlerinden oluşturulmuştur. Kadın edebiyatçıların eserlerine dair tür dağılımı ve dağılımın yıllara göre değişimi görselleştirmeler aracılığıyla incelenmiştir. Görselleştirmeler "Flourish" veri görselleştirme programı (ücretli paket program) üzerinden oluşturulmuştur. Yazarların dağılımını görmek için radyal (radial bar) görselleştirme türü kullanılmıştır.

¹² Highnam, Arulkumaran, Hanif ve Jennings'in (2021) “BETH Dataset: Real Cybersecurity Data for Unsupervised Anomaly Detection Research” adlı çalışmalarında kullanılan veri seti.

¹³ Principle Component Analysis (PCA): Temel bileşen analizi veya PCA, istatistikte büyük bir değişken kümesini, büyük kümedeki bilgilerin çoğunu içeren daha küçük bir değişken kümesine dönüştürerek, büyük veri kümelerinin boyutluluğunu azaltmak için sıklıkla kullanılan bir boyut azaltma yöntemidir (Jaadi, 2021).

¹⁴ KDD (Knowledge Discovery and Data Mining): Üçüncü Uluslararası Bilgi Keşfi ve Veri Madenciliği Araçları Yarışması için kullanılan veri setidir (Kaynak: <https://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html>).

Er ve Akman'ın (2022) "Kurumsal kaynak planlama ve veri görselleştirme teknolojisi: Müşteri ilişki analizi" adlı çalışmalarında kurumsal kaynak planlama sistemlerinin etkili bir şekilde kullanımı ve bu aşamada veri görselleştirme teknolojilerinin uygulanmasının önemine vurgu yapılmıştır. Çalışmada görselleştirmelerin uygulanacağı veri seti, bir üretim firmasının geçmiş dönem müşteri ve ürün verilerinden oluşturulmuş ve müşteri satış verilerinin gösterge paneli (dashboard) üzerinden görselleştirmelerle takip edilmesinin önemi anlatılmıştır. Bu sayede müşteriler arasındaki anlamlı ilişkilerin istatistiksel olarak incelenebileceği ifade edilmiştir. Çalışmada ayrıca yaygın kullanılan görselleştirme türleri kısaca açıklanmış ve örnek gösterge panelleri için donut grafikleri (donut chart) kullanılmıştır.

Günümüzde, genellikle büyük veri olarak adlandırılan çok miktarda veriden değerli içgörüler elde etmenin önemi, çeşitli sektörlerde giderek daha fazla kabul görmektedir. Büyük verinin titizlikle analiz edilmesi ve anlaşılır görsel temsillere dönüştürülmesi süreci, karar vericiler ve planlayıcılar için önemli hale gelmiştir. Bu yaklaşım, yalnızca verilerde yer alan temel örüntülerin, eğilimlerin ve korelasyonların daha kapsamlı anlaşılmasını kolaylaştırmakla kalmaz, aynı zamanda bu veriye dayalı içgörülere dayanarak bilinçli kararlar alınmasına da olanak tanımaktadır. Büyük verinin göze çarpan yönlerini görselleştirmeler yoluyla doğru bir şekilde yorumlama ve etkili bir şekilde iletme isteği, günümüzün veri odaklı ortamında hayati bir olgu haline gelmiştir.

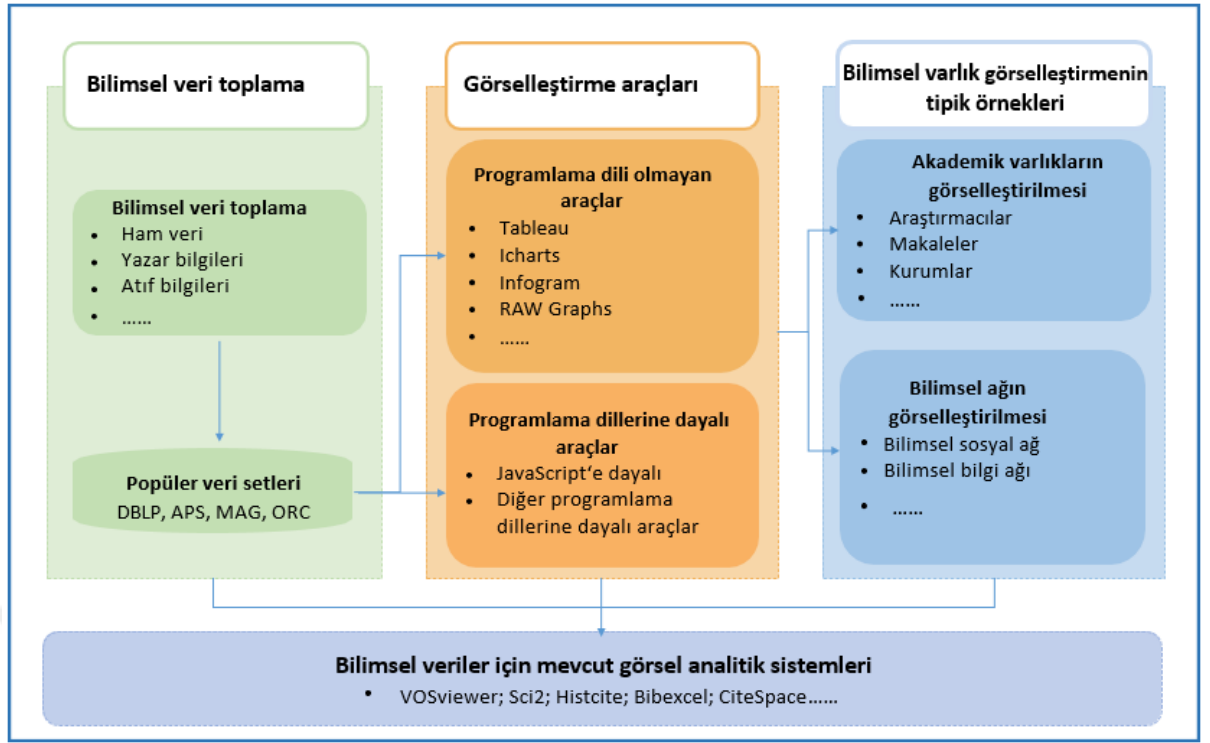
Veri görselleştirme, kütüphanecilik disiplininde, özellikle de akademik kütüphanelerde önemli bir rol oynamaktadır. Kütüphanelerde araştırma faaliyetlerini desteklemeye veri görselleştirmenin nasıl kullanıldığına ışık tutan ve konunun önemine dikkat çeken birçok çalışma yapılmıştır. Görselleştirme teknolojileri ve bunların kütüphane ortamındaki uygulamalarına ilişkin çalışmalar incelendiğinde, veri görselleştirme tekniklerinin genellikle; kütüphane hizmetlerini ve performansını değerlendirmek ve desteklemek, kütüphane koleksiyonunun gelişimini izlemek, bibliyografik verileri sunmak ve bilim disiplinlerinin araştırma etkisini ölçmek gibi çeşitli amaçlar için kullanıldığı görülmüştür.

Ruhland, Bioletti ve Sedlmair (2009) "LibViz: Data visualisation of the Old Library" adlı çalışmalarında, 1732 yılında inşa edilen, uluslararası üne sahip bir araştırma kütüphanesi olan Dublin Trinity College Kütüphanesinde koleksiyonu koruma amacıyla (s.178) bir veri görselleştirme yazılımı -LibViz- geliştirildiğinden söz etmektedirler. Koleksiyonda bulunan Kells Kitabı (The Book Of Kells/Columba Book), Yeni Ahit'in içerdiği dört incil ile önsöz niteliğindeki bazı metinleri ve kanon tablolarını içeren işlemeli el yazması kitap, kütüphanenin önemini oldukça arttırmaktadır. Bu kitap son yıllarda turist çektiği için ziyaretçi sayısında artış

olmuştur. Ziyaretçi sayısındaki artışın bu tür eserler için bir tehdit oluşturması ve bu durumun kontrol altına alınması için kütüphanedeki toz dağılımı, sıcaklık, bağıl nem seviyesi ve ziyaretçi sayılarına ilişkin toplanan verilerin görselleştirilmesini sağlayan "LibViz" yazılımının geliştirildiği belirtilmektedir. Çalışmada ayrıca kütüphane hizmetlerinin ve koleksiyonlarının değerlendirilmesi için veri görselleştirme yazılımlarının benimsenmesi gerektiğine de vurgu yapılmıştır.

Akademik verilerin nasıl görselleştirileceği ile ilgili bilgiler sunan "A survey of scholarly data visualization" adlı çalışmada (Liu, Tang, Wang, Xu, Kong ve Xia, 2018); bilimsel veri görselleştirme alanındaki son araştırmaları gözden geçirmek, bu alandaki gelişmelere ve zorluklara ilişkin kapsamlı bir anlayış sağlamak ve ortaya çıkan bu disiplinin geleceğini belirleyecek önemli konuları bulmak amaçlanmıştır. Bu çalışma, akademik verilerin görselleştirilmesine yönelik çeşitli yöntemleri ve araçları inceleyen kapsamlı bir araştırma sunmaktadır. Öncelikle veri görselleştirme ile ilgili temel kavramlar tanıtılmakta ve akademik verilerin (scholarly data)¹⁵ nasıl toplanacağı, analizi ve görselleştirilmesi üzerine odaklanılmaktadır. Başlıca konular arasında, bilimsel iş birliği ağlarının, atıf ağlarının ve yazar iş birliği ağlarının görselleştirilmesi yer almaktadır. Akademik veri görselleştirmesi için kullanılan mevcut görsel analitik sistemler detaylandırılmakta ve bu sistemlerin işlevselliği, etkinliği ve uygulama alanları tartışılmaktadır. Bu kapsamda çalışmada, bilimsel verilerin neler olabileceği, mevcut popüler açık veri setlerinin, görselleştirme araçlarının ve bilimsel veriler için mevcut görsel analitik sistemlerinin gruplanarak verildiği Şekil 15'teki "bilimsel veri görselleştirme çerçevesi" oluşturulmuştur.

¹⁵ Bilimsel veri (Scholarly data): Bilimsel veriler; bilimsel belgeler (makaleler, kitaplar, patentler ve bilimsel raporlar vb.) gibi bol miktarda akademik kaynağın yanı sıra yazarların bilgileri, atıflar, şekiller, tablolar vb. dahil ilgili verileri içerir (Liu ve diğerleri, 2018, s.1).



Şekil 15: Bilimsel veri görselleştirme çerçevesi (Liu ve diğerleri, 2016, s. 3'ten Türkçe'ye uyarlanmıştır)

Şekil 15'te görüleceği üzere yazarlar akademik veri görselleştirmeyi kategorize etmişlerdir. Çalışmada, çevrimiçi dijital kütüphanelerden veya akademik arama motorlarından toplanan ham veriler sağlanacak ilk veri kümesi olarak adlandırılmakta, bu ilk veri kümesinin; yazarları, yayın başlığını, özetleri, anahtar kelimeleri, yayın yerini, yayıncıyı, sayfa numaralarını, yayın tarihini ve DOI vb. verileri kapsadığı belirtilmektedir. Akademik veri görselleştirme için yazar bilgilerinin derlenmesinde makalelerin kullanılabileceği ve yazar bilgileri kullanarak bağlı olunan kurum, araştırma ilgisi ve ortak yazarlık durumlarının saptanabileceği vurgulanmaktadır. Ayrıca makalelerdeki atıfların derlenmesi için ParsCit, FLUX-CiM ve CRF tabanlı sistemlerin kullanılabileceğine vurgu yapılmaktadır. Yazarlar tarafından, görselleştirme araçları olarak programlama dili olmayan; Tableau, Icharts, Infogram ve RAW Graphs uygulamaları ile programlama dillerine dayalı JavaScript gibi dillerin kullanılabileceği belirtilmiştir. Çalışmada son olarak karmaşık bilimsel veri görselleştirmesine yönelik yeni çözümler konusunda öneriler tartışılmıştır.

"Use of data in collections work: An exploratory survey" adlı çalışmada Brown ve Stowers (2013), kütüphanelerin hizmet ve koleksiyonlarının değerini göstermeye çalışmışlardır. Bu süreçte, verilerin toplanması ve kullanılmasının son derece önemli olduğunu belirtmişlerdir. Koleksiyonlara dair verilerle ilgili kullanılan uygulamaları belirlemek

amacıyla, başta akademik kütüphaneler olmak üzere halk kütüphaneleri ve özel kütüphanelerde çalışan 200'den fazla kütüphaneciye bir anket uygulanmıştır. Anket, koleksiyon verilerinin kullanılma yollarının yanı sıra, eğitim, kütüphaneci değerlendirmeleri ve veri değerlendirme bileşenlerinin varlığı da dâhil olmak üzere veri kullanımını teşvik etmeye hizmet edebilecek kurumsal uygulamalar ve tutumları araştırmaktadır.

Çalışmada, kütüphanelerin sağladıkları hizmetlerin ve koleksiyonların değerini gösterebilmek için hizmetlere dair veri toplama ve kullanmanın önemi vurgulanmıştır. Kütüphaneciler arasında veri kullanımına yönelik uygulamalar ve bu uygulamaları teşvik eden ya da engelleyen faktörler belirlenmeye çalışılmıştır. Ankete katılanların demografik profilinin büyük ölçüde doktora düzeyindeki kurumlarda çalışan kütüphanecilerden oluştuğu ve %48'inin koleksiyon işlerinde on yıldan fazla deneyime sahip kişiler olduğu tespit edilmiştir. Anket ayrıca veri üretimini, araçlar ve uygulamalar açısından da incelemiştir. Ankete katılanların %73,1'i ticari koleksiyon analiz araçlarına sahip olmadıklarını bildirmiştir. En yaygın kullanılan analiz aracının "OCLC WorldCat Collection Analysis Tool" olduğu tespit edilmiştir. Anket sonuçlarına göre:

- Birçok kütüphaneci yıllık olarak kütüphane sistemlerinde raporlar üretmekte, başkalarının ürettiği raporları incelemekte ve rapor taleplerinde bulunmaktadır.
- Kütüphaneler, veri üretimi ve kullanımı açısından farklılık göstermektedir.
- Ankete katılanlar, veri kullanımlarının beş yıl öncesine göre arttığını belirtmiştir.
- Veri kullanımındaki en yaygın amaçlar arasında harcamaları izlemek, koleksiyon ayıklama ve sürekli yayın incelemeleri yapmak yer almaktadır.
- Birçok kütüphaneci koleksiyonlara dair veri kullanımı ve analizi konusunda resmi bir eğitim almamıştır.

Anket sonuçlarına dayanarak yazarlar; kütüphane hizmetiçi eğitim müfredatına daha geniş tabanlı ve tutarlı eğitim programlarının entegre edilmesi gerektiğini önermektedirler. Ayrıca veri toplama ve analizinin kütüphane iş tanımlarına dâhil edilmesinin, kütüphanelerin değerlerini arttıracaklarını belirtmektedirler. Çalışmada fiziksel kütüphane koleksiyonunun görselleştirilmesinin, kütüphane koleksiyonu geliştirmede karar vermeyi desteklemek açısından faydalı olduğu sonucuna varılmıştır.

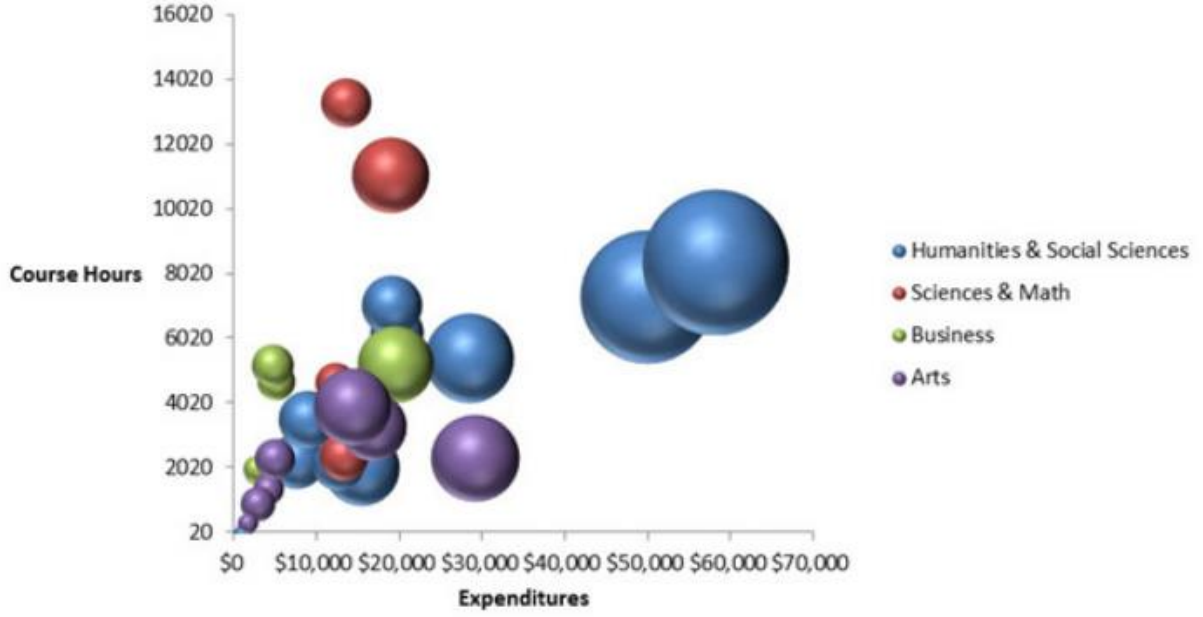
Murphy, veri görselleştirme ve veri analiz yazılımı olan Tableau'yu anlattığı çalışmasında (2013), Ohio State Üniversitesi (OSU) Kütüphaneleri tarafından Tableau Desktop adlı veri görselleştirme ve analiz yazılımının nasıl kullanıldığını anlatmaktadır. Yazar, Tableau

yazılımının, farklı veri kaynaklarından toplanan verileri birleştirerek kütüphanecilerin verileri daha etkili ve verimli bir şekilde yönetmelerini, keşfetmelerini ve sunmalarını sağladığını belirtmektedir. Makalede, OSU Kütüphaneleri'nin Tableau'yu kullanarak gerçekleştirdiği iki veri analizi projesinin (işlem günlükleri – transaction logs ve Google Analytics verileri) görselleştirmeleri paylaşılmakta ve çoklu veri kaynaklarından gelen verilerin birleştirilmesinin faydaları vurgulanmaktadır. Yazar, işlem günlüklerinin, kütüphanecilere kütüphane sistemlerindeki belirli işlemler hakkında ayrıntılı bilgi sağladığını, örneğin; bu günlükler sayesinde, kullanıcıların belirli bir süre içinde hangi kütüphane kaynaklarını ve hizmetlerini kullandığının takip edilebildiğini belirtmektedir. Kütüphaneler, bu verileri kullanarak kullanıcı davranışlarını analiz edebilir, hangi hizmetlerin daha popüler olduğunu belirleyebilir ve kaynakların etkinliğini değerlendirebilir. Çalışmada, kütüphane web sitelerinin trafiğini, kullanıcıların hangi sayfaları ziyaret ettiğini, hangi dillerde içerik görüntülediklerini ve hangi cihazlardan erişim sağladıklarını takip etmek için Google Analytics'in kullanılabileceği belirtilmektedir. Ayrıca bu verilerin, kütüphanelerin web sitelerinin performansını iyileştirmesine ve kullanıcı deneyimini optimize etmesine yardımcı olacağı ifade edilmektedir. Yazar, veri analiz yöntemini; sırasıyla veri temizleme ve hazırlama, Tableau kullanarak görselleştirme ve raporlama olarak ele almaktadır. Çalışmada, Tableau'nun, kullanıcıların çeşitli veri kaynaklarını bir araya getirerek etkileşimli panolar (dashboards) oluşturmaya olanak tanıdığı, bu panoların kütüphane yönetimi ve personeline, operasyonel kararlar alırken faydalı bilgiler sağlayacağı belirtilmektedir. Yazar, Tableau'nun OSU Kütüphaneleri'nde veri analizini nasıl daha erişilebilir ve etkili hale getirdiğini ve gelecekte kütüphanelerin Tableau'yu kullanarak daha fazla veri kaynağını entegre etmesi ve analiz yeteneklerini genişletmesi gerektiğine vurgu yapmaktadır.

Tableau'nun, kütüphane verilerinin karar vericiler için anlamlı ve erişilebilir hale getirilmesi için kullanılabileceğini belirten diğer bir çalışma Hoelscher ve Mortimer'e (2018) aittir. Çalışmada verileri anlamak ve verilerin karar verme sürecini geliştirebilecek bilgilere nasıl dönüştürülebileceğine yardımcı olmak için veri görselleştirme yazılımlarının kullanımı ve veri analizinin önemi vurgulanmıştır. Buhler ve diğerleri (2016) ise Tableau'yu kullanarak bir kütüphanenin hem görselleştirmeleri özelleştirmek hem de bağlam sağlamak için filtreler ve açıklamalarla tamamlanan esnek, derinlemesine çevrimiçi kontrol panelleri üretebileceğini belirtmişlerdir. Çalışmada kütüphanecilere yöneltilen üç soruya yanıt aranmıştır. Bu sorular; “Kütüphaneniz Tableau'yu değerlendirme programına nasıl dâhil etmiştir?”, “Tableau'nun büyük veri kümelerini anlamlandırma, verileri erişilebilir hale getirme ve paydaş iletişimini

geliştirme beceriniz üzerinde nasıl bir etkisi olmuştur?” ve “Tableau, kütüphanenizin veri stratejisinin hangi aşamalarına uymaktadır?” şeklindedir.

Finch ve Flenner (2016) çalışmalarında; College of Charleston Kütüphanesi’ndeki kitap koleksiyonunun konu bazlı bölümlerini, bu bölümlere dair harcamaları, ana dal ve yan dallardaki öğrenci kayıtlarını ve ders sayısını karşılaştırmak için çeşitli veri görselleştirmeler yapmışlardır. Kurumun “Kurumsal Araştırma Ofisi” ve “Koleksiyon Geliştirme” departmanından toplanan veriler şu şekildedir; ana dallarda ve yan dallarda sunulan ders sayıları, her konu alanında yapılan harcamalar, her konu alanındaki fiziksel koleksiyon büyüklükleri, öğrenci kayıtları ve bir yıllık kütüphane dolaşım (ödünç/iade) verileridir. Veriler, özel programlama becerileri gerektirmeyen Microsoft Excel’de derlenmiştir. Görselleştirmeler, öncelikle Microsoft Excel ve Tableau Public kullanılarak oluşturulan dairesel grafikler ve veri baloncuklarından oluşmaktadır. Görselleştirme uygulaması ile amaçlananlar; koleksiyon geliştirmedeki dengesizlikleri belirlemek, bütçe tahsisini ve gerekçelendirmeyi desteklemek ve ayıklama veya daha fazla geliştirme için potansiyel alanlar önermek şeklindedir. Yazarlar, görselleştirmelerin, koleksiyonlardaki olası dengesizlikler hakkında içgörüler sunduğunu ve verilerin analizi ile üniversitedeki ana ve yan dalları desteklemek için gerekenden daha gelişmiş veya daha az gelişmiş alanların tespit edilebildiğini belirtmişlerdir. Çalışmada, kullanılan metodolojinin, koleksiyonlarını incelemek isteyen diğer kurumlar için izlenecek bir şablon niteliğinde olduğu, harcamaları ayarlamak, hizmet fırsatları önermek veya koleksiyonun grafiksel analizden önce tespit edilememiş noktaların gün yüzüne çıkarılabilmesinin önemine vurgu yapılmıştır. Çalışmada ayrıca; İnsan ve Toplum Bilimleri Fakültesi, Fen ve Matematik Fakültesi, İşletme Fakültesi ve Sanat Okulu için ders saatlerinin ve satın alınan kitap sayısının ve satın alma için yapılan harcamaların görselleştirildiği bir örnek oluşturulmuştur (Şekil 16).

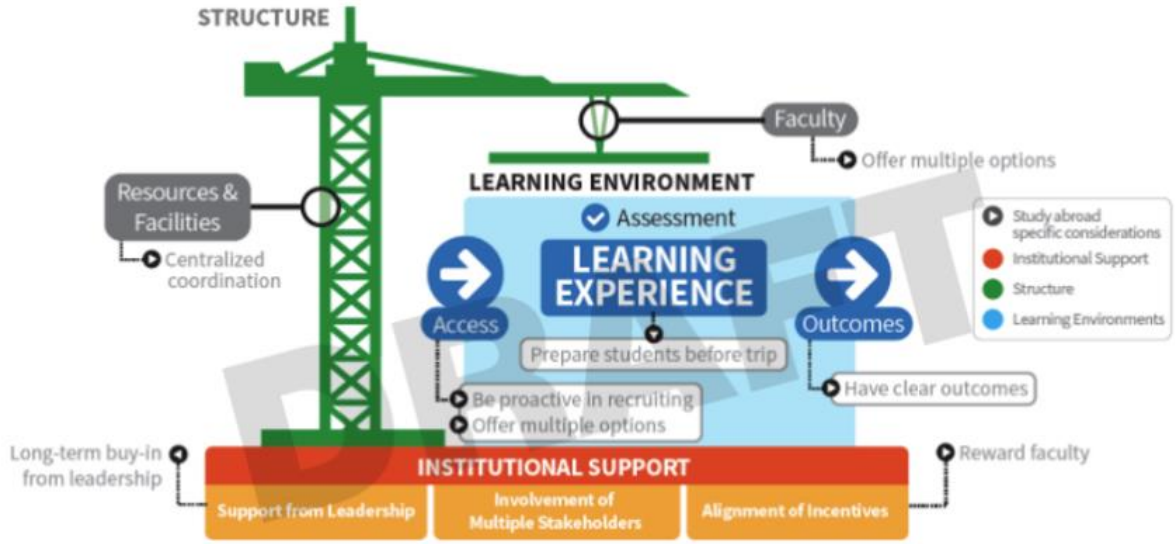


Şekil 16: College of Charleston ders saatleri ve kitap harcama oranları (Finch ve Flenner, 2016)

Şekil 16'ya göre "Humanities & Social Sciences" alanlarındaki harcamalar diğer alanlara göre daha fazladır. Ancak ders saatleri daha fazla olan "Sciences & Math" alanında yapılan satın almaların yeterli düzeyde olmadığı görülmektedir.

Ogier ve Stamper (2018) çalışmalarında; veri görselleştirmeyi bir kütüphane hizmeti olarak ele almakta ve bu hizmetin araştırma yaşam döngüsüne nasıl entegre edilebileceğini tartışmaktadırlar. Bu bağlamda; Virginia Tech'te üniversite kütüphaneleri tarafından bir danışmanlık hizmeti olarak sunulan veri görselleştirmenin, araştırma yaşam döngüsünün çeşitli aşamalarındaki kullanımına dikkat çekmişlerdir. 2016 yılında, Virginia Tech Üniversite Kütüphaneleri, araştırmacıların verilerini bulma, dönüştürme, modelleme, görselleştirme ve paylaşma yeteneklerini geliştirmek amacıyla bir "Veri Görselleştirme Tasarımcısı" istihdam etmiştir. Kurum dâhilindeki kütüphaneler, yalnızca bilgi okuryazarlığı modeliyle sınırlı kalmamak ve araştırma gruplarına veri hizmetleri sunmak amacıyla veri görselleştirme danışmanlık hizmetleri geliştirmektedir. Danışmanlık hizmetleri; veri türü ve formatının belirlenmesi, veri toplama yöntemleri, analiz prosedürleri, veri organizasyon ilkeleri ve tasarım süreçleri gibi veri görselleştirme ihtiyaçlarının belirli aşamalarında yararlı olmaktadır. Hizmetlerin amacı, kullanıcılara sadece görselleştirme sağlamak değil, aynı zamanda onları eğitmek ve kendi başlarına görselleştirme yapabilme becerilerini geliştirmektir. Ogier ve Stamper (2018), Virginia Tech Üniversite Kütüphaneleri tarafından sunulan veri görselleştirme danışmanlık hizmetlerinden elde edilen dersleri ve iki genel vaka çalışmasını ele almaktadır.

İlk vaka çalışmasında araştırma verilerinin analizi için görselleştirmeler kullanılmış, ikinci vaka çalışmasında ise, verilerin açık ve öz bir şekilde iletilmesi amaçlanmıştır. Sonuç olarak görselleştirmelerin tasarım seçimlerinin, araştırmanın başarısı üzerinde önemli bir etkiye sahip olduğu belirlenmiştir. Çalışmada veri görselleştirme danışmanlık hizmetlerinin; net beklentiler belirlemek, özel ve açık kaynak teknolojilerini dikkate almak ve danışmanlık deneyiminin aynı zamanda bir öğrenme deneyimi olmasını sağlamak da dâhil olmak üzere birkaç benzersiz hususa sahip olduğu sonucuna varılmıştır. Çalışmada görselleştirme örneği olarak verilen, 2017 yılında Amy Hermundstad Nave ve Denise R. Simmons tarafından Adobe Illustrator yazılımı kullanılarak oluşturulan, bir öğretme ve öğrenme metodolojisini özetleyen süreç diyagramı Şekil 17'de verilmiştir.



Şekil 17: Bir öğretme ve öğrenme metodolojisini özetleyen süreç diyagramı (Ogier ve Stamper, 2018, s.9)

Eaton (2017) “Seeing library data: A prototype data visualization application for librarians” adlı çalışmasında Kingsborough Community College Kütüphanesi (City University of New York) için kitap ve e-kitap koleksiyonlarını görselleştiren SeeCollections adında bir web uygulaması geliştirildiğini belirtmektedir. Yazar, SeeCollections'ın Kingsborough keşif katmanı API'sinden (Application Programming Interface-Uygulama Programlama Arayüzü) nasıl veri toplandığını ve bu verilerin web için görselleştirmeler oluşturmak üzere nasıl dönüştürüldüğünü anlatmaktadır. Çalışmada, API verilerine genellikle akademik kütüphanelerden erişim sağlandığı, SeeCollections keşif katmanı verilerine görsel bir arayüz sağlanarak veri kaynaklarının Kingsborough kütüphanecilerine anlık olarak gösterildiği belirtilmektedir.

Kitap ve e-kitap koleksiyonlarının (üst verilerinin) görselleştirilmesi sayesinde kütüphaneciler, büyük veri setlerini daha kolay analiz edebilir ve bu verilerdeki önemli eğilimleri ve kalıpları hızla tespit edebilir. Görselleştirme araçları, kütüphane verilerini daha erişilebilir ve anlaşılır hale getirir, bu da karar alma süreçlerini iyileştirebilir. Kütüphaneciler, kütüphane kaynaklarının kullanımını ve dağılımını daha iyi izleyerek, kütüphane koleksiyonlarının yönetimini optimize edebilirler. Bu görselleştirmeler sayesinde, kütüphanelerin verimliliği artırılarak, kütüphane hizmetlerinin kalitesi ve kullanıcı memnuniyeti yükseltilebilir.

2020 yılında LaPolla tarafından yapılan bir anket (telefon ve e-posta yoluyla yarı yapılandırılmış görüşme) çalışmasında akademik sağlık bilimleri kütüphanelerinde (sekiz kurum) kütüphanecilere sağlanan Excel odaklı veri görselleştirme atölyeleri incelenmiş ve bu konudaki deneyimlerine ilişkin bilgiler sentezlenmiştir. Ayrıca veri görselleştirme, bibliyografik verileri analiz etmek ve CiteSpace¹⁶ gibi görselleştirme araçlarına dayalı atıf ağları oluşturmak için bibliyometrik göstergelere odaklanan çalışmalar da bir yaklaşım olarak kullanılmıştır (ör. C. Chen, 2005; Su, Li ve Kang, Y. 2019; Wu, Kennedy, Qiu ve Smyth, 2019; Zhang ve Eichmann-Kalwara, 2019).

Zakaria (2021); üniversite kütüphanelerinde araştırma desteği olarak çevrimiçi sunulan veri görselleştirme hizmetlerini incelediği çalışmasında QS Dünya Üniversiteleri Sıralamasına göre ilk 100 içindeki üniversitelerden otuzunun kütüphane web sitesini esas almıştır. Web sayfalarından elde ettiği sonuçlar beş unsur açısından değerlendirilmiştir. Çalışmada veri görselleştirme hizmeti bağlamında kütüphanelerin %93,3 oranında Tableau, R, Excel, Gephi ve Plotly gibi veri görselleştirme yazılımları sunduğu, 26 kütüphanenin (%86,6) ders kitapları, veri kaynakları, e-kaynaklara harici bağlantılar ve bloglar gibi veri görselleştirmeyle ilgili bilgi kaynakları sağladığı, ayrıca kütüphanelerin %80'inin kendi web sayfaları aracılığıyla üniversite topluluğuna veri görselleştirme hizmeti sunduğu belirtilmiştir.

Kütüphane ve bilgi bilim alanı bağlamında incelenen literatürde genellikle veri görselleştirme için hazır yazılımların (Tableau, CiteSpace vb.) kullanıldığı, python'ın veri analiz kütüphanesi Pandas kütüphanesi ve görselleştirme paketleri olan “matplotlib, seaborn ve ployly.express” kütüphanelerine vurgunun yapılmadığı tespit edilmiştir. Ayrıca Türkiyede alanımızda veri görselleştirmeye yönelik çalışmaların kısıtlı olduğu, var olan çalışmaların

¹⁶ CiteSpace, bilimsel literatürdeki eğilimleri ve örüntüleri görselleştirmek ve analiz etmek için ücretsiz olarak kullanılabilen bir Java uygulamasıdır. Aşamalı bilgi alanı görselleştirmesi için bir araç olarak tasarlanmıştır (Chen, 2004).

çoğunlukla bibliyometrik analizler (örneğin; Al, Sezen ve Soydal, 2012; Doğan ve Taşkın, 2019; Doğru, 2021; Çuhadar, Ongun ve Topsakal, 2022) bağlamında ele alındığı ve görselleştirmeler için genellikle CiteSpace veya VosViewer gibi yazılımların kullanıldığı tespit edilmiştir.

Literatür incelemesi değerlendirildiğinde; veri görselleştirmenin, kütüphanecilik ve bilgi bilim alanlarında önemli bir araç olarak ele alındığı görülmektedir. Veri görselleştirmenin; koleksiyonun kullanımı ve geliştirilmesi, kütüphane içindeki iklim koşullarının kontrolü, araştırmacılara destek hizmeti, kullanıcı davranışlarının belirlenmesi, bibliyografik kimlikler ve bütçe kontrolü gibi süreçlerde kullanıldığı ve kütüphanecilerin Tableau, Excel gibi hazır yazılımları kullanmayı tercih ettikleri tespit edilmiştir. Kütüphanecilik ve bilgi bilim alanlarında veri görselleştirmenin; büyük ve karmaşık veri kümelerinin anlaşılmasını kolaylaştırmak, bilgiye dayalı karar verme süreçlerini desteklemek ve araştırma sonuçlarının etkili bir şekilde sunulmasını sağlamak amacıyla kullanım potansiyeli bulunmaktadır. Bu teknoloji, karmaşık veri setlerinin daha anlaşılır ve erişilebilir hale getirilmesine yardımcı olarak, kütüphane kaynaklarının yönetimi, kullanıcı etkileşimlerinin analizi ve bilgi paylaşımı süreçlerinde kritik bir rol oynamaktadır. Kütüphaneler, koleksiyon yönetimi, kullanıcı davranış analizi ve kaynak tahsisi gibi konularda veri görselleştirmeyi kullanarak daha verimli operasyonel stratejiler geliştirebilirler.

2.3 Veri Görselleştirme Yöntem ve Araçları

Günümüz veri dünyasında, büyük miktarda veriyi analiz etmek ve veriye dayalı kararlar vermek için veri görselleştirme yöntem ve araçları artık vazgeçilmez teknolojiler haline gelmiştir. Bu teknolojiler, verilerdeki eğilimlerin, aykırı değerlerin ve örüntülerin tespit edilmesine ve anlaşılmasına olanak tanır; bu da değerli içgörülerin elde edilmesi ve bilinçli iş kararlarının alınması açısından önemlidir.

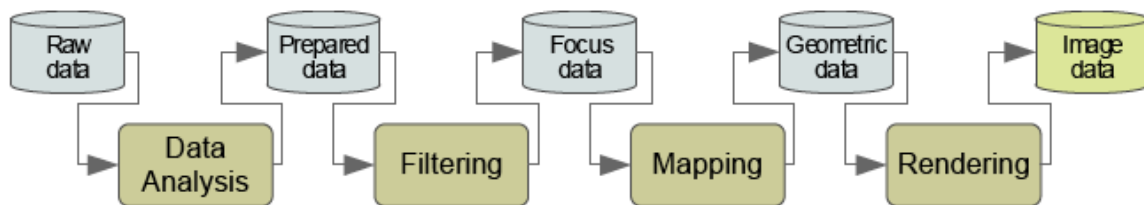
Kuruluşlar bilgi sistemleri ile çalışarak veriyi öz varlıkları olarak değerlendirmekte ve etkileşimli web sayfaları aracılığı ile veri sağlayabilmektedirler. Örneğin üniversite kütüphaneleri özelinde değerlendirilecek olursa, kütüphaneler, çeşitli amaçlar için; hizmet sunma, derme geliştirme gibi konularda kullanabilecekleri verilere sahiptirler. Veri görselleştirme, hem kullanıcılarına kendilerini daha iyi tanıtmalarına yardımcı olacak hem de politika geliştirmek için kullanılabilecek bir yöntem olma potansiyeline sahiptir.

Bu bölümde veri görselleştirme yöntem ve araçları detaylı olarak değerlendirilmiştir.

2.3.1 Veri Görselleştirme Yöntemleri

Web of Science üzerinden yapılan literatür taramasında, veri görselleştirmeyle ilgili yayınlanmış bilimsel makalelerin büyük bir çoğunluğunun bilgisayar bilimleri, mühendislik, biyoloji, matematik ve telekomünikasyon alanlarında (bkz. Şekil 13) ele alındığı ve konu ile ilgili üretilen yayın sayısında ise sürekli bir artışın olduğu görülmektedir (bkz. Şekil 14). Bu alanlarda yapılan çalışmalar, veri görselleştirmenin önemini ve yaygın kullanımını göstermektedir.

Veri görselleştirme, karmaşık veri kümelerinden içgörülerin, örüntülerin ve eğilimlerin saptanmasını kolaylaştırdığı için veri analizinde sıkça kullanılan bir yöntemdir. Veri görselleştirme için daha önce belirlenmiş bir model kullanmak, özellikle büyük veri kaynaklarıyla uğraşırken görselleştirmelerin etkinliğini önemli ölçüde artırabilir. Bir model kullanımı; veri toplamadan ön analize, keşfedici incelemeye, filtrelemeye ve görselleştirmeye kadar bütün sürecin sistematize edilmesini sağlar. Kullanılacak model, veri görselleştirmenin tüm adımlarını açıkça belirtmeli, görselleştirme sürecinin tekrarlanabilir, yorumlanabilir ve içgörü aktarmada etkili olmasını sağlamalıdır. Aşağıda örnek bir veri görselleştirme modeli verilmiştir.



Şekil 18: Veri görselleştirme modeli (Dos Santos ve Brodlie, 2004; Tominski, 2006)

Şekil 18’de sunulan süreç odaklı veri görselleştirme modeli, aynı zamanda süreçte izlenmesi gereken adımları da göstermektedir; adımlar aşağıda açıklanmıştır:

1. Veri Analizi (data analysis): Verilerin genellikle bilgisayar merkezli, çok az veya hiç kullanıcı etkileşimi olmadan görselleştirme için hazırlanmasıdır (örneğin yumuşatma¹⁷

¹⁷ Yumuşatma filtresi: Görüntü yumuşatma, bilgisayarlı görüntüleme uygulamalarının temel parçalarından biridir. Görüntüdeki istenmeyen detayların kaldırılması, bir diğer deyişle gürültüyü azaltmak için kullanılan bir yöntemdir (Asan, 2017).

işlemi uygulayarak, eksik değerleri interpolasyon¹⁸ yaparak veya hatalı ölçümleri düzelterek ham verinin hazırlanması)

2. Filtreleme (filtreleme): Görselleştirilecek veri bölümlerinin seçilmesidir. Genellikle kullanıcı merkezli bir aşamadır.
3. Eşleştirme (mapping): Odak verilerin basit geometrik sembollerle (örneğin noktalar, çizgiler) ve bunların nitelikleri ile (örneğin renk, konum, boyut) eşleştirilmesidir; etkileycilik ve etkililiğe ulaşmak için en kritik adımdır.
4. Oluşturma-görselleştirme (rendering): Geometrik verilerin görüntü verilerine dönüştürülmesidir.

Bir görselleştirme projesinde verilerle çalışırken ilk adım, verilerin sağlanmasıdır. Kirk (2019, ss.96-97), veri sağlama sürecinde kullanılacak çeşitli veri toplama tekniklerini aşağıdaki gibi ele almıştır;

- **Hazır veri:** Veri elde etmenin en basit yöntemidir. Bir başka proje için oluşturulmuş ve hazır halde ulaşılabilen veri setleridir.
- **Sistem(ler)den indirme (downloading):** Kurumlarda, kurumsal raporlama araçlarından veya sistemlerinden dışa aktarılan ve kullanıcıların indirme yoluyla yararlandığı verilerdir. Kurumlar, web üzerinden, ilgili kullanıcıların bazen ayrıntılı sorgular oluşturmalarına ve ihtiyaçlarına göre özelleştirilmiş yapılandırılmış veri indirmesine izin veren arayüzler sunabilmektedir.
- **Web kazıma (Web scraping):** Web kazıma, web sayfalarında yayınlanan yapılandırılmış ve yapılandırılmamış veri öğelerinin özel programların veya araçların kullanılması ile toplanmasını ifade eder. Bu program ve araçlar ile ilgili web sitesinin birden çok sayfasından veri çıkarmak için genellikle rutinler (tekrar eden HTML etiketleri veya etiketlerde yer alan class veya ID isimleri) harekete geçirilir. Bu, web taraması olarak da bilinir. Herhangi bir web kazıma veya tarama faaliyetinde, uygulanacak olan işlemin web sitesinin erişim kurallarına yasal olarak uygun olması gerekmektedir.
- **Uygulama programlama arayüzü (Application programming interface-API):** Belirli web siteleri, insanların veri akışlarına erişmesini sağlamak için bir API sunar.

¹⁸ İnterpolasyon: İlk kez Uygulamalı Matematik biliminin bir alt kategorisi olan Sayısal Analiz yöntemlerinde tanımlanan ve elde var olan değer noktalarından yola çıkarak bu noktalar arasında, farklı bir yerde ve değeri bilinmeyen bir noktadaki olası değeri bulmaya/tahmin etmeye yarayan yöntemlerin tümüne verilen genel isimdir (İnterpolasyon, 2024).

Örneğin, meteoroloji veya borsa bilgileri, belirli bölgelerdeki trafik yoğunluk durumları veya güncel haber içerikleri API'ler aracılığıyla eşzamanlı olarak paylaşılabilir.

- **Birincil kaynaktan veri sağlama (Primary collection):** İhtiyaç duyulan verilerin dijital ortamda mevcut olmaması durumunda bu yöntem kullanılır. Bu yöntem şunları içerebilir;
 - Duygu ve/veya retorik örüntülerini keşfetmek için yapılandırılmamış video veya ses kayıtlarından içerik sağlamak.
 - Bir araştırma kapsamında veri toplamak için bir katılımcı anketi uygulamak.
 - Belirli bir süre boyunca sağlık bilgilerini izlemek için ölçüm cihazları kullanmak.
- **Veri arama (Data foraging):** İhtiyaç duyulan verilerin uygun tek bir formda veya yerde mevcut olmaması durumunda farklı ortamlarda farklı teknikler ile veri arama işleminin yapılması gerekir. Bu genellikle, nispeten küçük miktarlarda farklı veya dağınık veri değerlerinin manuel olarak tedarik edilmesini içerir.

Derlenen veriler, analiz ve görselleştirme yapılacak sistem için uygun bir formata dönüştürülerek sisteme aktarılır (Excel, CSV, SQL, XML vb.). Budiu (2019) “arama” (foraging) kavramını “bilgi arama” (information foraging) kapsamında ele aldığı çalışmada, bu kavramın, hayvanların yiyecek için nasıl yem aradığına dair hayvan davranış teorilerinden ilham alınarak belirlendiğini (Pirolli ve Card, 1999; Budiu, 2019) ifade etmektedir. Hayvanların yiyecek araması ile insanların bilgi araması sürecindeki davranışlarının ortak terminoloji ile açıklanabileceğini (Budiu, 2019) belirten görsel Şekil 19’da verilmiştir.

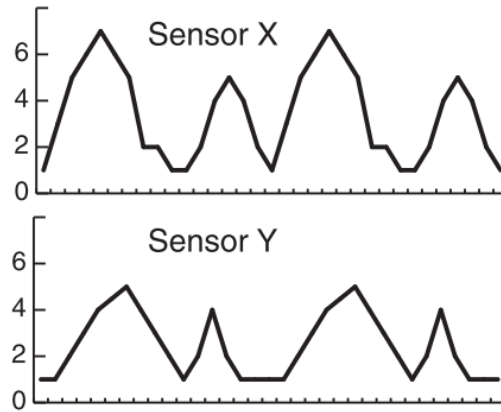
Hayvan (yem arama)		İnsan (bilgi arama)	
 Yiyecek	Amaç	Bilgi 	
 Bir veya daha fazla potansiyel gıda kaynağı içeren bir alan	Arama yapılacak alan	Bir web sitesi (veya başka bir bilgi kaynağı) 	
 Yiyecek ara	Arama	Bilgi arayın 	
 Hayvanın belirli bir alanın yiyecek sağlama ihtimalinin ne kadar olduğuna ilişkin değerlendirmesi	Sezme	Potansiyel bir bilgi kaynağının kullanıcıya ne kadar umut verici görüldüğü 	
 Bir hayvanın açlığını gidermek için göz önünde bulundurabileceği yiyecek türlerinin toplamı	Beslenme	Bir kullanıcının bilgi ihtiyacını karşılamak için dikkate alabileceği bilgi kaynaklarının toplamı 	

Şekil 19: Hayvanların yiyecek araması ile insanların bilgi arama davranışları (Budiu, 2019)

Veri analizi genellikle, farklı veri tiplerine sahip, bütünlükli işlenebilir nesnelerin karşılaştırılmasını içerir. Verilerin sürekli artan miktarları ve karmaşıklığıyla birlikte, bu karşılaştırmalara yardımcı olacak sistemlere olan talep de artmaktadır. Bilgi görselleştirme araçları, her bir veri nesnesinin ayrı ayrı incelenmesine olanak sağladığı gibi, karşılaştırmalar yapılmasını da desteklemektedir. Bilim insanları, analistler giderek daha büyük ve karmaşık veri kümeleriyle çalışmaktadırlar. Görselleştirme araçları, verileri anlamak, analiz etmek ve iletmek için bir gereklilik haline gelmiştir. Oluşturulacak görselleştirmelerin işlevselliği ve anlaşılabilirliği, üzerinde durulması gereken konulardır. Örneğin; farklı veri tiplerine sahip verilerle karşılaştırmalar yapmak genellikle yoğun hafıza aktiviteleri gerektirir (Gleicher, Albers, Walker, Jusufi, Hansen ve Roberts, J, 2011, s.289). İyi tasarlanmış görselleştirmeler, bu aktiviteleri azaltmaya yardımcı olur ve karşılaştırma gerektiren verilerden kısa sürede öngörüler elde edilmesini sağlar (Gleicher ve diğerleri, 2011, s.291).

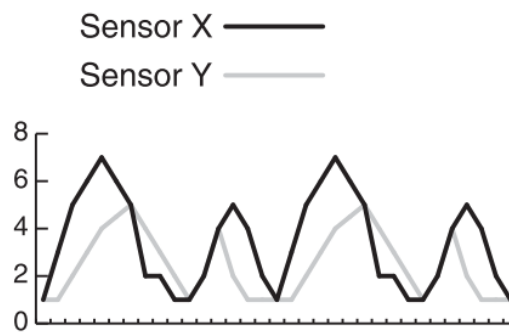
Gleicher ve diğerleri (2011, s.291), karşılaştırmaya açıkça yardımcı olan görsel tasarımları üç kategoriye ayırmaktadırlar:

- **Yan yana / veya ayırma (Juxtaposition) tasarımları**, her bir nesneyi ayrı ayrı (yani zaman veya mekânda yan yana) sunar. Yan yana tasarımlar, karşılaştırılacak nesneleri ayrı ayrı gösterir. Bu ayrılık zaman veya mekânda gerçekleşebilir. Bir yan yana tasarımdaki anahtar unsur, tek tek nesnelerin bağımsız olarak gösterilmesidir.



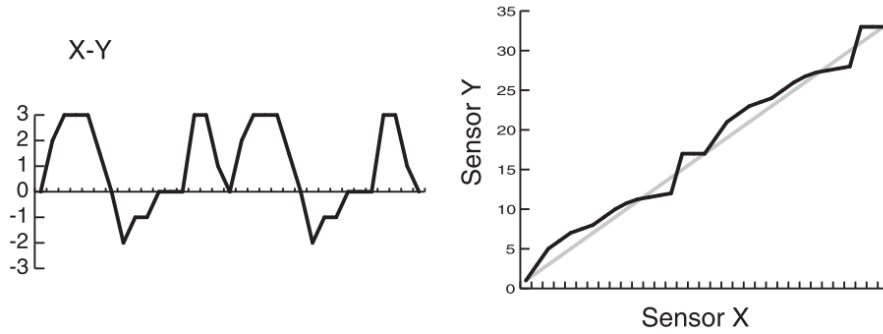
Şekil 20: Yan yana tasarımların karşılaştırılması (Gleicher ve diğerleri, 2011, s.291)

- **Üst üste bindirme / veya bindirme (Superposition) tasarımları**, aynı koordinat sisteminde (birbirinin üstünde) birden fazla nesne sunar. Bindirme tasarımları, karşılaştırılacak nesneleri aynı uzamda gösterir. Bu tür tasarımlar, genellikle bir nesnenin diğerinin üzerine bindirilmesini içerdiğinden, örtü (overlay) tasarımları olarak adlandırılabilir. Bu, bir nesneyi yarı saydam yapmak, hatta bir nesnenin diğerini kısmen gizlemesine izin vermek şeklinde olabilir.



Şekil 21: Üst üste bindirme tasarımların karşılaştırılması (Gleicher ve diğerleri, 2011, s.291)

- **Açık temsil / ilişkilerin açık kodlaması (Explicit encoding)**, üçüncü bir tasarım kategorisi, nesneler arasındaki bağlantıları görsel olarak doğrudan kodlayan ilişkilerin açık temsidir. Bu kategorilerin basit bir örneği Şekil 11'de gösterilmektedir. Açık kodlama kategorisi, nesneler arasındaki ilişkilerin görsel bir kodlama sağlayarak açıkça gösterildiği tasarımları içerir. Böyle bir tasarım, tanımı gereği, nesneler arasındaki ilişkilerin bilinmesini gerektirir. Bu, hangi ilişkilerin ilgi çekici olabileceğine dair önceden varsayılan bir anlayış ve bunları açıkça hesaplamak için bir mekanizma gerektirir. Buna karşılık, diğer temel tasarımların nesneler arasındaki ilişkileri bilmesine gerek yoktur: nesneler doğrudan gösterilir ve kullanıcı, nesneler arasındaki ilişkileri bulur.



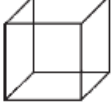
Şekil 22: Açık temsil tasarımı şeklinde iki zaman serisinin karşılaştırılması (Gleicher ve diğerleri, 2011, s. 291)

Bu üç kategori, nesneler arasında bağlantı kurmak için kullanılan temel mekanizma ile ayırt edilebilir: yan yana yerleştirme, izleyicinin hafızasını kullanır; üst üste yerleştirme görsel sistemi kullanır ve açık temsil, ilişkileri belirlemek için hesaplamayı kullanır (Gleicher ve diğerleri, 2011, s.292).

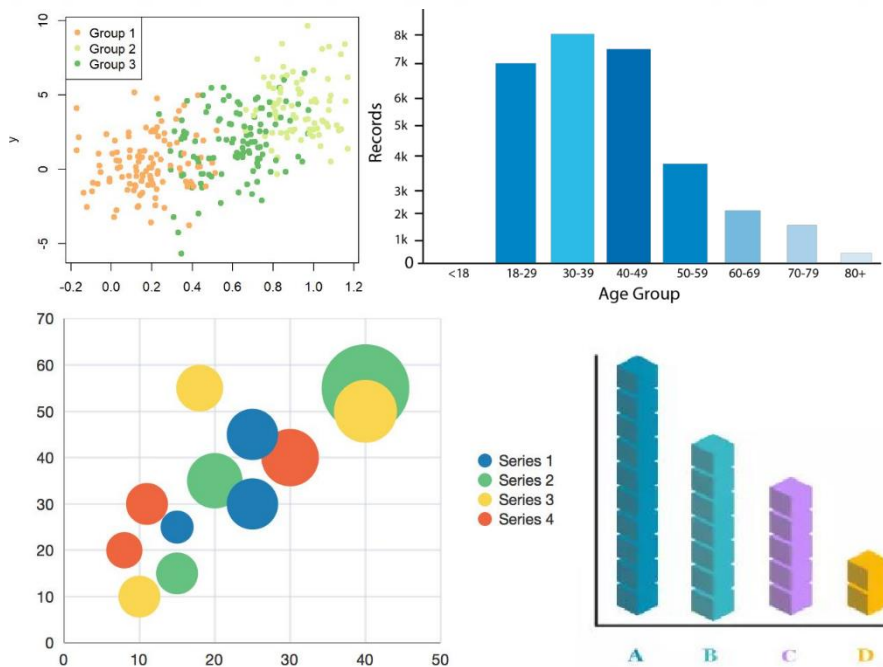
2.3.1.1 Veri Görselleştirmede Kodlama

Verileri görsel olarak temsil etmek, görsel kodlama (encoding) eylemini gerektirir. Görselleştirmede veriler iki ana görsel özellik olan; işaretler ve nitelikler kullanılarak kodlanmaktadır. İşaretler, tablolanmanın biçimine bağlı olarak farklı kayıtlar veya ayrı gruplamalar gibi veri öğelerini temsil eden görsel yer tutuculardır (Kirk, 2019, s.135). Bunlar, Tablo 1'de gösterildiği gibi dört ana işaret türü olarak verilmiştir.

Tablo 1: İşaret kodlama sınıflandırması (Kirk, 2019, s.136)

İşaretler (Marks)	Örnek	Tanım
Nokta (point)		Nokta işareti yaygın olarak, bir ölçekte konum aracılığıyla nicel değerleri temsil etmek için bir işaretleyici olarak kullanılır. (Örneğin dağılım grafiğinde (scatter plot) kullanılır.)
Çizgi (line)		Çizgi işareti, boyut (uzunluk) varyasyonu yoluyla nicel değerleri temsil etmek için yaygın olarak kullanılır. (Örneğin çubuk grafiğinde (bar chart) kullanılır.)
Şekil (shape)		Şekil işareti, yaygın olarak, boyut ve konumdaki varyasyon yoluyla nicel değerleri temsil etmek için kullanılır. (Örneğin kabarcık grafiğinde (bubble plot) kullanılır.)
Form		Form işareti, boyut (hacim) varyasyonu yoluyla nicel değerleri temsil etmek için kullanılan 3d sunumlardır.

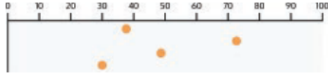
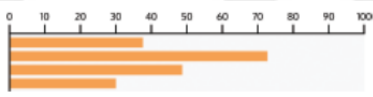
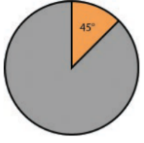
Şekil 23'te, işaretlerin (nokta, çizgi, şekil ve form) kullanıldığı grafik örnekleri verilmiştir.



Şekil 23: İşaretlerin (marks) kullanıldığı örnek görselleştirmeler

Nitelikler, her bir veri ögesiyle ilişkili değerleri temsil etmek için işaretlerin görsel görünümündeki varyasyonlardır (Kirk, 2019, s.135). Ana nitelikler, Tablo 2’de verilenleri kapsamaktadır.

Tablo 2: Nitelik kodlama sınıflandırması (Kirk, 2019, s.136-137)

Nitelikler (attributes)	Örnek	Tanım
Konum		Bir ölçek boyunca konumdaki değişiklik, genellikle bir nokta işareti kullanılarak nicel bir değeri belirtmek için kullanılır.
Boyut: uzunluk		Boyut (uzunluk) varyasyonu, daha büyük boyutların daha büyük miktarlar anlamına geldiği orantısız ölçeklere dayalı nicel değerleri temsil etmek için kullanılır. Çizgi işaretinin tek bir "doğrusal" uzamsal boyutu vardır, yani miktarları yükseklik veya genişlik yoluyla gösterir, ancak ikisini birden göstermez.
Boyut: alan		Boyut (alan) varyasyonu, daha büyük boyutların daha büyük miktarlar anlamına geldiği orantısız ölçeklere dayalı nicel değerleri temsil etmek için kullanılır. Şekil işaretinin iki ('kuadratik') uzamsal boyutu vardır, yani hem yükseklik hem de genişlik kombinasyonu yoluyla miktarları gösterir.
Boyut: hacim		Boyut (hacim) varyasyonu, daha büyük boyutların daha büyük miktarlar anlamına geldiği orantısız ölçeklere dayalı nicel değerleri temsil etmek için kullanılır. Form işaretinin üç ('kübik') uzamsal boyutu vardır, yani miktarları yükseklik, genişlik ve derinlik kombinasyonu yoluyla gösterir.
Açı		Bir açının boyutundaki değişiklik, daha büyük açıların daha büyük miktarlar veya daha spesifik olarak bir bütünün daha büyük parçaları anlamına geldiği nicel değerleri temsil etmek için kullanılır.

Nicelik		Bir dizi nokta işaretinin (semboller gibi) miktarındaki değişiklik, tek veya toplu bir nicel değeri temsil etmek için kullanılabilir.
Renk: tonu (hue)		Renk tonundaki değişiklik, tipik olarak kategorik veri değerlerini ayırt etmek için kullanılır.
Renk: doygunluk (saturation)		Renk doygunluğundaki varyasyon, sıralı ölçekleri temsil etmek için genellikle diğer renk özellikleriyle bağlantılı olarak kullanılabilir; tipik olarak, doygunluk ne kadar büyükse, hiyerarşik vurgu da o kadar büyük olur.
Renk: hafiflik (lightness)		Rengin açıklığındaki değişiklik, nicel ölçekleri temsil etmek için kullanılabilir; tipik olarak, renk ne kadar koyu olursa, miktar o kadar yüksek olur.
Örüntü (pattern)		Örüntüdeki varyasyon (bazen örüntü dokusu veya yoğunluğu olarak da tanımlanır), sıralı ölçekleri temsil etmek veya kategorik değerleri ayırt etmek için kullanılabilir.
Simge (symbol)		Sembollerdeki varyasyon, kategorik veri değerlerini ayırt etmek için yaygın olarak kullanılır. Bu özelliğin kapsamı, veri değerlerini açıkça temsil eden resimlere ve çizimlere kadar uzanabilir.
Bağlantı (connection)		Bağlantı (kenar olarak da bilinir), bir bağlantı hattı tarafından kurulan iki düğüm arasındaki ilişkiyi belirtir. Bağlantının şekli ve boyutu genellikle anlamsızdır, ancak bazen ilişkideki bir yön kavramını kodlamak için oklar veya çizgi kalınlığındaki değişiklikler kullanılabilir.
Kapsama (containment)		Kapsama, ilgili bir "ana" kategori grubuna ait kategoriler arasındaki hiyerarşik bir ilişkiyi kodlamanın bir yoludur.

Kirk (2019, s.138), işaretler ve nitelikleri bir yemeğin bileşenleri, görselleştirme türlerini ise bu bileşenlerin oluşturduğu yemek tarifi olarak nitelemektedir. Görselleştirme türlerinin her biri, işaret ve nitelik kombinasyonlarından oluşur ve verileri temsil etmenin farklı yollarını sunar.

2.3.1.2 Veri Görselleştirme Türleri

Veri görselleştirme temsilleri, verileri hızlı bir şekilde anlamamıza yardımcı olur. Veri görselleştirmenin temel işlevleri, verileri görüntülemek ve bir konunun daha fazla araştırılması için kişiyi teşvik etmektir. Veri analizi sürecinde uygun görselleştirme türlerini kullanmak; hem veri noktaları arasındaki önemli ilişkileri veya örüntüleri anlamayı ve sunmayı, hem de yeni fikirler için öngörüler elde etmeyi sağlar. Bunun için en uygun görselleştirme türünü seçmek önemlidir. Günümüz bilgisayar teknolojisi ve veri görselleştirme programları büyük miktarda veri görselleştirme türüne olanak vermektedir. Bu bölümde, veri görselleştirme için kullanılan çeşitli veri görselleştirme türleri ele alınacaktır. Bu alanda kaynak belirtilerek kullanılan görseller, literatür taraması sonucunda elde edilmiş olan ve bilindik veri setleri ile oluşturulan görselleştirmelerdir.

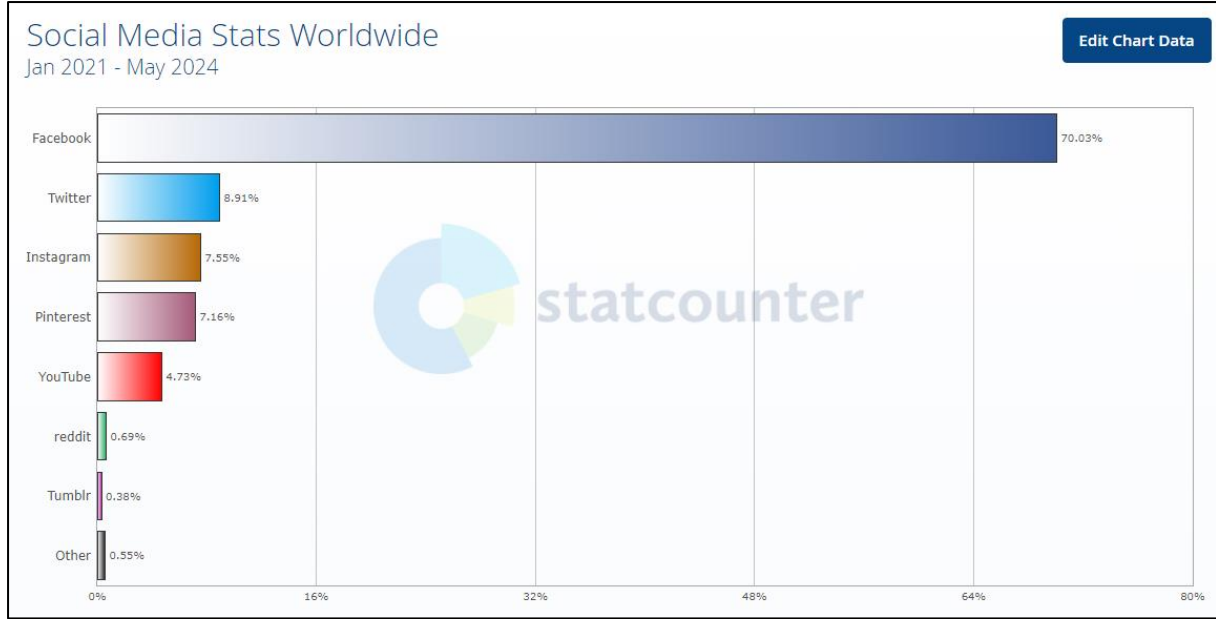
Etiket Bulutu (Word cloud): Etiket bulutu, bir metinsel veri içindeki kelimelerin sıklığını gösterir. Her öge kelimelerle temsil edilir ve daha sonra her birinin yazı tipi boyutu, kullanım sıklığına göre ölçeklenir (Kirk, 2019, s.149). Web 2.0’ın internete kazandırdığı yeniliklerden biridir. Temel amaç, bir sitedeki her bir içeriği etiketleyerek başkalarının bu içeriği bulmasını kolaylaştırmaktır (Etiket bulutu, 2023).



Şekil 24: “Data visualisation: a handbook for data driven design” kitabının ilk bölümünde kullanılan kelimelerin sıklığı (Kirk, 2019, s.149)

Çubuk Grafik (Bar chart):

Çubuk grafikler en yaygın veri görselleştirmelerinden biridir. Bu grafikler; kategoriler arasında verileri hızlı bir şekilde karşılaştırmak, farklılıkları vurgulamak, eğilimleri ve aykırı değerleri göstermek ve bir bakışta en yüksek ve en düşük değerleri ortaya çıkarmak için kullanılabilir. Çubuk grafikler, özellikle birden çok kategoriye ayrılabilen veriler olduğunda etkilidir (Hardin, Hom, Perez ve Williams, 2012, s.2).



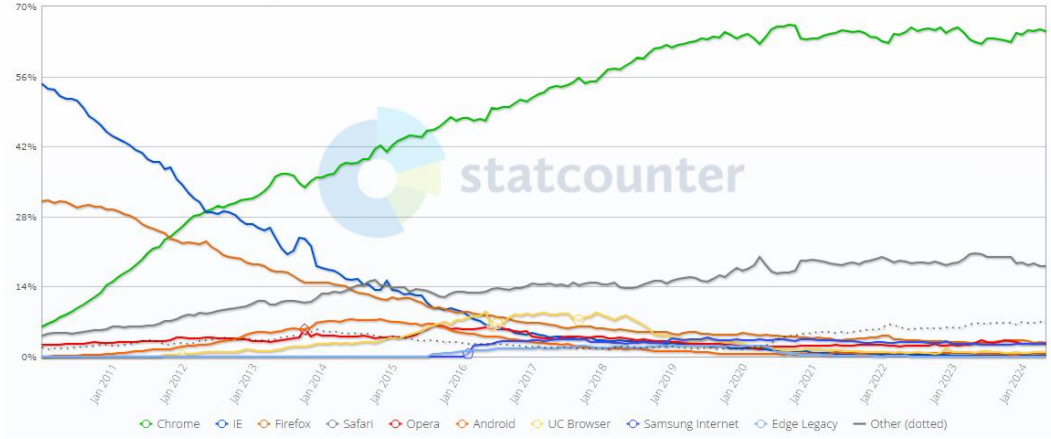
Şekil 25: Sosyal medya platformlarının 2021-2024 arası dünyadaki kullanım oranları
(Statcounter Global Stats, 2024a)

Çizgi Grafik (Line chart, or line graph):

Çizgi grafiği, birkaç farklı veri noktasını birbirine bağlayarak onları tek bir sürekli eğilim olarak sunar. Verilerdeki, genellikle zaman içindeki eğilimleri görüntülemek için genellikle çizgi grafikler kullanılır. Bir değerdeki değişikliklerin diğer değerlere göre görselleştirilmesini sağlar (Hardin ve diğerleri, 2012, s.3).

Browser Market Share Worldwide
Jan 2010 - May 2024

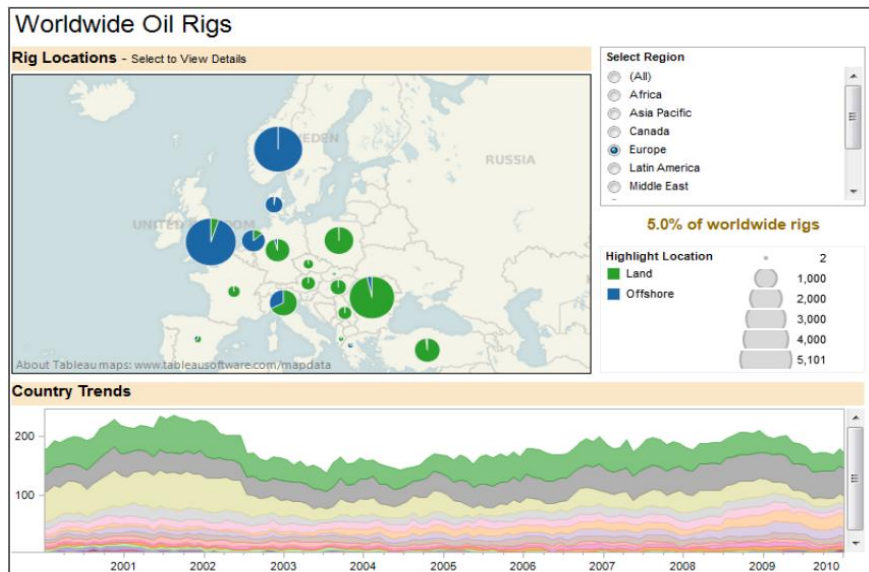
Edit Chart Data



Şekil 26: Dünya çapındaki internet tarayıcı pazar payı (Statcounter Global Stats, 2024b)

Pasta Grafik (Pie chart):

Pasta grafikler, genellikle diğer görselleştirmelere ayrıntı eklemek için kullanılabilir. Tek başına bir pasta grafik, izleyiciye bilgileri hızlı ve doğru bir şekilde karşılaştırması için etkili olmayabilir. İzleyicinin kendi bağlamını oluşturması gerektiğinden, verilerdeki önemli noktaları gözden kaırması mümkündür. Bir pasta grafiğini, oluşturulan görselleştirmenin odak noktası yapmak yerine, görselleştirmelerin detayına inmek için kullanılabilir, ek bilgi olarak kullanılması daha yararlı olur (Rodger, t.y.). Pasta grafikler, izleyicilere orantılı verileri anlamaları için hızlı bir yol sağlar. Şekil 27'de harita görselinin detaylarını göstermek ve ek bilgi olarak kullanılan bir pasta grafik örneği verilmiştir.

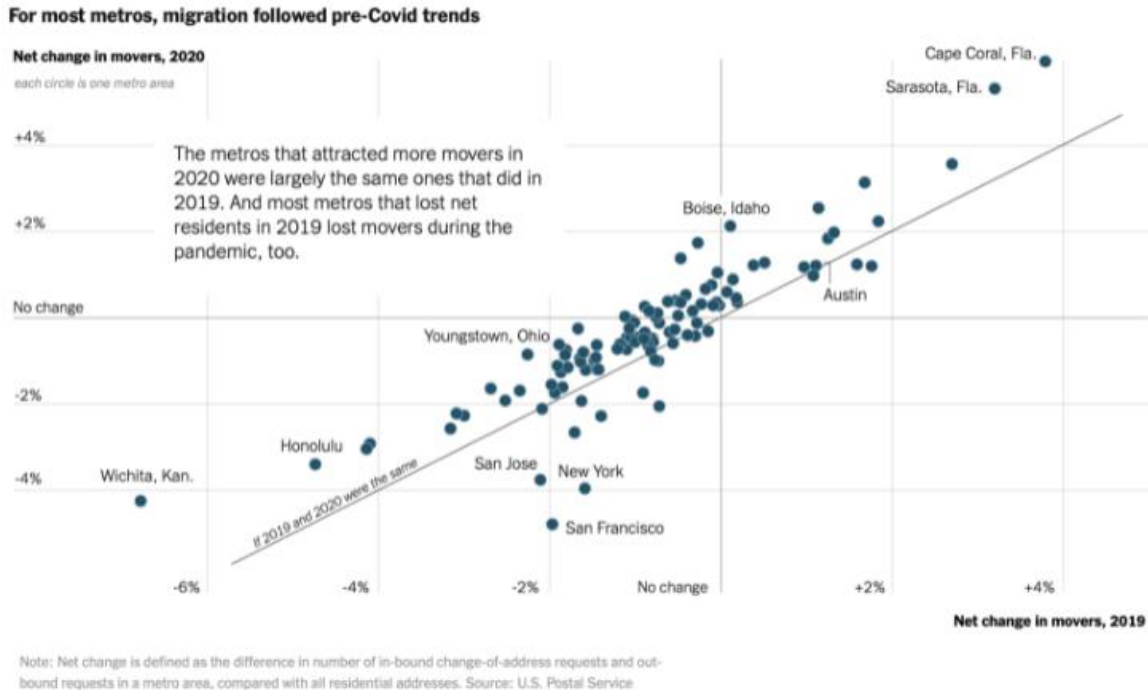


Şekil 27: Petrol kulelerinin dünya çapındaki dağılımları (Hardin ve diğerleri, 2012, s.6)

Dağılım Grafiği (Scatter plot):

Dağılım grafiği, her biri bir dizi veriyi temsil eden iki eksenle oluşur (Data visualization, 2021). İki eksenle konumlanmış noktaları kullanarak iki sayısal değişken üzerindeki değerleri görüntüler. Dağılım grafikleri, çizilen değişkenler arasındaki ilişkinin korelasyonunun güçlü veya zayıf, pozitif veya negatif, doğrusal veya değil olması gibi- çok yönlü bir gösterimdir. Dağılım grafikleri, verilerdeki aykırı noktaları ve olası boşlukları belirlemek için de başvurulmuş görselleştirme türüdür (Yi ve Sapountzis, 2024). Şekil 28'de, COVID-19 pandemi kısıtlamaları nedeniyle, daha fazla açık alan için şehirlerden ayrılan insanların oranını tespit etmek amacıyla “The New York Times” gazetesinin The Upshot eki için “Amerika Birleşik Devletleri Posta Servisi-USPS” verileri kullanılarak oluşturulan dağılım grafiği örnek olarak verilmiştir. Yazarlar, pandeminin kısıtlamaları nedeniyle alışılmadık derecede büyük bir insan dalgasının daha fazla açık alan için şehirleri terk edeceğinin düşünülebileceğini, ancak “Posta Servisi” adres değişikliği verilerinin, durumun pek de böyle olmadığını gösterdiğini belirtmişlerdir (Kolko, Badger ve Bui, 2021).

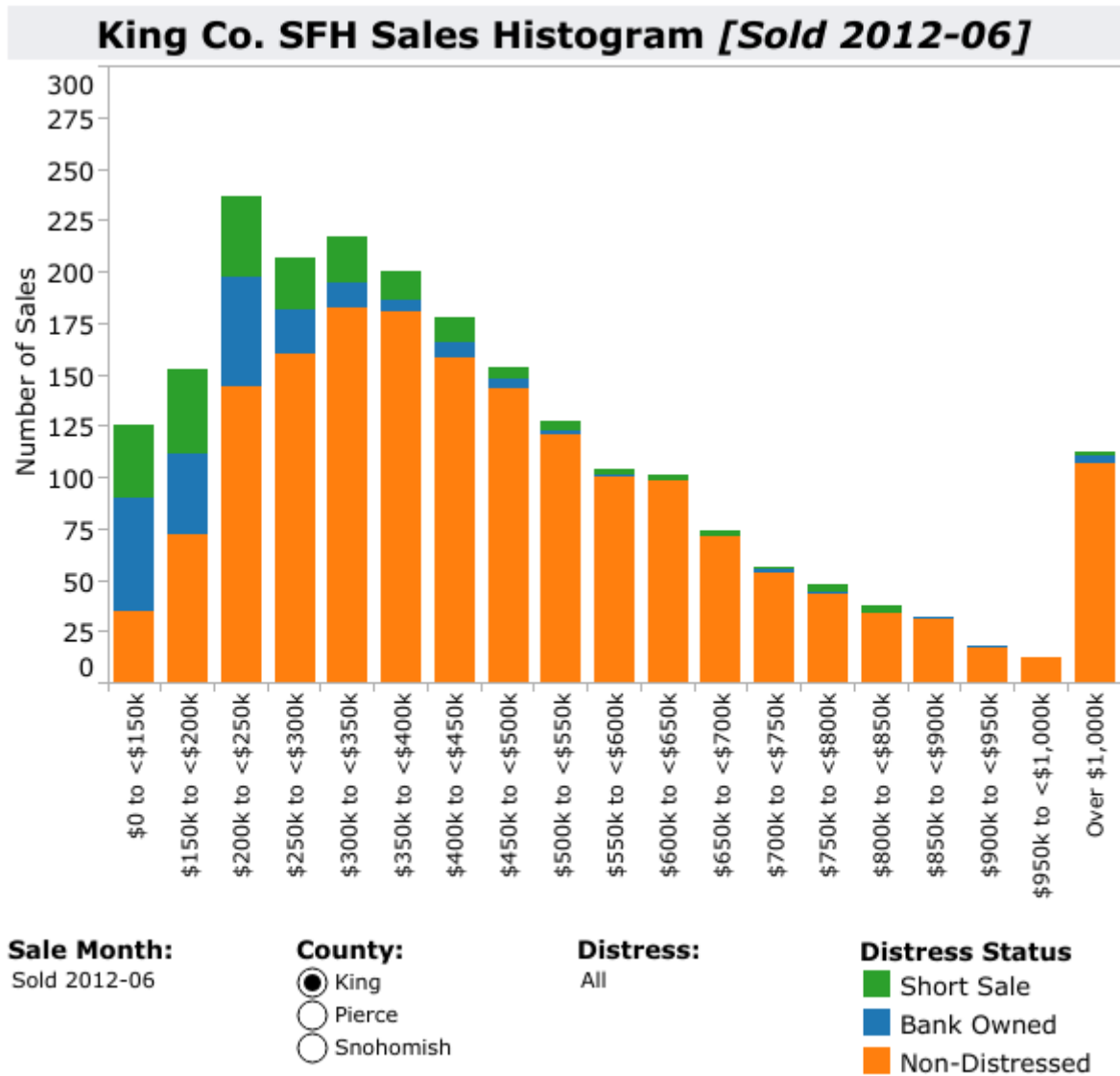
Pandemic migrations



Şekil 28: COVID-19 pandemisi göçü (Flowingdata, 2021)

Histogram:

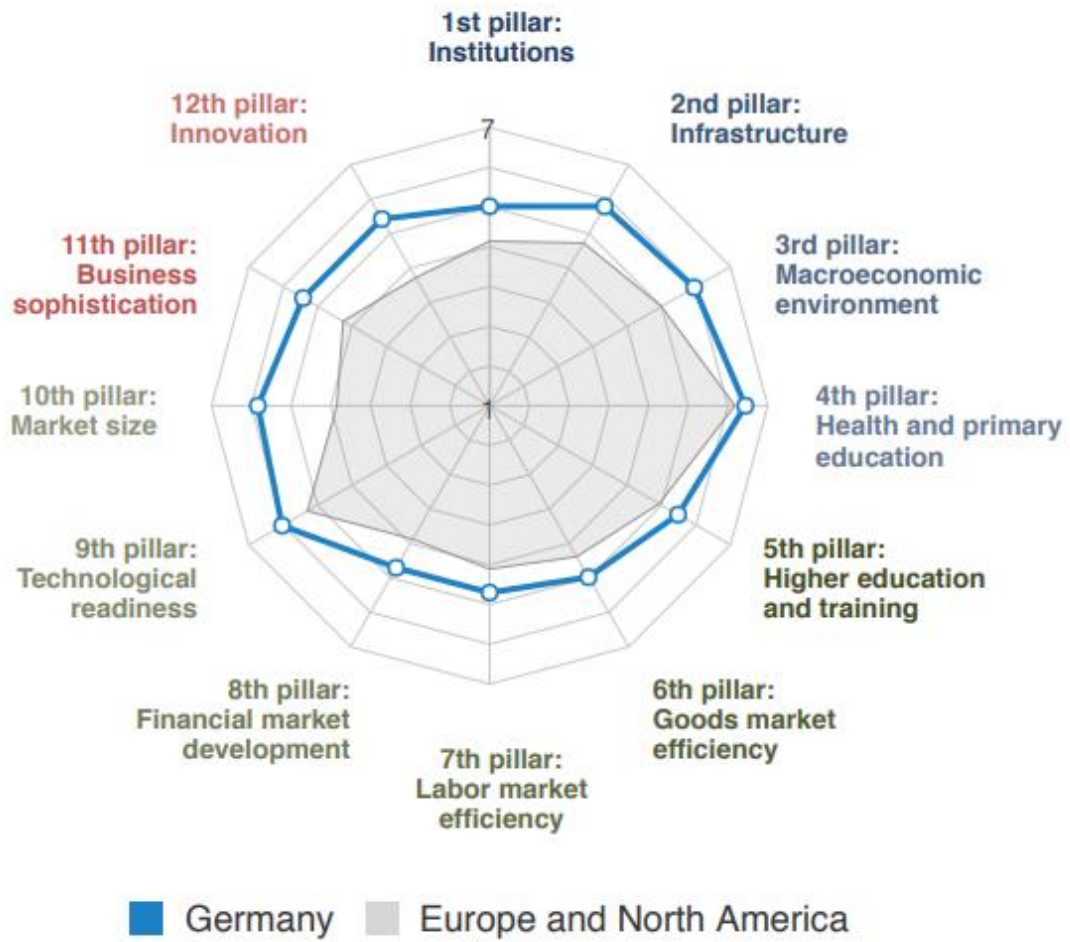
Histogram, sayısal verileri ve frekans dağılımını sunan özel bir dikey çubuk grafik türüdür. Histogramdaki çubuk uzunlukları tipik olarak veri noktalarının sayısına karşılık gelir ve bunların örüntüleri, verilerdeki değişkenlerin dağılımını gösterir (Yi ve Sapountzis, 2024). Verilerin gruplar arasında nasıl dağıtıldığı incelenmek istendiğinde histogram kullanılabilir (bkz. Şekil 29).



Şekil 29: ABD- King County'deki müstakıl ev satışlarının düştüğünü gösteren histogram örneği (Seattle Bubble, 2010)

Radar Grafik (Radar chart):

Radar Grafikler, çoklu nicel değişkenleri karşılaştırmanın bir yoludur. Hangi değişkenlerin benzer değerlere sahip olduğunu veya her değişken arasında herhangi bir aykırı değer olup olmadığını görmek için ideal bir görsellik sağlar. Radar Grafikleri, bir veri kümesinde hangi değişkenlerin yüksek veya düşük puan aldığını görmek için de kullanışlıdır. Şekil 30'da, Dünya Ekonomi Forumu tarafından hazırlanan “2017-2018 Küresel Rekabet Edebilirlik Raporu” nda yer alan, on iki temel gösterge kullanılarak, Almanya'nın Avrupa ve Kuzey Amerika'ya karşı küresel rekabet edebilirlik puanlarının karşılaştırması verilmiştir.



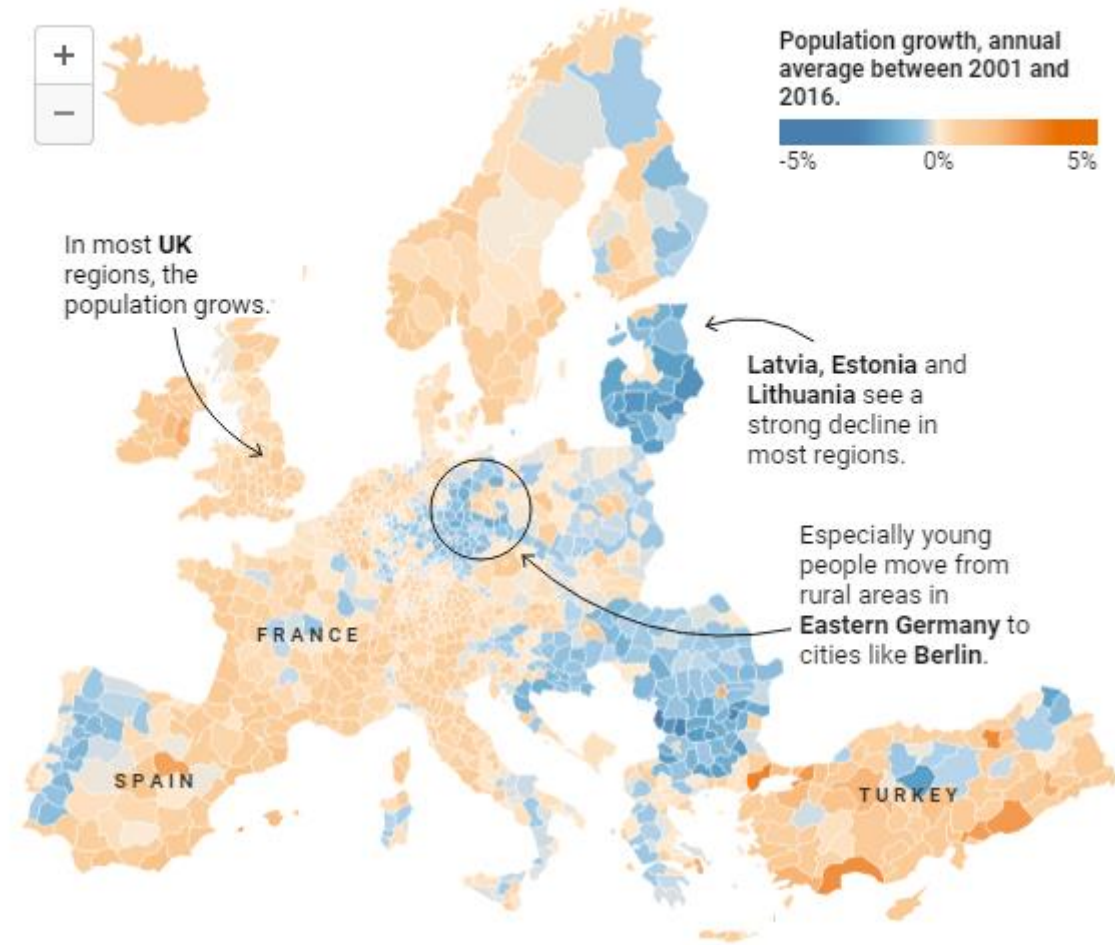
Şekil 30: Almanya'nın Avrupa ve Kuzey Amerika'ya karşı küresel rekabet edebilirlik puanlarının karşılaştırması (World Economic Forum, 2017)

Haritalar (Maps):

Haritalar; posta kodları, eyalet veya şehir kısaltmaları, ülke adları veya belli bir disiplini için kullanılan özel coğrafi kodlamaları konum bilgileri referans alınarak yapılan bir

görselleştirme türüdür. Analiz edilen verilerde, ilişkili coğrafi bilgiler mevcut olduğu takdirde, haritalar, konumun verilerdeki eğilimlerle nasıl ilişkili olduğunu kolayca gösterebilir. Şekil 31’de, 2001-2016 yılları arasında Avrupa nüfusunun arttığı ve azaldığı yerler harita üzerinde gösterilmiştir.

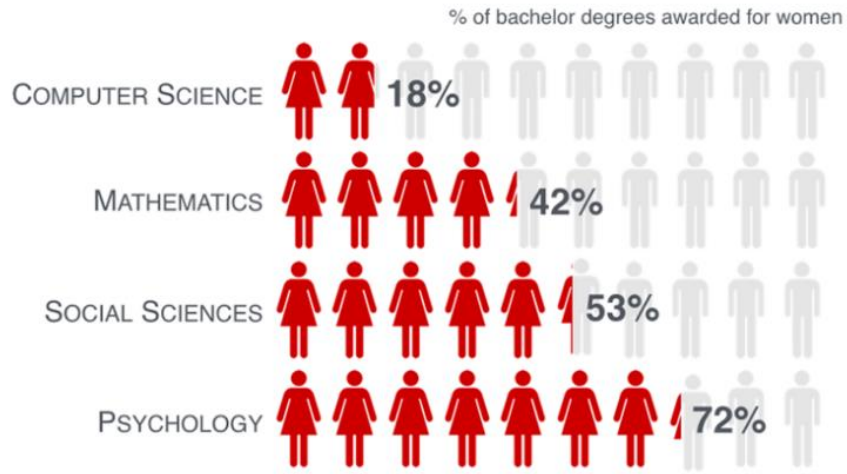
Where the population of Europe is growing, and where it's declining



Şekil 31: Avrupa nüfusunun arttığı ve azaldığı yerler (Datawrapper, t.y.)

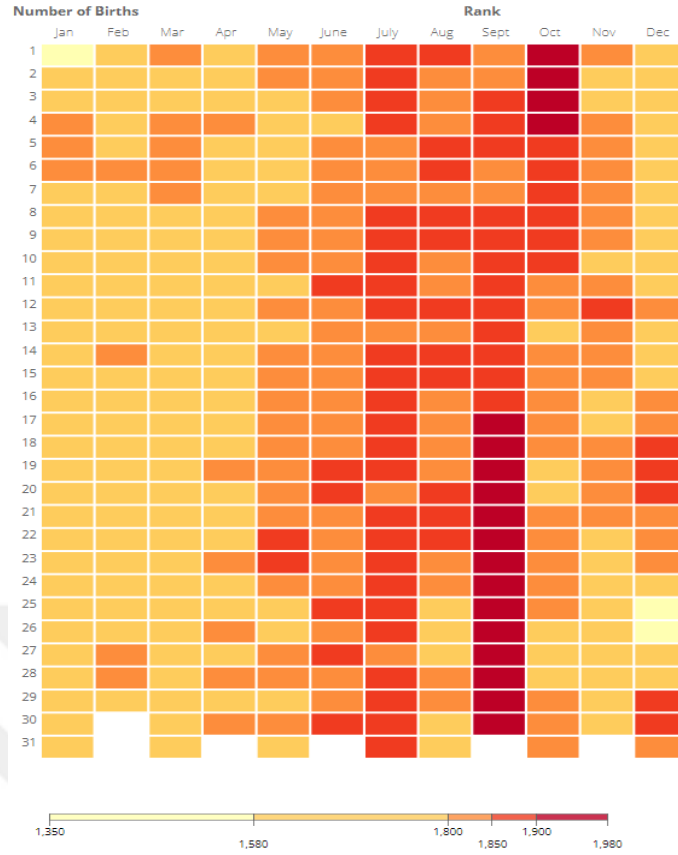
Piktogramlar (Pictograms): Piktogram grafikleri, küçük ayırık veri kümelerinin daha ilgi çekici bir genel görünümünü vermek için simgeleri kullanır. Tipik olarak simgeler, verilerin konusunu veya kategorisini temsil eder; örneğin, nüfusla ilgili veriler, insan simgesini kullanır. Her simge bir birimi veya herhangi bir sayıda birimi temsil edebilir (her bir simgenin 10 birimi temsil etmesi gibi). Piktogram grafiklerde kullanılan işaretlerin en temel özelliği, evrensel

nitelik taşıyıcılarıdır (The Data Visualisation Catalogue, t.y.). Piktogram grafik örneği Şekil 32’de verilmiştir.



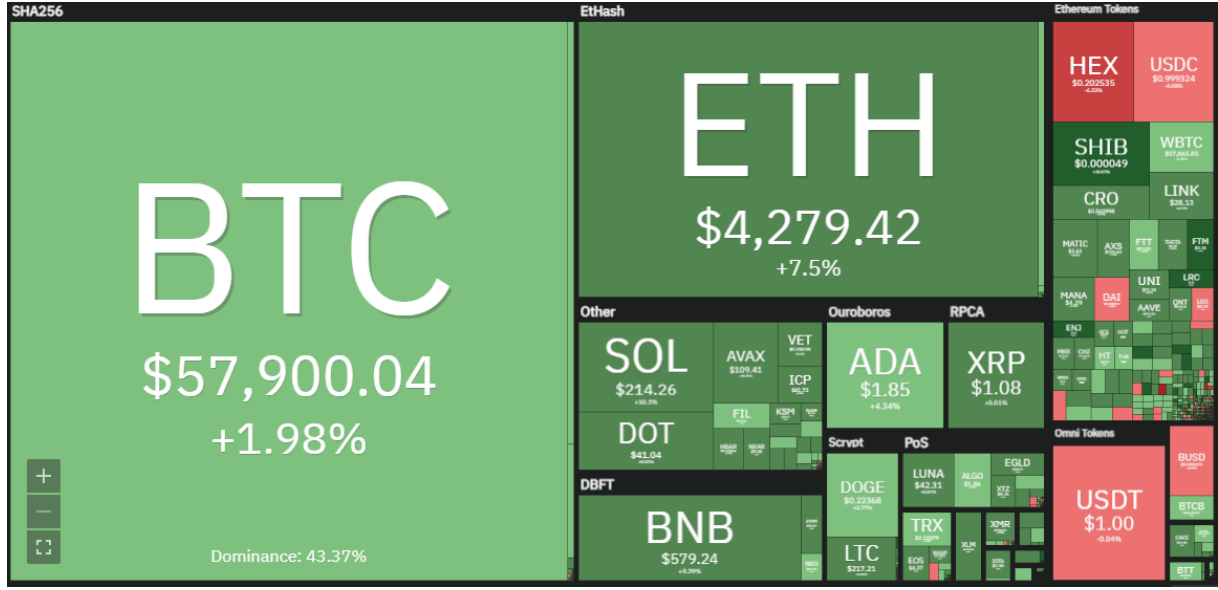
Şekil 32: ABD’de kadınların bölümlere göre üniversite mezuniyet oranları (Conner-Simons, 2015)

Isı Haritası (Heat Map): Isı haritası, “bir veri kümesi içindeki bireysel değerlerin büyüklüğünü renk olarak temsil eden 2 boyutlu bir veri görselleştirme tekniğidir. Renkteki değişiklik renk tonuna veya yoğunluğa göre olabilir. Suç analitiği veya web sitesi tıklama takibi gibi bazı uygulamalarda renk, her noktayla ilişkili bir değer yerine veri noktalarının yoğunluğunu temsil etmek için kullanılır” (Heat Map, 2024). Şekil 33’te 1995-2014 yılları arası İngiltere ve Galler’deki ortalama günlük doğum oranlarını gösteren ısı haritası örneği verilmiştir. Görseldedeki İngiltere ve Galler’e ilişkin doğum verileri, son yirmi yılda en çok ve en az popüler olan doğum tarihlerindeki eğilimleri görmek için kullanılmıştır. Görsel, en fazla doğumun eylül ayı sonunda gerçekleştiğini ortaya koymaktadır.



Şekil 33: 1995-2014 yılları arası İngiltere ve Galler'deki ortalama günlük doğum oranları
(Office for National Statistics, 2015)

Ağaç Haritası (Treemap): Ağaç haritası, farklı kurucu kategorik parçalar için niceliksel değerlerin bir bütünü nasıl oluşturduğunu gösteren hiyerarşik bir görüntü sağlayan bir kapsam diyagramıdır. Toplamın yüzde yüzünü temsil eden, genellikle kareleştirilmiş olarak adlandırılan, kapalı bir dikdörtgen düzen kullanır (Kirk, 2019, s.162). Şekil 34'te <https://coin360.com> web sitesindeki kripto paraların hacmini gösteren örnek bir ağaç haritası görseli verilmiştir.



Şekil 34: 20 Kasım 2021 tarihli kripto para hacimleri ve fiyat durumları (Kaynak: <https://coin360.com/>)

Veri görselleştirme için kullanılan diğer çeşitli görselleştirme türleri aşağıda Tablo 3'te sıralanmıştır.

Tablo 3: Veri görselleştirme için kullanılan çeşitli görselleştirme türleri (Kirk, 2019, ss.140-188)

Clustered Bar Chart	Beeswarm Plot	Venn Diagram	Instance Chart
Bullet Chart	Density Plot	Bubble Plot	Choropleth Map
Waterfall Chart	Box-And-Whisker Plot	Network Diagram	Isarithmic Map
Polar Chart	Waffle Chart	Sankey Diagram	Proportional Symbol Map
Connected Dot	Stacked Bar Chart	Chord Diagram	Prism Map
Plot Pictogram	Diverging Bar Chart	Bump Chart Slope	Also Dot Map
Proportional Symbol Chart	Marimekko Chart	Area Chart	Flow Map
Heat Map Matrix Chart	Sunburst Chart	Stream Graph	Area Cartogram
Dot Plot	Dendrogram	Gantt Chart	Grid Map

Veri görselleştirme sürecinde kritik adımlardan biri uygun görselleştirme yönteminin seçimidir. Bu seçimi yapabilmek için çeşitli görselleştirme tekniklerini bilmenin yanı sıra bunların güçlü yönleri ve sınırlılıkları hakkında önceden bilgi sahibi olmak gerekmektedir. Seçilen yöntem, verileri doğru bir şekilde temsil etmeli ve belirli bir hedef kitleye ve bağlama uygun olmalıdır. Bir görselleştirme yöntemi seçildikten sonra, görselleştirmenin

oluşturulmasında kullanılan kodlama veya algoritmalar için net açıklamalar sağlanması da bir o kadar önemlidir. Ayrıca, hangi tür görselleştirmenin amaçlanan mesajı en iyi şekilde ileteceğine karar vermek gerekir. Farklı görselleştirme türleri, farklı veri ve mesaj türleri için daha uygundur. Örneğin, kategorik verileri karşılaştırmak için çubuk grafikler daha etkili olabilirken, zaman içindeki eğilimleri göstermek için çizgi grafikler daha uygun olabilmektedir. Görselleştirme türünün seçimi, görselleştirmenin hedeflerine ve izleyicinin ihtiyaçlarına göre belirlenmelidir.

İzleyen bölümde veri görselleştirme için kullanılabilecek araçlar (programlar, yazılım dilleri vb.) detaylı olarak incelenmiştir.

2.3.2. Görselleştirme araçları

Günümüz dünyasında, eğitim ve öğretim kaynaklarından sosyal etkileşim için oluşturulan içeriklere, internette bilgi arama davranışlarından tıbbi kayıtlara kadar hemen hemen her şey dijital ortamda üretilmekte ve her gün petabaytlarca veri işlenmektedir. Bu verilerin işlenmesi ve analiz edilmesi sürecinde veri görselleştirme önemli bir yere sahiptir.

Verilere dair istatistikleri tablo veya düz metin formatında sunmaktan ziyade görsel formatta sunmak daha avantajlıdır. İyi tasarlanmış bir veri görselleştirme, örüntü(lerin) ve ilişkilerin tanınmasını kolaylaştırarak bilginin daha doğru ve çekici bir şekilde iletilmesini, karmaşık verilerin daha kolay anlaşılmasını sağlar.

Yazılım üreticileri, işletmelerin karmaşık veri kümelerinden öngörüler elde etmesini sağlayan yenilikçi çözümler sunarak pazarı önemli ölçüde etkilemektedirler. Bu üreticiler, kullanıcı dostu arayüzler ve sezgisel özellikler sunarak veri analizinin demokratikleşmesine katkıda bulunmakta, farklı düzeylerde teknik uzmanlığa sahip kullanıcıların ilgi çekici görselleştirmeler oluşturmalarına olanak tanımaktadırlar. Dahası, üreticiler arasındaki rekabet işlevsellik, performans ve uyumlulukta sürekli gelişmelere yol açarak veri görselleştirme uygulamalarına yönelik pazarın büyümesini ve benimsenmesini teşvik etmektedirler (Linkedin, 2024).

Veri görselleştirme araçlarının ve yazılım pazarının büyümesi, çeşitli sektörlerdeki işletmelerin ürettiği verilerin artan hacmi ve karmaşıklığı da dâhil olmak üzere çeşitli faktörlerden kaynaklanmaktadır. Ek olarak, kuruluşlar arasında veriye dayalı karar vermenin önemi konusunda artan farkındalık, ham verileri eyleme geçirilebilir içgörülere

dönüştürebilecek araçlara olan talebin artmasına yol açmıştır. Ayrıca, büyük veri analitiği, yapay zekâ ve bulut bilişim gibi teknolojideki ilerlemeler, veri görselleştirme araçlarının yeteneklerini genişleterek pazarın büyümesini daha da arttırmıştır (Linkedin, 2024).

Veri görselleştirme için gerek ticari gerekse açık kaynak kodlu olarak kullanılabilir birçok uygulama mevcuttur. Avrupa Birliği Açık Veri Portalı (EU Open Data Portal) "görselleştirmeler" (visualizations) arayüzünde açıklamaları ile verilmiş olan görselleştirme araçları tabloya dönüştürülerek Tablo 4'te verilmiştir. Bu portal üzerinden derlenen veriler Tablo 4'te "uygulama, görselleştirme türleri, işlenen veri, girdi formatı, görsel etkileşim, lisans durumu ve yazılım türü" sütunları ile kategorize edilmiştir.



Tablo 4: Veri görselleştirme araçları (EU Open Data Portal, 2021).

Uygulama	Görselleştirme türleri	İşlenen veri	Girdi formatı (input)	Görsel etkileşim	Lisans durumu	Yazılım türü	Not
Rstudio	Çubuklar, noktalar / baloncuklar, çizgiler, haritalar, ilişkiler, şekiller / oranlar, etiket bulutu	Jeo-uzamsal (geospatial) veriler, ağ verileri, sayısal veriler, metin verileri	Veri dosyaları, veri tabanları	Statik	Açık kaynak (ücretsiz)	Masaüstü ve web uygulaması	RStudio, R programa dili için entegre bir geliştirme ortamıdır.
Tableau	Çubuk grafik, kutu-bıyık grafiği, madde işareti grafiği, nokta grafiği, histogram, piramit, arka arkaya grafik, aralık grafiği, yığın çubuk grafik, şelale grafiği	Jeo-uzamsal veriler, sayısal veriler, metin veriler	Veri dosyaları, veri tabanları	Dinamik	Lisanslı (Tableau Public adında ücretsiz sürümü mevcuttur)	Masaüstü ve web uygulaması	Sürükle ve bırak mantığıyla çalışan arayüzü ile herhangi bir veriyi görselleştirmek, farklı görünümeleri keşfetmek ve hatta birden fazla veri tabanını kolayca birleştirmek için kullanılabilecek bir görselleştirme uygulamasıdır.
D3.js	Çubuklar, noktalar / baloncuklar, çizgiler, haritalar, ilişkiler, şekiller / oranlar	Jeo-uzamsal (geospatial) veriler, ağ verileri, sayısal veriler, metin verileri	Veri dosyaları	Dinamik	Açık kaynak (ücretsiz)	Kod kütüphanesi	JavaScript, HTML ve CSS kullanarak her türlü görselleştirmeyi çizmeye izin veren bir JavaScript kitaplığıdır.
Chart.js	Çubuk grafikler ve çizgi grafikler	Sayısal veri	JavaScript	Dinamik	Ücretsiz	Kod kütüphanesi, framework	Web sitelerine ücretsiz olarak animasyonlu, etkileşimli grafikler eklemeye izin veren bir JavaScript kitaplığıdır.
Qlik Sense	İnfoğrafikler, çubuklar, noktalar / baloncuklar, çizgiler, haritalar, ilişkiler, şekiller / oranlar	Sayısal veri	Veri dosyaları, veri tabanları	Dinamik	Lisanslı	Web uygulaması	Esnek ve etkileşimli görselleştirmeler oluşturmaya olanak tanır. Sürükle bırak işlemleri yapılabilir.
PowerBI (Microsoft)	Çubuklar, noktalar / baloncuklar, çizgiler, haritalar, şekiller / oranlar	Jeo-uzamsal veriler, sayısal veriler ve metin verileri	CSV, XLS, veri tabanları	Dinamik	Lisanslı	Masaüstü, sunucu, web, mobil uygulama	Farklı veri kaynaklarından gelen verileri etkileşimli panolara ve raporlara dönüştüren bir veri görselleştirme ve iş zekâsı aracıdır.
Webtools Maps	Haritalar	Jeo-uzamsal veriler ve sayısal veriler	GeoJSON, JSON, XML	Dinamik	Açık kaynak (ücretsiz)	Web uygulaması	Avrupa Birliği kurumları tarafından üretilen, web sitelerine etkileşimli haritalar entegre etmeye olanak tanıyan bir görselleştirme programıdır.
Google-Visualization	Çubuklar, ızgaralar, ilişkiler, şekiller / oranlar	Ağ verisi	Google e-tablolar, CSV, Google Füzyon tablolar	Dinamik	Ücretsiz	Web uygulaması	Web tarayıcılar ve mobil cihazlar için etkileşimli grafikler oluşturulabilen bir web uygulamasıdır.

Raw	Çubuklar, noktalar / baloncuklar, haritalar, ilişkiler	Sayısal veriler	XLS, Google Dokümanlar, OpenRefine, TextEdit, CSV, TSV, CORS özellikli uç noktalar	Dinamik	Ücretsiz	Kod kütüphanesi, framework	Vektör tabanlı görselleştirmeler oluşturulabilen bir uygulamadır.
ArcGIS Online (Esri)	Haritalar	Jeo-uzamsal veriler	XLS, CSV, TSV, veri tabanları	Dinamik	Lisanslı	Web uygulaması	Haritalar oluşturmaya ve paylaşmaya izin veren bulut tabanlı bir harita platformudur.
Microsoft Excel	Çubuklar, noktalar/ kabarcıklar, çizgiler, ilişkiler/bağımlılıklar	Sayısal veriler	CSV, TSV, TXT, Access veri tabanı	Dinamik	Lisanslı	Masaüstü, sunucu, web uygulaması	Microsoft Excel, Microsoft tarafından Windows işletim sistemi için geliştirilmiş bir elektronik hesaplama aracıdır. Veri hesaplamaya, grafik araçlarına, pivot tablolara ve makro programlamaya izin verir.
MATLAB	Çubuklar, noktalar/ kabarcıklar, ızgaralar, çizgiler, haritalar, ilişkiler/ bağımlılıklar, şekiller/ orantılar	Jeo-uzamsal veriler, ağ verileri, sayısal veriler, metin verileri	Veri dosyaları, veri tabanları	Statik	Lisanslı	Masaüstü ve sunucu uygulaması	MATLAB, istatistik analizi, makine öğrenimi ve finans gibi belirli uygulamalar için araçlar sunar.
Leaflet.js	Haritalar	Jeo-uzamsal veriler, sayısal veriler	Veri dosyaları	Dinamik	Açık kaynak (ücretsiz)	Kod kütüphanesi, framework	Mobil uyumlu etkileşimli haritalar için kullanılan JavaScript kitaplığıdır.
IBM SPSS Modeler	Çubuklar, haritalar	Sayısal veri	XLS, CSV, TSV, veri tabanları	Dinamik	Lisanslı	Masaüstü uygulama	IBM SPSS Modeler, veri modellerini belirlemeye ve veriye dayalı karar vermeye olanak tanıyan tahmine dayalı analitik sağlar. Araç, metin analitiği, varlık analitiği, karar yönetimi ve optimizasyon dahil olmak üzere bir dizi gelişmiş algoritma ve teknik sağlar.
Gephi	Ağ grafikleri	Ağ verisi	GEXF, GDF, GML, GraphML, Pajek NET, GraphViz DOT, CSV, UCINET DL, Tulip TPL, Netdraw VNA, Elektronik Tablo	Dinamik	Ücretsiz	Masaüstü uygulama	Her türlü grafik ve ağ için kullanılabilen bir görselleştirme ve keşif uygulamasıdır.

Yoğun bir şekilde kullanılan CiteSpace programı, Taleau ve veri görselleştirme için geliştiriciler ve tasarımcılar tarafından kullanılan; matplotlib, pandas ve numpy vb. kod kütüphanelerini barındıran Python programlama dilinden bahsetmekte yarar vardır.

CiteSpace

CiteSpace, bilimsel literatürdeki eğilimleri ve örüntüleri görselleştirmek ve analiz etmek için ücretsiz olarak kullanılabilen bir Java uygulamasıdır. Aşamalı bilgi alanı görselleştirmesi için bir araç olarak tasarlanmıştır (C. Chen, 2004). CiteSpace, yayınlara dair ağ örüntülerinin anlaşılmasını ve yorumlanmasını kolaylaştırmak için çeşitli işlevler sağlar. Bu işlevler arasında hızlı büyüyen konu alanlarının belirlenmesi, incelenen yayın grubunda atıf sıcak noktalarının bulunması, bir ağın kümelere ayrıştırılması, kümelerin alıntı yapılan makalelerden gelen terimlerle otomatik olarak etiketlenmesi, coğrafi iş birliği modelleri ve benzersiz uluslararası iş birliği alanları yer almaktadır (CiteSpace, 2021).

CiteSpace, iş birliği ağları, yazar ortak atıf ağları ve belge ortak atıf ağları dâhil olmak üzere bilimsel yayınlardan türetilen çeşitli ağların yapısal ve zamansal analizlerini destekleyen bir programdır. Ayrıca terimler, kurumlar ve ülkeler gibi hibrit düğüm türlerinin ağlarını ve ortak atıf, ortak oluşum ve yönlendirilmiş atıf bağlantıları gibi karma bağlantı türlerini de destekler. CiteSpace için birincil girdi verisi kaynağı Web of Science'dır. CiteSpace ayrıca PubMed, arXiv, ADS ve NSF Ödül Özetlerinden veri elde etmek için bazı basit arayüzler de sağlamaktadır (CiteSpace, 2021).

Tableau

Tableau, kullanıcıların verileri çok çeşitli formatlarda keşfetmesine, analiz etmesine ve sunmasına olanak tanıyan bir veri görselleştirme yazılımıdır (Deardorff, 2016, s.182). Program, tablolar, çubuk grafikler, ısı haritaları, ağaç haritaları, histogramlar, kabarcık grafikleri, dünya haritaları ve daha fazlasını içeren çeşitli görselleştirme türleriyle kullanıcı dostu olacak şekilde tasarlanmıştır. Kullanıcılar tarafından renk, şekil ve ölçü gibi boyutlar eklenerek görselleştirmeler özelleştirilebilmektedir (Deardorff, 2016, s.182). Tableau hem Windows hem de Mac için "Tableau Public" adı verilen ücretsiz, bulut tabanlı bir sürümün yanı sıra ücretli masaüstü, sunucu ve çevrimiçi sürümleriyle de mevcuttur.

Python programlama dili

Açık kaynak kodlu Python programlama dili hem statik hem de dinamik görselleştirmeler oluşturmak için de kullanılan popüler bir araçtır. Veri görselleştirme için

geliştiriciler ve tasarımcılar tarafından kullanılan en popüler kod kütüphaneleri "matplotlib, pandas ve numpy" kütüphaneleridir (Patnaik, 2019, s.26).

Görselleştirme araçları; çeşitli ölçeklerdeki verileri analiz etmek, öngörüler elde etmek, karar destek süreçlerine rehberlik edecek çıkarımlarda bulunmak vb. ihtiyaçlar için tasarlanmış araçlardır. Bu araçların araştırmacılara ve görselleştirme uzmanlarına enformasyonu açığa çıkaran görseller üretebilmesi, teknolojik gelişmelere paralel olarak ortaya çıkabilecek yeni veri tipleri ve dosya formatlarını da kapsayacak şekilde esnek olması önemlidir. Ali, Gupta, Nayak ve Lenka (2016, s.657) bu araçların etkileşimliliği üzerinde durmakta ve bu araçların taşınması gereken bazı temel fonksiyonları şu şekilde sıralamaktadırlar: Bu araçların: etkileşimli olması, yani kullanıcının görselleştirme ile etkileşime girebilmesi, üretilen görselleştirmelerin üzerine gelindiğinde ilgili bilgileri göstermesi, yakınlaştırma ve uzaklaştırma paneli olması, veri alt kümesi veya üst kümesi seçildiğinde görselleştirmenin çalışma zamanında kendini uyarlaması gerekmektedir.

Pavlopoulos, Paez-Espino, Kyrpides ve Iliopoulos (2017) sağlık alanı ile ilgili çalışmalarında, büyük ölçekli ağ analizi yapabilen görselleştirme araçlarından olan Gephi (Version 0.9.1), Tulip (Version 4.10.0), Cytoscape (Version 3.5.1) ve Pajek (Version 5.01) uygulamalarını aşağıdaki göstergelere göre karşılaştırmışlardır:

- Ölçeklenebilirlik (Scalability)
- Düzen tasarımları (Layouts)
- Görselleştirme sonrası analiz (Post visualization analysis)
- Düzenleme (Editing)
- Dosya formatları (File Formats)
- Elde edilebilirlik, kullanılabilirlik (Availability)

Pavlopoulos ve diğerlerinin (2017) bu göstergelere göre yaptıkları karşılaştırma sonucunda ilgili görselleştirme araçlarına dair tespit ettikleri zayıf ve güçlü özellikler Tablo 5'te verilmiştir.

Tablo 5: Etkileşimli ağ görselleştirme araçlarının değerlendirilmesi (Pavlopoulos ve diğerleri, 2017, s.6)

	Cytoscape	Tulip	Gephi	Pajek
Scalability	**	*	***	****
User friendliness	**	****	***	*
Visual styles	***	**	***	*
Edge bundling	***	****	**	—
Relevance to biology	****	**	***	*
Memory efficiency	*	**	***	****
Clustering	****	***	*	**
Manual node/edge editing	***	****	***	*
Layouts	***	**	****	*
Network profiling	****	**	***	*
File formats	**	***	****	*
Plugins	****	**	***	*
Stability	***	*	****	***
Speed	**	*	***	****
Documentation	***	*	**	***

* = weaker; ** = medium; *** = good; **** = strongest.

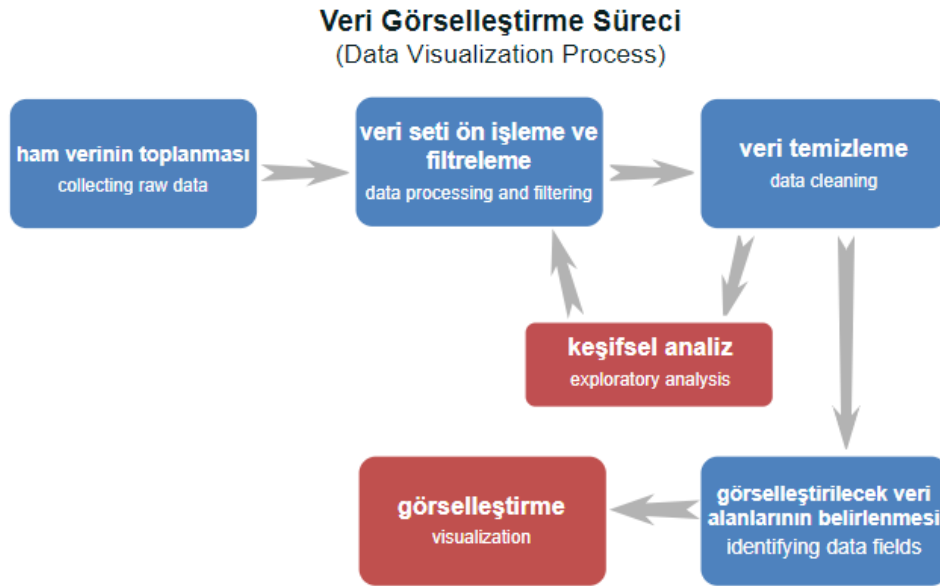
Tablo 5 incelendiğinde, Crtoscape sisteminin görsel stil, biyoloji ile alakalılığı, kümeleme, ağ profillemeye, eklenti desteği ve dokümantasyon başlıklarında daha etkili olduğu görülmektedir. Tulip sisteminin kullanıcı dostu olma, ilişki ağı oluşturma ve not ekleme başlıklarında etkiliyken, Gephi sisteminin yerleşim planı, daha çok dosya formatı destekleme ve istikrar başlıklarında daha etkili olduğu görülmektedir. Pajek ise ölçeklenebilirlik, bellek verimliliği ve hız başlıklarında diğer sistemlere göre daha etkili bir sistem olarak belirlenmiştir.

İzleyen bölümde çalışmanın yöntemine dair içerik, kullanılan veri görselleştirme modeli, bu model bağlamında oluşturulan veri görselleştirme uygulaması ayrıntılı olarak ele alınmıştır.

3. BÖLÜM: YÖNTEM

Bilimsel bilgi üretiminin can damarı olan araştırma; çözüm bulunmak istenen probleme/lere dair; hipotez/ler oluşturulması, çalışmada kullanılacak verilerin tespit edilmesi, çeşitli yöntem ve araçlarla verilerin derlenmesi ve sistematik hale getirilmesi, elde edilen verilerin analiz edilmesi, görselleştirilmesi, değerlendirilmesi ve yayına dönüştürülmesi süreçlerinden oluşmaktadır. Görüleceği üzere veriler bir araştırmanın çekirdeğini oluşturmaktadır ve araştırma için gerekli verilerin saptanmasından, analizinden, bulguların elde edilmesine değin tüm veri işlemlerinin doğru biçimde gerçekleştirilmesi, sonuçların doğruluğu ile ne kadar ilişkili ise bir veri görselleştirme uygulaması sonucunda sunulacak bilginin doğruluğu ile de o kadar ilişkilidir.

Bu bölümde araştırmanın amacı doğrultusunda bir veri görselleştirme uygulaması geliştirme, tüm adımları ile ayrıntılı bir biçimde ve bir modele dayalı olarak ele alınmaktadır. Söz konusu model, Andrews ve diğerleri (2023) tarafından geliştirilen ve Şekil 35'te sunulan veri görselleştirme modelidir. Model uyarınca araştırmacı tarafından geliştirilen ve çalışmanın özgün yanını oluşturan görselleştirme uygulamasında aynı zamanda karşılaşılan sorunlara da değinilmiş ve çözüm önerilerinde bulunulmuştur.



Şekil 35: Veri görselleştirme süreci (Andrews ve diğerleri, 2023, s. 6'dan uyarlanmıştır.)

Ham verinin toplanması: Veri görselleştirme sürecinde kullanılacak veriler, belirli araçlar yardımıyla çeşitli kaynaklardan derlenir. Dolayısıyla öncelikle verilerin nereden yani hangi veri kaynaklarından (web sayfaları, dâhili veritabanları, harici API'ler veya genel veri havuzları)

derleneceđi belirlenir, uygun aralar ile veriler toplanır ve veri zerinde yapılacak alıřma dikkate alınarak uygun bir dosyada (csv, xlsx, veri tabanı dosyası vb.) depolanır.

Veri seti n iřleme ve fitreleme: Toplanan ham verilerin n iřleme srecine tabi tutulması veri setine hkimiyeti sađlayan bir iřlemdir. Veri yapısı, veri miktarı ve deđiřkenler zerinde eřitli n analizler yapılarak veri seti, veri temizleme srecine hazır hale getirilir.

Veri temizleme: Bu iřlem; genellikle dađınık olan, eksik veya hatalı veri ierebilen ham verinin grselleřtirilmeden nce temizlenmesi, aykırı deđerlerin tespit edilmesi, yinelenen satırların ayıklanması ve veri tr dnřtrme iřlemlerinin yapılmasını ieren sretir.

Keřifsel veri analiz: Keřifsel veri analizi (Exploratory data analysis), veri seti n iřleme ve veri temizleme srelerinde uygulanan bir iřlemdir. Keřifsel veri analizinin temel amacı, bir veri setini sistematik ve grsel bir řekilde incelemek ve anlamaktır. Analiz srecinin kritik bir bařlangı adımı olup, temel hedef; veriyi anlama, rntleri ve iliřkileri tespit etme, aykırı deđerleri ve anormallikleri belirleme ve veriyi temizlemektir. Ayrıca, verileri daha derinlemesine anlamak, igrleri ortaya ıkarmak ve verileri daha ileri analiz ve modellemeler iin hazırlar. Daha geliřmiř analizlerin temelini oluřturduđu ve veriye dayalı kararların dođru ve anlamlı bilgilere dayanmasını sađlamaya yardımcı olduđu iin veri bilimi iř akıřında ok nemli bir yere sahiptir.

Grselleřtirilecek veri alanlarının belirlenmesi: Grselleřtirme gereksinimlerine bađlı olarak, daha odaklanılmıř bir veri kmesi oluřturmak iin eřitli sorgular yapılarak verilerin filtrelenmesi veya birleřtirilmesi iřlemdir. Bu srete, karřılařtırma yapıldıđında anlamlı sonular elde edilebilecek deđerlerden alt veri setleri oluřturulur ve grselleřtirmeye hazır hale getirilir.

Grselleřtirme: Veri iliřkilerinin grafikler, izelgeler, řekiller veya haritalar gibi grsel formatlarla ve renklerle temsil edilmesidir. Bylelikle, veri iliřkileri daha kolay algılanabilir ve temsil ettiđi bilgi daha anlařılır hale gelir. Grsel temsil, rntlerin ve igrlerin daha kolay tanımlanabilmesini ve daha rasyonel kararlar verilmesini sađlar.

Bu dođrultuda alıřmada veri grselleřtirme uygulaması iin TR Dizin'deki 2017-2021 arası (veri seti 2022 yılında oluřturulmuřtur) Ankara niversitesi adresli bilimsel makalelerin st verileri kullanılmıřtır.

TR Dizin; ulusal nitelikli ve denetimli ilk kurumsal veri tabanıdır. TR Dizin, 1992 yılında ulusal akademik dergilerin kalitesinin artırılması ve ulusal bilgiye eriřim amacıyla oluřturulmuřtur. Bařlangıcı TUBİTAK-ULAKBİM bnyesinde yrtlen Ulusal Tıp Veri

tabanı / Türk Tıp Dizini çalışmalarına dayanmaktadır. Bu çalışmalar, 2013 yılından itibaren TR Dizin olarak adlandırılmıştır (Aslan, 2019, s.1). Kalkınma Bakanlığı desteği ile ULAKBİM çatısı altında yürütülen altyapı ve yazılımları tamamlanan TR Dizin, ulusal atıf dizini özelliği kazanarak ilk arayüzü ile Temmuz 2018 tarihinde kullanıma açılmıştır (TUBİTAK ULAKBİM TV, 2019).

TR Dizin’de, dergi değerlendirme kriterlerine bağlı olarak seçilen ulusal, hakemli akademik dergiler dizinlenmektedir¹⁹. Bu dergilerdeki makalelere ait bibliyografik kimlikler “Fen Bilimleri” ve “Sosyal Bilimler” ana konu kategorilerinde gruplanmaktadır. Fen Bilimleri kategorisinde “Dış Hekimliği, Eczacılık, Mühendislik, Temel Bilimler, Sağlık Bilimleri ve Veterinerlik” konuları bulunurken, Sosyal Bilimler kategorisinde “Sosyal ve Beşerî Bilimler”e ait konular yer almaktadır (TR Dizin, t.y). Sosyal Bilimler kategorisindeki alt başlıkların, Fen Bilimlerinde olduğu gibi detaylı olarak ayrıntılandırılmadığı, sadece “Sosyal” ve “Beşerî Bilimler” şeklinde iki alt grupta toplandığı gözlemlenmiştir²⁰.

Dergi makalelerinin bibliyografik kimliklerinin yanı sıra, ULAKBİM Müdürlüğü ile dergi editörlükleri arasında imzalanan katılım izni sözleşmesine bağlı olarak, makalelerin tam metinlerine de platform üzerinden erişim sağlanmaktadır. Ayrıca, TÜBİTAK Araştırma Destek Programları Başkanlığı (ARDEB) bünyesindeki **1000-Araştırma ve Geliştirme Potansiyelinin Artırılmasına Yönelik Destek Programı** kapsamında desteklenen projelere de TR Dizin’de yer verilmektedir (TR Dizin, t.y).

TR Dizin sunduğu platform ile Türkiye’de üretilen akademik yayınlara erişimi kolaylaştırmakta ve dizinlenen yayınların görünürlüğünü artırmaktadır. Ayrıca Yüksek Öğretim Kurulu (YÖK) ile Üniversitelerarası Kurul (ÜAK)’un atama ve yükseltme kriterlerinde kabul ettiği bir dizindir (TR Dizin, t.y).

ULAKBİM’in TR Dizin ile bağlantılı bir diğer hizmeti DergiPark’tır. Akademik süreli yayıncılığın, Türkiye’de kaliteli ve standartlara uygun olarak gelişmesini sağlamak amacıyla oluşturulan DergiPark Platformu, Türkiye adresli, akademik ve hakemli olduğunu beyan eden tüm dergileri elektronik ortamda bulunduran ve editöryal süreç yönetimi sağlayan bir alt yapı hizmetidir. DergiPark kapsamında yer alan dergiler arasında TR Dizin, Web of Science,

¹⁹ TR Dizin’in kapsamını oluşturan ulusal bilimsel dergiler, TR Dizin uzmanları ile konu alanı uzman ve akademisyenlerden oluşan komiteler tarafından TR Dizin Dergi Değerlendirme Kriterlerine bağlı olarak seçilip dizinlenmektedir (Kaynak: <https://trdizin.gov.tr/tr/tr-dizin-dergi-degerlendirme-kriterleri/>).

²⁰ Kaynak: <https://trdizin.gov.tr/yaritim/>

Scopus, DOAJ ve diğ er ulusal/uluslararası dizinlerde tarananların yanı sıra hen z hi bir dizinde taranmayan, yayın hayatına yeni ba lamı  dergiler de bulunmaktadır (DergiPark, 2022).

TR Dizin’de makalelere tam metin eri im, DergiPark ve TR Dizin koordinasyonu ile saėlanmaktadır. TR Dizin, eėer makale DergiPark’ta yer alıyorsa DergiPark baėlantısıyla tam metne eri im saėlamaktadır. DergiPark’ta yer almayan tam metin i eriklere eri mek i in ise, ilgili derginin web sayfasına y nlendirme yapılmaktadır. DergiPark’ın ama larından biri de aynı zamanda bir atıf dizini olan TR Dizin i in  l  lebilir, yapılandırılmı  veri saėlamaktır (DergiPark, 2022).

TR Dizin; ulusal atıf dizini olması, makalelerin tam metin eri imine izin vermesi, makalelere ait  st verilerin eri ilebilir olması sebebiyle bu  alı ma i in verinin toplanacaėı platform olarak tercih edilmi tir. Bu ara tırmanın veri setini, TR Dizin  zerinden eri ilebilen Ankara  niversitesi adresli bilimsel makalelere ait  st veriler olu turmaktadır.  st veri (metadata), belge i eriėinin anla ılmasına ve bunlara teknolojik olarak eri ilebilmesine olanak saėlayan bir ara  olup,  st veri kullanımının elektronik bilgi kaynaklarını d zenleme, bilgiye eri im ve sistemler arası i lerliėi saėlama olmak  zere    ana yararı vardır (Bayter, 2008, s.3).

 alı manın bu b l m n n alt ba lıkları  ekil 35’teki veri g rselle tirme adımları doėrultusunda olu turulmu tur. Ancak bu adımlara ge meden  nce  alı ma kapsamında hangi veri kaynaėı kategorisi ve veri t r n n kullanılacaėı belirlenmi tir (bkz. B l m 3.1.1). Bunun yanı sıra TR Dizin web sayfasında kar ıla ılan sorunlar ve kullanılan alternatif veri toplama   z m  ele alınmı , bir veri toplama uygulaması geli tirilmi  ve bu  er evede kullanılan ara  ve y ntemler tanıtılmı tır. Veri toplandıktan sonra olu an veri seti  zerinde veri temizleme ve analiz i lemleri ger ekle tirilmi , ke ifsel veri analizi yapılmı  ve g rselle tirilecek veri alanları belirlenerek g rselle tirmeler olu turulmu tur.

3.1 Veri Toplama

Bir veri g rselle tirme uygulaması ger ekle tirebilmek veya geli tirebilmek ancak veriye sahip olmakla m mk nd r. Veri g rselle tirmeyi ger ekle tirmeden  nce doėru veri kaynaėı kategorisinin belirlenmesi, veri alanlarının/deėi kenlerin tanımlanması ve veri alanları i in kullanılacak veri t rlerinin atanması son derece  nemlidir.

3.1.1 Veri Kaynağı Kategorisi, Veri Alanları ve Veri Türlerinin Belirlenmesi

“California Digital Library” (Sources of Data, 2022) veri kaynaklarını dört ana kategoride toplamaktadır. Bunlar; gözlemsel, deneysel, simülasyon ve türetilmiş/derlenmiş verilerdir.

Gözlemsel: Gözlemsel veriler, olayların doğal seyrinin veya geçmiş olayların izlenmesiyle elde edilen verileri ifade eder. Gerçek zamanlı olarak, genellikle laboratuvar dışında elde edilir. Örnek olarak; sensörlerden elde edilen veriler, telemetri, anket sonuçları, görüntüler vb.

Deneysel: Deneysel veriler kontrollü deneyler veya testlerle elde edilir ve neden-sonuç ilişkilerini incelemek için kullanılır. Genellikle tekrarlanabilir, ancak maliyetli veya zaman alıcı olabilir. Örnek olarak; gen dizileri, kromatogramlar, manyetik alan okumaları verilebilir.

Simülasyon: Bu tür veriler, makineler tarafından, tasarlanmış test modelleri üzerinden üretilir. Model ve girdiler korunursa tekrarlanabilir olması muhtemeldir. Örnek olarak; iklim modelleri, ekonomi modelleri verilebilir.

Türetilmiş / Derlenmiş: Genellikle birden fazla kaynaktan toplanan veya mevcut verilerin analizi sonucunda elde edilen verilerdir. Örnek olarak; metin ve veri madenciliği, derlenmiş veri tabanı, 3D modeller verilebilir (Sources of Data, 2022).

Bu sınıflandırma, veri analizi çalışmalarında veri kaynağını, kullanılacak veri alanlarını ve atanacak veri türlerini anlamak için temel bir çerçeve sağlar ve doğru analiz yöntemlerinin ve görselleştirme tekniklerinin seçilmesine yardımcı olur. Özetle, veri kaynaklarını Gözlemsel, Deneysel, Simülasyon ve Türetilmiş/Derlenmiş kategorilere ayırmak, uygun veri toplama ve analiz yöntemlerinin seçilmesi ve veri kalitesinin değerlendirilmesi açısından önemlidir. Ayrıca araştırmacılara veri toplama yaklaşımlarını belirlemede rehberlik eder ve yaklaşımın araştırma hedefleri ile uyumlu olmasını sağlar.

Çalışmada veri görselleştirme uygulaması için TR Dizin’deki 2017-2021 yılları arası Ankara Üniversitesi adresli bilimsel makalelerin üst verileri kullanılmıştır. Bu amaçla TR Dizin web sayfasından ilgili makalelerin üst verileri çekilmiştir. “California Digital Library”nin yaptığı, yukarıda belirtilen, sınıflamaya göre bu çalışmada kullanılan veri kaynağı “türetilmiş / derlenmiş” kategorisine girmektedir.

Yukarıda bahsedilen veri kaynaklarından verileri toplamak ve onu yapılandırılmış ve yönetilebilir bir veri setine dönüştürmek önceden planlanması gereken bir süreçtir. Veri seti

içerikleri, yapılan çalışma boyunca sabitlenebilir veya değişiklik gerektirebilir. Analiz amacıyla oluşturulabilecek veri seti kategorileri şunlardır (California Digital Library, t.y.):

Sabit veri setleri: Toplandıktan veya oluşturulduktan sonra asla değişmez.

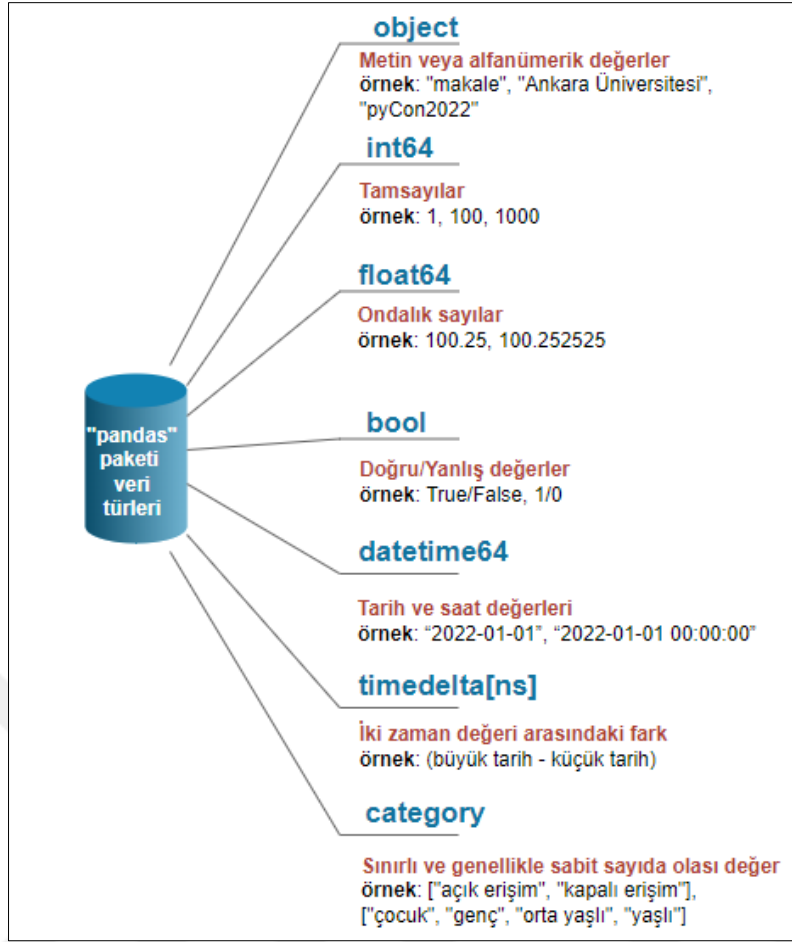
Büyüyen veri setleri: Yeni veriler eklenebilir, ancak eski veriler asla değiştirilmez veya silinmez.

Revize edilebilir veri setleri: Yeni veriler eklenebilir ve eski veriler değiştirilebilir veya silinebilir.

Bu araştırmadaki verilerin derlenmesi ve veri setine dönüştürülmesi işlemleri “Revize edilebilir veri seti” kategorisine uymaktadır.

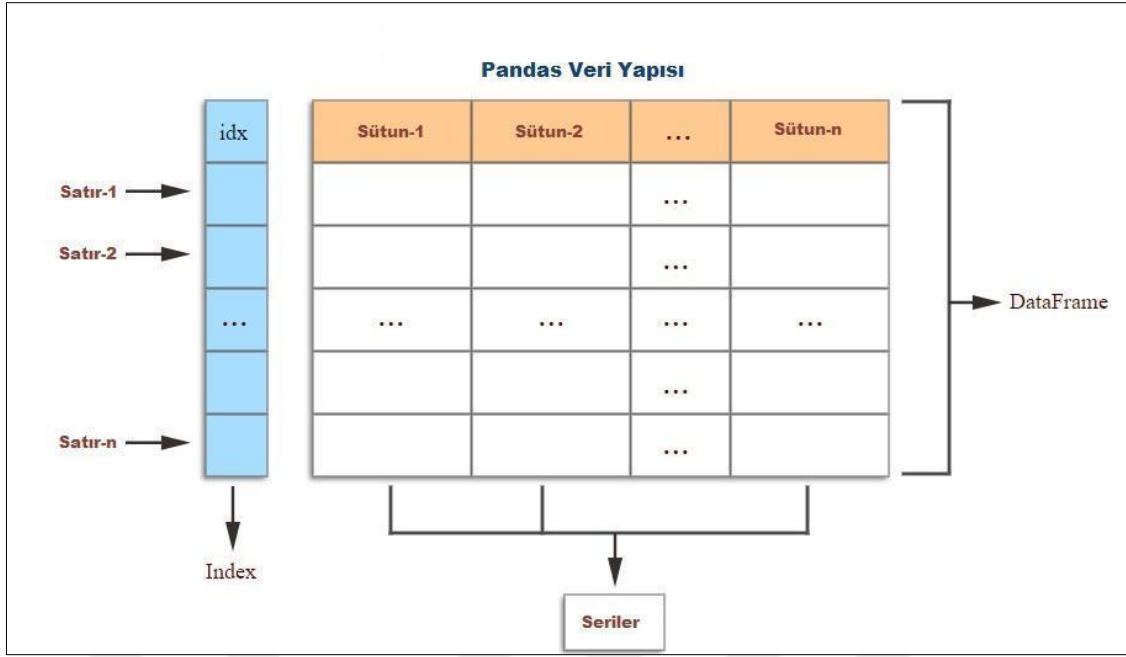
Hem herhangi bir araştırmaya başlarken hem de bir görselleştirme sürecinde hangi değişkenlerin kullanılacağı ve bu değişkenlerin hangi veri türleri ile temsil edileceği yürütülen çalışmanın sonraki aşamaları için son derece önemlidir. Her sütunda yer alacak değişkene göre uygun veri türünün belirlenmesi, analiz aşamasında karşılaşılabilecek olası hataların ve beklenmeyen sonuçların önüne geçecektir. Çalışmada görselleştirme uygulamasında kullanılacak değişkenlerin veri türlerini belirlemede Python programlama dilinin “pandas²¹” kod kütüphanesi kullanılmıştır. Söz konusu veri türleri (data types) Şekil 36’da örnekleri ile verilmiştir (Pandas, t.y. “Intro to data structures”).

²¹ Pandas: pandas, veri işlemesi ve analizi için Python programlama dilinde yazılmış olan bir kod kütüphanesidir (paket). (Kaynak: <https://tr.wikipedia.org/wiki/Pandas>)



Şekil 36: Python programlama dili “pandas” kod kütüphanesi veri türleri

Veri alanlarını/değişkenleri denetlemek ve düzenlemek için veri setinin pandas'ın yapılandırılmış tablo veri yapısı olan DataFrame'e dönüştürülmesi gerekmektedir. DataFrame, satır ve sütunlardan oluşan, değişkenleri (sütun) ve buna karşılık gelen verileri (satır) içeren iki boyutlu bir veri yapısıdır (Stojiljković, 2024). DataFrame örnek görseli Şekil 37' de verilmiştir.



Şekil 37: Pandas veri yapısı örneği (w3resource, 2024)

DataFrame her ne kadar veri setindeki birçok değişkene varsayılan veri türü atasa da veri seti temizleme sürecinde bu işlemin manuel olarak tekrar düzenlenmesi gerekmektedir (Moffitt, 2018). Veri türü, temelde bir programlama dilinin verileri nasıl depolayacağını ve değiştireceğini anlamak için kullandığı dâhili bir yapıdır. Örneğin, bir programın $5+10 = 15$ gibi bir hesaplama yapabilmesi için 5 ile 10'nun birer sayı olduğunu tanıması gerekir (Moffitt, 2018). Aksi halde verilen değerlerin toplamını 510 şeklinde metin olarak algılayacaktır veya tam sayı olan 2022 ile bir tarihin parçası olan 2022'nin ayırt edilmesi, ancak ilgili değişkenlere doğru veri türünün atanması ile mümkündür. Pandas DataFrame üzerinden her bir veri türüne dönüştürme işlemi Kod 1'deki örnek kod satırları ile uygulanabilir.

```
DataFrame["Sütun adı"].astype('object')
DataFrame["Sütun adı"].astype('int64')
DataFrame["Sütun adı"].astype('float64')
DataFrame["Sütun adı"].astype('bool')
DataFrame["Sütun adı"].astype('datetime64')
DataFrame["Sütun adı"].astype('category')
```

Kod 1: Veri türü dönüştürme

Çalışmada kullanılan veriler ile ilgili veri türü dönüştürme işlemleri veri temizleme bölümünde detaylı olarak ele alınmıştır.

3.1.2 Veri Toplama Aşamasında Kullanılan Araçlar / Yöntemler

Araştırma kapsamında, kurum olarak Ankara Üniversitesi'nin 2017-2021 yılları arasında yayınlanmış makale üst verileri ile çalışılmıştır. Makalelere ait üst veri alanları TR Dizin resmî web sayfasından elde edilmiştir. “Proje” üst verileri kapsama alınmamıştır. Derlenen veriler, üzerinde herhangi bir değişiklik yapılmadan CSV dosya formatında bir veri seti olarak Google Drive depolama alanına kaydedilmiştir.

Veri derleme/toplama aşamasında bazı sorunlarla karşılaşmıştır. Bu sorunlar aşağıda detaylandırılmıştır:

1. Aramadan kaynaklanan sorunlar:

Şekil 38: TR Dizin Basit Arama Çubuğu (Kaynak: <https://trdizin.gov.tr/>)

Şekil 38’de görüleceği üzere arama arayüzünde bulunan “yayın arama” seçeneğinin kullanıcıya yalnız yayın adından arama yapabileceği; yıl, yazar, kurum adı vb. kısıtlamaları kullanamayacağı izlenimi vermektedir. “Yardım” sayfasında “Gelişmiş arama” alanı incelendiğinde, çeşitli üst veri alanlarından ne şekilde arama yapılacağı belirtilmiş olsa da çalışmamız için kullanacağımız sorgu parametrelerinden “yıl aralığı” ve “kurum ismi” terimlerinin filtreleme olarak nasıl kullanılabileceği ile ilgili bir bilgiye rastlanmamıştır.

2. TR Dizin’in sunduğu veri seti ile ilgili sorunlar:

Şekil 39: TR Dizin Dışarı Aktarma Arayüzü (Kaynak: <https://trdizin.gov.tr/>)

TR Dizin web sayfasında yapılan sorgu sonuçları Şekil 39’da görüldüğü gibi belirlenen üst veri alanları seçilerek 100’er kayıt şeklinde “xlsx”, “csv” veya “ods” formatlarında indirilebilmektedir. Bu üst veri alanları haricinde, çalışmamızda kullanılacak olan “Makale TR Dizin URL’si, makale konu kategorisi, erişim türü, makale anahtar kelimeleri, makale konusu, makale web sayfası linki” ve dergi üst veri alanlarının (“dergi dili, yıllık yayın sayısı, konu kategorisi, konu alanları, makale sayısı, derginin aldığı atıf sayısı, kendine atıf sayısı, atıf alan makale sayısı, atıf ortalaması ve atıf oranı”) indirilebilen dosyada olmadığı tespit edilmiştir.

3. TR Dizin’deki bazı üst veri alanlarında veri eksikliğinin olması:

TR Dizin web sayfasında, yayınlara ait “konu” ve “anahtar kelimeler” alanlarında verilerin eksik olduğu tespit edilmiştir. DOI adresi DergiPark olan kayıtlar için ilgili yayının DergiPark sayfasından bu alanlara ait verilerin toplanması ihtiyacı doğmuştur.

TR Dizin web arayüzünde yapılan sorgularda “yıl aralığı” ve “kurum ismi” terimleri kullanılarak birlikte filtreleme yapılamaması, sorgu sonucunda indirilebilir üst veri alanlarının yapılacak çalışma için ihtiyacı karşılamaması ve bazı üst veri alanlarının DergiPark platformundan toplanması ihtiyacı, veri toplama süreci için farklı bir yöntemin kullanılmasını gerektirmiştir. İhtiyaç duyulan bütün üst veri alanlarına ait bir API hizmetinin olmaması veya indirilebilir dosyanın, Web of Science gibi bütün üst veri alanlarını kapsamaması nedeniyle üst verilerin toplanması için farklı bir veri derleme yöntemi olan web kazıma (web scraping) yöntemi uygulanmıştır.



Şekil 40: Web kazıma için kullanılan araçlar

Şekil 40’ta yer alan, web kazıma sürecinde kullanılacak araçlar aşağıda sırasıyla açıklanmıştır.

Web kazıma: Bir programlama dili kullanılarak bir ya da birden çok kaynaktan verimli ve etkili bir şekilde veri toplamaya olanak veren bir yöntemdir. Web kazıma yöntemi (web hasat veya web veri çekimi olarak da bilinmektedir) web sitelerinden bilgi toplamayı sağlayan bir yazılımdır. Bu yazılımların insanların web üzerindeki araştırma davranışlarını simüle edebilme özelliği de vardır (“Web kazıma”, 2024).

Bu çalışmada Web kazıma işlemi için Python programlama dili ve onun bazı kod kütüphaneleri kullanılmıştır.

Python programlama dili: Yorumlamalı, etkileşimli, nesne yönelimli ve açık kaynak kodlu bir programlama dilidir (Python, t.y.). Python, veriyle ilgili işlemlerde ve veri bilimi alanında yaygın olarak kullanılan bir programlama dilidir. Terra (2024), Python programlama dilinin veri analizi için tercih edilmesinin temel sebeplerini; dilin esnek olması, öğrenilmesinin kolay olması, açık kaynak olması, geniş bir topluluğa sahip olması ve bu yolla geniş bir kesim tarafından desteklenmesi şeklinde ifade etmiştir. Ayrıca kodlamasının okunabilirliği ve basitliği, çok yönlü bir yapıya sahip olması, disiplinlerarası entegrasyon esnekliği, büyük veri araçlarıyla entegrasyonu, hızlı modellemeye imkân vermesi, piyasa talebi ve veri bilimiyle ilgili kapsamlı kod kütüphanelerini barındırması diğer tercih nedenleridir.

Pandas: Verilerin işlemlerini ve analizini sağlayan Python programlama dilinde yazılmış olan bir kod kütüphanesidir (<https://pandas.pydata.org/>).

Numpy: Sayısal hesaplama işlemleri yapabilen açık kaynak kodlu bir Python programlama dili kod kütüphanesidir (Numpy, 2023).

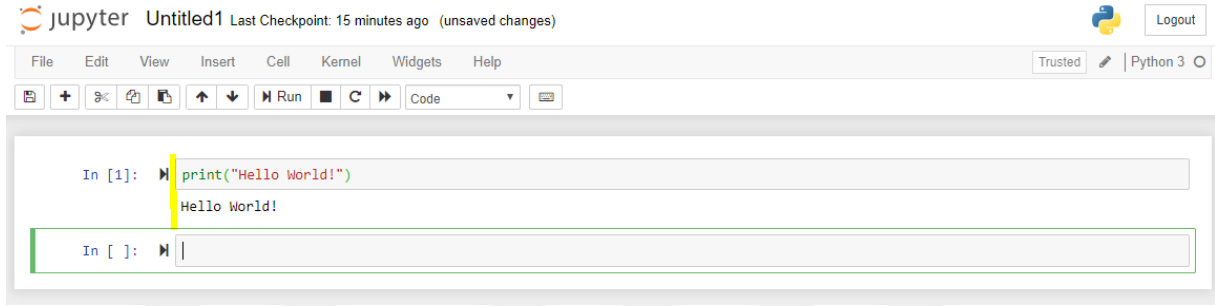
Beautifulsoup: HTML ve XML belgelerini ayrıştırmak için kullanılan bir Python kod kütüphanesidir (Beautifulsoup, 2024).

re (regular expression): Belirli bir örüntü doğrultusunda metin içerisinde kelime, kelime grubu veya karakter gibi ifadeleri tespit etmek için kullanılan bir kod kütüphanesidir (Python, t.y.).

Bunların yanı sıra çalışmada Google Research'ün bir ürünü olan Google Colab kullanılmıştır. Colaboratory veya kısaca "Colab", herkesin web tarayıcısı aracılığıyla python kodu yazmasına ve yürütmesine izin veren, kodların yazılması uygulanması için Jupyter Notebook uygulaması ile çalışma imkânı sunan bir platformdur; özellikle makine öğrenimi, veri analizi ve eğitim için elverişli bir ortam sağlamaktadır (Google, t.y.).

requests: "Requests" kütüphanesi Python'da HTTP istekleri yapmak için kullanılan bir kütüphanedir. Get, Post vb. gibi HTTP işlemleriyle etkileşim kurmak için basit bir API sağlar (Python Requests Library, 2023).

Jupyter Notebook: Jupyter Notebook (eski adıyla IPython Notebook), kod yazmak ve uygulamak için kullanılan web tabanlı, açık kaynak kodlu etkileşimli bir geliştirme ortamıdır (Jupyter, 2022).



Şekil 41: Jupyter Notebook’da “print” fonksiyonunun nasıl çalıştığını gösteren örnek ekran görüntüsü

Google Drive: Google tarafından oluşturulmuş olan ve yönetilen bir dosya depolama, yedekleme ve eşitleme (senkronizasyon) ortamıdır (https://www.google.com/intl/tr_tr/drive/). Google Drive; kullanıcıların bulutta doküman depolayabilmelerini, dosya paylaşabilmelerini ve iş birliği yaptıkları kişilerle dokümanları düzenleyebilmelerini sağlayan bir hizmettir (<https://support.google.com/drive/>).

İzleyen bölümde TR Dizin ve DergiPark web sayfalarından veri toplarken kullanılan araç ve yöntemler tanıtılmıştır ve uygulama adımları anlatılmaktadır.

3.1.3 Web Kazıma ile Veri Seti Oluşturma

Web kazıma adımları sırasıyla aşağıda sıralanmıştır.

1. Adım: Web sayfası izin kontrolü

Çalışmada öncelikle TR Dizin kök dizininde bulunan, bir web sayfasındaki kullanıcı izinlerini belirleyen “robot.txt²²” dosyası incelenmiş ve web kazıma ile ilgili herhangi bir kısıtlamanın olmadığı görülmüştür (<https://trdizin.gov.tr/robots.txt>). TR Dizin kök dizininde yer alan, 20.12.2022 tarihli “robots.txt” içeriği aşağıdaki gibidir.

²² Kaynak: <https://trdizin.gov.tr/robots.txt>

```
# XML Sitemap & Google News version 5.3.3 -  
https://status301.net/wordpress-plugins/xml-sitemap-feed/  
Sitemap: https://trdizin.gov.tr/sitemap.xml  
Sitemap: https://trdizin.gov.tr/sitemap-news.xml  
  
User-agent: *  
Disallow: /wp-admin/  
Allow: /wp-admin/admin-ajax.php
```

Şekil 42: TR Dizin robots.txt dosyası (Kaynak: TR Dizin)

2. Adım: Çekilecek üst veri alanları ve çalışmada kullanılacak değişken adlarının atanması.

Web kazıma işlemine geçmeden önce, TR Dizin web sayfasından toplanmasına karar verilen değişken adları belirlenmiştir. Makalelere ve makalelerin yayınlandığı dergilere ait üst veri alanları Tablo 6 ve Tablo 7’de görülmektedir. Söz konusu üst veri alanlarına karşılık gelecek şekilde çalışmanın veri setinde kullanılmak üzere değişken adları/sütun başlıklarına da ikinci sütunda yer verilmiştir.

Tablo 6: TR Dizin Makale üst veri alanları ve veri seti değişken adları

No	TR Dizin makale üst veri alanları	Veri seti için atanan değişken adı
1	Makale adı	makale_Adi
2	Makale konu kategorisi	makale_konu_Kategorisi
3	Makale yılı	makale_Yil
4	Makale atıf sayısı	atif_Sayisi
5	Makale görüntüleme sayısı	makale_Goruntuleme
6	Makale indirme sayısı	makale_Indirme
7	Makale TR Dizin URL’si	makale_URL
8	Yazar adı	yazar_Adi
9	Yazar kurum bilgisi	yazar_Kurum
10	Makale dili	makale_Dil
11	Makale türü	makale_Turu
12	Erişim türü	erisim_Turu
13	Makale DOI bilgisi	DOI_Durumu
14	Makale anahtar kelimeleri	makale_Keywords

15	Makale konusu	makale_Konu
16	Makale web sayfası linki	makale_web_Sayfasi

Tablo 7: TR Dizin Dergi veri alanları ve veri seti değişken adları

No	TR Dizin dergi üst veri alanları	Veri seti için atanan değişken adı
1	Dergi adı	dergi_Adi
2	Dergi TR Dizin URL'si	dergi_URL
3	Dergi dili	der_dili
4	Dergi yayına başlama yılı	der_yayin_yili
5	Derginin dizinlendiği yıllar	der_dizin_yillari
6	Dergi yıllık yayın sayısı	der_yillikYayin_sayisi
7	Dergi konu kategorisi	der_konu_kategorisi
8	Derginin yayınlandığı aylar	der_yayin_aylari
9	Dergi konu alanları	der_konu_alanlari
10	Dergi toplam makale sayısı	der_dergi_makaleSayisi
11	Derginin aldığı atıf sayısı	der_aldigi_atifSayisi
12	Derginin kendine atıf sayısı	der_kendine_atifSayisi
13	Atıf alan makale sayısı	der_atifAlan_makaleSayisi
14	Derginin atıf ortalaması	der_dergi_atifOrt
15	Kendine atıf oranı	der_kendine_atifOrani

TR Dizin makale URL'si, makale adı, dergi adı, makale konu kategorisi, TR Dizin dergi URL'si, makale yılı ve atıf sayısı alanları makale özet sayfasından elde edilmiştir. Örnek makale özet sayfa görseli Şekil 43'te verilmiştir.



Şekil 43: TR Dizin makale özet sayfası üst veri alanları

Yazar adı/adları, yazar kurum bilgisi, makale dili, makale türü, erişim türü, makale DOI bilgisi ise Şekil 44’te örnek bir görseli bulunan makale detay alanından elde edilmiştir.

Halk Kütüphaneleri Personeli İçin Sürekli Mesleki Eğitim ve Gelişim Modeli Önerisi

✓ *Recep IŞIK, (Ankara Üniversitesi, Bilgi ve Belge Yönetimi Bölümü, Ankara, Türkiye)*
✓ *Tülay OĞUZ (Ankara Üniversitesi, Bilgi ve Belge Yönetimi Bölümü, Ankara, Türkiye)*

Bilgi Yönetimi

Yıl: 2020 Cilt: 3 Sayı: 1 Sayfa Aralığı: 65 - 89 Metin Dili: Türkçe DOI: 10.33721/by.727458

Türkçe İngilizce

Halk Kütüphaneleri Personeli İçin Bir Sürekli Mesleki Eğitim ve Gelişim Modeli Önerisi

Öz:
Bilgi ve iletişim teknolojilerinin çalışma yaşamında yarattığı değişimin de etkisiyle geleneksel hizmet içi eğitim

Anahtar Kelime:

Belge Türü: Makale Makale Türü: Araştırma Makalesi Erişim Türü: Erişime Açık

Şekil 44: TR Dizin makale detay sayfası üst veri alanları

Makalelerin yayınlandığı dergilere ait üst veri alanları (bkz. Tablo 7) ise, TR Dizin web sayfası dergi detay sayfasından toplanmıştır. Dergi detay sayfasının örnek bir görseli Şekil 45’te verilmiştir.

Bilgi Dünyası



EISSN: 2148-354X

Web

Editör: Üniversite ve Araştırma Kütüphanecileri Derneği

Yayın Formatı: Elektronik

Yayına Başladığı Yıl: 2000

Yıllık Yayın Sayısı: 2

Yayın Yapması Gereken Aylar: Haziran, Aralık

Konu Alanları: Bilgi, Belge Yönetimi

Yayıncı: Üniversite ve Araştırma Kütüphanecileri Derneği

Yayın Dili: Türkçe, İngilizce

Dizinlendiği Yıllar: 2002 - 2022 (Sosyal)

Konu Kategorisi: Sosyal > Sosyal

Makale Sayısı	Atıf Sayısı	Kendine Atıf Sayısı	Atıf Alan Makale Sayısı	Atıf Ortalaması	Kendine Atıf Oranı
365	1046	107	183	2,87	%0,1

Şekil 45: TR Dizin dergi detay bilgi alanı

Çalışmada kullanılmak üzere Tablo 6 ve Tablo 7’de belirlenen verilerin toplanması için daha önce de belirtildiği gibi açık kaynak kodlu bir programlama dili olan Python programlama dili kullanılmıştır. Kullanımının ücretsiz olması, açık kaynak kodlu olması, veri toplama, temizleme, analiz ve görselleştirmeye imkân veren çeşitli kod kütüphanelerinin bulunması nedeniyle bu programlama dili tercih edilmiştir. Program kodlarının üretilmesi için yine bulut ortamında bulunan, Jupyter Notebook ile çalışma imkânı sunan Google Colab platformu ve veri seti depolama alanı olarak Google Drive ortamı kullanılmıştır.

TR Dizin web sayfası üzerinden Ankara Üniversitesi akademik yayınlarına dair veriler derlenirken, Tablo 6 ve Tablo 7’de belirlenen her bir veri alanı için uygulanan web kazıma adımları aşağıda sırasıyla detaylandırılmıştır.

3. Adım: Google Drive yetki işlemlerinin tanımlanması

Öncelikle; Google Drive’dan dosya çekmek/kaydetmek için gerekli olan güvenlik izinlerinin tanımlanması gerekmektedir. Bu izinler Kod 2 ile tanımlanmıştır.

```
from google.colab import drive
drive.mount('/content/drive/')
```

Kod 2: Google Drive için izinlerin tanımlanması

Kod 2, veri setinin Google Colab üzerinden Google Drive depolama alanına aktarılması ve Google Drive’daki dosyalar üzerinde işlem yapılması için gerekli olan izinleri sağlamaktadır.

4. Adım: Web kazıma için kod kütüphanelerinin aktifleştirilmesi

TR Dizin web sayfası üzerinde web kazıma işlemi için kullanılacak “BeautifulSoup, requests ve pandas” adlı Python kütüphaneleri ve metin içerisinde belli kriterleri taşıyan kelime, kelime grubu veya karakter gibi ifadeleri tespit etmek için kullanılacak “re” (regular expression) kütüphanesi Kod 3 ile aktifleştirilmiştir.

```
from bs4 import BeautifulSoup
import requests
import pandas as pd
import re
```

Kod 3: Web kazıma için kütüphanelerin aktifleştirilmesi

5. Adım: Sorgu alanlarının belirlenmesi ve sorgu fonksiyonu oluşturulması

Çalışmanın 3.1 veri toplama bölümünde değinilen veri toplama sorunlarını aşmak için aşağıdaki alanları baz alan Kod 4’teki fonksiyon oluşturulmuştur. Makalelere ait URL’ler bu fonksiyon aracılığıyla toplanmıştır.

- Yayın türü: Makale
- Yayın yıl aralığı: 2017-2021
- Kurum: Ankara Üniversitesi
- Yayın sayfalama sayısı: 1000
- Sıralama ölçüsü: Yayın adı

Sorgu için oluşturulan fonksiyon aşağıdaki gibidir.

```
def sorgu(sayfaSayisi):
    sayfaUrl=[]
    for i in range(sayfaSayisi):
        sayfaUrl.append("https://search.trdizin.gov.tr/tr/yayin/ara?q=*"+
            "order=title-ASC&facet-documentType=paper"+
            "facet-publication_year=2021&facet-publication_year=2020"+
            "facet-publication_year=2019"+
            "facet-publication_year=2018"+
            "facet-publication_year=2017"+
            "facet-authorInstitution=ANKARA+C3%9CN%C4%B0VERS%C4%B0TES%C4%B0"+
            "page="+str(i+1)+"&limit=1000")
    return sayfaUrl

trDizinUrl = sorgu(8)
```

Kod 4: TR Dizin sayfası için oluşturulan sorgu fonksiyonu

Bu sorgu doğrultusunda, her seferinde 1000 kayıt olmak üzere sorgu çalıştırılmış ve Ankara Üniversitesi adresli 7.436 makalenin özet bilgisine ait HTML verisi elde edilmiştir.

Bundan sonraki adımlarda uygulanacak süreçler aşağıdaki gibidir:

- Liste değişkenlerin tanımlanması,
- HTML etiketleri içerisinde verilerin toplanması ve liste değişkenlere atanması,
- Liste değişkenlerin DataFrame’e değişken (sütun) olarak eklenmesi.

Bu adımlar DataFrame’e eklenecek sütun sayısının tamamlanmasına kadar tekrar edecektir.

6. Adım: “Makale özet sayfa” verilerinin toplanması

Bu adımda toplanacak üst veri alanları; “makale URL, makale adı, dergi adı, makale konu kategorisi, dergi URL ve atıf sayısı” alanlarıdır. Kod 4’teki sorgu sonucunda çekilen HTML

etiketleri içerisinde Kod 5 kullanılarak bu TR Dizin üst veri alanları için liste değişkenleri (array) tanımlanmıştır.

```
makaleURL = []
makaleAdi = []
dergiAdi = []
makaleKonuKategorisi = []
dergiURL = []
atifSayisi = []
```

Kod 5: Makale özet alanı için liste değişkenlerinin tanımlanması

Kod 5'teki her bir liste değişkeni, TR Dizin'deki ilgili üst veri alanlarındaki verilerin geçici olarak saklanması için kullanılmıştır. Kod 6 kullanılarak TR Dizin'den toplanan veriler ilgili liste değişkenlerine aktarılmıştır.

```
for i in range(len(trDizinUrl)):
    URL = str(trDizinUrl[i])
    page = requests.get(URL)
    soup = BeautifulSoup(page.content, "html.parser")
    makaleLink = soup.findAll('a', class_="search-item")
    dergiLink = soup.findAll('div', class_="result-journal")
    yil = soup.findAll('div', class_="result-detail")
    atif = soup.findAll('div', class_="result-type ms-2")
    print(len(atif))
    for j in range(len(makaleLink)):
        makaleURL.append("https://search.trdizin.gov.tr"+str(makaleLink[j].get('href')))
        makaleAdi.append(makaleLink[j].text.strip())
        dergiAdi.append(dergiLink[j].a.text.strip())
        makaleKonuKategorisi.append(dergiLink[j].span.text.strip())
        dergiURL.append("https://search.trdizin.gov.tr"+str(dergiLink[j].a.get('href')))
        if "Atıf" in atif[j].text:
            atifSayisi.append(atif[j].text.replace('Atıf', '').strip())
        else:
            atifSayisi.append("0")
```

Kod 6: Makale özet alanı verilerinin toplanması

Liste değişkenlerine geçici olarak aktarılan verileri tabloya dönüştürmek için bir DataFrame (Python “pandas” kütüphanesinin veri yapısı) oluşturulmuş ve makale özet veri liste değişkenleri Kod 7 ile bu DataFrame'e (df) değişken (sütun) olarak aktarılmıştır.

```
df = pd.DataFrame()
df['makale_Adi'] = makaleAdi
df['dergi_Adi'] = dergiAdi
df['makale_konu_Kategorisi'] = makaleKonuKategorisi
df['atif_Sayisi'] = atifSayisi
df['makale_URL'] = makaleURL
df['dergi_URL'] = dergiURL
```

Kod 7: Makale özet alanı liste değişkenlerinin DataFrame'e aktarılması

İlerleyen satırlarda, web kazıma ile elde edilen her veri alanı, oluşturulan “df” adlı DataFrame’e değişken (sütun) olarak eklenmiştir.

7. Adım: “makale detay sayfası” verilerinin toplanması

Bu adımda toplanacak TR Dizin üst veri alanları; “yazar adı, yazar kurum adı, makale yılı, makale dili, makale türü, makale erişim türü, makale görüntüleme sayısı, makale indirme sayısı ve DOI”dir. TR Dizin makale detay alanındaki veri alanlarını toplamak için, daha önce DataFrame’e eklenen “makale_URL” (her makaleye ait URL bilgisi) değişkeni referans olarak kullanılmıştır. Bu alanlar aşağıdaki kod blokları ile toplanmıştır.

Öncelikle yazar adı verisinin çekilmesi ve DataFrame’e eklenmesi için Kod 8 ile “yazarAdi” liste değişkeni tanımlanmıştır.

```
▶ yazarAdi = []
```

Kod 8: Yazar verilerinin toplanması için liste değişken tanımlanması

Kod 9 ile TR Dizin’den yazar adı verileri toplanarak “yazarAdi” liste değişkenine aktarılmıştır.

```
▶ for i in range(len(df['makale_URL'])):
    URL = df['makale_URL'][i]
    page = requests.get(URL)
    soup = BeautifulSoup(page.content, "html.parser")
    yazar = soup.findAll('div', class_="new")
    b=""
    for k in range(len(yazar)):
        b = b + yazar[k].a.text.strip()
    yazarAdi.append(b)
```

Kod 9: Yazar adlarının toplanması ve “yazarAdi” liste değişkenine aktarılması

Yazar adı verilerini taşıyan liste değişkeni Kod 10 ile DataFrame’e değişken olarak eklenmiştir. Ortak yazarlı makalelerde, her makale için yazar adları DataFrame’de aralarında noktalı virgül kullanılarak kaydedilmiştir.

```
▶ df['yazar_Adi'] = yazarAdi
```

Kod 10: “yazarAdi” liste değişkeninin DataFrame’e değişken olarak eklenmesi

TR Dizin’den yazar kurum adı verisinin toplanması için öncelikle Kod 11 ile “yazar_Kurum” adlı bir liste değişkeni tanımlanmıştır.

```
▶ yazar_Kurum = []
```

Kod 11: “yazar_Kurum” liste değişkeninin tanımlanması

Kod12 ile TR Dizin üzerindeki yazar kurum verileri toplanarak “yazar_Kurum” liste değişkenine aktarılmıştır.

```
for i in range(len(df.makale_URL)):
    URL = str(df.makale_URL[i])
    page = requests.get(URL)
    soup = BeautifulSoup(page.content, "html.parser")
    label = soup.findAll('span', class_='author-institution')
    k = ""
    for j in range(len(label)):
        # print(label[j])
        if label[j].get('title') != "":
            k += label[j].get('title')+";"
        else:
            k += "Tanımlanmamış Kurum"+";"
    k = k[:-1]
    yazar_Kurum.append(k)
```

Kod 12: Yazar kurum verilerinin toplanması ve “yazar_Kurum” liste değişkenine aktarılması

Yazar kurum verilerini taşıyan “yazar_Kurum” liste değişkeni Kod 13 ile DataFrame’e değişken olarak eklenmiştir. Ortak yazarlı makalelerde, her makale için kurum adları DataFrame’de aralarında noktalı virgül kullanılarak kaydedilmiştir.

```
df['yazar_kurum'] = yazar_Kurum
```

Kod 13: “yazar_Kurum” geçici liste değişkeninin DataFrame’e değişken olarak aktarılması

Makale detay alanında “makale yılı, makale dili, makale türü, makale erişim türü” verilerinin toplanarak DataFrame’e eklenmesi için öncelikle Kod 14 ile her bir alan için aşağıdaki gibi liste değişkenler tanımlanmıştır.

```
makaleYil = []
makaleDil = []
makaleTuru = []
erisimTuru = []
```

Kod 14: Makale detay alanı için liste değişkenlerin oluşturulması

Daha sonra HTML verisi içerisinde ilgili alanlara ait bilgiler Kod 15 ile toplanarak Kod 14’teki geçici liste değişkenlerine aktarılmıştır.

```

for i in range(len(df['makale_URL'])):
    URL = df['makale_URL'][i]
    page = requests.get(URL)
    soup = BeautifulSoup(page.content, "html.parser")
    spans=soup.findAll('span', class_="me-3")

    for j in range(len(spans)):
        if "Yıl:" in spans[j].text:
            makaleYil.append(spans[j].text[5:])
        if "Metin Dili:" in spans[j].text:
            makaleDil.append(spans[j].text[13:-2])
        if "Makale Türü:" in spans[j].text:
            makaleTuru.append(spans[j].text[12:-4].strip())
        if "Erişim Türü:" in spans[j].text:
            erisimTuru.append(spans[j].text[14:-2])

```

Kod 15: Makale detay alanı verilerinin toplanması ve liste değişkenlerine aktarılması

Kod 16 ile makale detay alanı için oluşturulan geçici liste değişkenler mevcut DataFrame’e değişken olarak eklenmiştir.

```

df['makale_Yil'] = makaleYil
df['makale_Dil'] = makaleDil
df['makale_Turu'] = makaleTuru
df['erisim_Turu'] = erisimTuru

```

Kod 16: Makale detay alanı liste değişkenlerinin DataFrame’e değişken olarak eklenmesi

TR Dizin makale detay alanında bulunan diğer üst veri alanlarından “makale görüntüleme sayısı ve makale indirme sayısı” verilerinin HTML içerisinden çekilmesi için Kod 17 ile liste değişkenler tanımlanmıştır.

```

goruntuleme = []
indirme = []

```

Kod 17: Görüntülenme ve indirme sayısı verileri için liste değişkenlerinin oluşturulması

Makale görüntülenme ve makale indirme sayıları Kod 18 ile HTML verisi içerisinden taplanarak bu iki liste değişkenine aktarılmıştır.

```

for i in range(len(df['makale_URL'])):
    if not validators.url(str(df['makale_URL'][i])):
        görüntuleme.append("0")
        indirme.append("0")
        continue
    URL = str(df['makale_URL'][i])
    try:
        page = requests.get(URL, timeout=30, verify=True)
        page.raise_for_status() # Check if the request was successful
    except requests.exceptions.RequestException as e:
        görüntuleme.append("Hata")
        indirme.append("Hata")
        continue
    soup = BeautifulSoup(page.content, "html.parser")
    görüntuleme_span = soup.find('span', class_='me-2')
    görüntuleme_sayisi = görüntuleme_span.text if görüntuleme_span else "Veri Yok"
    indirme_span = soup.find('span', class_='me-5')
    indirme_sayisi = indirme_span.text if indirme_span else "Veri Yok"
    görüntuleme.append(görüntuleme_sayisi)
    indirme.append(indirme_sayisi)

```

Kod 18: Görüntülenme ve indirme verilerinin toplanması ve liste değişkenlerine aktarılması

Kod 19 ile de makale görüntüleme ve indirme liste değişkenleri mevcut DataFrame'e değişken olarak eklenmiştir.

```

df['makale_Görüntuleme'] = görüntuleme
df['makale_Indirme'] = indirme

```

Kod 19: Görüntülenme ve indirme liste değişkenlerinin DataFrame'e aktarılması

TR Dizin makale detay alanında bulunan DOI linklerinin derlenmesinin amacı; yayınlara ait TR Dizin'de olmayan anahtar kelimeleri ve makale konu alanlarını DergiPark platformu üzerinden toplamak için referans olarak kullanmaktır. TR Dizindeki DOI linki aracılığıyla makalelerin DergiPark kaydına gidilmiştir. Kod 20 ile makale detay alanında bulunan DOI üst veri alanı için liste değişkeni oluşturulmuştur.

```

makaleDoi = []

```

Kod 20: DOI verisi için liste değişkeni oluşturulması

Makale detay sayfası HTML verisi içerisinde DOI verileri Kod 21 ile derlenerek "makaleDoi" liste değişkenine aktarılmıştır.

```

for i in range(len(df['makale_URL'])):
    URL = df['makale_URL'][i]
    page = requests.get(URL)
    soup = BeautifulSoup(page.content, "html.parser")
    doi = soup.find('a', attrs={'href': re.compile("^http://doi.org/")})
    if doi is not None:
        makaleDoi.append(str(doi.get('href')))
    else:
        makaleDoi.append("0")

```

Kod 21: DOI adreslerinin toplanması ve liste değişkenine aktarılması

Kod 22 ile DOI verileri için tanımlanan “makaleDoi” liste değişkeni DataFrame’e değişken olarak eklenmiştir.

```

df['DOI_Durumu'] = makaleDoi

```

Kod 22: “makaleDoi” geçici liste değişkeninin DataFrame’e değişken olarak eklenmesi

8. Adım: DergiPark platformundan anahtar kelimeler ve konu alanları verilerinin toplanması

TR Dizin’in makale özet ve makale detay sayfalarında kimi makalelerin anahtar kelime, makale konu alanı ve DOI gibi üst veri alanlarının eksik olduğu saptanmıştır (bkz. Şekil: 43, 44). TR Dizin makale detay sayfasında DOI verisi bulunan ve DOI adresi DergiPark olan makaleler için anahtar kelimeler ve konu alanı verileri DergiPark’tan toplanmıştır (bkz. Şekil 46). DOI linki bulunmayan veya DOI linki olup da erişim adresi DergiPark olmayan yayınlara ait anahtar kelime ve makale konu verileri toplanamamıştır. Bunun nedeni; DOI linki ile erişim sağlanan DergiPark haricindeki dergi web sayfalarında üst veri alanlarının standart bir yapıda olmamasıdır.



Şekil 46: Anahtar kelime ve konu alanı verilerinin toplandığı DergiPark arayüzü

TR Dizin makale detay sayfasında DOI linki bulunan ve erişim adresi DergiPark olan makalelere ait anahtar kelimelerin toplanması için öncelikle “keywords” adlı bir liste değişkeni tanımlanmıştır.

```
keywords = []
```

Kod 23: Anahtar kelimeler için liste değişken tanımlanması

Daha sonra TR Dizin üzerinde DOI adresi olan makaleleri tespit etmek için Python programlama dili “validators²³” paketi aktif hale getirilmiştir.

```
!pip install validators
import validators
import requests
```

Kod 24: Validators paketinin yüklenmesi ve aktifleştirilmesi

Kod 25 ile tespit edilen DOI adreslerinin içeriğinde “dergipark” ifadesi geçen makaleler için DergiPark platformuna gidilmiş ve her makale için “anahtar kelimeler” üst veri alanından veriler toplanarak “keywords” liste değişkenine aktarılmıştır.

²³ Python "validators" paketi, URL'yi doğrulayan, güvenli olup olmadığını test eden ve geçersiz anahtar sözcükleri ve karakterleri arayan bir kod kütüphanesidir (Zeeshan, 2022).

```

for i in range(len(df['makale_Doi'])):
    if str(df['makale_Doi'][i]) == "0":
        keywords.append("")
    else:
        URL = df['makale_Doi'][i]
        try:
            page = requests.get(URL, timeout = 20, verify=False)
            if "/pub/" in page.url:
                soup = BeautifulSoup(page.content, "html.parser")
                keys = soup.find('div', class_='article-keywords data-section')
                if keys is not None:
                    a = keys.findAll('a')
                    kelimeler = ""
                    for j in range(len(a)):
                        kelimeler += a[j].text+";"

                    keywords.append(kelimeler[:-1])
                else:
                    keywords.append("DP_değil")
            else:
                keywords.append("DP_değil")
        except requests.exceptions.Timeout:
            keywords.append("DP_değil")

```

Kod 25: DergiPark’taki anahtar kelimelerin toplanması ve “keywords” liste değişkenine aktarılması

Son olarak Kod 26 ile anahtar kelime verilerini barındıran “keywords” liste değişkeni DataFrame’e değişken olarak eklenmiştir.

```

df['makale_Keywords'] = keywords

```

Kod 26: Anahtar kelimeler verilerini taşıyan liste değişkenin DataFrame’e eklenmesi

DergiPark platformu üzerinden makale konu verilerinin toplanması için anahtar kelimelerin toplanması adımlarının aynısı kullanılmıştır. İlgili kod blokları aşağıdaki gibidir.

Makale konu alanı ve makalelerin erişim sayfası bağlantı verileri için Kod 27 ile “topic” ve “page_url” liste değişkenleri tanımlanmıştır.

```

topics = []
page_url = []

```

Kod 27: DergiPark makale konu alanları için liste değişkenlerin tanımlanması

Daha sonra HTML verisi içerisinde ilgili alanlara ait veriler Kod 28 ile derlenerek “topic” ve “page_url” liste değişkenlerine aktarılmıştır. TR Dizin sayfası makale detay sayfasında DOI linki olup erişim adresi DergiPark olmayan yayınlar için “topic” liste değişkenine “DP_değil” (DergiPark değil) şeklinde bilgi eklenmiştir.

```

for i in range(len(df['makale_URL'])):
    if validators.url(str(df['makale_Doi'][i])) is not True:
        topics.append("0")
        page_url.append("0")
    else:
        URL = str(df['makale_Doi'][i])
        try:
            page = requests.get(URL, timeout = 30, verify=False)
            page_url.append(page.url)
        except requests.exceptions.ConnectionError:
            requests.status_code = "Connection refused"
            topics.append("Connection refused")
            page_url.append(page.url)
            continue
        if "/pub/" in page.url:
            soup = BeautifulSoup(page.content, "html.parser")
            keys = soup.find('table', class_='record_properties table')
            a = keys.findAll('tr')
            topics.append(a[1].td.text.strip())
        else:
            topics.append("DP_değil")

```

Kod 28: DergiPark makale konu alanlarının ve URL adreslerinin toplanması

Son olarak “topic” ve “page_url” liste değişkenleri Kod 29 ile DataFrame’e değişken olarak eklenmiştir.

```

df['makale_Konu'] = topics
df['makale_web_Sayfasi'] = page_url

```

Kod 29: “topics ve page_url” liste değişkenlerinin DataFrame’e değişken olarak eklenmesi

TR Dizin’de indekslenen dergilerin DergiPark platformunu kullanmayı tercih etmeleri veya indekslenen dergi içeriğindeki bütün veri alanlarının TR Dizin’de ilgili makale üst veri alanlarına eksiksiz yansıtılması, verisi eksik olan üst veri alanlarına da erişimi sağlayacaktır.

9. Adım: Dergilerin Detay Sayfalarından Verilerin Toplanması

Ankara Üniversitesi adresli makalelerin yayınlandığı dergilere ulaşmak amacıyla daha önce makale özet alanından elde edilip DataFrame’e eklenen dergi URL (“dergi_URL”) değişkeni referans olarak kullanılmıştır. TR Dizin dergi detay sayfasındaki üst veri alanları aşağıdaki gibidir:

- Dergi dili
- Yayına başlama yılı
- TR Dizin’de indekslenme yılı
- Yıllık yayın sayısı
- Dergi konu kategorisi

- Yayınlanan aylar
- Dergi konu alanları
- Dergi makale sayısı
- Atıf sayısı
- Kendine atıf sayısı
- Atıf alan makale sayısı
- Dergi atıf ortalaması
- Kendine atıf oranı

Bu alanlar aşağıdaki kod blokları kullanılarak dergi detay arayüzünden toplanmıştır. Öncelikle HTML içerisinden derlenecek olan; “dergi dili, yayın yılı, dizin yılları, dergi yıllık yayın sayısı, dergi konu kategorisi, dergi yayın ayları, dergi konu alanı, dergi makale sayısı, derginin aldığı atıf sayısı, derginin kendine yaptığı atıf sayısı, atıf alan makale sayısı, dergi atıf ortalaması ve derginin kendine atıf oranı” veri alanları için Kod 30 ile liste değişkenleri oluşturulmuştur.

```

dergi_dili = []
yayin_yili = []
dizin_yillari = []
yillikYayin_sayisi = []
konu_kategorisi = []
yayin_aylari = []
konu_alani = []
dergi_makaleSayisi = []
aldigi_atifSayisi = []
kendine_atifSayisi = []
atifAlan_makaleSayisi = []
dergi_atifOrt = []
kendine_atifOrani = []

```

Kod 30: TR Dizin dergi sayfası için liste değişkenlerinin tanımlanması

Kod 31 ile her bir makalenin yayınlandığı dergi URL adresleri için bir bağlantı oluşturulmuş ve her dergiye ait detay sayfasının HTML verisi alınmıştır.

```

for i in range(len(df['dergi_URL'])):
    URL = df['dergi_URL'][i]
    page = requests.get(URL)
    soup = BeautifulSoup(page.content, "html.parser")
    try:
        s = soup.findAll("div", class_='col-xs-12 col-md-6 col-lg-6')
    except (TypeError, AttributeError):
        s="-----"
    dil = ""
    yıl = ""
    dizin=""
    yillikyayin = ""
    konuKat = ""
    yayinAylari = ""
    konuAlani = ""

```

Kod 31: TR Dizin dergi sayfası verilerinin toplanması (1)

Kod 32, 33 ve 34 ile daha önce elde edilen dergi detay sayfası HTML verisi içerisinde üst veri alanlarına ait veriler toplanarak Kod 30 ile tanımlanan liste değişkenlerine aktarılmıştır.

```

for j in range(len(s)):
    try:
        if "Yayın Dili" in s[j].text:
            dil = s[j].span.text.strip()
    except (TypeError, AttributeError):
        dil=""
    try:
        if "Yayına Başladığı Yıl" in s[j].text:
            yıl = s[j].span.text.strip()
    except (TypeError, AttributeError):
        yıl=""
    try:
        if "Dizinlendiği Yıllar" in s[j].text:
            dizin = s[j].span.text.strip()
    except (TypeError, AttributeError):
        dizin=""
    try:
        if "Yıllık Yayın Sayısı" in s[j].text:
            yillikyayin = s[j].span.text.strip()
    except (TypeError, AttributeError):
        yillikyayin=""
    try:
        if "Konu Kategorisi" in s[j].text:
            konuKat = s[j].span.text.strip()
    except (TypeError, AttributeError):
        konuKat=""

```

Kod 32: TR Dizin dergi sayfası verilerinin toplanması (2)

```

try:
    s = soup.findAll("div", class_='col-xs-12 col-md-12 col-lg-12')
except (TypeError, AttributeError):
    s="-----"
for j in range(len(s)):
    try:
        if "Yayın Yapması Gereken Aylar" in s[j].text:
            yayınAylari = s[j].span.text.strip()
    except (TypeError, AttributeError):
        yayınAylari=""
    try:
        if "Konu Alanları" in s[j].text:
            item = s[j].findAll('item', class_="badge subjects")
            for k in range(len(item)):
                konuAlani += item[k].text.strip()+";"
    except (TypeError, AttributeError):
        item=""

dergi_dili.append(dil)
yayin_yili.append(yil)
dizin_yillari.append(dizin)
yillikYayin_sayisi.append(yillikyayin)
konu_kategorisi.append(konuKat)
yayin_aylari.append(yayinAylari)
konu_alani.append(konuAlani[:-1])

```

Kod 33: TR Dizin dergi sayfası verilerinin toplanması (3)

```

x = soup.find('div', class_="row justify-content-around numeric-container")
if x is not None:
    if "Makale Sayısı" in x.text:
        drg_istatistik = x.findAll("div")

        dergi_makaleSayisi.append(drg_istatistik[0].text.strip("Makale Sayısı").strip())
        aldigi_atifSayisi.append(drg_istatistik[1].text.strip("Atıf Sayısı").strip())
        kendine_atifSayisi.append(drg_istatistik[2].text.strip("Kendine Atıf Sayısı").strip())
        atifAlan_makaleSayisi.append(drg_istatistik[3].text.strip("Atıf Alan Makale Sayısı").strip())
        dergi_atifOrt.append(drg_istatistik[4].text.strip("Atıf Ortalaması").strip())
        kendine_atifOrani.append(drg_istatistik[5].text.strip("Kendine Atıf Oranı").strip()[:])
    else:
        dergi_makaleSayisi.append("0")
        aldigi_atifSayisi.append("0")
        kendine_atifSayisi.append("0")
        atifAlan_makaleSayisi.append("0")
        dergi_atifOrt.append("0")
        kendine_atifOrani.append("0")

```

Kod 34: TR Dizin dergi sayfası verilerinin toplanması (4)

Kod 35 ile liste değişkenler birer değişken olacak şekilde DataFrame'e aktarılmıştır. Dergi detay sayfasından toplanan veriler DataFrame'de; "der_dili, der_yayin_yili, der_dizin_yillari, 'der_yillikYayin_sayisi, der_konu_kategorisi, der_yayin_aylari, der_konu_alanlari, der_makaleSayisi, der_aldigi_atifSayisi, der_kendine_atifSayisi, der_atifAlan_makaleSayisi, der_dergi_atifOrt, der_kendine_atifOrani" değişken adları ile kaydedilmiştir.

```

df['der_dili'] = dergi_dili
df['der_yayin_yili'] = yayin_yili
df['der_dizin_yillari'] = dizin_yillari
df['der_yillikYayin_sayisi'] = yillikYayin_sayisi
df['der_konu_kategorisi'] = konu_kategorisi
df['der_yayin_aylari'] = yayin_aylari
df['der_konu_alanlari'] = konu_alani
df['der_makaleSayisi'] = dergi_makaleSayisi
df['der_aldigi_atifSayisi'] = aldigi_atifSayisi
df['der_kendine_atifSayisi'] = kendine_atifSayisi
df['der_atifAlan_makaleSayisi'] = atifAlan_makaleSayisi
df['der_dergi_atifOrt'] = dergi_atifOrt
df['der_kendine_atifOrani'] = kendine_atifOrani

```

Kod 35: Liste değişkenlerinin DataFrame'e değişken olarak aktarılması

10. Adım: DataFrame'in kalıcı olarak CSV dosya türünde Google Drive'a kaydedilmesi

Geçici olarak DataFrame'de bulunan toplanmış bütün değişkenler, Kod 36 ile "CSV" formatında kalıcı olarak Google Drive depolama alanına kaydedilmiştir.

```

!pip install xlswriter
df.to_excel('/content/drive/MyDrive/Colab Notebooks/tez/demo_dataSet.csv',
            index=False, engine='xlswriter')

```

Kod 36: TR Dizin'den toplanan verilerin dosya olarak Google Drive'a kaydedilmesi

Ham verilerin toplanmasından sonraki adım, veri setinin ön analiz ve temizleme için hazır hale getirilmesidir. Oluşturulan veri seti ile ilgili ön analiz ve veri temizleme adımları bir sonraki başlıkta ele alınmıştır.

3.2 Veri Ön işleme (Data preprocessing)

Ön işleme süreci; verilerin önizlenmesi, temizlenmesi, değişkenlerin dönüştürülmesi, veri alanlarındaki eksik değerlerin tamamlanması veya varsayılan değerlerin atanması ve aykırı değerlerin ortalama ile değiştirilmesi işlemlerinden oluşmaktadır.

Yukarıdaki işlemlere ek olarak verileri keşfetmeye çalışırken, verileri kolay bir şekilde anlamak için değişkenlere veri görselleştirme uygulanabilir. Veri ön işleme sürecinde değişkenler arasındaki ilişkileri tespit etmek, aykırı değerleri bulmak, eksik verileri tespit etmek

gibi veriye dair bazı teknik işlemlerde çeşitli görselleştirme yöntemleri uygulanır. Bu süreç aynı zamanda keşifsel veri analizi²⁴ olarak da adlandırılmaktadır.

3.2.1 Veri Seti Ön İzleme

Öncelikle Google Drive depolama alanında bulunan ve bir önceki bölümde oluşturulan veri setine erişim sağlamak, bu veri setini Google Colab kodlama platformuna aktarmak ve tekrar bir DataFrame'e dönüştürmek gerekmektedir. Bu amaç için gerekli olan güvenlik izinleri için kod kütüphaneleri Kod 37 ile aktif hale getirilmiştir.

```
from google.colab import drive
drive.mount('/content/drive/')
```

Kod 37: Google Drive için erişim izinlerinin tanımlanması

Ayrıca verileri incelerken keşifsel veri analizi yapmak ve görselleştirmeleri oluşturmak için Kod 38 ile Google Colab platformunda “pandas”, “numpy” “matplotlib”, “pyplot” ve “seaborn” kütüphaneleri aktifleştirilmiş ve kod blokları arasında görselleştirme yapabilmek için “%matplotlib inline” kod satırı uygulanmıştır.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

Kod 38: Veri seti ön analiz işlemleri için kod kütüphanelerinin aktifleştirilmesi

İlgili kod kütüphaneleri aktif hale getirildikten sonra, Google Drive ortamındaki veri seti Kod 39 ile Google Colab platformuna yüklenerek DataFrame'e (iki boyutlu tablo yapısına) dönüştürülmüştür.

```
df = pd.read_excel("/content/drive/MyDrive/Colab Notebooks/tez/trDizin_veriSeti.xlsx")
```

Kod 39: Ön analiz işlemleri yapılacak veri setinin yüklenmesi ve DataFrame'e dönüştürülmesi

Veri setindeki satır (kayıt/makale) ve sütun (değişken) sayısını tespit etmek için analiz yapılacak DataFrame'e atanan “df” değişkeninin “shape” özelliği uygulanarak (“df.shape”), TR

²⁴ Keşifsel Veri Analizi: Keşifsel veri analizi (Exploratory Data Analysis-EDA), veri bilimcileri tarafından veri görselleştirme yöntemlerini kullanarak veri kümelerini analiz etmek için kullanılan istatistiksel bir yöntemdir (Kaynak: <https://www.ibm.com/tr-tr/cloud/learn/exploratory-data-analysis>).

Dizin veri setinin 7.436 satır ve 31 sütundan oluştuğu tespit edilmiştir (Kod 40). “shape”, veri setinde bulunan satır ve sütun sayısını gösteren bir DataFrame özelliğidir (property).

```
df.shape
(7436, 31)
```

Kod 40: Veri seti satır ve sütun sayısının görüntülenmesi

Veri seti dosyasındaki değişkenler (sütun adları), sütunlardaki dolu kayıt sayısı ve sütunların veri türünü özet olarak görebilmek için Kod 41 ile DataFrame’in bir fonksiyonu olan “info()” metodu uygulanmıştır.

```
df.info()
```

Kod 41: Veri seti özet bilgilerinin görüntülenmesi

Tablo 8: Veri alanları (değişken) özellikleri

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7436 entries, 0 to 7435
Data columns (total 31 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   makale_Adi                           7436 non-null   object
1   makale_konu_Kategorisi                7436 non-null   object
2   makale_Dil                           7436 non-null   object
3   makale_Konu                          1609 non-null   object
4   makale_Yil                           7436 non-null   int64
5   makale_Goruntuleme                   7436 non-null   int64
6   makale_Indirme                       7436 non-null   int64
7   atif_Sayisi                          2265 non-null   float64
8   makale_Keywords                      2161 non-null   object
9   DOI_Durumu                          7436 non-null   object
10  makale_web_Sayfasi                   4332 non-null   object
11  makale_URL                          7436 non-null   object
12  yazar_Kurum                         7436 non-null   object
13  yazar_Adi                           7436 non-null   object
14  makale_Turu                         7436 non-null   object
15  acik_Erisim                         7436 non-null   object
16  dergi_adi                           7436 non-null   object
17  der_dili                             7293 non-null   object
18  der_yayin_yili                      7436 non-null   int64
19  der_dizin_yillari                   7428 non-null   object
20  der_yillikYayin_sayisi              7436 non-null   int64
21  der_konu_kategorisi                 7436 non-null   object
22  der_yayin_aylari                   7428 non-null   object
23  der_konu_alanlari                  7436 non-null   object
24  der_makaleSayisi                   7436 non-null   int64
25  der_aldigi_atifSayisi               7436 non-null   int64
26  der_kendine_atifSayisi              7436 non-null   int64
27  der_atifAlan_makaleSayisi           7436 non-null   int64
28  der_dergi_atifOrt                   7436 non-null   float64
29  der_kendine_atifOrani               7436 non-null   float64
30  der_URL                            7436 non-null   object
dtypes: float64(3), int64(9), object(19)
memory usage: 1.8+ MB
```

Tablo 8’e göre veri setinde toplam 31 değişken bulunmaktadır. “Non-Null Count” alanı, her bir sütunun ne kadar veriye sahip olduğunu göstermektedir. Örneğin “makale_Adi” sütunundaki tüm satırlar (7.436) veri içerirken “makale_Konu” sütununun 1.609 satırında veri bulunduğu, geri kalanın veri içermediği görülmektedir. “Dtype” sütununda ise sütunların taşıdığı veri türü bilgisi yer almaktadır.

Tabloya göre 3 değişken float64 türünde, 9 değişken “int64” türünde ve 19 değişken ise “object” türünde veri barındırmaktadır. Bu değerlerin gruplanmış haline Kod 42 ile erişilebilmektedir.

```
df.dtypes.value_counts()

object      19
int64       9
float64     3
dtype: int64
```

Kod 42: Değişkenlerin veri türüne göre gruplanması

Veri setinde var olan değişkenleri görmek için DataFrame’in “DataFrame.columns” özelliği kullanılmaktadır. Bu özellik sayesinde bütün değişkenler bir arada görülebilmektedir. Kod 43 ile görüntülenen değişken adları listesi aşağıdaki gibidir:

```
df.columns

Index(['makale_Adi', 'makale_konu_Kategorisi', 'makale_Dil', 'makale_Konu',
      'makale_Yil', 'makale_Goruntuleme', 'makale_Indirme', 'atif_Sayisi',
      'makale_Keywords', 'DOI_Durumu', 'makale_web_Sayfasi', 'makale_URL',
      'yazar_Kurum', 'yazar_Adi', 'makale_Turu', 'acik_Erisim', 'dergi_adi',
      'der_dili', 'der_yayin_yili', 'der_dizin_yillari',
      'der_yillikYayin_sayisi', 'der_konu_kategorisi', 'der_yayin_aylari',
      'der_konu_alanlari', 'der_makaleSayisi', 'der_aldigi_atifSayisi',
      'der_kendine_atifSayisi', 'der_atifAlan_makaleSayisi',
      'der_dergi_atifOrt', 'der_kendine_atifOrani', 'der_URL'],
      dtype='object')
```

Kod 43: Veri setinde yer alan değişkenlerin (sütunların) görüntülenmesi

DataFrame’in “head()”metodu ile veri setindeki ilk beş satırın (Kod 44), “tail()” metodu ile de son beş satırın (Kod 45) içerdiği veriler değişken adları ile beraber görüntülenebilmektedir. Böylece değişkenlerin içerdiği veriler için ön izleme yapılarak veri seti ile ilgili genel bir fikir edinilebilmektedir.

```
df.head()
```

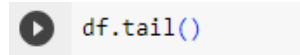
Kod 44: Veri setinde yer alan ilk beş satırın görüntülenmesi

Kod 44'ün uygulanması sonucunda görüntülenen ilk beş satırın verileri Tablo 9'daki gibidir:

Tablo 9: Veri setinin ilk beş satırının görünümü

	makale_Adi	makale_konu_Kategorisi	makale_Konu	makale_Yil	atif_Sayisi	makale_Keywords	dergiPark	makale_Doi	yazar_Kurum
0	Temporomandibular eklem disfonksiyonu açısında...	Fen	Radyolojik, Nükleer Tıbbi Görüntüleme	2021	0	Temporomandibular eklem disfonksiyonu; osteoar...	DergiPark	1	GAZİ ÜNİVERSİTESİ; ANKARA ÜNİVERSİTESİ; GAZİ ÜNİ...
1	Öğrencilerin Sınav Kavramına Yönelik Metaforla...	Sosyal	NaN	2017	4	NaN	Doi Yok	0	ALANYA ALAADDİN KEYKUBAT ÜNİVERSİTESİ; HAKKARİ ...
2	Cehennemde Evlilik: Türkiye'de Eğitim Denetimi...	Sosyal	NaN	2017	2	NaN	Doi Yok	0	ANKARA ÜNİVERSİTESİ; ANKARA ÜNİVERSİTESİ
3	'BİZ DE, KENDİ ÇAPIMIZDA AVRUPAYIZ': KÜRESELLE...	Sosyal	NaN	2018	2	NaN	Doi Yok	0	ANKARA ÜNİVERSİTESİ
4	"GEÇMİŞTEN GÜMÜMÜZE TÜRK HUKUKUNDA	Sosyal	NaN	2017	3	NaN	Doi Yok	0	ANKARA ÜNİVERSİTESİ; ANKARA ÜNİVERSİTESİ

"df.tail()" metodunun uygulanması (Kod 45) sonucunda ise veri setinin son beş satırı görüntülenmektedir.



Kod 45: Veri setinde yer alan son beş satırın görüntülenmesi

Kod 45'in uygulanması sonucunda görüntülenen son beş satırın verileri Tablo 10'daki gibidir.

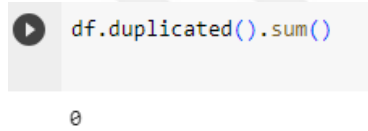
Tablo 10: Veri setinin son beş satırının görünümü

	makale_Adi	makale_konu_Kategorisi	makale_Konu	makale_Yil	atif_Sayisi	makale_Keywords	dergiPark	makale_Doi	yazar_Kurum
7431	"SADAKALLAHU'L-AZİM" DEDİĞİMİZDE NE DEMEK İSTİ...	Sosyal	Sosyal	2020	0	Qur'an; Qir'aat; Tilawat; Sidq; Human-Qur'an Commu...	DergiPark	1	ANKARA ÜNİVERSİTESİ
7432	"Senin derdin neydi ey Sâmîrî?" (20/Tâ-Hâ 95)	Sosyal	NaN	2017	1	NaN	Doi Yok	0	ANKARA ÜNİVERSİTESİ
7433	"Suyula Arınma"dan Modern Zamanların "Gelin Ham...	Sosyal	NaN	2020	1	NaN	DergiPark Değil	1	ANKARA ÜNİVERSİTESİ
7434	"UTILİTARIANİZM" VE "PRAGMATİZM": FARKLAR, ORT...	Sosyal	NaN	2019	1	NaN	Doi Yok	0	ANKARA ÜNİVERSİTESİ; SAKARYA ÜNİVERSİTESİ
7435	"Yetişkin Yeterliklerinin Uluslararası Değerle...	Sosyal	NaN	2018	1	NaN	DergiPark Değil	1	ANKARA ÜNİVERSİTESİ; ANKARA ÜNİVERSİTESİ; ANKARA...

Değişken adları belirtilmeden "head()" ve "tail()" metodları çalıştırıldığında, varsayılan olarak veri setinin belli sayıdaki sütunu ve beş satırı görüntülenebilmektedir. Görüntülenecek

satır sayısı ihtiyaca göre güncellenebilmektedir. Tablo 10 incelendiğinde, “makale_Konu” alanındaki boş kayıtların “NaN” olarak atandığı görülmektedir. “dergiPark” sütunu, ilgili yayının DergiPark platformunda kayıtlı olup olmadığının kontrol edilmesi için oluşturulmuştur. Eğer TR Dizin sayfasında makaleye ait “DOI” verisi yoksa “Doi Yok” ifadesi kullanılmıştır. Makaleye ait “DOI” kaydı varsa, ilgili linkin DergiPark sayfası olup olmadığı kontrol edilmiştir. İlgili link DergiPark’a gidiyorsa sütun verisi “DergiPark”, gitmiyorsa “DergiPark değil” şeklinde kaydedilmiştir.

Veri setinde bulunan kayıtlarda herhangi bir mükerrerliğin olup olmadığını tespit etmek için DataFrame’in “df.duplicated()” metodu kullanılmıştır. Kod 46’da belirtildiği gibi, uygulanan bu metod ile toplam 7.436 satırda herhangi bir mükerrerliğin olmadığı tespit edilmiştir.



```
df.duplicated().sum()
0
```

Kod 46: Veri setinde yer alan mükerrer kayıt sayısının görüntülenmesi

Görüleceği üzere veri ön izleme, diğer ön işlemleri gerçekleştirmeden önce veriyi keşfetmek için ilk adımdır. Ham verilerin yapısı ve özellikleri hakkında ön bilgi edinmek için veri setinin başlangıç incelemesidir. Veri ön işleme sürecinin diğer bir adımı olan “veri temizleme” takip eden başlıkta detaylı olarak incelenmiştir.

3.2.2 Veri Temizleme

Veri temizleme, veri setlerindeki hataların ve tutarsızlıkların analiz için kullanılabilecek şekilde belirlenmesi ve düzeltilmesi işlemidir. Veri ön işlemede önemli bir adım olup analiz için kullanılan verilerin kalitesini ve güvenilirliğini sağlamak için gereklidir (Roldós, 2021; Kowieski, 2022). Bu süreç, kopyaların, alakasız verilerin kaldırılması, formatların standartlaştırılması, veri setlerinin birleştirilmesi, veri bütünlüğünün kontrol edilmesi ve verilerin güvenli bir şekilde saklanması gibi adımları içerir (Kowieski, 2022).

Temiz olmayan veriler yanıltıcı öngörülere ve yanlış kararlara yol açarak kurumlar için önemli mali kayıplara neden olabilir (Roldós, 2021). Veri temizliği, insan hatalarından, tutarsız formatlardan, mükerrer girişlerden veya veri seti içindeki alakasız bilgilerden kaynaklanan yanlışlıkları önlemek için gereklidir. Bunun yanı sıra veri kalitesini artırması, kuruluşlarda karar alma için doğru ve güvenilir bilgiler sağlaması bakımından veri yönetimi için de kritik bir

adımdır (Stedman, t.y.). Bu çalışmada gerçekleştirilen temizleme adımları izleyen başlıklarda açıklanmıştır.

3.2.2.1 Ayıklama İşlemi

Bir önceki bölümde veri toplama aşamasında, gerek veri kontrolü (makale_URL, der_URL, makale_web_Sayfasi) ve veri doğrulama, gerekse başka bir üst veri alanındaki içeriği elde etmek amacıyla (makale_Adi, der_dizin_yillari ve der_yayin_aylari) ve referans öge olarak kullanılmak üzere bazı alanlar veri setine dâhil edilmişti. Bu haliyle veri seti 31 değişkenden (sütun) oluşmaktaydı. Ancak anılan değişkenler, yapılacak analizlerde ve görselleştirmelerde kullanılmayacağı için bu aşamada veri setinden ayıklanmıştır. Veri setinden ayıklanan değişken listesi aşağıda verilmiştir:

Tablo 11: Veri setinden ayıklanan değişken listesi

No	TR Dizin makale veri alanları	Veri seti için atanan değişken adı
1	Makale adı	makale_Adi
2	Makale TR Dizin URL'si	makale_URL
3	Makale web sayfası linki	makale_web_Sayfasi
4	Dergi TR Dizin URL'si	der_URL
5	Derginin dizinlendiği yıllar	der_dizin_yillari
6	Derginin yayınlandığı aylar	der_yayin_aylari

Analiz ve görselleştirme aşamasında kullanılmayacak değişkenler “DataFrame.drop()” metodu kullanılarak DataFrame’den silinmiştir. Tablo 3’teki değişkenlerin veri setinden ayıklanması için Kod 47 kullanılmıştır.

```
df.drop('makale_Adi',axis=1,inplace=True)
df.drop('makale_URL',axis=1,inplace=True)
df.drop('makale_web_Sayfasi',axis=1,inplace=True)
df.drop('der_URL',axis=1,inplace=True)
df.drop('der_dizin_yillari',axis=1,inplace=True)
df.drop('der_yayin_aylari',axis=1,inplace=True)
df.columns
```

Kod 47: Analizlerde kullanılmayacak değişkenlerin ayıklanması

”df.drop()” methodu için üç adet parametre kullanılmıştır. İlk parametre değişken adını, ikinci parametre (axis=1) silinecek alanın sütun olduğunu ve “inplace=True” parametresi ise

işlemin kalıcı olarak yapıldığını belirtmektedir. İlgili değişkenler ayıklandıktan sonra veri setindeki güncel değişken durumu Kod 48 ile aşağıdaki gibi görüntülenmiştir:

```
df.columns

Index(['makale_konu_Kategorisi', 'makale_Dil', 'makale_Konu', 'makale_Yil',
      'makale_Goruntuleme', 'makale_Indirme', 'atif_Sayisi',
      'makale_Keywords', 'DOI_Durumu', 'yazar_Kurum', 'yazar_Adi',
      'makale_Turu', 'acik_Erisim', 'dergi_adi', 'der_dili', 'der_yayin_yili',
      'der_yillikYayin_sayisi', 'der_konu_kategorisi', 'der_konu_alanlari',
      'der_makaleSayisi', 'der_aldigi_atifSayisi', 'der_kendine_atifSayisi',
      'der_atifAlan_makaleSayisi', 'der_dergi_atifOrt',
      'der_kendine_atifOrani'],
      dtype='object')
```

Kod 48: Veri seti değişkenleri (sütun adları)

Analizlerde kullanılmayacak değişkenlerin ayıklanmasından sonra veri setinde toplam 25 değişken kalmıştır. Veri setinde “yazar_Kurum” değişkenindeki verilerin bir kısmı küçük karakter bir kısmı ise büyük karakterlerden oluşmaktadır. Analizlerde beklenmedik sonuçların önüne geçmek için bu değişkendeki veriler Kod 49 ile başlık formatına dönüştürülerek standart hale getirilmiştir.

```
df.yazar_Kurum=df.yazar_Kurum.str.replace('İ','i').str.replace('ı','i').str.lower().str.title()
```

Kod 49: Yazar kurum verilerinin başlık formatına dönüştürülmesi

Kod 49, karakterlerden oluşan ve başlık formatına dönüştürülmek istenen her değişkene uygulanabilmektedir.

3.2.2.2 Veri Türlerinin Dönüştürülmesi

Bir veri setindeki veri türlerini dönüştürmek, veri görselleştirmenin, yorumlaması ve analizi açısından önemli bir adımdır. Bu adım, verileri çeşitli analitik görevlere daha uygun hale getirmek için verilerin formatının veya yapısının değiştirilmesini içerir. Veri setini temel alan anlamlı görselleştirmeler ve tablolar oluşturmak için tür dönüştürme yoluyla veri tutarlılığı ve doğruluğu sağlanır (Linkedin, 2024). Ayrıca veri türü dönüştürme, veri kalitesinin iyileştirilmesinde, verilerin erişilebilir, tutarlı ve analiz araçları ve yöntemleriyle uyumlu olmasını sağlamada önemli bir rol oynar.

Görselleştirmeler oluşturulurken, değişkenlerin veri türlerinin, oluşturulan görselleştirme türüyle uyumlu olması gerekmektedir. Örneğin, görselleştirme olarak çubuk grafik oluşturulacaksa ilgili değişkenin kategorik veri türünde olması gerekir²⁵.

DataFrame'in "dtypes" özelliği değişkenlerin hangi veri türünde olduğunu görmemizi sağlar. Veri setine Kod 50'nin uygulanması sonucunda DataFrame'in ilk aşamada değişkenlere varsayılan olarak atadığı veri türleri gözlemlenmiştir.

```
df.dtypes
```

Kod 50: Değişkenlerin veri türlerinin toplu olarak görüntülenmesi

"df.dtypes" özelliğinin uygulanması sonucunda DataFrame'in ilk aşamada varsayılan olarak her değişken için atadığı veri türleri Tablo 12'de görüntülenmiştir.

Tablo 12: Değişken veri türleri görünümü

makale_konu_Kategorisi	object
makale_Dil	object
makale_Konu	object
makale_Yil	int64
makale_Goruntuleme	int64
makale_Indirme	int64
atif_Sayisi	float64
makale_Keywords	object
DOI_Durumu	object
yazar_Kurum	object
yazar_Adi	object
makale_Turu	object
acik_Erisim	object
dergi_adi	object
der_dili	object
der_yayin_yili	int64
der_yillikYayin_sayisi	int64
der_konu_kategorisi	object
der_konu_alanlari	object
der_makaleSayisi	int64
der_aldigi_atifSayisi	int64
der_kendine_atifSayisi	int64
der_atifAlan_makaleSayisi	int64
der_dergi_atifOrt	float64
der_kendine_atifOrani	float64

Keşifsel veri analizi ve görselleştirmelerin hatasız olarak yapılabilmesi için dönüştürülecek olan değişken veri türleri Tablo 13'te listelenmiştir.

²⁵ Veri türü dönüştürme, görselleştirme sürecinde hataların veya beklenmeyen sonuçların önlenmesini sağlar. Örneğin bir sütun, bir dize (string) veri türü olarak temsil ediliyorsa, sayısal değerleri hesaplamak için kullanılamaz ve hesaplamalarda hatalara veya görselleştirmede beklenmeyen sonuçlara yol açabilir.

Tablo 13: Veri türü dönüştürülecek değişkenler

Sütun başlığı (veri alanı)	DataFrame'in atadığı veri türü	Dönüştürülen veri türü
makale_konu_Kategorisi	object	category
makale_Konu_	object	category
makale_Yil	Int64	category
atif_Sayisi	float64	Int64
makale_Dil	object	category
makale_Turu	object	category
acik_Erisim	object	category
der_dili	object	category
der_yayin_yili	Int64	category

Veri türünün float64 veya Int64 olarak değiştirilebilmesi için ilgili değişkende boş değer bulunmaması gerekmektedir. Boş alanlara varsayılan değerler atandıktan sonra dönüştürme işlemi yapılabilir. Tablo 13' te yer alan “atif_Sayisi” değişkeni için “Numpy²⁶ kod kütüphanesi kullanılarak boş kayıtlar için Kod 51 ile “0” değeri varsayılan olarak atanmıştır.

```
import numpy as np
df.atif_Sayisi.replace(np.nan, 0, inplace=True)
```

Kod 51: “atif_Sayisi” değişkenindeki boş kayıtlara varsayılan değer atama işlemi

İlgili sütuna varsayılan değerler atandıktan sonra veri türü dönüştürme işlemi yapılmıştır. Bunun için Tablo 13'teki veri alanlarına dair veri türü dönüştürme işlemi için, Kod 52'de gösterildiği gibi “astype()” metodu uygulanmıştır. “astype()” metodu, değişkenlerdeki verilerin veri türünü dönüştürmek için kullanılan bir DataFrame metodudur.

²⁶ Numpy, Python programlama dili ile sayısal hesaplamayı mümkün kılan açık kaynaklı bir kod kütüphanesidir (Numpy, 2023).

```
df = df.astype({
    "makale_konu_Kategorisi" : "category",
    "makale_Konu" : "category",
    "makale_Yil" : "category",
    "atif_Sayisi" : "int64",
    "makale_Dil" : "category",
    "makale_Turu" : "category",
    "acik_Erisim" : "category",
    "der_dili" : "category",
    "der_yayin_yili" : "category"
})
```

Kod 52: Değişkenlere ait veri türlerinin dönüştürülmesi

Veri türü dönüştürme işlemi yapıldıktan sonra değişkenlerin veri tipleri “df.dtypes” kodu ile Tablo 14’te görüntülenmiştir.

Tablo 14: Dönüştürme işlemi sonrası değişken veri türleri görünümü

makale_konu_Kategorisi	category
makale_Dil	category
makale_Konu	category
makale_Yil	category
makale_Goruntuleme	int64
makale_Indirme	int64
atif_Sayisi	int64
makale_Keywords	object
DOI_Durumu	object
yazar_Kurum	object
yazar_Adi	object
makale_Turu	category
acik_Erisim	category
dergi_adi	object
der_dili	category
der_yayin_yili	category
der_yillikYayin_sayisi	int64
der_konu_kategorisi	object
der_konu_alanlari	object
der_makaleSayisi	int64
der_aldigi_atifSayisi	int64
der_kendine_atifSayisi	int64
der_atifAlan_makaleSayisi	int64
der_dergi_atifOrt	float64
der_kendine_atifOrani	float64

Veri dönüştürme işlemi sonucunda, “df.dtypes” kodu ile üretilen Tablo 14 incelendiğinde, veri setinde 8 category, 2 float64, 8 int64 ve 7 object veri türünde olmak üzere toplam 25 değişkenin bulunduğu görülmektedir.

3.2.2.3 Veri Alanlarında Boş Değerlerin Tespit Edilmesi

Değişkenlere ait kayıp değerleri (missing values)²⁷ bulmak için DataFrame’e ait “isna()”²⁸ metodu uygulanarak (Kod 53) Tablo 15 elde edilmiştir.

```
df.isna().sum()
```

Kod 53: Değişkenlere ait boş kayıt sayısının görüntülenmesi

DataFrame’e Kod 53 uygulanarak değişkenlere ait boş değer olup olmadığı bilgisi Tablo 15’te verilmiştir. İlgili tablo incelendiğinde, makale konusu ve makale anahtar kelime alanlarının büyük oranda eksik olduğu tespit edilmiştir.

Tablo 15: Değişkenlere ait kayıp değer sayıları görünümü

```
makale_konu_Kategorisi    0
makale_Dil                0
makale_Konu              5827
makale_Yil                0
makale_Goruntuleme        0
makale_Indirme            0
atif_Sayisi              0
makale_Keywords          5275
DOI_Durumu                0
yazar_Kurum              0
yazar_Adi                0
makale_Turu              0
acik_Erisim              0
dergi_adi                0
der_dili                 143
der_yayin_yili           0
der_yillikYayin_sayisi   0
der_konu_kategorisi      0
der_konu_alanlari        0
der_makaleSayisi         0
der_aldigi_atifSayisi     0
der_kendine_atifSayisi    0
der_atifAlan_makaleSayisi 0
der_dergi_atifOrt         0
der_kendine_atifOrani     0
dtype: int64
```

Kayıp değerlerin görüntülenmesi için “seaborn” kod kütüphanesi kullanılmıştır. Seaborn, matplotlib tabanlı bir Python veri görselleştirme kod kütüphanesidir. Görsel olarak çekici ve bilgilendirici istatistiksel grafikler oluşturmak için çok sayıda fonksiyona sahip bir kütüphanedir (Waskom, 2021). Seaborn, Pandas veri yapısıyla bütünleşik çalışacak şekilde

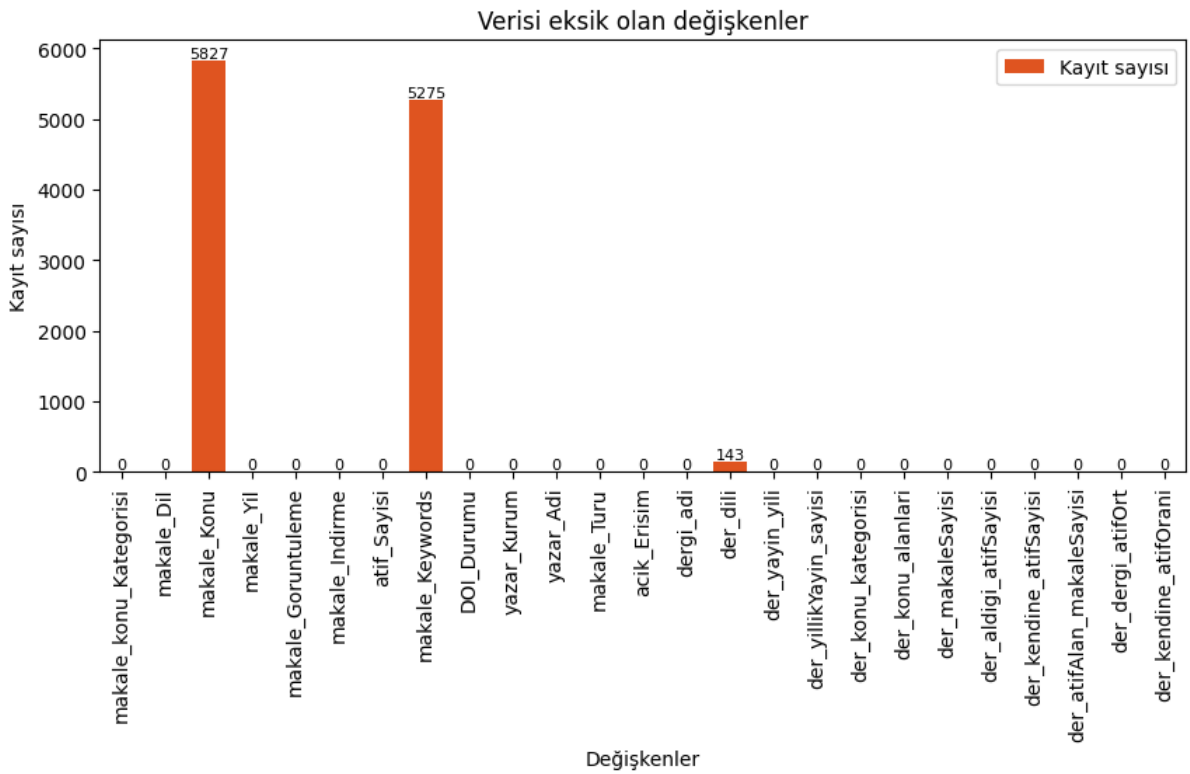
²⁷ Pandas’da kayıp değerler genellikle NaN (Not a Number) özel değeriyle temsil edilir. NaN değerleri eksik veya tanımsız değerleri temsil etmek için kullanılır (Pandas, 2023).

²⁸ “isna()”, bir değişkene ait boş kayıt sayısını gösteren bir DataFrame metodudur. Kayıt eğer boş ise “True” boş değil ise “False” şeklinde sonuç döndürür ve “True” olanların sayısını verir.

tasarlanmıştır. Veri seti değişkenlerinin ne kadarının boş değer içerdiğinin görselleştirilmesi için Kod 54 ile Şekil 47 oluşturulmuştur.

```
fig, ax = plt.subplots(1,1,figsize = (10,4))
sns.barplot(x=df.isna().sum().index,y=df.isna().sum().values, color='orangered', label='Kayıt sayısı')
for index,data in enumerate(df.isna().sum().values):
    ax.annotate(data, xy=(index,data+100), ha='center', va='center',fontsize=8)
plt.title("Verisi eksik olan değişkenler")
plt.ylabel("Kayıt sayısı")
plt.xlabel("Değişkenler")
plt.legend(fontsize=10)
plt.xticks(rotation = 90)
plt.show()
```

Kod 54: Değişkenlerin eksik kayıt sayılarının görselleştirilmesi



Şekil 47: Eksik verisi olan değişkenler

Veri setinde toplam 7.436 kayıt bulunmaktadır. Eksik verisi olan sütunlar sırasıyla makale konusu (5.827 kayıt), makale anahtar kelimeler (5.275 kayıt) ve dergi dili (143 kayıt) değişkenleridir. TR Dizin’de, akademik yayınlara ait makale konuları ile anahtar kelimeler ya yansıtılmamış ya da eksik yansıtılmıştır. Bu nedenle erişim adresi DergiPark olan yayınlar için ilgili alanlar DergiPark veri tabanından ayrıca toplanmıştır. Bu çalışmada kullanılan makale “DOI_Durumu” değişkeni, DergiPark sayfasında akademik yayınlara ait konuları ve anahtar kelimeleri toplamak için kullanılmış bir değişkendir.

3.2.2.4 Mükerrer Kayıtların Tespit Edilmesi

Aşağıda Kod 55 ile “Pandas” paketinin “duplicated()” metodu ile veri setinde mükerrer satır olup olmadığı kontrol edilmiştir. Uygulanan bu kod sonucunda veri setinde mükerrer satırların olmadığı tespit edilmiştir. Mükerrer kayıt olması durumunda aşağıda belirtildiği gibi ilgili satırlar “drop_duplicates()” metodu ile veri setinden ayıklanabilir.

```
df.duplicated().sum()  
df.drop_duplicates(keep='first', inplace=True)  
  
0
```

Kod 55: Mükerrer satırların tespit edilmesi ve ayıklanması

Pandas'taki duplicated() fonksiyonu, bir DataFrame içindeki yinelenen satırları tanımlamak için kullanılmaktadır. Bu fonksiyon, veri girişlerinin tekrarlandığı örnekleri vurgulayarak, gereksiz bilgilerin tespit edilmesine ve kaldırılmasına olanak tanıyarak veri kalitesi güvencesine yardımcı olur. Ayrıca fonksiyon, “subset” ve “keep” gibi parametreler aracılığıyla esneklik sağlayarak kullanıcıların kopyaları aramak için gereksinim duydukları sütunları belirlemelerine ve tekrarlanan bir satırın hangi örneğini benzersiz olarak kabul etmeleri gerektiğini belirlemelerine de olanak tanımaktadır.

Bu bölümde yöntem bağlamında veri toplama ve veri seti ön işleme süreçleri ele alınmıştır. Bu aşamada kullanılan araç ve yöntemlerin yanı sıra veri kaynağı kategorilerinin, alanlarının ve türlerinin belirlenmesini kapsayan veri toplama ayrıntılı bir şekilde incelenmiştir. Ayrıca web kazıma yoluyla bir veri kümesinin oluşturulması ele alınmıştır. Veri ön işleme kapsamında, verilerin analiz için hazırlanmasının önemi vurgulanmış ve ön analize tabi tutulmuştur. Veri kümesi ön işleme kapsamında veri setine; veri temizleme, tutarlılık için veri türlerini dönüştürme, bütünlük için veri alanlarındaki boş değerleri algılama ve veri doğruluğunu sağlamak için yinelenen kayıtları tanımlama gibi temel adımlar uygulanmıştır.

4.BÖLÜM: BULGULAR

4.1 Keşifsel Veri Analizi ve Görselleştirmeler

Bir veri setini oluşturan değişkenler üzerinde çeşitli görselleştirme türleri kullanılarak keşifsel veri analizi yapılabilmektedir. Keşifsel veri analizi (Exploratory data analysis-EDA), verilerin barındırdığı örüntüleri, aralarındaki ilişkileri ve dağılımlarını anlamaya odaklanan, bir veri analiz yaklaşımıdır. Bu istatistiksel yaklaşım 1977’de John Tukey tarafından geliştirilmiştir. Tukey keşifsel veri analizinin felsefesini “verilere bakarak, verilerin ne söylüyor olabileceğini görmek” şeklinde ifade etmiştir (Tukey 1977, s. v; Carranza, 2018, s. 1). Keşifsel veri analizi’nin amacı, daha fazla analiz veya karar verme konusunda bilgi sağlayabilecek veriler hakkında ön fikir edinmektir.

Bu çalışmada keşifsel veri analizi ve görselleştirmeler; değişkenlerin içerdiği verilerin özelliklerini anlamak, veri içerisindeki örüntüleri, eğilimleri tespit etmek, veri ön işleme bölümünde tespit edilen boş değer içeren değişkenlerin görselleştirmeye uygun olup olmadıkları ile ilgili değerlendirme yapmak ve verilerdeki hataları, aykırı değerleri ve tutarsızlıkları belirlemek için kullanılmıştır.

Bir önceki bölümde eldeki veri seti üzerinde veri ön işleme gerçekleştirilmiştir. Ayıklama, veri türlerinin dönüştürülmesi vb. temizleme işlemlerinden sonra veri setindeki değişkenler görselleştirmeler için hazır duruma getirilmiş ve veri setine Tablo 16’daki sorular yöneltilmiştir. Soruların bir kısmı; yayınların temel konu, dil ve tür gibi karakteristiklerini, akademisyenlerin tercih ettikleri dergileri ve yayınların etki durumu ile erişilebilirliklerini anlamaya, diğer bir kısmı ise; yazar, kurum ve konu başlıkları bazında iş birliği durumunu saptamak amacıyla sorulmuştur. Bu bölümde, keşifsel veri analizi ve görselleştirmeler ile bu sorular yanıtlanmaya çalışılmıştır.

Tablo 16: Veri setine yöneltilen sorular

Makale Demografik Bilgileri Bağlamındaki Sorular	
Soru 1	Makalelerin temel konu kategorilerine ²⁹ göre dağılımı nedir?
Soru 2	Makalelerin dillere göre dağılımı nedir? 2.1. Türkçe dışındaki dillerde yayınlanmış makaleler hangi konu kategorisinde yer almaktadır?
Soru 3	Makalelerin türlere göre atıf alma durumları nedir? 3.1 Türlerin konu ve yıla göre dağılımı nedir? 3.2 En çok atıf alan makale türleri hangileridir?
Soru 4	Makale sayıları ile atıf sayılarının yıllara göre dağılımı nedir? 4.1 Makalelerin yıllara göre görüntülenme ve indirilme sayıları nedir?
Soru 5	Ulaşılabilirlik açısından makalelerin erişim türüne (açık/kapalı) göre dağılımı nedir? 5.1 Hangi temel konulu makaleler erişime kapalıdır; bunlar yıllar içinde artış gösterme eğiliminde midir?
İş Birliği Durumunun Saptanmasına Yönelik Sorular	
Soru 6	Yazar iş birliği (tek yazarlı/ortak yazarlı) eğilimi ne düzeydedir? 6.1 Yazar sayısına göre yayın sayısı dağılımı nedir? 6.2 Makale başına düşen yazar sayısına göre atıf alma durumu nedir? 6.3 Tek yazarlı makaleler ile çok yazarlı makalelerin atıf oranları ne düzeydedir? 6.4 Kurumsal iş birliği ne düzeydedir? 6.5 Kurumsal iş birliğinin yıla göre dağılımı nedir? 6.6 Kamu iş birliği olan makalelerde ilk yazara göre yazar kurum dağılımı nedir?
Soru 7	Makalelerin konusal dağılımı nedir?

²⁹ TR Dizinde makalelere ait bibliyografik kimliklerin gruplandırıldıkları temel kategoriler.

	7.1 Türkçe makalelerin konusal olarak yıl ve genel konu kategorisi bazındaki dağılımı nedir? 7.2 İngilizce makalelerin konusal olarak yıl ve genel konu kategorisi bazındaki dağılımı nedir? 7.3 Türkçe anahtar kelimelerin yıllara göre dağılımı nedir? 7.4 İngilizce anahtar kelimelerin yıllara göre dağılımı nedir?
Tercih Edilen Dergiler Bağlamındaki Sorular	
Soru 8	Makalelerin yayınlandığı dergi sayılarının yıllara göre dağılımı nedir? 8.1 Dergi sayısı artış oranı ile yayın sayısı artış oranı arasında bir korelasyon var mıdır? 8.2 Yıllara göre dergi başına düşen makale sayıları nedir? 8.3 AÜ yazarları hangi dergileri tercih etmektedir? 8.4 AÜ makalelerinin yayınlandığı dergilerin yayın hayatına başlama tarihine göre dağılımı nedir? 8.5 AÜ makalelerinin yayınlandığı dergilerin yıllık yayın periyodları, makale sayıları, aldığı atıf sayıları ve kendine atıf sayıları nedir?

Keşifsel veri analizi ve görselleştirmeler oluşturmak için Python programlama dilinin “pandas, numpy, matplotlib, seaborn ve plotly” kod kütüphaneleri kullanılmıştır. Tablo 16’daki soruların yanıtlarını bulmak için veri setinde bulunan değişkenler aracılığıyla sorgular oluşturulmuştur. Ayrıca daha ayrıntılı görseller üretme ihtiyacının olduğu sorularda var olan değişkenlerdeki veriler kullanılarak yeni değişkenler eklenmiş ve filtrelemeler yapılarak yeni alt veri setleri oluşturulmuştur. Örneğin; “Yazar” veri alanında yazar iş birliği durumunun “Kurum” veri alanında ise kurum iş birliği durumunun tespit edilebilmesi ve görselleştirilmeye hazır hale getirilmesi için ilgili veri alanlarından nümerik değerlerin olduğu yazar-sayısı ve kurum-sayısı gibi değişkenler oluşturulmuştur. Bu yeni değişkenler ve veri setinde var olan diğer değişkenler bir araya getirilerek yeni alt veri setleri yaratılmış ve görselleştirilmiştir. Alt veri seti/setleri oluşturmak; büyük veri kümelerini daha yönetilebilir hale getirmeyi ve veri setinde yer alan belirli değişkenlere veya alt kümelere odaklanmayı kolaylaştıran bir işlemdir. Oluşturulan filtrelemeler, alt veri setleri ve görselleştirmeler için kullanılan kod blokları ilgili soruların altında açıklamaları ile verilmiştir.

4.1.1 Soru 1: Makalelerin temel konu kategorilerine göre dağılımı nedir?

Yıllara göre temel konu kategorisi dağılımını görmek için Kod 56 ile veri setindeki “makale_konu_Kategorisi” ve “makale_Yil” değişkenleri gruplanmıştır.

```
df.groupby('makale_Yil')['makale_konu_Kategorisi'].value_counts().index
```

```
MultiIndex([(2017, 'Sosyal'),  
            (2017, 'Fen'),  
            (2018, 'Fen'),  
            (2018, 'Sosyal'),  
            (2019, 'Fen'),  
            (2019, 'Sosyal'),  
            (2020, 'Fen'),  
            (2020, 'Sosyal'),  
            (2021, 'Fen'),  
            (2021, 'Sosyal')],  
            names=['makale_Yil', 'makale_konu_Kategorisi'])
```

Kod 56: Makale temel konu kategorisi ögeleri

TR Dizin’deki makalelerin “Temel konu kategorisi” üst verisinin “Fen ve Sosyal” olarak iki temel grupta toplandığı tespit edilmiştir. Makalelerin yıllara göre sayısal dağılımını görmek için ise Kod 57 ile aynı değişkenler süzülerek “yil_df” adlı yeni bir alt veri seti oluşturulmuştur.

```
yil_df = pd.crosstab(df.makale_Yil, df.makale_konu_Kategorisi)  
yil_df = yil_df.div(yil_df.sum(axis=1), axis=0)  
yil_df
```

Kod 57: Makale yılı ve makale konu kategorisinden oluşan “yil_df” alt veri setinin oluşturulması

Kod 57’de, “yil_df” alt veri setini oluşturmak için, DataFrame’in “crosstab()” fonksiyonu ile “makale_yil” değişkeninde bulunan veriler tekile indirgenerek indekse dönüştürülmüş ve “makale_konu_Kategorisi” değişkeni yıllara göre gruplandırılmıştır. Ayrıca oluşturulacak görselde verilerin oran olarak gösterilmesi için “makale_konu_Kategorisi” verileri “Fen ve Sosyal” şeklinde iki ayrı değişkene dönüştürülerek verilerin yüzdelik oranları satır bazında hesaplanmıştır. Tablo 17(a) “df” veri setinin ilk beş kaydını, Tablo 17(b) ise “yil_df” alt veri setinin görünümünü vermektedir.

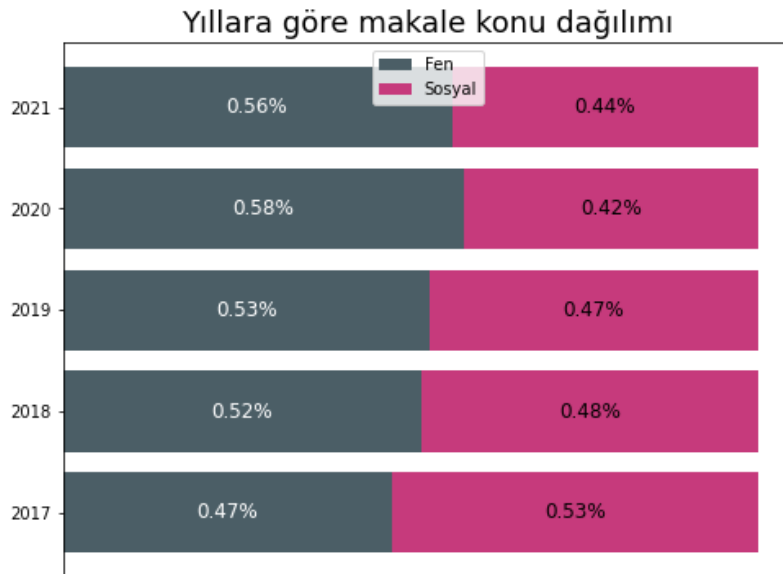
Tablo 17: Makale yılı ve makale konu kategorisinden oluşan alt veri seti (yil_df)

(a) “df” veri seti görüntüsü			(b) “yil_df” alt veri seti görüntüsü		
makale_Yil		makale_konu_Kategorisi		makale_konu_Kategorisi	
				Fen Sosyal	
				makale_Yil	
0	2021	Fen		2017	0.474699 0.525301
1	2017	Sosyal		2018	0.517799 0.482201
2	2017	Sosyal		2019	0.530839 0.469161
3	2018	Sosyal		2020	0.577307 0.422693
4	2017	Sosyal		2021	0.563477 0.436523

Tablo 17 (b)’deki "yil_df" alt veri seti aşağıda sunulan Kod 57 ile Şekil 48’de görselleştirilmiştir.

```
fig,ax=plt.subplots(figsize=(8,6))
ax.barh(yil_df.index,yil_df['Fen'],color='#4b5e66',label='Fen')
ax.barh(yil_df.index,yil_df['Sosyal'],color='#c63a7c',label='Sosyal',left=yil_df['Fen'])
for index in yil_df.index:
    ax.annotate((str(round(yil_df['Fen'].loc[index],2)))+'%',xy=(yil_df['Fen'].loc[index]/2,index),
                ha='center',va='center',fontsize=12,color='ffffff')
    ax.annotate((str(round(yil_df['Sosyal'].loc[index],2)))+'%',
                xy=(yil_df['Fen'].loc[index]+yil_df['Sosyal'].loc[index]/2,index),
                ha='center',va='center',fontsize=12,color='000000')
ax.set_xticks([])
ax.set_label(yil_df.index)
plt.title('Yıllara göre makale konu dağılımı', fontsize=18)
plt.legend(loc='upper center')
plt.show()
```

Kod 58: “yil_df” alt veri setinin görselleştirilmesi



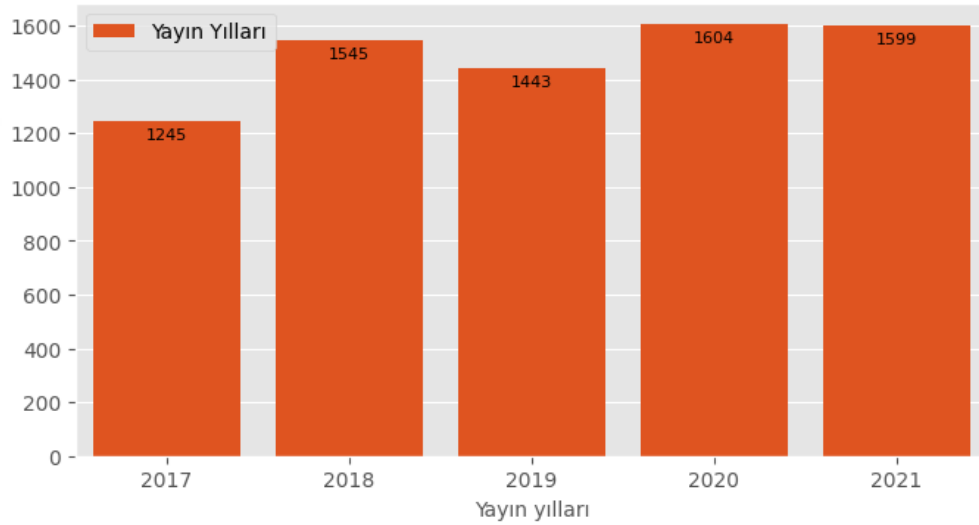
Şekil 48: Yıllara göre makale konu kategorisi dağılımı

TR Dizin veri tabanındaki makalelere ait bibliyografik kimlikler “Fen Bilimleri” ve “Sosyal Bilimler” ana konu kategorilerinde gruplanmaktadır. Şekil 48’e bakıldığında, 2017 yılı hariç diğer yıllarda Fen Bilimleri ana konu kategorisindeki makale sayısının daha fazla olduğu görülmektedir.

Makalelerin yıllara göre sayısal dağılımı ayrıca görselleştirilmiştir. “makale_Yıl” değişkeninin görselleştirilmesi için “seaborn” görselleştirme paketinin barplot() (çubuk grafik) türü kullanılmıştır. Seborn ile oluşturulan görselleştirme için uygulanan kod bloğu (Kod 59) ve oluşturulan görsel (Şekil 49) aşağıdaki gibidir:

```
fig, ax1=plt.subplots(1,1,figsize=(8,4))
sns.barplot(x=df.makale_Yil.value_counts().index,y=df.makale_Yil.value_counts().values,
            color='orangered', label='Yayın Yılları')
for index, value in enumerate(df.makale_Yil.value_counts().sort_index().values):
    ax1.annotate(value, xy=(index,value-50), ha='center', va='center',fontsize=8)
plt.legend(fontsize=10)
plt.xlabel('Yayın yılları', fontsize=10)
plt.show()
```

Kod 59: Yıllara göre yayın sayısı dağılımı



Şekil 49: Yıllara göre yayın sayısı dağılımı

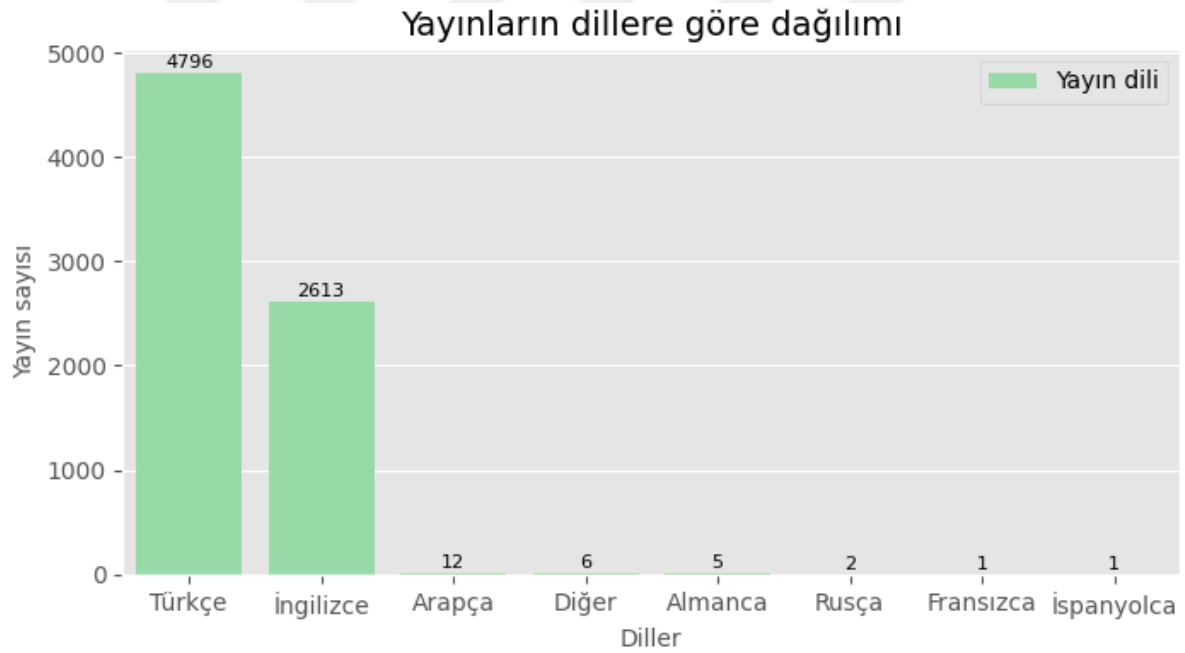
Şekil 49’daki veriler, yıllara göre makale sayılarında düzenli bir artışın ya da azalışın olmadığını göstermektedir. 2017 yılında 1.245 makale yayımlanmış, 2018 yılında bu sayı 300 artışla 1.545’i bulmuştur, 2019 yılında makale sayısında bir azalma yaşanmış, 2020 yılında artış olmuş ve 2021 yılında makale sayısında tekrar bir azalma gözlenmiştir. 2020 yılı 1.604 makale ile makale sayısının en fazla olduğu yıl olmuştur. Yıllara göre makale sayısında doğrusal bir artış eğiliminin olmadığı görülmektedir.

4.1.2 Soru 2: Makalelerin dillere göre dağılımı nedir?

Makalelerin yayınlandığı dile göre dağılımını görmek için veri setindeki “makale_konu_Kategorisi” ve “makale_Dil” değişkenleri kullanılmıştır. Görselleştirme için ise Kod 60 ile “seaborn” un “barplot” grafiği kullanılarak Şekil 50 elde edilmiştir.

```
import seaborn as sns
fig, ax1=plt.subplots(1,1,figsize=(8,4))
sns.barplot(x=df.makale_Dil.value_counts().index,y=df.makale_Dil.value_counts().values,
            color='#8de5a1', label='Yayın dili')
for index, value in enumerate(df.makale_Dil.value_counts()):
    ax1.annotate(value, xy=(index,value+100), ha='center', va='center',fontsize=8)
plt.legend(fontsize=10)
plt.xlabel('Diller', fontsize=10)
plt.ylabel('Yayın sayısı', fontsize=10)
plt.title('Yayınların dillere göre dağılımı', fontsize=14)
plt.show()
```

Kod 60: Yayınların dilleri göre dağılımı



Şekil 50: Makalelerin dillere göre dağılımı

Grafiğe göre Türkçe dilinden başka yoğunluklu olarak İngilizce dilinde makale yapıldığı ve İngilizceden sonra makalelerin en çok Arapça dilinde olduğu görülmektedir. “Diğer” ögesinde hangi dillerin olduğu tespit edilememiştir.

4.1.2.1 Soru 2.1 Türkçe dışındaki dillerde yayınlanmış makaleler hangi konu kategorisinde yer almaktadır?

Türkçe dışındaki dillerde, hangi temel konu kategorisinde makale yapma eğilimi olduğunu saptayabilmek için Kod 61 ile “makale_Yil, makale_konu_Kategorisi ve makale_Dil” değişkenlerinden oluşan gruplanmış “dil_df” adında bir alt veri seti oluşturulmuştur.

```
dil_df = df.groupby(["makale_Yil", "makale_konu_Kategorisi", "makale_Dil"]).size().reset_index(name='sayi')
dil_df = dil_df.loc[dil_df['makale_Dil'] != "Türkçe"].reset_index()
dil_df = dil_df.drop('index', axis=1)
```

Kod 61: “dil_df” alt veri setinin oluşturulması

29 satırdan³⁰ oluşan “dil_df” alt veri setinin son beş satırı “dil_df.tail()” fonksiyonu kullanılarak aşağıdaki gibi ekrana yansıtılmıştır. “tail()” fonksiyonu kullanılarak hem sütun ve satırların hangi verileri içerdiği hem de veri setinin kaç satırdan oluştuğu bilgisi görüntülenebilmektedir. Alt veri setinde “makale_Yil, makale_konu_Kategorisi ve makale_Dil” değişkenleri gruplanmış ve bunların sayısal değerleri “sayı” değişkeninde verilmiştir.

Tablo 18: “dil_df” alt veri seti son beş kayıt görüntüsü

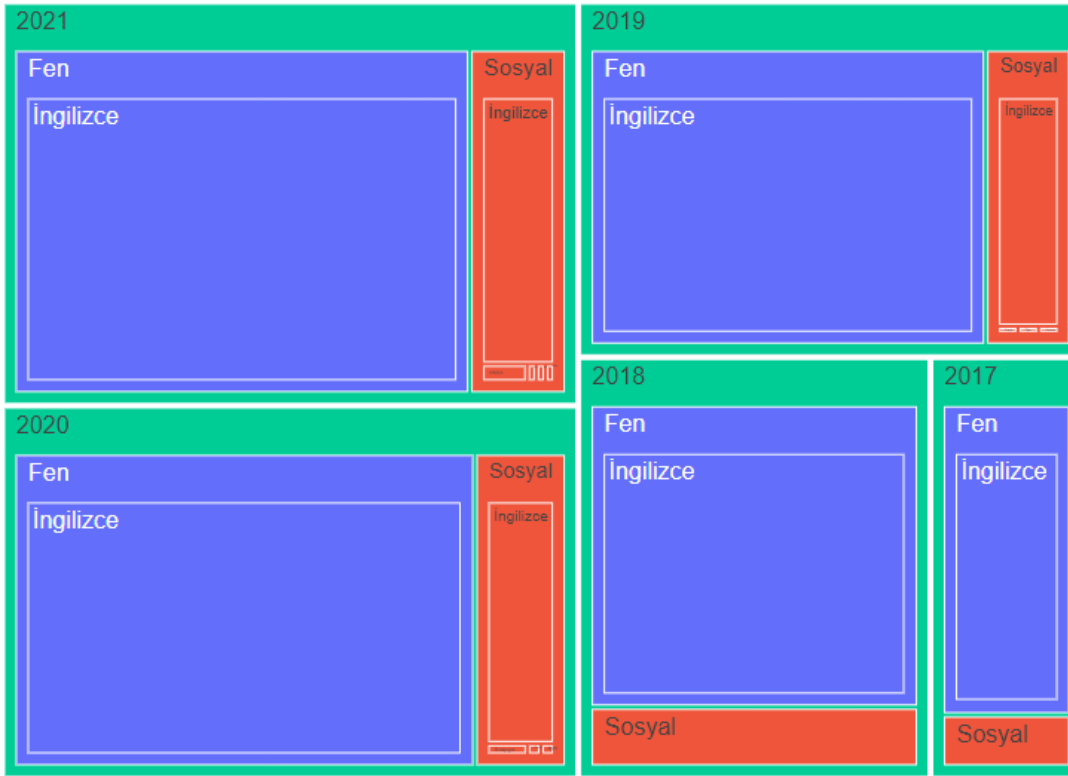
	makale_Yil	makale_konu_Kategorisi	makale_Dil	sayı
24	2021	Sosyal	Almanca	1
25	2021	Sosyal	Arapça	5
26	2021	Sosyal	Diğer	1
27	2021	Sosyal	İngilizce	120
28	2021	Sosyal	İspanyolca	1

“dil_df” alt veri seti Kod 62 kullanılarak Şekil 51’deki etkileşimli “Treemap” grafiğine dönüştürülmüştür. Treemap grafiği, Python’ın “plotly.express” kütüphanesinden yararlanılarak elde edilmiştir.

```
import plotly.express as px
px.treemap(dil, color='makale_konu_Kategorisi', values="sayı",
           path=['makale_Yil', 'makale_konu_Kategorisi', 'makale_Dil'], height=700, width=1000)
```

Kod 62: “Dil” alt veri setinin “Treemap” olarak görselleştirilmesi

³⁰ DataFrame’lerde index 0’dan başladığı için son index numarasının 28 olduğu görülmektedir.



Şekil 51: Dil alt veri setinin “Treemap” olarak görselleştirilmesi

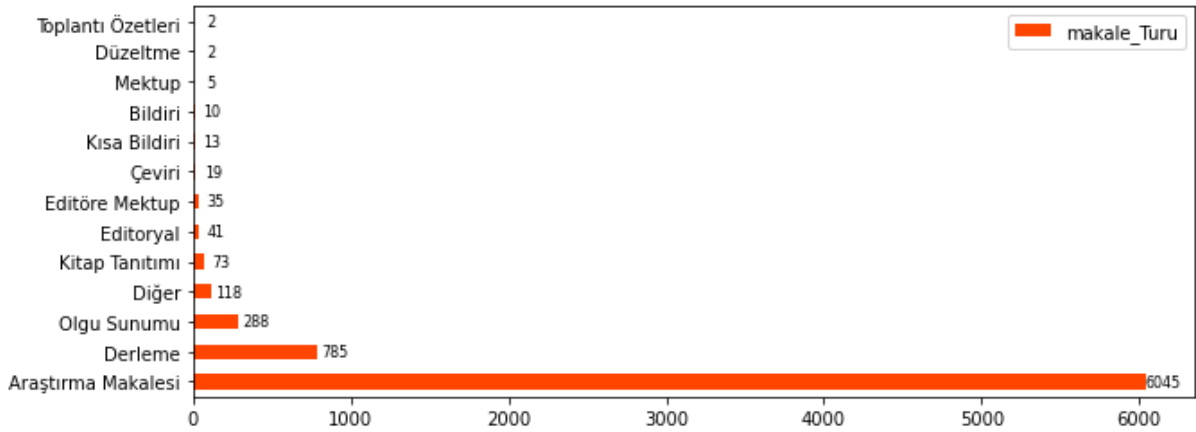
Etkileşimli grafikte imlecin yıl alanları üzerinde gezdirilmesi ile yabancı dilde yayınlanan makalelerin sayısı görülebilmektedir; en çok yayının 2021 yılında (732 adet), en az yayının ise 2017 yılında (205 adet) yayınlandığı tespit edilmiştir. Bu yayınların büyük oranda İngilizce dilinde olduğu, Arapça (12), Almanca (5), Rusça (2), Fransızca (1) ve İspanyolca (1) dillerinde nadir de olsa yayın yapıldığı görülmektedir. Ayrıca Şekil 48’e bakıldığında, makale dili Türkçe olmayan yayınların ağırlıklı olarak “Fen” genel konu kategorisinde olduğu tespit edilmiştir.

4.1.3 Soru 3: Makalelerin türlere göre atıf alma durumları nedir?

Makale türü değişkeninde bulunan tür adlarını ve bu türlerdeki makale sayısını elde etmek için Kod 63 ile Şekil 52’deki yatay çubuk grafik oluşturulmuştur. Yatay çubuk grafik, matplotlib görselleştirme kütüphanesinin “barh” grafik türü kullanılarak oluşturulmuştur.

```
df.makale_Turu.value_counts().plot(kind='barh',figsize=(10, 4), color = "orangered")
for index, value in enumerate(df.makale_Turu.value_counts()):
    plt.annotate(value, xy=(110+value,index), ha='center', va='center',fontsize=8)
```

Kod 63: Makale türüne göre dağılım



Şekil 52: Yayınların makale türüne göre dağılımı

Şekil 52 incelendiğinde, “toplantı özeti, mektup, bildiri, kısa bildiri, editöre mektup, editoryal, kitap tanıtımı, diğer” gibi bilimsel makale olmayan türlerin de bulunduğu gözlemlenmektedir. Makale türü değişkenine ait veriler, TR Dizin sayfasında yayın detay alanında “Makale Türü” üst verisinden toplandığı için TR Dizin’in bir dergi içerisindeki yayınların hakemli olup olmamasına bakmaksızın bütün içeriği “Makale türü” etiketi alanında dizinlendiği düşünülmektedir. Bu ayrımın yapılabilmesi için TR Dizin sayfasında ilgili alanda “hakemli yayınlar” ve “diğer yayınlar” öğelerini içeren bir veri alanını tanımlanabilir. Böylece hakemli yayınlara daha kolay erişim sağlanacaktır.

4.1.3.1 Soru 3.1 Türlerin konu ve yıla göre dağılımı nedir?

Kod 64 ile “makale_Yil ve makale_Turu” değişkenlerinden oluşan “tur_df” alt veri seti oluşturulmuş ve Tablo 19’da alt veri seti görüntülenmiştir. Alt veri setinde “makale_Yil” değişkeni indekse dönüştürülerek “makale_Turu” değişkenindeki öğeler yıla göre gruplandırılmıştır. Böylece her gruba ait yıllık veriler incelenebilmektedir.

```

tur_dict={}
for yil in df['makale_Turu'].value_counts().index:
    tur_dict[yil]=df[df['makale_Turu']==yil]['makale_Yil'].value_counts()

tur_df=pd.DataFrame(tur_dict)
tur_df.index.rename('index',inplace=True)
display(tur_df)

```

Kod 64: “tur_df” alt veri setinin oluşturulması

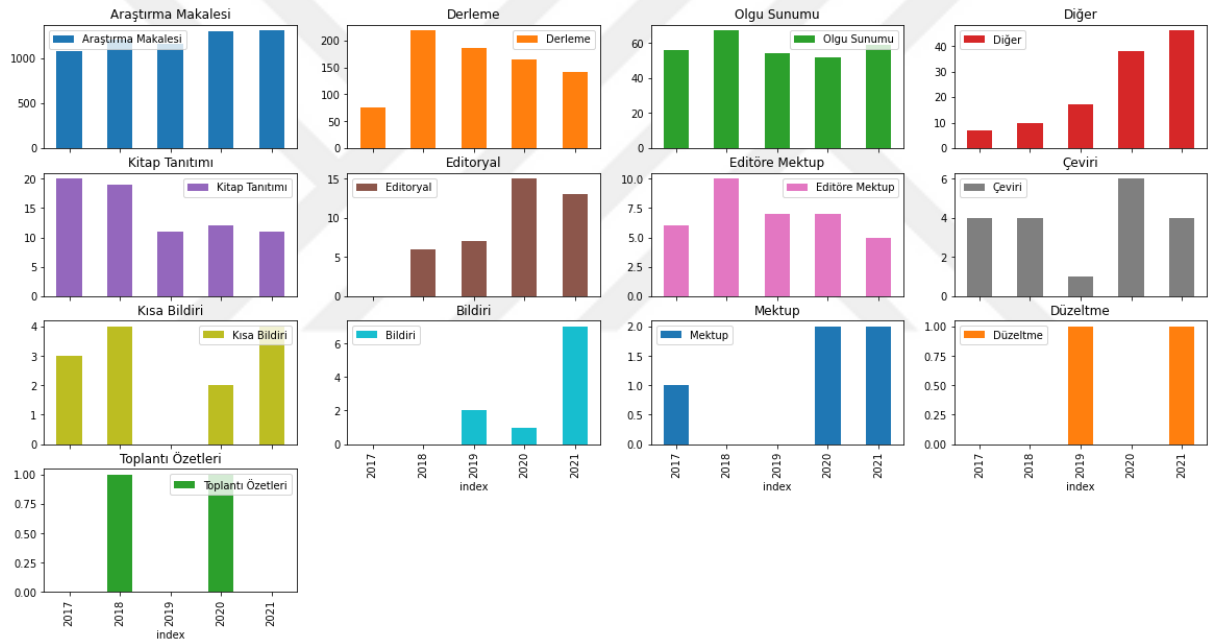
Tablo 19: “tur_df” alt veri seti görünümü

	Araştırma Makalesi	Derleme	Olgu Sunumu	Diğer	Kitap Tanıtımı	Editoryal	Editöre Mektup	Çeviri	Kısa Bildiri	Bildiri	Mektup	Toplantı Özetleri	Düzeltilme
index													
2017	1073	75	56	7	20	NaN	6	4	3.0	NaN	1.0	NaN	NaN
2018	1206	218	67	10	19	6.0	10	4	4.0	NaN	NaN	1.0	NaN
2019	1158	185	54	17	11	7.0	7	1	NaN	2.0	NaN	NaN	1.0
2020	1303	165	52	38	12	15.0	7	6	2.0	1.0	2.0	1.0	NaN
2021	1305	142	59	46	11	13.0	5	4	4.0	7.0	2.0	NaN	1.0

“tur_df” alt veri seti, Kod 65 ile matplotlib görselleştirme paketinde yer alan “plot()” metodunun “subplot=True” parametresi kullanılarak Şekil 53 oluşturulmuştur. “subplot=True” ile ilgili görselde her makale türü ögesi için yıllara göre dağılım verilmiştir.

```
tur_df.plot(kind='bar', subplots=True, layout=(4,4), figsize=(20,10))
```

Kod 65: “tur_df” alt veri setinin görselleştirilmesi



Şekil 53: Makale türlerinin yıllara göre dağılımı

Şekil 53 incelendiğinde; araştırma makalelerinin çoğunlukta olduğu, bunu sırasıyla derleme ve olgu sunumu türlerinin takip ettiği, “araştırma makalesi” ve “diğer” türlerinde yıllara göre artış olduğu, “derleme” ve “mektup” türlerinde ise 2018 yılından itibaren düşüş olduğu görülmektedir. En az makalenin ise sırasıyla “toplantı özetleri, düzeltilme ve mektup” türlerinde olduğu görülmektedir. Yıllara göre yayın sayısı bulunmayan makale türleri incelendiğinde; 2017 yılında “editoryal”, “bildiri”, “toplantı özetleri” ve “düzeltilme” türleri, 2018 yılında “bildiri”, “mektup” ve “düzeltilme” türleri, 2019 yılında “kısa bildiri”, “mektup”

ve “toplantı özetleri” türleri ve son olarak 2021 yılında “toplantı özetleri” türüne dair makale bulunmamaktadır.

Makalelerin konu kategorisine göre kümelenecek, her küme altındaki makale türlerinin dağılımını görmek için daha önce oluşturulan “tur_df” (bkz. Tablo 19) alt veri setine “temel_konu_Kategorisi” değişkeni eklenmiştir. “tur_df” alt veri seti Kod 66 ile “temel_konu_Kategorisi” ve “makale_Turu” değişkenlerine göre gruplanarak güncellenmiş ve ilk beş satırlık görünüm Tablo 20’de verilmiştir.

```
tur_df = df.groupby(["makale_konu_Kategorisi", "makale_Turu"]).size().reset_index(name='sayi')  
tur_df.head()
```

Kod 66: “tur_df” alt veri setinin güncellenmesi

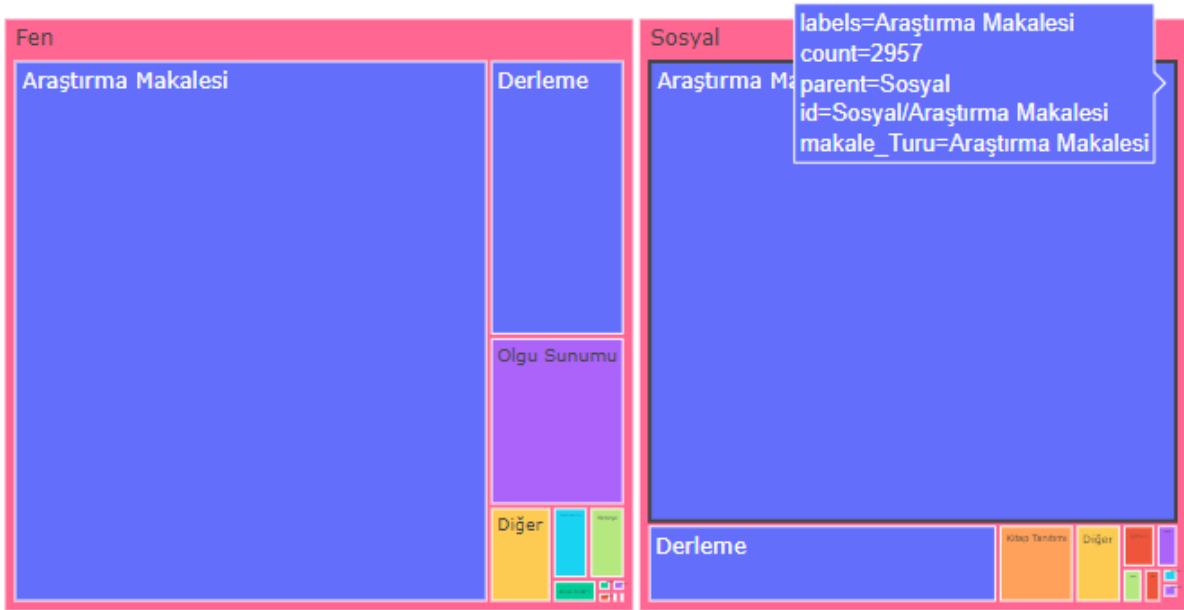
Tablo 20: Güncel “tur_df” alt veri setinin ilk beş kaydı

index	makale_konu_Kategorisi	makale_Turu	sayi
0	Fen	Araştırma Makalesi	3.097
1	Fen	Bildiri	2
2	Fen	Derleme	451
3	Fen	Diğer	73
4	Fen	Düzeltilme	1

“tur_df” alt veri seti Kod 67 ile “plotly.express” kütüphanesindeki “treemap” grafik türü kullanılarak görselleştirilmiştir. Bu görsel türü bilgisayar ortamında mouse kullanılarak etkileşimli bir şekilde incelenebilmektedir.

```
import plotly.express as px  
px.treemap(tur_df, color='makale_Turu', values=tur_df.sayi,  
           path=['makale_konu_Kategorisi', 'makale_Turu'], height=700)
```

Kod 67: Temel konu kategorilerine göre makale türlerinin dağılımı

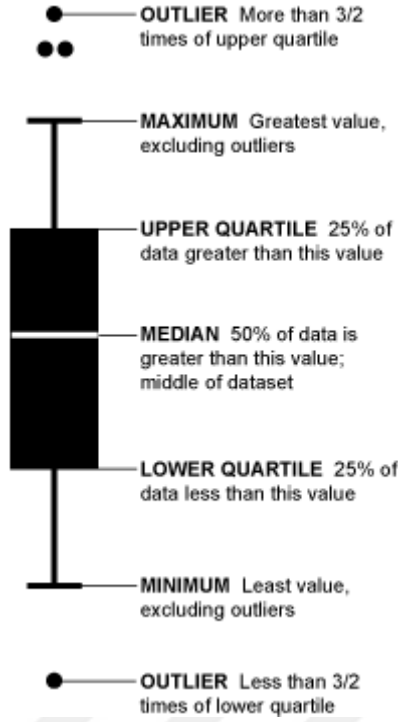


Şekil 54: Konu kategorisine göre kümelenmiş makale türü dağılımı

Şekil 54 incelendiğinde hem sosyal hem de fen temel konu kategorilerinde “Araştırma makalesi” türünün çoğunlukta olduğu, fen kategorisindeki “derleme ve olgu sunumu” türlerinde yapılan yayınların sosyal kategorisindeki “derleme ve olgu sunumu” türlerinden daha fazla olduğu görülmektedir.

4.1.3.2 Soru 3.2 En çok atıf alan makale türleri hangileridir?

En çok atıf yapılan makale türlerinin atıf dağılımlarını tespit etmek için “makale_Turu ve atıf” değişkenleri incelenmiş ve görselleştirme için seaborn paketindeki “boxplot” grafik türü kullanılmıştır. Kutu ve bıyık grafikleri olarak da bilinen boxplot, medyanı, çeyrekleri ve aykırı değerleri gösteren bir veri kümesinin dağılımını çizmek için kullanılır. Bunlar özellikle aykırı değerler için kullanışlıdır ve verilerin dağılımını etkili bir şekilde temsil ederler (Enago Academy, 2019).



Şekil 55: Boxplot grafik türünün özellikleri (Yau, 2008)

Şekil 55'teki boxplot (kutu grafik) görselinde kullanılan önemli parametrelerin özellikleri aşağıda açıklanmıştır³¹:

- Minimum ve Maksimum Değerler: Boxplot'un alt ve üst uçlarındaki çizgiler, veri setinin minimum ve maksimum değerlerini temsil eder. Aykırı değerler genellikle bu çizgilerin ötesinde yer alır.
- Quartiles (çeyreklikler): Boxplot, veri setini dört eşit parçaya ayırarak çeyreklikleri gösterir. Bu çeyreklikler şu şekildedir:
 - Lower Quartile (alt çeyreklik): Veri setinin alt yarısının ortalamasını temsil eder.
 - Medyan: Veri setinin ortanca değerini temsil eder. Veri setini iki eşit parçaya böler.
 - Upper Quartile (üst çeyreklik): Veri setinin üst yarısının ortalamasını temsil eder.
 - IQR (Interquartile Range - çeyreklikler arası aralık): IQR, Q3'ten Q1'i çıkartarak elde edilen bir değerdir. Bu değer, veri setinin ortasındaki 50% lik bir kısmını temsil eder ve aykırı değerleri belirlemekte kullanılır.

³¹ Boxplot: Veri setinin merkezi eğilimi, dağılımı ve aykırı değerleri, minimum ve maksimum değerleri hakkında bilgi sunan grafik türüdür (matplotlib, 2023; Mcleod, 2023).

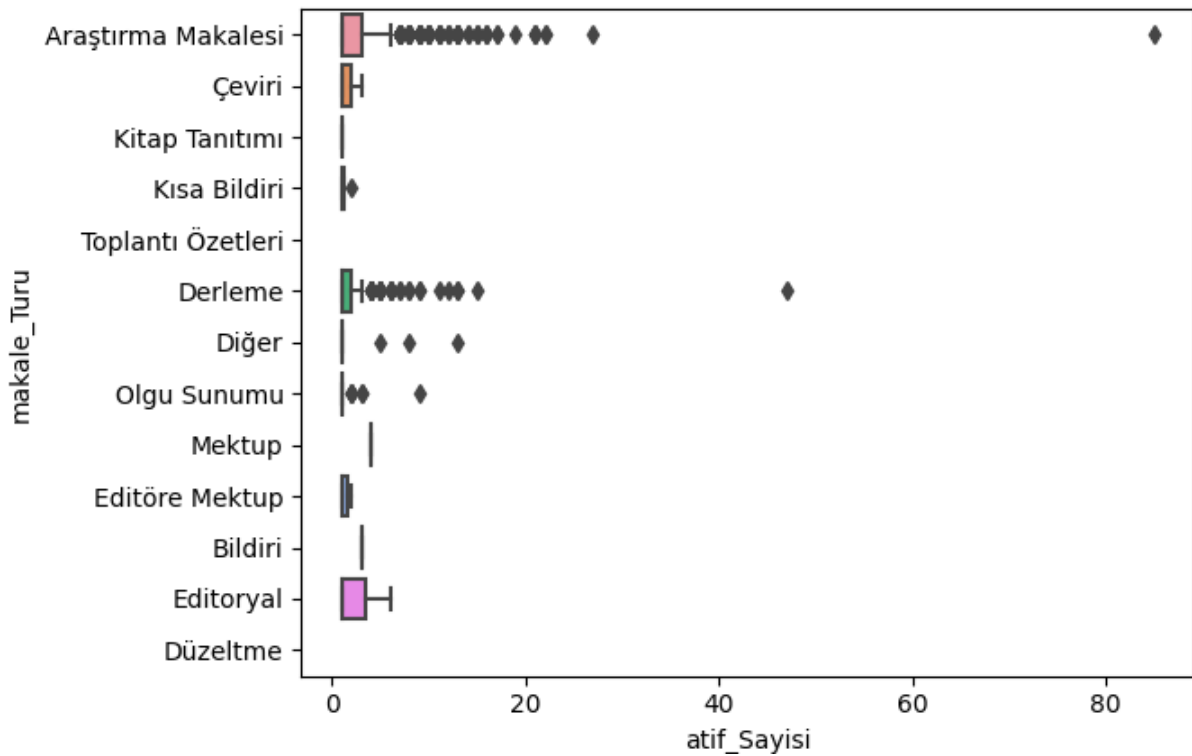
- Outliers (aykırı değerler): Aykırı değerler, genellikle veri setinin üstünde veya altında yer alan noktalardır. Boxplot, aykırı değerleri belirlemek için genellikle $Q1 - 1.5 * IQR$ ve $Q3 + 1.5 * IQR$ sınırlarını kullanır.

Whisker (bıyık): Boxplot'ta kutunun üst ve alt kenarlarından çıkarak uzanan çizgilere denir. Bıyıklar genellikle minimum ve maksimum değerlerin ötesindeki aykırı değerleri gösterir.

“makale_Turu” değişkenindeki ögelerin atıf dağılımlarını ve aykırı değerleri incelemek için Kod 68 ile Şekil 56'daki boxplot görseli oluşturulmuştur.

```
import seaborn as sns
sns.boxplot(data=df, x="atıf_Sayısı", y="makale_Turu")
plt.show()
```

Kod 68: Makale türlerinin atıf sayılarına göre dağılımı



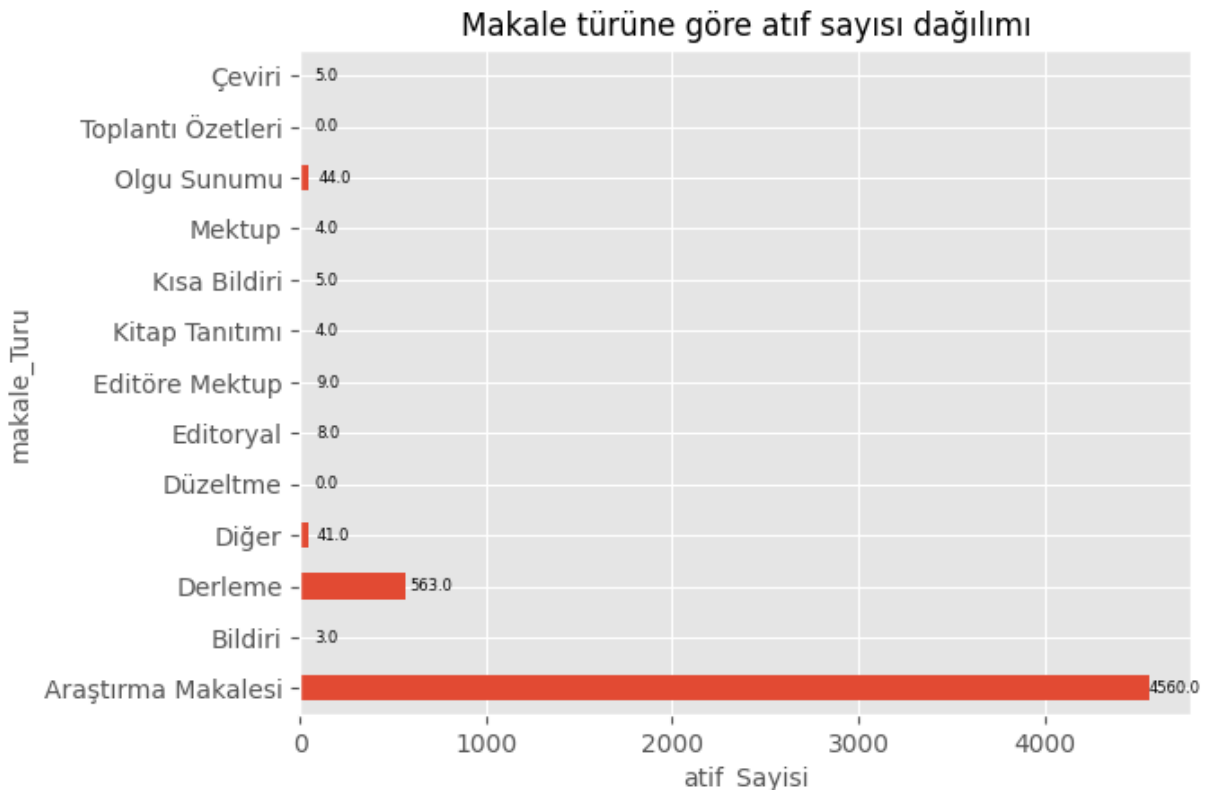
Şekil 56: Makale türü kategorilerinin aldıkları atıf sayılarına göre dağılımı

Şekil 56 incelendiğinde; aykırı atıf sayısı değerinin araştırma makalesinde olduğu, bir makalenin ortalamanın çok üzerinde, 80’den fazla atıf aldığı görülmektedir. İkinci aykırı atıf sayısı değeri derleme türünde ortaya çıkmıştır ve bir derleme makalesinin ortalamanın üzerinde, 40’tan fazla atıf aldığı görülmektedir. “Toplantı özetleri” ve “düzeltme” makale türlerinin hiç atıf almadığı tespit edilmiştir.

Boxplot grafik türü toplam atıf sayılarını göstermemektedir. Makale türlerinin aldığı toplam atıf sayıları için Kod 69 ile matplotlib paketinin “barh plot” grafik türü kullanılarak Şekil 57’deki grafik oluşturulmuştur. “barh plot”, kategorik verilerin oranlarını veya sayımlarını görüntülemek için uygun bir görselleştirme türüdür.

```
df.groupby("makale_Turu")["atif_Sayisi"].sum().plot(kind="barh", y="atif_Sayisi")
plt.xlabel('atif_Sayisi', fontsize=10)
plt.ylabel('makale_Turu', fontsize=10)
```

Kod 69: Makale türlerinin atıf sayılarına göre dağılımı



Şekil 57: Makale türü kategorilerinin aldıkları atıf sayılarına göre dağılımı

Makale türlerinin atıf sayısına göre dağılımını veren Şekil 57 incelendiğinde en yüksek atıf sayılarının sırasıyla; araştırma makalesi (4.560), derleme (563), olgu sunumu (44) ve diğer (41) ögelerinde olduğu, düzeltme ve toplantı özetleri ögelerinde atıf olmadığı gözlemlenmiştir. Bildiri ve olgu sunumları da sonuç/bulgu ortaya koydukları için bilimsel yayın niteliği taşıyan türlerdir. “Araştırmacılar bulguları değerlendirmekten çok literatür taraması bağlamında mı atıf yapma eğilimindedir?” sorusu bu görsel bağlamında sorulabilecek başka bir soru olabilir. Ama bu soru başka bir sorgulama/çalışma kapsamında derinleştirilebilir.

4.1.4 Soru 4: Makale sayıları ile atıf sayılarının yıllara göre dağılımı nedir?

TR Dizin'in kendi içindeki makalelerin yıllara göre toplam sayılarını ve makalelerin ilgili yıl için aldıkları toplam atıfları (bu sistem içindeki) değerlendirebilmek için Kod 70 ile “makale_atif_sayisi” alt veri seti oluşturulmuştur.

```
makaleSayisi = df.makale_Yil.value_counts().sort_index()
atifSayisi = df.groupby('makale_Yil').atif_Sayisi.sum().sort_index()
makale_atif_sayisi = pd.DataFrame({'Makale_sayısı': makaleSayisi, 'Atıf_sayısı': atifSayisi})
makale_atif_sayisi
```

Kod 70: “makale_atif_sayisi” alt veri setinin oluşturulması

Atıf sayılarının makale yılına göre gruplanarak oluşturulan “makale_atif_sayisi” verileri Tablo 21’de verilmiştir.

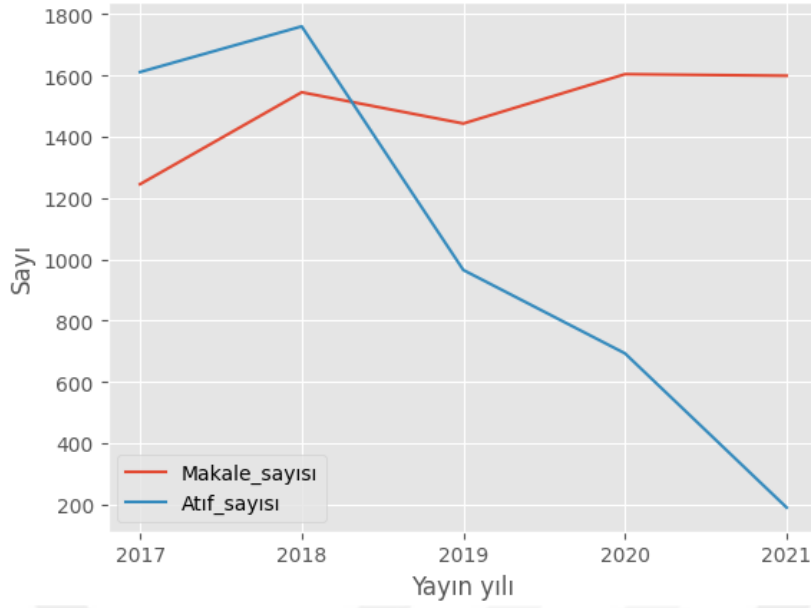
Tablo 21: “makale_atif_sayisi” alt veri seti

index	Makale_sayısı	Atıf_sayısı
2017	1245	1617
2018	1545	1761
2019	1443	966
2020	1604	695
2021	1599	207

Kod 71’de matplotlib görselleştirme kütüphanesi kullanılarak Şekil 58’deki çizgi grafik oluşturulmuştur.

```
plt.style.use('ggplot')
makale_atif_sayisi.plot(kind="line")
plt.xticks(makale_atif_sayisi.index, makale_atif_sayisi.index)
plt.xlabel("Yayın yılı")
plt.ylabel("Sayı")
```

Kod 71: Yıllık toplam makale sayısı ile yıllık toplam atıf sayılarının karşılaştırılması



Şekil 58: Yıllık toplam yayın sayısı ile yıllık toplam atıf sayılarının karşılaştırılması

Yıllık toplam makale sayısı ve makalelerin atıf sayıları incelendiğinde, en çok atfın 2018 yılında olduğu, en az atfın ise 2021 yılında olduğu görülmektedir. Veri setindeki atıf sayıları 20.12.2022 tarihli verilerden oluşmaktadır. Güncel atıf sayıları ile yıllık toplam atıf sayısı oranları değişiklik gösterecektir.

4.1.4.1 Soru 4.1 Makalelerin yıllara göre görüntülenme ve indirilme sayıları nedir?

Makalelerin TR Dizin platformu üzerinden görüntülenme ve indirilme sayılarını değerlendirmek için “makale_Yıl, makale_Goruntuleme ve makale_Indirme” değişkenleri kullanılarak Kod 72 ile “view_download_df” adlı bir alt veri seti oluşturulmuştur. Yıllık toplam değerleri elde etmek için veri setindeki görüntüleme ve indirme sayıları “makale_Yıl” değişkenine göre gruplanmıştır.

```
views_downloads_df = df.groupby("makale_Yıl")[["makale_Goruntuleme", "makale_Indirme"]].sum().reset_index()
views_downloads_df
```

Kod 72: “view_downloads_df” alt veri setinin oluşturulması

“makale_Yıl, makale_Goruntuleme ve makale_Indirme” sayılarının “makale_Yıl” değişkenine göre gruplanarak oluşturulan “view_download_df” verileri Tablo 22’de verilmiştir.

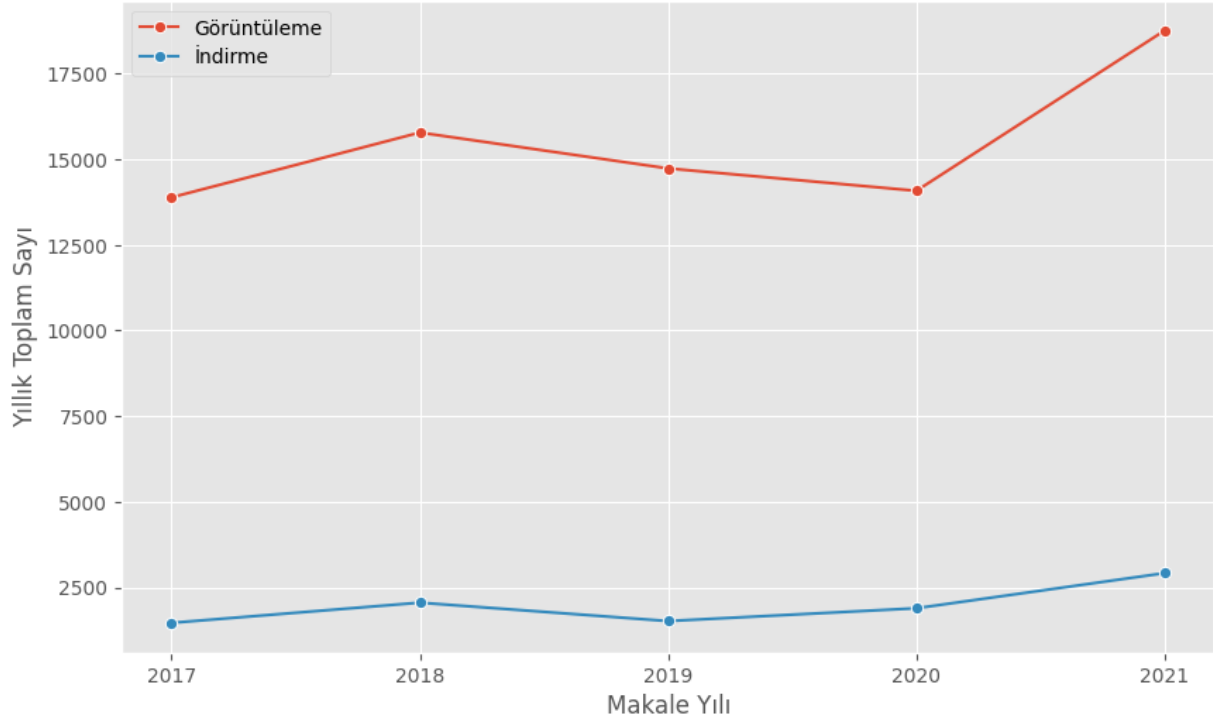
Tablo 22: Makale yılına göre toplam makale görüntüleme ve indirme sayıları

makale_Yil	makale_Goruntuleme	makale_Indirme
2017	13887	1468
2018	15780	2058
2019	14733	1521
2020	14080	1895
2021	18760	2920

“view_downloads_df” alt veri setinin görselleştirilmesi için Python’nın seaborn kütüphanesi kullanıldığı Kod 73 ile Şekil 59’daki çizgi grafik oluşturulmuştur.

```
import seaborn as sns
unique_years = views_downloads_df['makale_Yil'].unique()
plt.style.use('ggplot')
plt.figure(figsize=(10, 6))
sns.lineplot(data=views_downloads_df, x="makale_Yil", y="makale_Goruntuleme", marker='o', label="Görüntüleme")
sns.lineplot(data=views_downloads_df, x="makale_Yil", y="makale_Indirme", marker='o', label="İndirme")
plt.xlabel('Makale Yılı')
plt.ylabel('Yıllık Toplam Sayı')
plt.xticks(unique_years)
plt.legend()
```

Kod 73: Makale yılına göre makale görüntülenme ve indirme sayıları dağılımı



Şekil 59: Makale yılına göre makale görüntülenme ve indirilme sayıları dağılımı

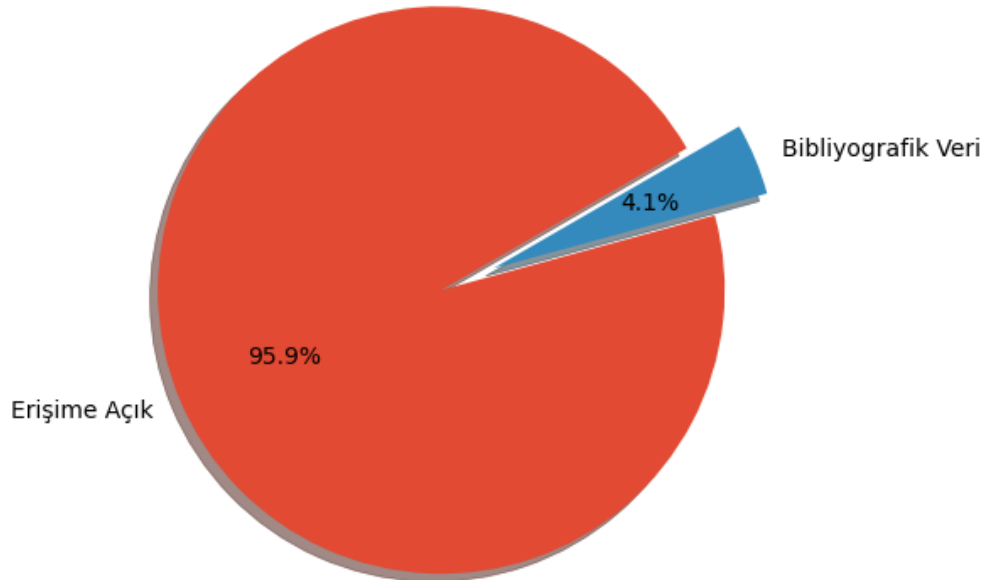
Şekil 59 incelendiğinde; makale yılına göre görüntülenme ve indirme sayılarındaki artış ve azalışların doğru orantılı olduğu görülmektedir. Her iki değişkende de 2018’den 2020’ye kadar bir düşüş olmuş ve 2020’den itibaren yükselme başlamıştır.

4.1.5 Soru 5: Ulaşılabilirlik açısından makalelerin erişim türüne (açık/kapalı) göre dağılımı nedir?

TR Dizin’deki makalelerin erişim türü, tam metne erişim ve bibliyografik veriye erişim şeklinde iki kategoride kaydedilmiştir. Veri setindeki makalelerin erişim türüne göre dağılımını görmek için Kod 74 ile matplotlib görselleştirme paketi kullanılarak Şekil 60’taki “pie chart” (pasta grafik) oluşturulmuştur.

```
explode = (0.1, 0.1)
fig, ax = plt.subplots()
ax.pie(df.acik_Erisim.value_counts(),
      explode=explode,
      labels=df.acik_Erisim.value_counts().index,
      autopct='%1.1f%%',
      shadow=True,
      startangle=30)
ax.axis('equal')
```

Kod 74: Yayınların erişim türüne göre dağılımı



Şekil 60: Yayınların erişim türüne göre dağılımı

Grafiğe göre makalelerin % 95,9’una (7.132) tam metin, % 4,1’ine (304) ise sadece bibliyografik veri olarak erişim sağlandığı tespit edilmiştir.

4.1.5.1 Soru 5.1 Hangi temel konulu makaleler erişime kapalıdır; bunlar yıllar içinde artış gösterme eğiliminde midir?

Hangi konu alanındaki makalelerin erişime kapalı olduğunu ve yıllara göre bir artış/düşüş eğilimi gösterip göstermediğini saptayabilmek için “makale yılı, makale konu kategorisi ve erişim türü” “bibliyografik veri” olan kayıtların gruplandığı “bibliyografik_df” adlı bir alt veri seti oluşturulmuştur ve “bibliyografik veri.head()” kodunun uygulanmasıyla görüntülenen ilk beş kayıt Tablo 23’te verilmiştir.

```
bibliyografik_df = df.groupby(['makale_Yil', 'makale_konu_Kategorisi', 'acik_Erisim']).size().reset_index(name='sayi')
bibliyografik_df = bibliyografik_df.loc[bibliyografik_df['acik_Erisim'] != "Erişime Açık"].reset_index()
bibliyografik_df = bibliyografik_df.drop('index', axis=1)
bibliyografik_df.head()
```

Kod 75: Erişim türü bibliyografik veri olan kayıtların alt veri seti

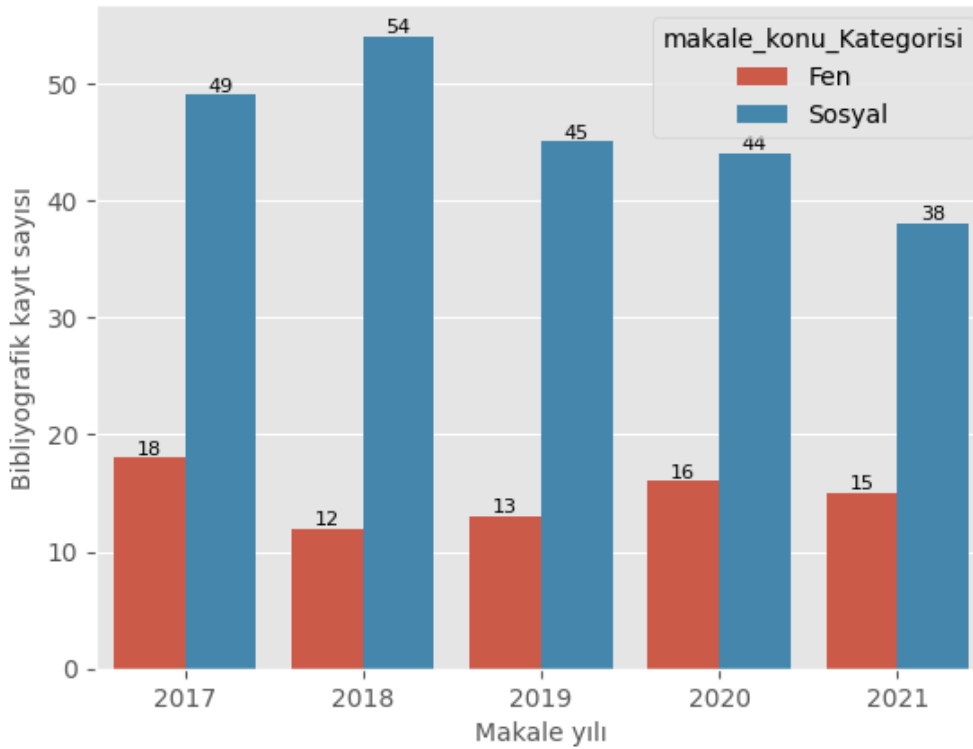
Tablo 23: “bibliyografik_df” veri seti ilk beş kayıt

	makale_Yil	makale_konu_Kategorisi	acik_Erisim	sayi
0	2017	Fen	Bibliyografik Veri	18
1	2017	Sosyal	Bibliyografik Veri	49
2	2018	Fen	Bibliyografik Veri	12
3	2018	Sosyal	Bibliyografik Veri	54
4	2019	Fen	Bibliyografik Veri	13

“bibliyografik_df” alt veri setindeki gruplanmış değişken verileri, Kod 76 ile seaborn paketinin barplot grafiği kullanılarak Şekil 61’deki gibi görselleştirilmiştir.

```
ax = sns.barplot(data=bibliyografik_df, x="makale_Yil", y="sayi", hue="makale_konu_Kategorisi")
ax.bar_label(ax.containers[0], fontsize=8)
ax.bar_label(ax.containers[1], fontsize=8)
plt.xlabel('Makale yılı', fontsize=10)
plt.ylabel('Bibliyografik kayıt sayısı', fontsize=10)
```

Kod 76: Sadece bibliyografik verilerine erişilen makalelerin yıllara göre konu dağılımı



Şekil 61: Erişim türü bibliyografik veri olan kayıtların temel konu kategorisine göre dağılımı

Grafik 61 incelendiğinde; erişime kapalı olan yayın sayısının en çok 2017 (67), en az ise 2021 (53) yılında olduğu görülmektedir. Sosyal konu kategorisindeki erişime kapalı yayınların her yıl için, fen konu kategorisinin iki katından fazla olduğu tespit edilmiştir.

4.1.6 Soru 6: Yazar iş birliği (tek yazarlı/ortak yazarlı) eğilimi ne düzeydedir?

TR Dizin’de “Makale Türü” üst veri alanında “araştırma makalesi, derleme ve olgu sunumu” gibi bilimsel makale türlerinin yanı sıra “toplantı özeti, mektup, bildiri, kısa bildiri, editöre mektup, editoryal, kitap tanıtımı, diğer” gibi diğer türlerin de bulunduğu gözlemlenmiştir. TR Dizin’de türe ilişkin veriler bir dergi içerisindeki yayınların hakemli olup olmamasına bakılmaksızın “Makale türü” etiketi alanında dizinlendiği düşünülmektedir. Bu bölümde araştırmacıların en çok ürettiği ve atıf verdiği makale türleri olan “araştırma makalesi, derleme ve olgu sunumu” türleri üzerinde görselleştirme yapılmıştır.

Veri setinde yazar isimleri “yazar_Adı” değişkeninde bulunmaktadır. Ortak yazarlı makalelerde “yazar_Adı” değişkeninde yazar isimlerinin noktalı virgül ile ayrıştırıldığı görülmektedir. Tablo 24’te örnek bir kesit verilmiştir.

Tablo 24: “yazar_Adi” değişkenindeki ilk beş kaydın görünümü

index	yazar_Adi
0	Umut P...; Bülent A....; İlkay P....
1	Duygu K...; Çilem GÜL D....; Emrah G., Ömay B...
2	Gül K....; Şakir Ç....
3	Sıla ŞENLEN G....
4	Gökçen T....; Belkıs K....

Tablo 24’ün “yazar_Adi” değişkeninin her bir hücresindeki yazar sayısını bulmak için Kod 77 ile “yazar_Sayisi” değişkeni oluşturulmuş ve yazar sayıları hesaplanarak bu değişkene atanmıştır. Ayrıca çeşitli analizler yapmak için aynı kod ile makale türü “araştırma makalesi, derleme ve olgu sunumu” olan makalelerin “yazar_Sayisi, atif, yıl, makale_Turu ve yazar_Kurum” değişkenlerini içeren ve Tablo 25’te ilk beş kaydı verilen “yazar_df” adlı bir DataFrame oluşturulmuştur.

```
yazar_df = pd.DataFrame()
yazar_df['yazar_Adi'] = df.yazar_Adi
yazar_df['yazar_Sayisi'] = df.yazar_Adi.str.count(',') + 1
yazar_df['atif'] = df.atif_Sayisi
yazar_df['yil'] = df.makale_Yil
yazar_df['makale_Turu'] = df.makale_Turu
yazar_df['yazar_Kurum'] = df.yazar_Kurum
yazar_df = yazar_df[yazar_df['makale_Turu'].isin(['Araştırma Makalesi', 'Derleme', 'Olgu Sunumu'])].reset_index()
```

Kod 77: “yazar_df” alt veri setinin oluşturulması

Tablo 25: “yazar_df” alt veri setinin ilk beş kaydı

index	yazar_Sayisi	atif	yil	makale_Turu	yazar_Kurum
0	3	NaN	2021	Araştırma Makalesi	GAZİ ÜNİVERSİTESİ; ANKARA ÜNİVERSİTESİ; GAZİ ÜNİVERSİTESİ
1	4	6.0	2017	Araştırma Makalesi	ALANYA ALAADDİN KEYKUBAT ÜNİVERSİTESİ; HAKKARİ ÜNİVERSİTESİ; HAKKARİ ÜNİVERSİTESİ; ANKARA ÜNİVERSİTESİ
2	2	5.0	2017	Araştırma Makalesi	ANKARA ÜNİVERSİTESİ; ANKARA ÜNİVERSİTESİ
3	1	2.0	2018	Araştırma Makalesi	ANKARA ÜNİVERSİTESİ
4	2	3.0	2017	Araştırma Makalesi	ANKARA ÜNİVERSİTESİ; ANKARA ÜNİVERSİTESİ

Pandas kütüphanesinin “describe()” metodu kullanılarak makalelere ait yazar sayılarının veri özeti görüntülenebilir. Describe() metodu ile elde edilen özet rapordaki ifadelerin açıklamaları şu şekildedir:

- **count:** Sütunda kaç adet kayıt olduğu,
- **mean:** Sütunda bulunan sayısal değerlerin ortalama değeri
- **std:** Standart sapma değeri
- **min:** Sütundaki en küçük değer
- **25%:** Sütundaki toplam kayıtların %25 'i nin ortalama değeri
- **50%:** Sütundaki toplam kayıtların %50'i nin ortalama değeri
- **75%:** Sütundaki toplam kayıtların %75 'i nin ortalama değeri
- **max:** Sütundaki en büyük değer

```
yazar_df.yazar_Sayisi.describe()

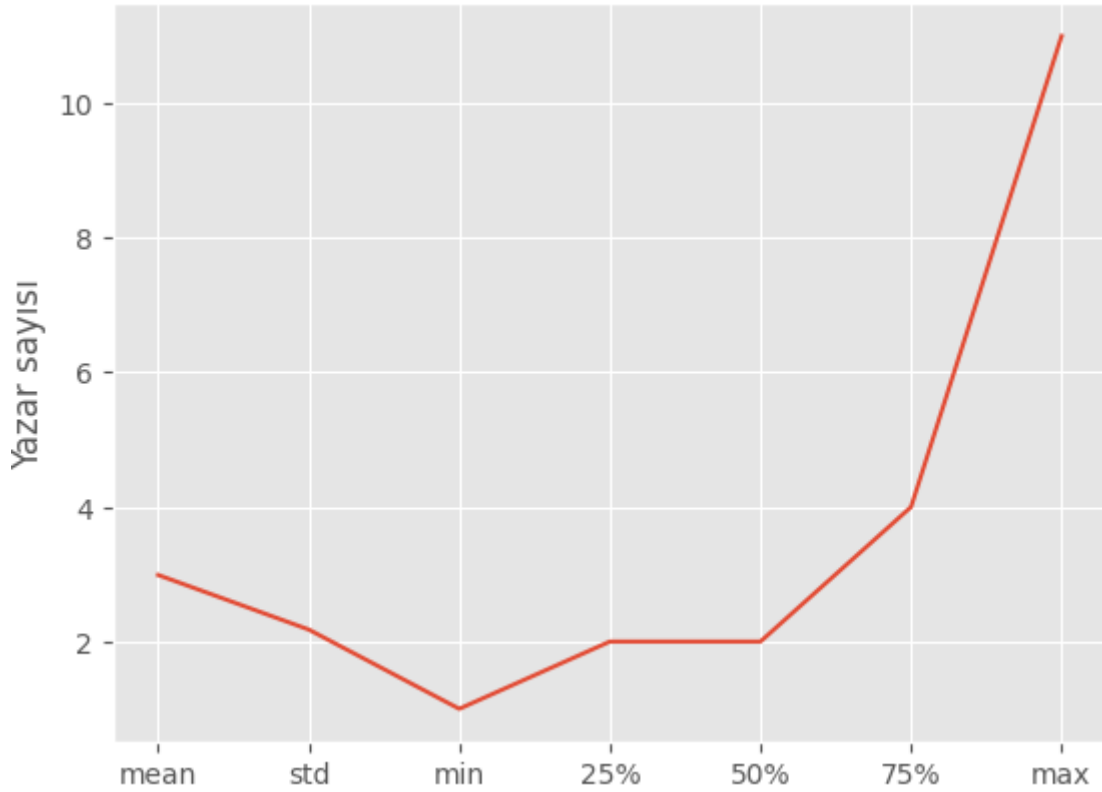
count    7118.000000
mean      2.989604
std       2.178336
min       1.000000
25%       2.000000
50%       2.000000
75%       4.000000
max      11.000000
Name: yazar_Sayisi, dtype: float64
```

Kod 78: Yazar sayısı sütununun “describe” metodu ile incelenmesi

describe() metodu ile yapılan incelemede “yazar_df” alt veri setinde makale türü “araştırma makalesi, derleme ve olgu sunumu” olan toplam makale sayısı 7.118’dir.

```
plt.style.use('ggplot')
yazar_df.yazar_Sayisi.describe().loc[['mean', 'std', 'min', '25%', '50%', '75%', 'max']].plot()
plt.ylabel('Yazar sayısı')
```

Kod 79: Yazar sayısı sütununun “describe” metodu ile incelenmesi



Şekil 62: Yayın başına yazar sayısının “describe()” metodu ile incelenmesi

Grafik 62’ye göre makalelerin ortalama(mean) yazar sayısının 2,98 olduğu görülmektedir. Makale başına düşen maksimum yazar sayısı ise 11’dir. Bu durum yazar iş birliği eğilimi olduğunu göstermektedir.

4.1.6.1 Soru 6.1 Yazar sayısına göre yayın sayısı dağılımı nedir?

Tek yazarlı yayınların ortak yazarlı yayınlara oranını tespit etmek için Kod 80 ile “yazar_df” alt veri setinde “durum” adlı bir değişken oluşturulmuş ve makalelerin tek yazarlı veya ortak yazarlı olma bilgisi bu sütuna kaydedilmiştir.

```
yazar_df['durum'] = yazar_df['yazar_Sayisi']  
yazar_df.loc[yazar_df['yazar_Sayisi'] == 1, 'durum'] = "Tek yazarlı"  
yazar_df.loc[yazar_df['yazar_Sayisi'] != 1, 'durum'] = "Ortak yazarlı"
```

Kod 80: “yazar_df” alt veri setine “durum” değişkeninin eklenmesi

Makalelerin tek yazarlılık/ortak yazarlılık durumunu incelemek için Kod 81 ile “yazar_df”deki “durum” değişkeni görselleştirilerek Şekil 59’daki grafik elde edilmiştir. Görselleştirme için matplotlib kütüphanenin pasta grafik türü kullanılmıştır.

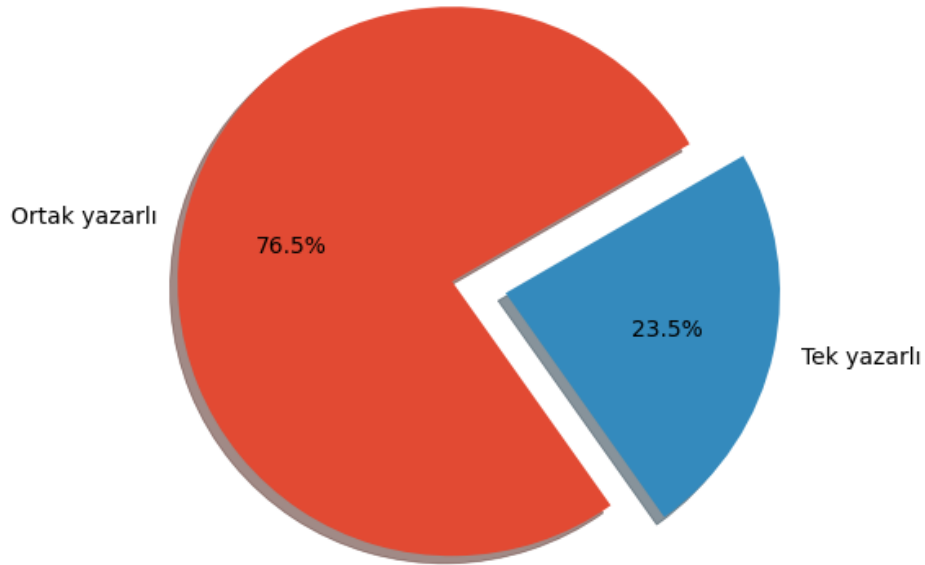
```

▶ explode = (0.1, 0.1)
fig, ax = plt.subplots()
ax.pie(yazar_df.durum.value_counts(),
      explode=explode,
      labels=yazar_df.durum.value_counts().index,
      autopct='%1.1f%%',
      shadow=True,
      startangle=30)
ax.axis('equal')
ax.set_title('Yazar sayısı durumu')

```

Kod 81: Tek yazarlı ve ortak yazarlı yayınların oranı

Bu kod ile “yazar_df” alt veri setine “durum” değişkeni eklenmiş ve yazar sayısı sütununda değeri 1 olan kayıtlar için durum değişkenine “Tek yazarlı”, değeri 1’den farklı olan kayıtlar için ise “Ortak yazarlı” verisi yansıtılmıştır. Daha sonra “durum” değişkeninde yer alan verilerin oransal durumları aşağıdaki pasta grafik ile görselleştirilmiştir.



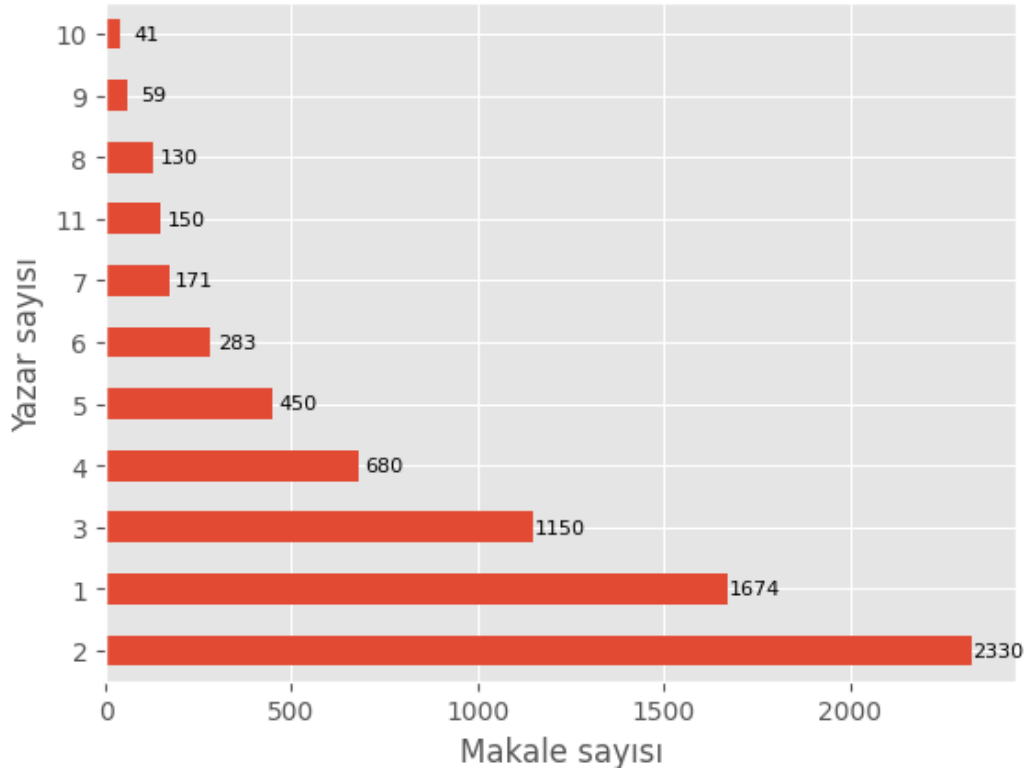
Şekil 63: Tek yazarlı ve ortak yazarlı yayınların oranı

Görsel incelendiğinde, yazar iş birliği eğiliminin güçlü olduğu, ortak yazarlı yayınların tek yazarlı yayınlara oranla üç kat fazla olduğu görülmektedir.

Tek ve ortak yazarlı makale sayılarını görebilmek için ise Kod 82 ile “yazar_df” alt veri setindeki “yazar_Sayisi” değişkeni görselleştirilerek Şekil 64’te verilmiştir. Görselleştirme için matplotlib kütüphanesinin “barh plot” türü kullanılmıştır.

```
plt.style.use('ggplot')
yazar_df.yazar_Sayisi.value_counts().plot(kind="barh")
for index, value in enumerate(yazar_df.yazar_Sayisi.value_counts()):
    plt.annotate(value, xy=(70+value,index), ha='center', va='center',fontsize=8)
plt.xlabel('Makale sayısı')
plt.ylabel('Yazar sayısı')
```

Kod 82: Yazar sayısına göre yayın sayısı dağılımı



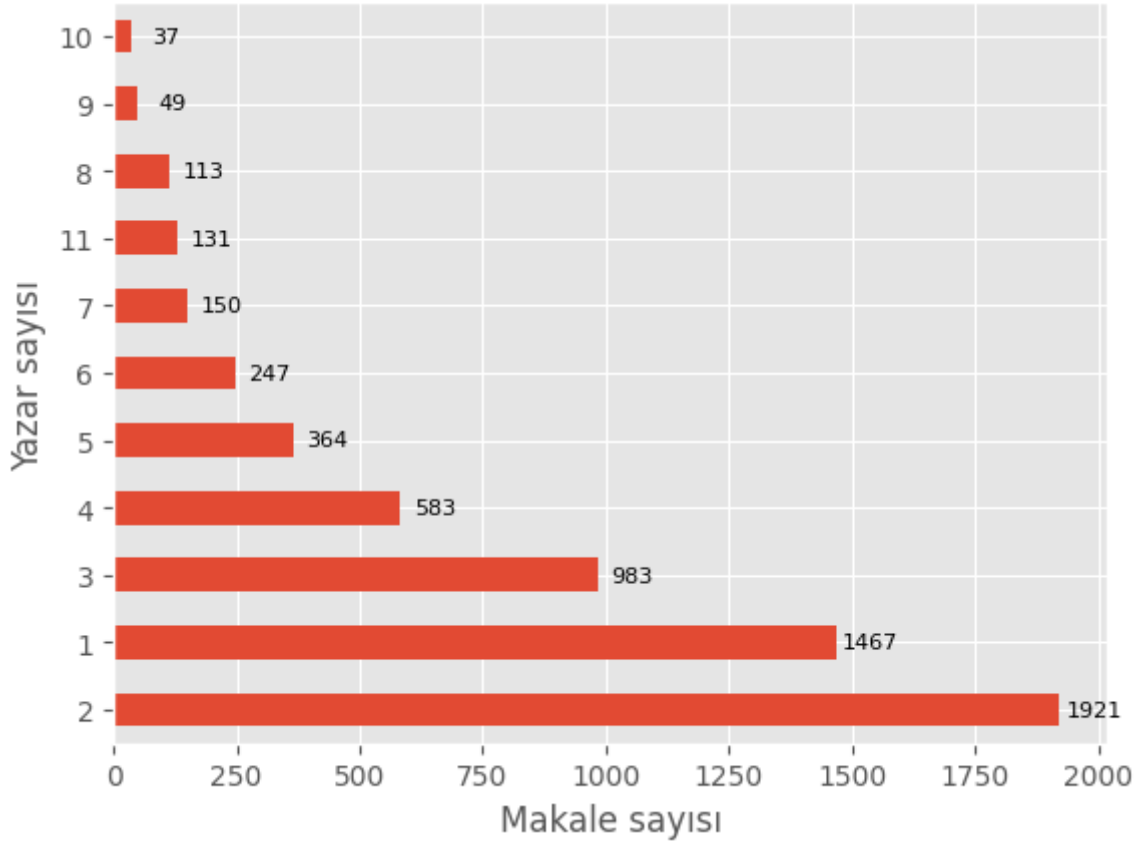
Şekil 64: Yazar sayısına göre makale sayısının “barh plot” grafiği ile incelenmesi

Makale türü “araştırma, derleme ve olgu sunumu” olan makalelerin “yazar_Sayisi” değişkeni ile karşılaştırıldığı Şekil 64’e bakıldığında sırasıyla; iki yazarlı makale sayısının 2.330, bir yazarlı makale sayısının 1.674, üç yazarlı makale sayısının 1.150 ve en az makale sayısının ise 41 ile on yazarlı makalelerde olduğu tespit edilmiştir.

Sadece “araştırma makalesi” (6.045 makale) türünde olan makale sayısı ve yazar sayısı karşılaştırması Kod 83 ile görselleştirilerek Şekil 65’te verilmiştir. Bu görselleştirme için de matplotlib kütüphanesinin “barh plot” türü kullanılmıştır.

```
plt.style.use('ggplot')
yazar_df = yazar_df[yazar_df.makale_Turu == "Araştırma Makalesi"]
yazar_df.yazar_Sayisi.value_counts().plot(kind="barh")
for index, value in enumerate(yazar_df.yazar_Sayisi.value_counts()):
    plt.annotate(value, xy=(70+value,index), ha='center', va='center', fontsize=8)
plt.xlabel('Makale sayısı')
plt.ylabel('Yazar sayısı')
```

Kod 83: Yazar sayısına göre makale sayısı dağılımı (araştırma makaleleri)



Şekil 65: Yazar sayısına göre makale sayısının “bar plot” grafiği ile incelenmesi (makale türü “araştırma makalesi” olan makaleler)

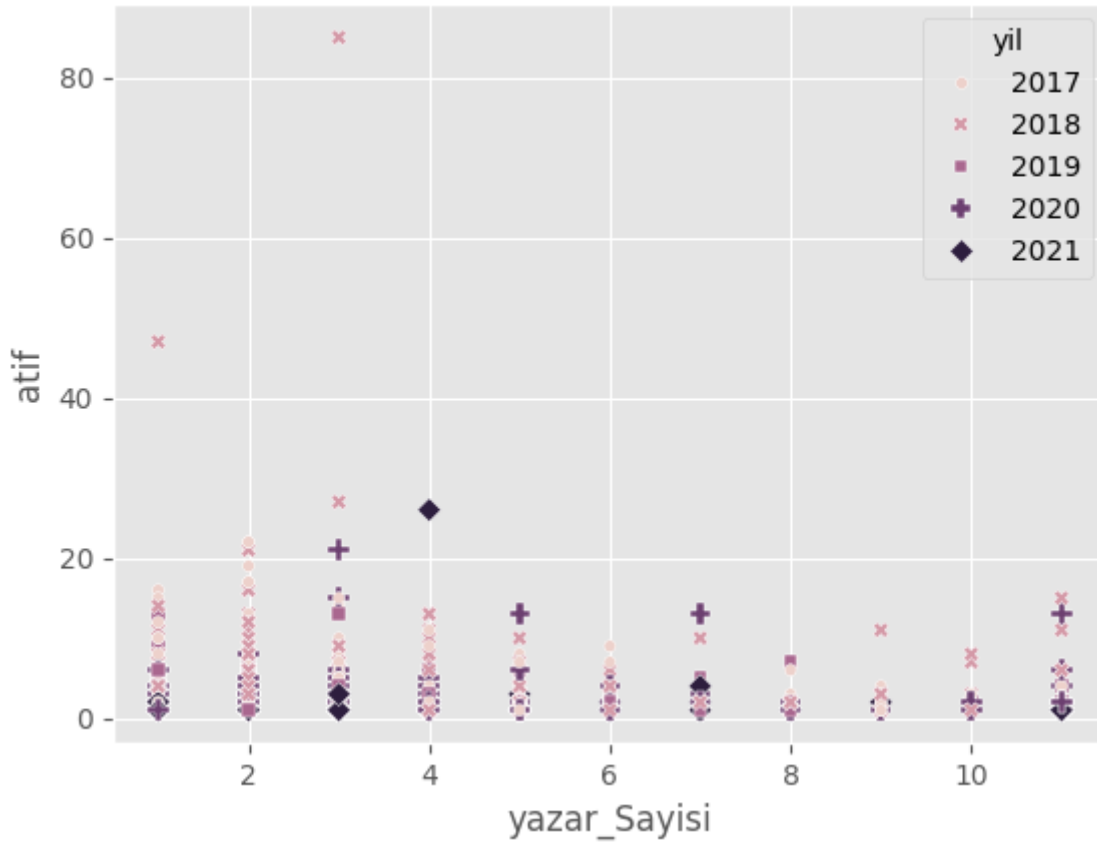
“yazar_df” alt veri setinde makale türü alanı “araştırma makalesi” olarak filtreledikten sonra, Şekil 64 ve 65’teki makale başına düşen yazar sayısı dağılımında bir değişiklik olmadığı gözlemlenmiştir. Bilimsel iş birliğine yatkınlık açısından değerlendirildiğinde, eğilimin 2 yazardan yana olduğu görülmektedir. Ayrıca Şekil 64 ve 65 incelendiğinde, araştırma makalelerinin sayısının (6.045) diğer makale türlerini (1.391) domine ettiği söylenebilir.

4.1.6.2 Soru 6.2 Makale başına düşen yazar sayısına göre atıf alma durumu nedir?

Yazar sayısına göre atıf alan makalelerin yıllara göre dağılımının incelenmesi için “yazar_df” alt veri setindeki “yazar_Sayisi, atıf ve yıl” değişkenleri kullanılmıştır. Bu üç değişken “seaborn” kütüphanesindeki “scatterplot (dağılım grafiği)” görsel türü kullanılarak Kod 84 ile görselleştirilmiş ve ilgili grafik Şekil 66’da verilmiştir.

```
import seaborn as sns
sns.scatterplot(data=yazar_df, x="yazar_Sayisi", y="atıf", hue="yıl", style="yıl", size="yıl")
```

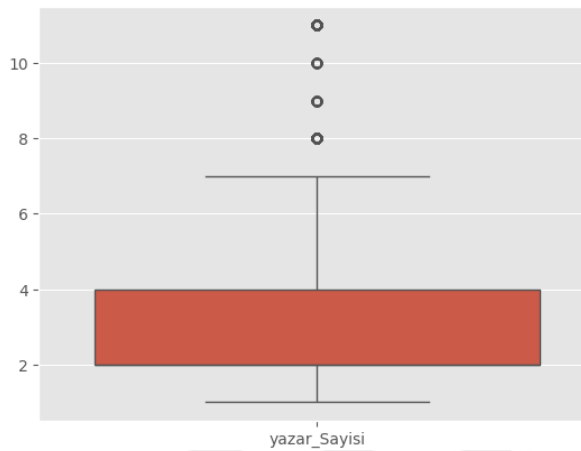
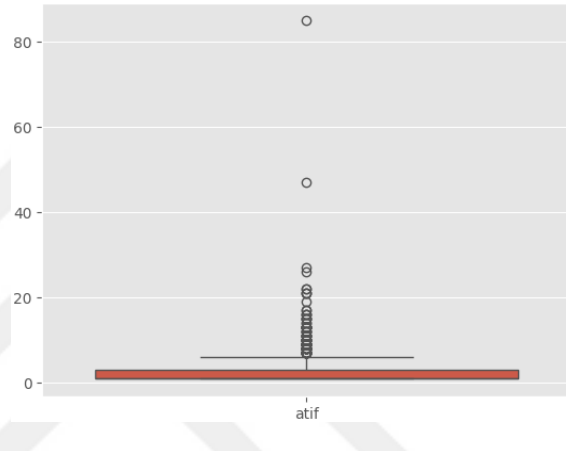
Kod 84: Makale başına düşen yazar sayısı ile atıf sayılarının yayın yılına göre incelenmesi



Şekil 66: Makale başına düşen yazar sayısı ile atıf sayılarının yayın yılına göre incelenmesi

Yıllara göre atıf durumuna baktığımızda en çok atfın 2017 ve 2018 yıllarına ait yayınlarda olduğu görülmektedir. İstisna olarak, 2018 yılına ait tek yazarlı bir makalede 40’ın üzerinde ve 2018 yılına ait 3 yazarlı bir makalede ise 80’in üzerinde atıf olduğu görülmektedir. Makalelerin çoğunluğunda atıf sayısı 20’nin altındadır.

Yazar sayısı ve atıf sayısı değişkenlerindeki dağılımları ve istisna kayıtları görebilmek için Kod 85 ve Kod 86 ile seaborn kütüphanesi kullanılarak Şekil 67 ve Şekil 68’de yer alan “boxplot” grafikleri (kutu grafik) oluşturulmuştur.

<pre>import seaborn as sns sns.boxplot(data=yazar_df[["yazar_Sayisi"]])</pre> Kod 85: Makale başına düşen yazar sayısının “box plot” grafiği ile incelenmesi	<pre>import seaborn as sns sns.boxplot(data=yazar_df[["atıf"]])</pre> Kod 86: Makale başına düşen atıf sayılarının “box plot” grafiği ile incelenmesi
 Şekil 67: Makale başına düşen yazar sayısının “box plot” grafiği ile incelenmesi	 Şekil 68: Makale başına düşen atıf sayısının “box plot” grafiği ile incelenmesi

Şekil 67’ye göre makale başına düşen ortalama yazar sayısı 2 ile 4 arasındadır. İstisna olarak 8 ile 12 arası yazar sayısı olan makaleler de bulunmaktadır. Şekil 68’e bakıldığında ise; ortalama atıf sayısının 10’un altında olduğu, istisna olarak 10’un üzerinde atıfların da olduğu görülmektedir. Hatta 80’in üzerinde atfın olduğu bir makalenin de olduğu görülebilmektedir.

4.1.6.3 Soru 6.3 Tek yazarlı makaleler ile çok yazarlı makalelerin atıf alma oranları ne düzeydedir?

Atıf dağılımlarının incelenmesi için makale türü “araştırma, derleme ve olgu sunumu” olan makaleler ele alınmıştır. Öncelikle tek yazarlı makaleler ile ortak yazarlı makalelerin almış oldukları toplam atıf sayılarını hesaplamak için Kod 87 ile “atıflar” adlı bir alt veri seti oluşturulmuştur. Aynı kod ile tek yazarlı ve ortak yazarlı makale atıflarının ayrı ayrı

toplamlarını içeren “atif_Sayisi” ve etiket isimlerini içeren “atif_Kategorisi” değişkenleri oluşturularak “atiflar” alt veri setine eklenmiştir.

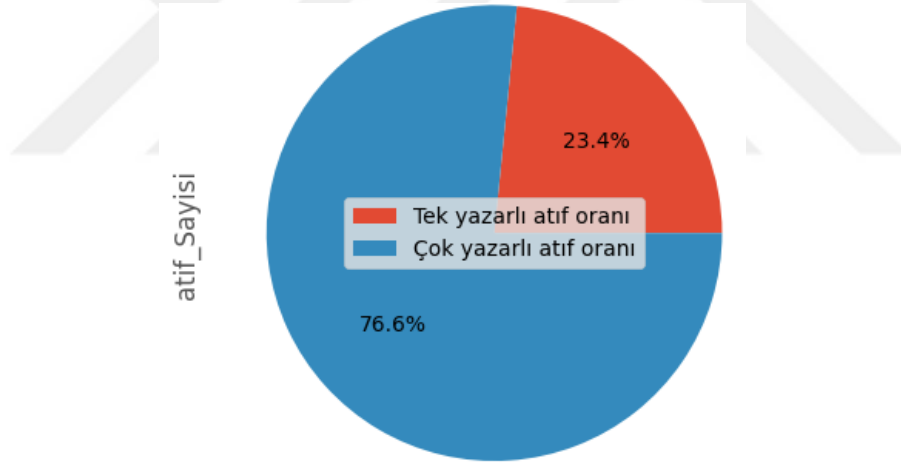
```
tek_yazarli = yazar_df.loc[yazar_df["yazar_Sayisi"] == 1, ("atif")].fillna("0").astype("int").sum()
cok_yazarli = yazar_df.loc[yazar_df["yazar_Sayisi"] != 1, ("atif")].fillna("0").astype("int").sum()
atiflar_df = pd.DataFrame({
    "atif_Kategorisi": ["Tek yazarlıların aldığı atıflar", "Ortak yazarlıların aldığı atıflar"],
    "atif_Sayisi": [tek_yazarli, cok_yazarli]
})
```

Kod 87: “atiflar_df” alt veri setinin oluşturulması (tek ve ortak yazarlı yayınlar için)

Daha sonra Kod 88 ile iki değişken matplotlib kütüphanesinin “pie chart (pasta grafik)” grafik türü kullanılarak Şekil 69’daki gibi görselleştirilmiştir.

```
ax = atiflar["atif_Sayisi"].plot(kind="pie", autopct='%1.1f%%', labels=None)
ax.legend(labels=['Tek yazarlı atıf oranı', 'Çok yazarlı atıf oranı'])
plt.show()
```

Kod 88: “atiflar_df” alt veri setinin görselleştirilmesi



Şekil 69: Tek ve ortak yazarlı yayınların atıf oranları

Şekil 69’a bakıldığında tek yazarlı yayınların toplam 5.167 atfın %23,4’ünü, ortak yazarlı yayınların ise %76,6’sını aldığı görülmektedir.

4.1.6.4 Soru 6.4 Kurumsal iş birliği ne düzeydedir?

AÜ’nün diğer kurumlarla iş birliğini incelemek için türü “araştırma, derleme ve olgu sunumu” olan ve yazar sayısı birden fazla olup en az bir yazarının AÜ haricinde bir kurumdan

olduğu makaleler ele alınmıştır. Bir makalede aynı kurumdan birden fazla yazarın olabileceği düşünülerek Kod 89 ile veri setindeki “yazar_Kurum” değişkenindeki kurum adları her makale için tekilleştirilmiştir.

```
yazar_df.yazar_Kurum=yazar_df.yazar_Kurum.str.replace('I','i').str.replace('İ','i').str.lower().str.title()
yazar_df = yazar_df.loc[yazar_df['yazar_Sayisi'] !=1]
yazar_df.reset_index(drop=True, inplace=True)
def deduplicate_institutions(institutions):
    unique_institutions = set(institutions)
    return ';'.join(unique_institutions)
yazar_df['yazar_Kurum'] = yazar_df['yazar_Kurum'].apply(lambda x: deduplicate_institutions(x.split(';')))
```

Kod 89: Makaleler için kurum adlarının tekilleştirilmesi

Daha sonra Kod 90 ile “yazar_df”teki “yazar_Kurum” değişkeni kullanılarak “kurum_df” adlı bir alt veri seti oluşturulmuş ve yazar sayıları “yazar_Kurum”a göre gruplandırılmıştır.

```
kurum = []
for i in range(len(yazar_df['yazar_Sayisi'])):
    for j in range(len(yazar_df['yazar_Kurum'][i].split(';'))):
        kurum.append(yazar_df['yazar_Kurum'][i].split(';')[j])
kurum_df = pd.DataFrame()
kurum_df['kurum'] = kurum
kurum
```

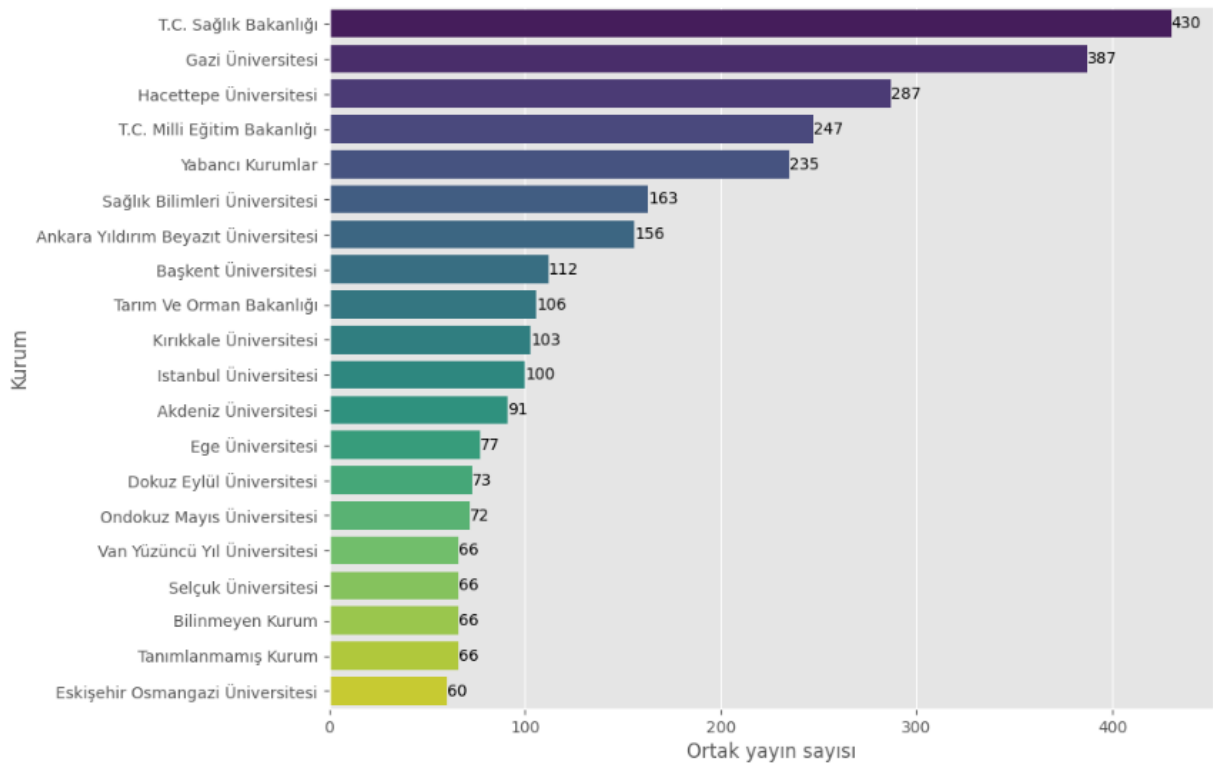
Kod 90: “kurum_df” alt veri setinin oluşturulması

Kod 91 ile “kurum_df” veri setindeki “kurum” değişkeninde her makale için kayıtlı olan “Ankara Üniversitesi” verisi silinerek sadece iş birliği yapılan diğer kurum adları kalacak şekilde görselleştirme yapılmıştır. Şekil 70’te seaborn kod kütüphanesinin yatay bar grafik türü kullanılarak AÜ’nün en çok iş birliği yaptığı 20 kurumun dağılımı verilmiştir.

```
import seaborn as sns
import matplotlib.pyplot as plt

filtered_counts = kurum_df.loc[kurum_df['kurum'] != 'Ankara Üniversitesi', 'kurum'].value_counts().head(20)
plt.figure(figsize=(10, 8))
sns.barplot(x=filtered_counts.values, y=filtered_counts.index, palette='viridis')
for i, count in enumerate(filtered_counts):
    plt.text(count, i, str(count), ha='left', va='center')
plt.xlabel('Ortak yayın sayısı')
plt.ylabel('Kurum')
plt.title('En çok iş birliği yapılan kurumlar (20 kurum)')
plt.show()
```

Kod 91: En çok iş birliği yapılan kurumlar (20 kurum)



Şekil 70: En çok iş birliği yapılan kurumlar (20 kurum)

```
▶ yazar_df.loc[yazar_df['yazar_kurum'] != 'Ankara Üniversitesi', 'yazar_kurum'].count()
3694
```

Kod 92: Diğer kurumlar ile iş birliği sayısı

```
▶ kurum_df.loc[kurum_df['kurum'] != 'Ankara Üniversitesi'].nunique()
kurum      383
```

Kod 93: İş birliği yapılan tekil kurum sayısı

Bir makalede bir kuruma ait birden fazla yazar bulunabilmektedir. AÜ tarafından iş birliği yapılan kurumların toplam tekil sayısını tespit edebilmek için Kod 93 uygulanmıştır.

```
▶ yazar_df.loc[yazar_df['yazar_kurum'] == 'Ankara Üniversitesi', 'yazar_kurum'].count()
1750
```

Kod 94: Kurum içi diğer akademisyenlerle iş birliği sayısı

Makale türü “araştırma, derleme ve olgu sunumu” ve ortak yazarlı olan 5.444 makalenin 3.694’ünde (Kod 92) 383 kurum (Kod 93) ile iş birliği yapıldığı, 1.750’sinde (Kod 94) ise Ankara Üniversitesi içerisinde iş birliği yapıldığı tespit edilmiştir. Şekil 70 incelendiğinde en çok iş birliği yapılan kurumların sırasıyla; T.C. Sağlık Bakanlığı (430), Gazi Üniversitesi (387), Hacettepe Üniversitesi (287), Millî Eğitim Bakanlığı (247), Yabancı ülke kurumları (235), Sağlık Bilimleri Üniversitesi (163) ve Yıldırım Beyazıt Üniversitesi (156) olduğu tespit edilmiştir. TR Dizin’de yazar üst veri alanında AÜ’nün iş birliği yaptığı yabancı ülke kurumları “yabancı kurum” şeklinde kaydedilmiştir. Kurum verilerinin toplandığı TR Dizin üst veri alanında ülke ve ülkelerin kurum isimleri olmadığı için yabancı kurumlara dair bir analiz yapılamamıştır.

Ortak yazarlı 5.444 makalede, iş birliğinin sadece kurum içinde mi yoksa diğer kurumlarla mı yapıldığını görmek için “yazar_df” alt veri setinin “yazar_Kurum” sütunu incelenmiştir. Öncelikle Kod 95 ile “yazar_Kurum” değişkeninde her makale için “Ankara Üniversitesi” verisinin yanında başka bir kurum adı olup olmadığı tespit edilmiştir. “Ankara Üniversitesi” verisinin yanında başka bir kurum adı varsa “diğer kurumlarla iş birliği” olduğu, yoksa kurum içi iş birliğinin olmadığı varsayılmıştır.

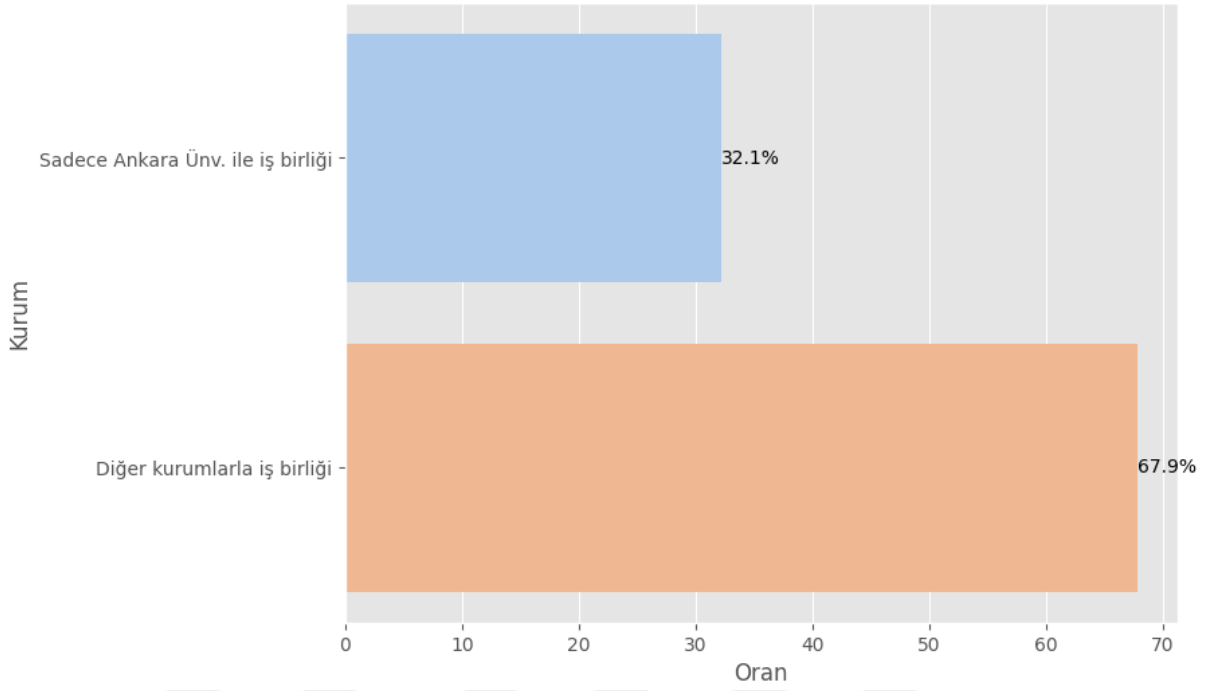
```
ankara_uni_count = yazar_df.loc[yazar_df['yazar_Kurum'] == 'Ankara Üniversitesi', 'yazar_Kurum'].count()
other_inst_count = yazar_df.loc[yazar_df['yazar_Kurum'] != 'Ankara Üniversitesi', 'yazar_Kurum'].count()
data = {'Institution': ['Sadece Ankara Üniv. ile iş birliği', 'Diğer kurumlarla iş birliği'],
        'Count': [ankara_uni_count, other_inst_count]}
ortak_yazarli_makaleler = pd.DataFrame(data)
total_count = ortak_yazarli_makaleler['Count'].sum()
ortak_yazarli_makaleler['Percent'] = ortak_yazarli_makaleler['Count'] / total_count * 100
```

Kod 95: Kurum iş birliği durumunun tespit edilmesi

Daha sonra seaborn kütüphanesinin barplot grafik türü kullanılarak Kod 96 ile Şekil 71’de iş birliği durumu görselleştirilmiştir.

```
import seaborn as sns
plt.figure(figsize=(8, 6))
sns.barplot(data=ortak_yazarli_makaleler, x="Percent", y="Institution", palette='pastel')
for i, percent in enumerate(ortak_yazarli_makaleler['Percent']):
    plt.text(percent, i, f"{percent:.1f}%", ha='left', va='center')
plt.xlabel('Oran')
plt.ylabel('Kurum')
plt.show()
```

Kod 96: Ortak yazarlı makalelerdeki kurum iş birliğinin görselleştirilmesi



Şekil 71: Ortak yazarlı makalelerde iş birliği oranı

Şekil 71’e bakıldığında, ortak yazarlı makalelerde diğer kurumlarla iş birliği oranının kurum içindeki iş birliğinden iki kat fazla olduğu görülmektedir. Diğer kurumlarla yapılan ortak yazarlı makaleler ile ilgili detaylı bilgi vermekte yarar vardır. Bu makalelerde en az bir yazar AÜ’den ve en az bir yazar diğer kurumlardan olacak şekilde filtreleme yapılmıştır.

4.1.6.5 Soru 6.5 Kurumsal iş birliğinin yıla göre dağılımı nedir?

Bütün makale türleri dâhil edilerek toplamda üretilen makaleler içerisinde (herhangi bir filtreleme yapmaksızın) kamu iş birliğinin güçlü mü yoksa zayıf mı olduğunu tespit edebilmek için “makale_Kurum” değişkeninde yer alan veriler kullanılmıştır. Veri setinde; makale kurum değişkenindeki veriler (kurum adları), eğer her yayın için kurum adı birden fazla ise kurum adları noktalı virgül ile ayrılarak kaydedilmiştir. Bir makalenin kurum iş birliği ile üretilip üretilmediği, ilgili kaydın “makale_Kurum” değişkeninde “Ankara Üniversitesi” değerinin yanında başka kurum adı/adlarının olup olmadığına bakılarak tespit edilmeye çalışılmıştır. Öncelikle kamu iş birliği olup olmadığı bilgisi Kod 97 ile daha önce oluşturulmuş olan “yazar_df” alt veri setine sütun olarak eklenmiştir.

```

isbirliigi = []
for i in range(len(yazar_df['yazar_Kurum'])):
    deger=""
    for j in range(len(yazar_df['yazar_Kurum'][i].split(';'))):
        if "Ankara Üniversitesi" in yazar_df['yazar_Kurum'][i].split(';')[j]:
            deger+="1"
        else:
            deger+="0"
    if "0" in deger:
        isbirliigi.append("Var")
    else:
        isbirliigi.append("Yok")
yazar_df['kamu_isbirliigi'] = isbirliigi

```

Kod 97: İş birliği durumunun veri setine sütun olarak eklenmesi

Daha sonra kamu iş birliği durumu oran olarak (Kod 98) ve yıllara göre dağılımı (Kod 99) Şekil 72 ve Şekil 73'teki gibi görselleştirilmiştir.

```

explode = (0.0, 0.1)
fig, ax = plt.subplots()
ax.pie(yazar_df.kamu_isbirliigi.value_counts(),
       explode=explode,
       labels=yazar_df.kamu_isbirliigi.value_counts().index,
       autopct='%1.1f%%',
       shadow=True,
       startangle=30)
ax.axis('equal')
ax.set_title('Kamu İşbirliği Durumu')

```

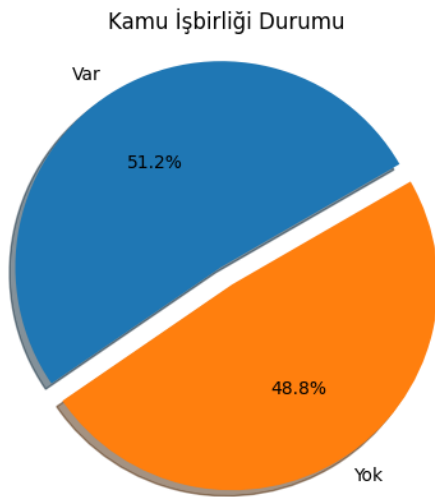
Kod 98: Kamu iş birliği durumu

```

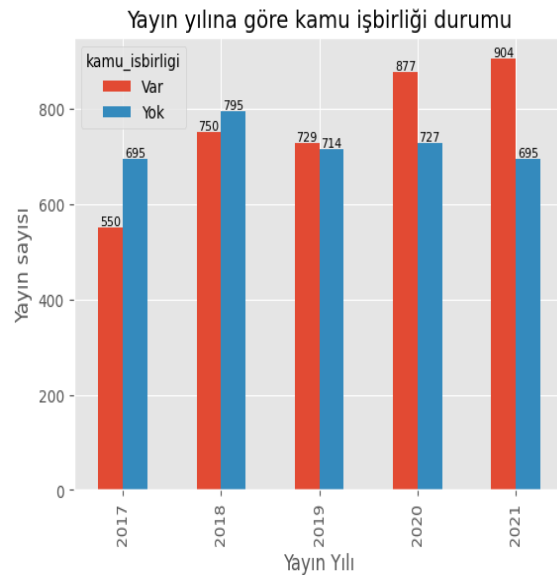
k_isbirliigi=yazar_df.groupby("yil")["kamu_isbirliigi"].value_counts()
k_isbirliigi=k_isbirliigi.reset_index(name = 'toplam')
fig,ax= plt.subplots()
plt.style.use('ggplot')
k_isbirliigi.pivot_table(index='yil', columns='kamu_isbirliigi',
                          values='toplam').plot(kind="bar", ax=ax)
plt.title("Yayın yılına göre kamu işbirliği durumu")
plt.xlabel("Yayın Yılı")
plt.ylabel("Yayın sayısı")
for container in ax.containers:
    ax.bar_label(container, fontsize=8)

```

Kod 99: Kamu iş birliği durumu (Yayın yılına göre)



Şekil 72: Kamu iş birliği durumu



Şekil 73: Kamu iş birliği durumu (Yayın yılına göre)

Şekil 72'ye bakıldığında, genel olarak beş yıllık kamu iş birliği oranının %51,2 olduğu görülmektedir. Şekil 73'te kırmızı renkteki çubuklar, AÜ yazarlarının en az bir diğer kurum yazarı/yazarları ile iş birliği yaptığı, mavi çubuklar ise sadece AÜ yazarlarının makale sayısını göstermektedir. 2017 ve 2018 yıllarındaki makalelerde kamu iş birliğinin diğer yıllara göre zayıf olduğu görülmektedir.

4.1.6.6 Soru 6.6 Kamu iş birliği olan makalelerde ilk yazara göre yazar kurum dağılımı nedir?

Başka kurumlarla ortak makalelerde ilk yazar olmak hem araştırmacı hem de çalışılan kurum için önemli bir başarı ve itibar unsuru olarak değerlendirilmektedir. İlk yazarlık genellikle çalışmanın en büyük katkısını sağlayan kişiyi temsil eder ve bu durum, üniversitenin araştırma ve akademik başarısını da vurgular. İlk yazarlık aynı zamanda araştırmanın liderliğini ve yönlendirilmesini ifade eder, bu da üniversitenin araştırma alanındaki etkinliğini ve liderliğini gösterir.

Bu görüşten yola çıkarak veri setimizi inceleyecek olursak, oluşturulan “yazar_df” alt veri setindeki “yazar_Kurum” sütununda, yazarların kurum bilgilerinin yer aldığı daha önce ifade edilmişti. Bu veri alanı aracılığıyla akademik yayınların tek kurum mu, yoksa birden fazla kurumun iş birliği ile mi üretildiği bilgisi tespit edilebilmektedir. Ayrıca bu veri alanında birden çok kurum ismi varsa (ortak yazarlı ise), ilk yazarın kurum bilgisi farklı bir değişkene atanarak ilk kurum tespit edilebilmektedir.

Öncelikle “yazar_df” alt veri setinde Kod 100 ile makale tür bilgisi “araştırma, derleme ve olgu sunumu” olan makaleler filtrelenmiştir.

```
yazar_df = pd.DataFrame()
yazar_df['yazar_Adi'] = df.yazar_Adi
yazar_df['yazar_Sayisi'] = df.yazar_Adi.str.count(',') + 1
yazar_df['atif'] = df.atif_Sayisi
yazar_df['yil'] = df.makale_Yil
yazar_df['makale_Turu'] = df.makale_Turu
yazar_df['yazar_Kurum'] = df.yazar_Kurum.str.replace('İ', 'i').str.replace('I', 'i').str.lower().str.title()
yazar_df = yazar_df[yazar_df['makale_Turu'].isin(['Araştırma Makalesi', 'Derleme', 'Olgu Sunumu'])].reset_index()
yazar_df.tail()
```

Kod 100: “araştırma, derleme ve olgu sunumu” türünde olan makalelerin filtrelenmesi

Kod 100 ile “yazar_df” veri setinden makale türü “araştırma, derleme ve olgu sunumu” olanlar haricindeki türler silindikten sonra “yazar_df” veri setinde 7.118 makale kalmıştır. Daha sonra Kod 101 ile “yazar_df” alt veri setinde ortak yazarlı makaleler filtrelenmiştir.

```

yazar_df.yazar_Kurum=yazar_df.yazar_Kurum.str.replace('I','i').str.replace('İ','i').str.lower().str.title()
yazar_df = yazar_df.loc[yazar_df['yazar_Sayisi'] !=1]
yazar_df.reset_index(drop=True, inplace=True)
if 'index' in yazar_df.columns:
    yazar_df.drop(['index'], axis=1, inplace=True)
yazar_df.tail()

```

Kod 101: “yazar_df” veri setindeki ortak yazarlı makalelerin filtrelenmesi

Kod 101 ile “yazar_df” veri setinden makale türü “araştırma, derleme ve olgu sunumu” olan makalelerden ortak yazarlı olanlar belirlenmiştir. Buna göre ortak yazarlı makale sayısı 5.444’tür.

Bu ortak yazarlı makalelere Kod 102 uygulanarak her bir makaledeki mükerrer kurum isimleri tekilleştirilmiş ve AÜ yazarlarının sadece kurum içinde yaptıkları iş birlikleri kayıtları “yazar_df” veri setinden silinmiştir. Sadece diğer kurumlarla yapılan iş birliği kayıtları kalacak şekilde veri seti güncellenmiştir.

```

def deduplicate_institutions(institutions):
    unique_institutions = []
    seen = set()
    for inst in institutions.split(';'):
        if inst not in seen:
            unique_institutions.append(inst)
            seen.add(inst)
    return ';'.join(unique_institutions)

yazar_df['yazar_Kurum'] = yazar_df['yazar_Kurum'].apply(lambda x: deduplicate_institutions(x))
yazar_df = yazar_df.loc[yazar_df['yazar_Kurum'] != 'Ankara Üniversitesi'].reset_index()
columns_to_drop = ['level_0', 'index']
if all(col in yazar_df.columns for col in columns_to_drop):
    yazar_df.drop(columns=columns_to_drop, axis=1, inplace=True)
yazar_df.tail()

```

Kod 102: AÜ yazarlarının diğer kurumlarla yaptıkları iş birliği sayısı

Kod 102 ile “yazar_df” veri setinde AÜ yazarlarının diğer kurumlarla iş birliği yaptıkları makaleler belirlenmiştir. Diğer kurumlarla yapılan iş birliği makale sayısı 3.694’tür.

Daha sonra Kod 103 ile kamu iş birliği olan makalelerdeki ilk kurum bilgisi “ilk_Kurum” adlı bir değişkene atanarak “yazar_df” alt veri setine eklenmiştir.

```

ilk_kurum = []
for i in range(len(yazar_df.yazar_Kurum.str.count(';')+1)):
    ilk_kurum.append(yazar_df.yazar_Kurum[i].split(';')[0])
yazar_df['ilk_Kurum'] = ilk_kurum
yazar_df.drop("yazar_Adi", axis='columns', inplace=True)
yazar_df.tail()

```

Kod 103: İlk kurum değişkeninin oluşturulması

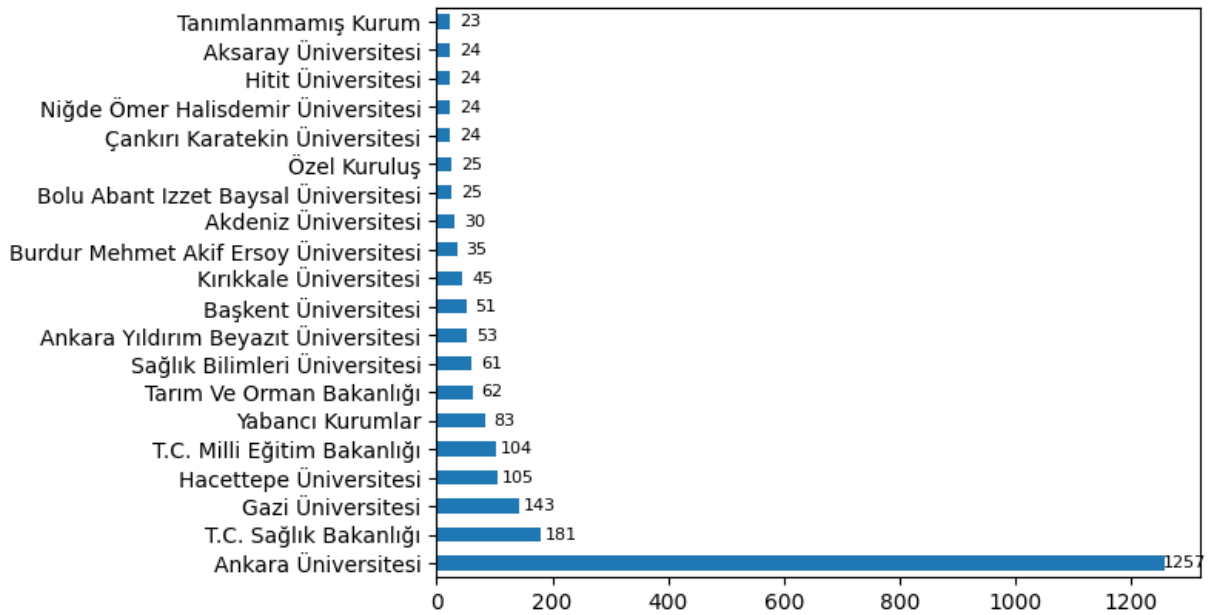
Tablo 25: “yazar_df” alt veri setine makale “ilk_Kurum” değişkeninin eklenmesi

yazar_Sayisi	atif	yil	makale_Turu	yazar_Kurum	kamu_isbirligi	ilk_Kurum
2	1.0	2018	Araştırma Makalesi	Gazi Üniversitesi;Ankara Üniversitesi	Var	Gazi Üniversitesi
7	NaN	2020	Araştırma Makalesi	T.C. Sağlık Bakanlığı;T.C. Sağlık Bakanlığı;T...	Var	T.C. Sağlık Bakanlığı
2	NaN	2021	Araştırma Makalesi	Zonguldak Bülent Ecevit Üniversitesi;Ankara Ün...	Var	Zonguldak Bülent Ecevit Üniversitesi
2	1.0	2019	Araştırma Makalesi	Ankara Üniversitesi;Sakarya Üniversitesi	Var	Ankara Üniversitesi
6	1.0	2018	Derleme	Ankara Üniversitesi;Ankara Üniversitesi;Ankara...	Var	Ankara Üniversitesi

Kod 104 ile AÜ yazarlarının diğer kurum/kurumlarla yaptıkları makalelerin ilk yazar kurumuna göre dağılımı Şekil 74’teki “barh plot” (yatay çubuk grafik) grafikte görselleştirilmiştir. Grafik matplotlib kütüphanesi kullanılarak oluşturulmuştur.

```
yazar_df.ilk_Kurum.value_counts().head(20).plot(kind="barh")
for index, value in enumerate(yazar_df.ilk_Kurum.value_counts()):
    plt.annotate(value, xy=(35+value,index), ha='center', va='center',fontsize=8)
```

Kod 104: Diğer kurumlarla iş birliği ile oluşturulan makalelerdeki ilk kurum durumu (20 kayıt)



Şekil 74: Diğer kurumlarla iş birliği yapılan makalelerdeki ilk kurum durumu (20 kayıt)

AÜ, beş yılda üretilen 7.436 makaleden kurum iş birliği ile yaptığı 3.694 makalenin %34,02’sinde (1.257) ilk kurum statüsündedir. İlk kurum statüsünde olan diğer kurumlar sırasıyla T.C. Sağlık Bakanlığı (181), Gazi Üniversitesi (143), Hacettepe Üniversitesi (105), T.C. Millî Eğitim Bakanlığı (104) ve Yabancı Kurumlar’dır (83).

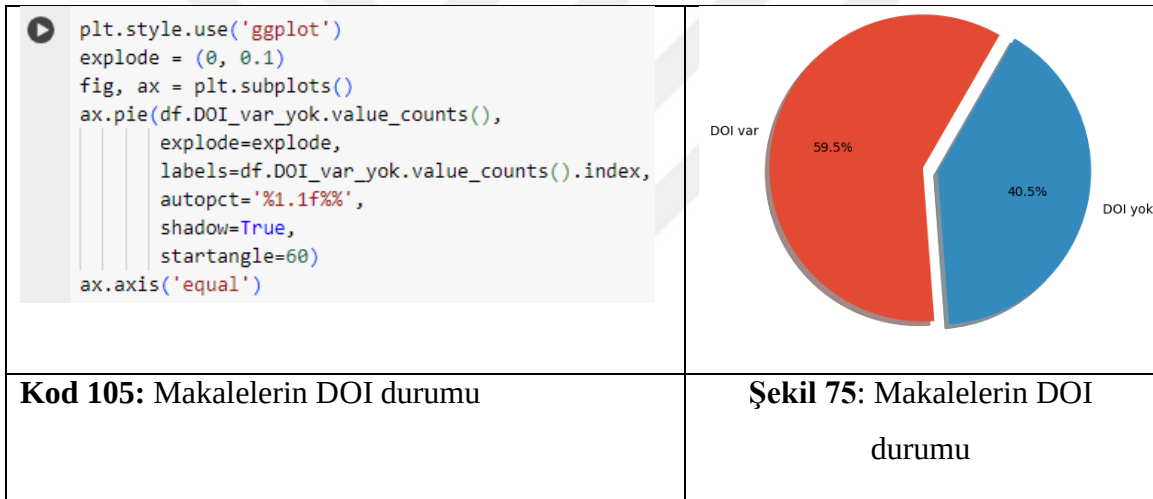
Ankara Üniversitesi toplam 383 tekil kurum (Kod 93) ile 3.694 makalede (Kod 92) iş birliğine gitmiş ve çalışmanın en büyük katkısını sağlayan kişiyi temsil eden ilk yazarlıkta

%34,02 oranında bir paya sahip olmuştur. İlk yazarlık aynı zamanda araştırmanın liderliğini ve yönlendirilmesini ifade eden bir gösterge niteliğindedir. Şekil 74’teki durum üniversitenin araştırma alanındaki etkinliğini göstermektedir.

4.1.7. Soru 7 Makalelerin konusal olarak dağılımı nedir?

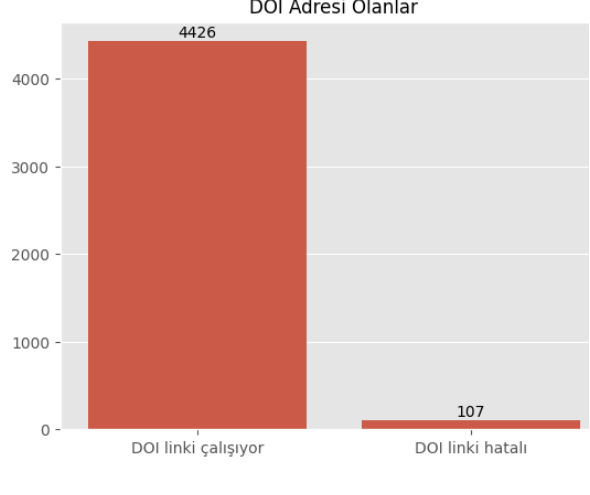
Makale konusu ile ilgili veriler, DOI adresi DergiPark olan makale üst veri alanlarından toplanmıştır. DOI, nesnelerin benzersiz bir şekilde tanımlanmasına ve güvenilir bir şekilde erişilmesine olanak tanıyan benzersiz bir erişim numarasıdır (DOI Foundation, 2022). Makale üst veri alanlarına erişilebilmek bağlamında DOI adresi büyük önem taşımaktadır.

Öncelikle TR Dizin’de AÜ adresli kaç makalenin DOI üst veri alanının dolu olduğu, veri setindeki “DOI_Durumu” değişkeni kullanılarak tespit edilmeye çalışılmıştır. Bunun için Kod 105 ile Şekil 75’teki pasta grafik oluşturulmuştur.

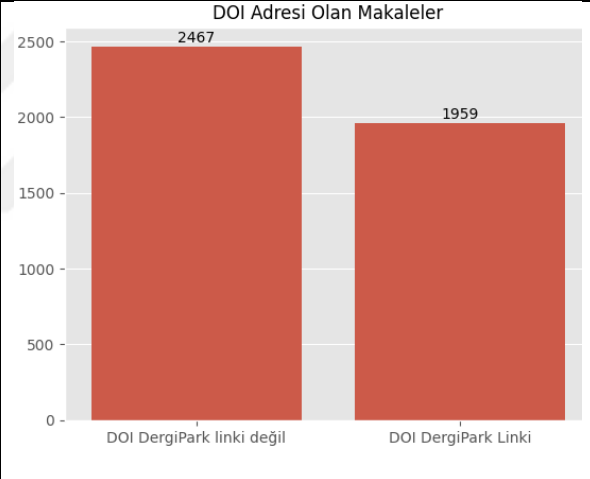


Şekil 75 incelendiğinde; TR Dizin’de bulunan beş yıllık toplam 7.436 AÜ yayınının %59,5’inde (4.533) DOI linki bulunduğu, %40.5’inde ise DOI linkinin bulunmadığı tespit edilmiştir.

DOI üst veri alanındaki linklerinin çalışıp çalışmadığının tespit edilmesi için de Kod 106 ile Şekil 76’daki çubuk grafik oluşturulmuştur. Şekil 76’ya göre 4.426 makalenin erişilebilir DOI linkleri barındırdığı, 107 tanesinin ise erişim linkinin çalışmadığı tespit edilmiştir.

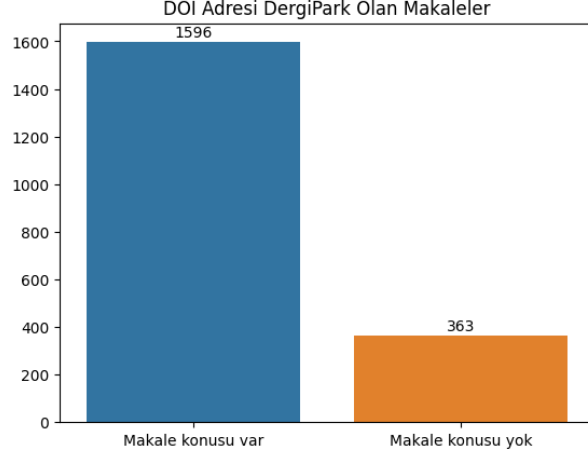
<pre> a=['DOI linki çalışıyor','DOI linki hatalı'] b=[4426,107] sns.barplot(x=a,y=b) for i, v in enumerate(b): plt.text(i, v + 10, str(v), ha='center', va='bottom', color='black') plt.title('DOI Adresi Olanlar', fontsize=12) plt.show() </pre>	 <table border="1"> <thead> <tr> <th>DOI Adresi Durumu</th> <th>Sayı</th> </tr> </thead> <tbody> <tr> <td>DOI linki çalışıyor</td> <td>4426</td> </tr> <tr> <td>DOI linki hatalı</td> <td>107</td> </tr> </tbody> </table>	DOI Adresi Durumu	Sayı	DOI linki çalışıyor	4426	DOI linki hatalı	107
DOI Adresi Durumu	Sayı						
DOI linki çalışıyor	4426						
DOI linki hatalı	107						
Kod 106: DOI linkinin çalışma durumu	Şekil 76: DOI linkinin çalışma durumu						

4.426 makalenin kaç tanesinin erişim adresinin DergiPark olduğunu tespit etmek için Kod 107 ile Şekil 77’deki çubuk grafik türü oluşturulmuştur.

<pre> a=['Makale konusu var','Makale konusu yok'] b=[1596,363] sns.barplot(x=a,y=b) for i, v in enumerate(b): plt.text(i, v + 10, str(v), ha='center', va='bottom', color='black') plt.title('DOI Adresi DergiPark Olan Makaleler', fontsize=12) plt.show() </pre>	 <table border="1"> <thead> <tr> <th>DOI Adresi Durumu</th> <th>Sayı</th> </tr> </thead> <tbody> <tr> <td>DOI DergiPark linki değil</td> <td>2467</td> </tr> <tr> <td>DOI DergiPark Linki</td> <td>1959</td> </tr> </tbody> </table>	DOI Adresi Durumu	Sayı	DOI DergiPark linki değil	2467	DOI DergiPark Linki	1959
DOI Adresi Durumu	Sayı						
DOI DergiPark linki değil	2467						
DOI DergiPark Linki	1959						
Kod 107: DOI hedef URL adresi belirleme	Şekil 77: DOI hedef URL adresi belirleme						

Şekil 77 incelendiğinde; 4.426 makalenin 1.959’unun hedef URL adresinin DergiPark olduğu, 2.467 tanesinin DergiPark haricinde, makalelerin yayınlandığı dergi web sayfalarına yönlendirildiği tespit edilmiştir.

DOI erişim adresi DergiPark olan 1.959 makalenin DergiPark sayfalarına gidilerek kaç tanesinin “makale konusu” üst veri alanında konu verisi olduğunu tespit etmek için Kod 108 ile Şekil 78’deki çubuk grafik oluşturulmuştur.

<pre> a=['DOI linki çalışıyor','DOI linki hatalı'] b=[4426,107] sns.barplot(x=a,y=b) for i, v in enumerate(b): plt.text(i, v + 10, str(v), ha='center', va='bottom', color='black') plt.title('DOI Adresi Olanların Durumu', fontsize=12) plt.show() </pre>	
Kod 108: DOI adresi DergiPark olan makaleler	Şekil 78: DOI adresi DergiPark olan makaleler

Şekil 78’de görüleceği üzere, DOI erişim adresi DergiPark olan 1.959 makalenin 1.596 tanesinde DergiPark makale üst veri sayfasındaki “makale konusu” alanında veri olduğu, 363 tanesinde ise “makale konusu” verisinin bulunmadığı tespit edilmiştir.

TR Dizin makale detay alanında bulunan DOI linklerinin kullanılmasının amacı; yayınlara ait TR Dizin’de olmayan makale konu alanlarını DergiPark platformu üzerinden toplamak için referans olarak kullanmaktır. TR Dizindeki DOI linki aracılığıyla makalelerin DergiPark kaydına gidililerek “makale_Konu” değişkenindeki veriler toplanmıştır.

DergiPark’tan toplanan makale konu verilerinin analizi için öncelikle Kod 109 ile “makale_Konu, makale_Yıl, makale_Dil ve makale_konu_Kategorisi” değişkenlerini içeren “konu_df” adlı bir alt veri seti oluşturulmuştur.

```

konu = []
yil = []
dil = []
kkategori = []
say = []
for i in range(len(df['makale_Konu'])):
    if pd.isnull(df['makale_Konu'][i]) == False:
        for j in range(len(df['makale_Konu'][i].split(';'))):
            konu.append(df['makale_Konu'][i].split(';')[j].strip())
            yil.append(df['makale_Yil'][i])
            dil.append(df['makale_Dil'][i])
            kkategori.append(df['makale_konu_Kategorisi'][i])
            say.append(1)
konu_df = pd.DataFrame()
konu_df['konu'] = konu
konu_df['yil'] = yil
konu_df['dil'] = dil
konu_df['kkategori'] = kkategori
konu_df['sayi'] = say

```

Kod 109: “konu_df” alt veri setinin oluşturulması

Makale konu alanı dolu olan 1.596 kaydın 127'sinde birden fazla konu başlığı girildiği tespit edilmiş ve “konu_df” alt veri seti her konu başlığı bir satırda olacak şekilde düzenlenmiştir. Böylece alt veri setindeki makale konu sayısı 1.839'a ulaşmıştır. “konu_df” alt veri setinin ilk beş kaydı “konu_df.head()” kodu ile Tablo 26'da verilmiştir.

Tablo 26: “konu_df” alt veri setinin ilk beş satırı

```
konu_df.head()
```

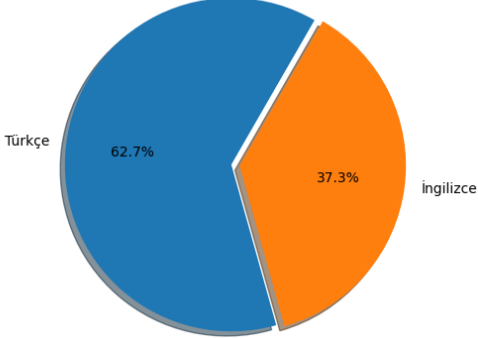
	konu	yil	dil	kkategori	sayi
0	Radyoloji ve Organ Görüntüleme	2021	Türkçe	Fen	1
1	Mimarlık	2020	Türkçe	Fen	1
2	Mathematical Sciences	2020	İngilizce	Fen	1
3	Mühendislik-Ortak Disiplinler	2021	Türkçe	Fen	1
4	Spor Hekimliği	2019	Türkçe	Sosyal	1

“konu_df” alt veri seti, herhangi bir filtreleme yapılmaksızın Kod 110 ile Şekil 79'daki “plotly.express” veri görselleştirme paketinin “sunburst³²” türü ile görselleştirilmiştir.

```
import plotly.express as px
px.sunburst(konu_df, color="konu", values="sayi", path=["yil", "dil", "kkategori", "konu"],
            hover_name="konu", height=850,
            color_discrete_sequence=px.colors.sequential.Agsunset_r)
```

Kod 110: “konu_df” alt veri setinin yıl, dil, konu kategorisi ve makale konusuna göre dağılımı (sunburst)

³² Sunburst chart: Halka grafiği ve radyal ağaç haritası gibi birden fazla adla bilinen Sunburst grafiği, hiyerarşik bir veri kümesini görselleştirmek için kullanılır. Hiyerarşiyi, her halkanın hiyerarşideki bir düzeye karşılık geldiği bir dizi eşmerkezli halka aracılığıyla gösterir. Her halka, kendisini oluşturan ayrıntıları temsil edecek şekilde orantılı olarak bölümlere ayrılmıştır. Halkadaki bir segmente odaklanmak, parçanın ana halka segmentiyle ilgili tüm ilişkisine dair bir fikir verir (FusionCharts, 2024a).

<pre> explode = (0, 0.05) fig, ax = plt.subplots() ax.pie(konu_df.dil.value_counts().head(2), explode=explode, labels=konu_df.dil.value_counts().head(2).index, autopct='%1.1f%%', shadow=True, startangle=60) ax.axis('equal') ax.set_title('Makale Konularının Dillere Göre Dağılımı') </pre>	Makale Konularının Dillere Göre Dağılımı (Türkçe,İngilizce) 
Kod 111: Konuların dillere göre dağılımı	Şekil 80: Konuların dillere göre dağılımı

Şekil 80 incelendiğinde; makale konusu verisi DergiPark’tan çekilen yayınların %62,7’sinin (1.144) konularının Türkçe, %37,3’ünün (681) ise İngilizce dilinde olduğu tespit edilmiştir.

Konu başlıkları yayın diline göre aynı anlama gelecek kayıtlar içermektedir. Örneğin Türkçe bir yayındaki “mühendislik” konu başlığı İngilizce bir yayında “engineering” olarak geçebilmektedir. Bu nedenle konu başlıkları alt veri setinde, Kod 112,113 ve 114 ile her konu başlığı dillere göre ayrıştırılmıştır.

Türkçe yayınlardaki konu başlığı sayısı Kod 112’de verilmiştir.

```

konu_df[konu_df.dil == 'Türkçe'].count()

```

konu 1144

Kod 112: Türkçe makale konuları sayısı

İngilizce yayınlardaki konu başlığı sayısı Kod 113’te verilmiştir.

```

konu_df[konu_df.dil == 'İngilizce'].count()

```

konu 681

Kod 113: İngilizce makale konuları sayısı

Türkçe ve İngilizce haricindeki yayınlar için atanan konu başlığı sayısı ise Kod 114’te verilmiştir.

```

▶ konu_df[(konu_df.dil != 'Türkçe') & (konu_df.dil != 'İngilizce')].count()
konu          14

```

Kod 114: Türkçe ve İngilizce olmayan makale konu sayısı

Şekil 79’deki görsele yansıtılmayan diğer dillerdeki konu başlıklarının toplam oranının %0,8’ini (14) oluşturduğu görülmektedir. Makale dili Türkçe ve İngilizce olan makalelere ait konular izleyen başlıklarda ayrı ayrı incelenmiştir.

4.1.7.1 Türkçe makalelerin konusal olarak yıl ve genel konu kategori bazındaki dağılımı nedir?

Türkçe makaleler için atanan konu başlıklarının “konu_df” alt veri setindeki “yıl” ve “kkategori”si (makale_konu_Kategorisi) değişkenlerine göre kümelenmiş durumları Kod 115 ile oluşturulan Şekil 81’deki “treemap³³” etkileşimli grafiğinde verilmiştir. Treemap grafiği, Python programlama dilinde veri görselleştirme için kullanılan bir diğer paket olan “plotly.express” kullanılarak elde oluşturulmuştur. Görselleştirmede kullanılan değişken hiyerarşisi; “yıl > kkategori > konu” şeklindedir.

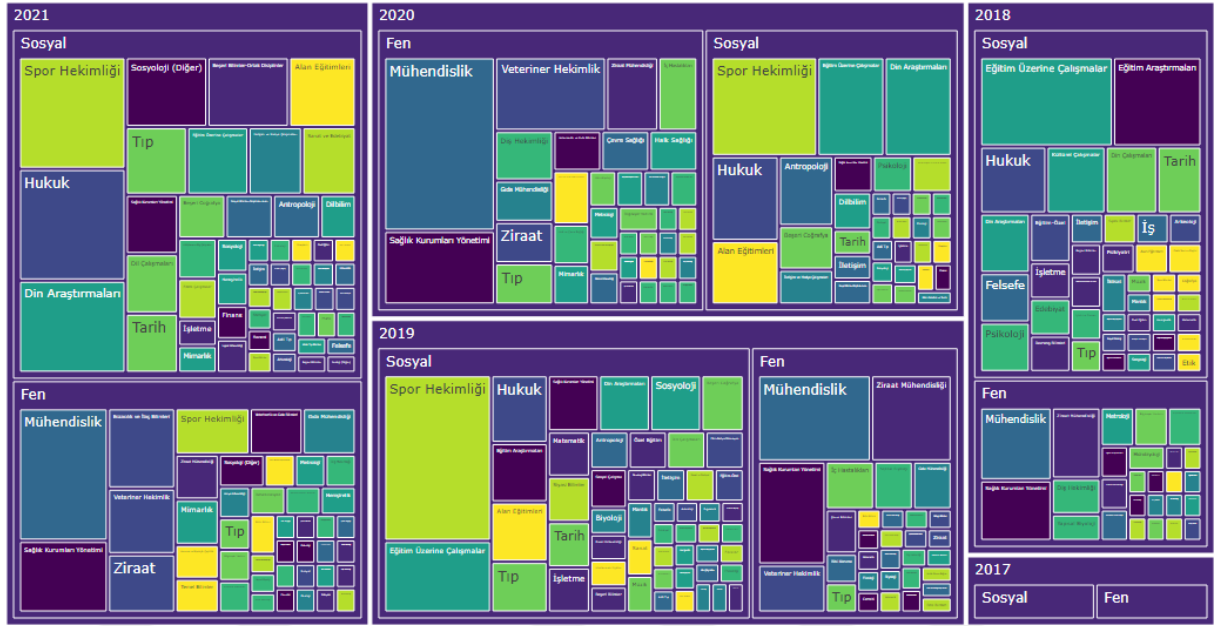
```

▶ import plotly.express as px
px.treemap(konu_df[konu_df.dil == 'Türkçe'], color='konu', values="sayi",
           path=['yıl', 'kkategori', 'konu'], height=700,width=1200,
           color_discrete_sequence=px.colors.sequential.Viridis_r)

```

Kod 115: Türkçe yayınların “konu_df” veri setindeki “yıl”, “kkategori” ve “konu” değişkenlerine göre dağılımı (treemap)

³³ Treemap, iç içe dikdörtgenler kullanarak hiyerarşik verileri görüntülemek için kullanılan bir veri görselleştirme tekniğidir; ağaç haritası grafiği bu veri görselleştirme tekniğine dayanarak oluşturulur. Hiyerarşik verileri ağaç benzeri bir yapıda temsil etmek için kullanılır. Dallar ve alt dallar olarak düzenlenen veriler, boyutları ve çizim renkleri her dikdörtgenle ilişkili niceliksel değişkenlere göre hesaplanan dikdörtgenler kullanılarak temsil edilir (FusionCharts, 2024b).



Şekil 81: Türkçe yayınların “konu_df” veri setindeki “yıl”, “kategorisi” ve “konu” değişkenlerine göre dağılımı (treemap)

Türkçe yayınlarda 2021 yılında Sosyal Bilimler alanında en çok yayın yapılan konular “spor hekimliği, hukuk, din araştırmaları, sosyoloji, beşerî bilimler-ortak disiplinler ve alan eğitimleri”dir. Fen alanında ise “mühendislik, sağlık kurumları yönetimi, eczacılık ve ilaç bilimleri ve veteriner hekimlik”tir.

2020 yılında Sosyal Bilimler alanında en çok yayın yapılan makale konuları sırasıyla “spor hekimliği, eğitim üzerine çalışmalar, din araştırmaları, hukuk ve alan eğitimleri” konularıdır. Fen temel konu kategorisinde ise “mühendislik, sağlık kurumları yönetimi, veteriner hekimlik, ziraat mühendisli ve iç hastalıklar” konularıdır.

2019 yılında Sosyal Bilimler alanında en çok yayın “spor hekimliği ve eğitim üzerine çalışmalar” konularında, Fen temel konu kategorisinde ise “mühendislik, ziraat mühendisliği, sağlık kurumları yönetimi ve veteriner hekimlik” konularında yapılmıştır.

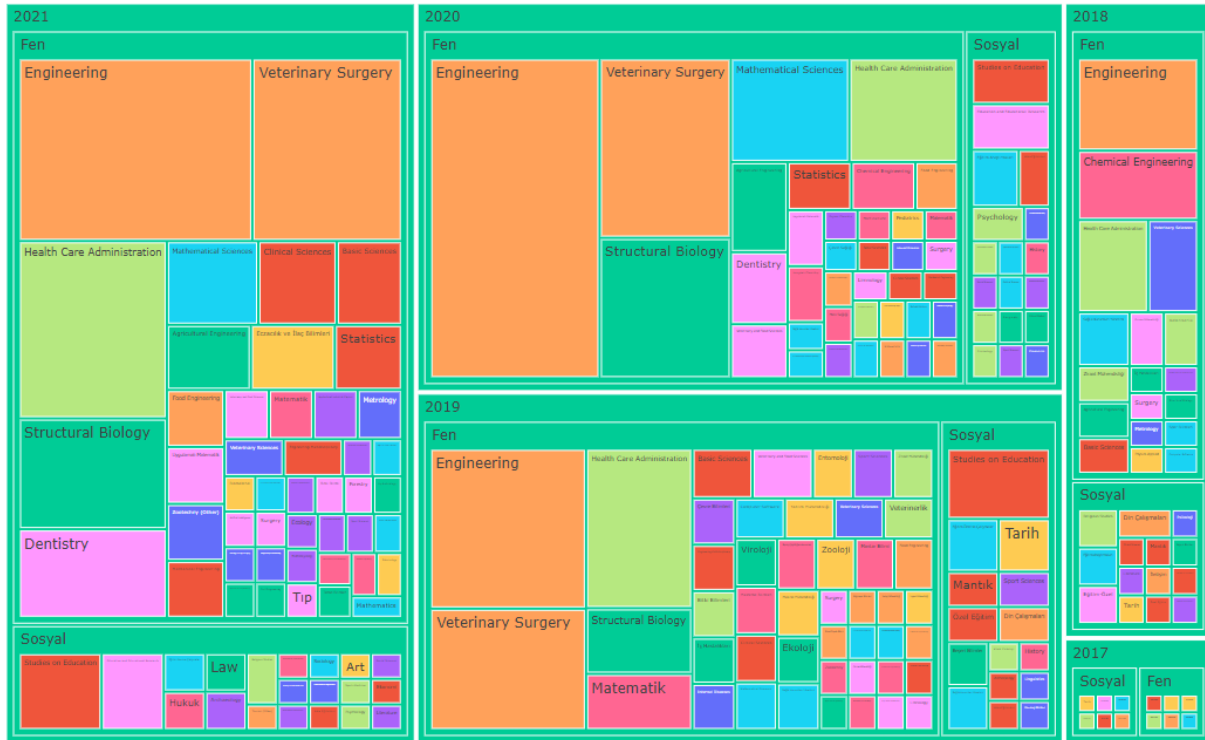
2018 yılında Sosyal Bilimler alanında en çok yayın “eğitim üzerine çalışmalar ve eğitim araştırmaları” konularında, Fen kategorisinde ise “mühendislik, sağlık kurumları yönetimi, ziraat mühendisliği, diş hekimliği ve yapısal biyoloji” konularında yayımlanmıştır. 2017 yılında ise makale konularındaki giriş yapılan kayıt sayısı azlığı nedeniyle (sosyal = 14 konu başlığı, fen = 13 konu başlığı) konular tespit edilememiştir.

4.1.7.2 İngilizce makalelerin konusal olarak yıl ve genel konu kategorisi bazındaki dağılımı nedir?

İngilizce makaleler için atanan konu başlıklarının “konu_df” alt veri setindeki “yıl” ve “kkategori”si (makale_konu_Kategorisi) değişkenlerine göre kümelenmiş durumları Kod 116 ile oluşturulan Şekil 82’deki “treemap” etkileşimli grafiğinde verilmiştir. Görselleştirmede kullanılan değişken hiyerarşisi; “yıl > kkategori > konu” şeklindedir.

```
import plotly.express as px
px.treemap(konu_df[konu_df.dil == 'İngilizce'], color='konu', values="sayi",
           path=['yıl', 'kkategori', 'konu'], height=700,width=1200)
```

Kod 116: İngilizce yayınların “konu_df” veri setindeki “yıl”, “kkategori” ve “konu” değişkenlerine göre dağılımı (treemap)



Şekil 82: İngilizce yayınların “konu_df” veri setindeki “yıl”, “kkategori” ve “konu” değişkenlerine göre dağılımı (treemap)

Mouse ile herhangi bir alanın üzerinde durulduğunda ilgili alana dair etiket bilgisi, yayın sayısı, üst kategori bilgisi vb. veriler görülebilmektedir. Alanların büyüklüğü yayın sayısı ile doğru orantılıdır.

İngilizce yayınlarda 2021 yılında Sosyal Bilgiler alanında en çok yayın yapılan konular “studies on education, education and education research”dır. Fen alanında ise “engineering, veterinary surgery, health care administration, structural biology ve dentistry”dir.

2020 yılında Sosyal Bilimler alanında en çok yayın yapılan makale konuları sırasıyla “studies on education, education and education research” konularıdır. Fen temel konu kategorisinde ise “engineering, veterinary surgery, structural biology ve mathematical sciences” konularıdır.

2019 yılında Sosyal Bilimler alanında en çok yayın “studies on education” konusunda, Fen temel konu kategorisinde ise “engineering, veterinary surgery, health care administration, structural biology” konularında yapılmıştır.

2018 yılında Sosyal Bilimler alanında en çok yayın “studies on education, religious studies, education” konularında, Fen kategorisinde ise “engineering, chemical engineering, health care administration ve veterinary sciences” konularında yayınlanmıştır. 2017 yılında ise makale konularındaki giriş yapılan kayıt sayısı azlığı nedeniyle (sosyal = 6 konu başlığı, fen = 7 konu başlığı) konular tespit edilememiştir.

4.1.7.3 Türkçe anahtar kelimelerin yıllara göre dağılımı nedir?

Yayınların anahtar kelime verileri, TR Dizin web sayfasından (202 yayın) ve DOI adresi DergiPark olan makalelerin (1.959 yayın) üst veri alanlarından derlenmiştir. Toplam 7.436 yayının 2.161’ine ait anahtar kelime verisine erişilebilmiştir. Bu sayıları elde etmek için Kod 117 ile filtreleme yapılmıştır.

```
df.loc[~df['makale_Keywords'].isnull(),"makale_Keywords"].count()

2161
```

Kod 117: Anahtar kelime alanı dolu olan makale sayısı (TR Dizin ve DergiPark)

DOI adresi DergiPark olan makale üst veri alanından toplanan anahtar kelimelerin sayısı ise Kod 118 ile filtreleme yapılarak tespit edilmiştir.

```
df.loc[(~df['makale_Keywords'].isnull()) &
((df['DOI_Durumu'] == "DOI DergiPark Linki") |
(df['DOI_Durumu'] == "DergiPark'ta Makale Konu alanı eklenmemiş")),
["makale_Keywords"].count()

1959
```

Kod 118: Anahtar kelime alanı dolu olan kayıtlar (DOI adresi DergiPark olanlar)

TR Dizin “anahtar kelimeler” üst veri alanından anahtar kelime toplanan makale sayısı Kod 119 ile filtreleme yapılarak tespit edilmiştir.

```
df.loc[(~df['makale_Keywords'].isnull()) &
        ((df['DOI_Durumu'] != "DOI DergiPark Linki") &
         (df['DOI_Durumu'] != "DergiPark'ta Makale Konu alanı eklenmemiş")),
        "makale_Keywords"].count()
```

202

Kod 119: Anahtar kelime alanı dolu olan kayıtlar (TR Dizin üst veri sayfası)

TR Dizin ve DergiPark üzerinden toplanarak veri setinin “makale_Keywords” değişkenine aktarılan makale anahtar kelimeleri, her makale için noktalı virgül ile ayrıştırılmıştır. “makale_Keywords” değişkeninde verisi dolu olan makaleler için Kod 120 ile “keywords_df” adlı yeni bir alt veri seti oluşturulmuş ve ana veri setindeki “makale_Dil, makale_konu_Kategorisi, makale_Konu ve makale_Yıl” değişkenleri alt veri setine değişken olarak dâhil eklenmiştir.

```
anahtar = []
say = []
yil = []
dil = []
konu = []
kkategori = []
for i in range(len(df['makale_Keywords'])):
    if pd.isnull(df['makale_Keywords'][i]) == False:
        for j in range(len(df['makale_Keywords'][i].split(';'))):
            if df['makale_Keywords'][i].split(';')[j].strip() != "":
                anahtar.append(df['makale_Keywords'][i].split(';')[j].strip())
                say.append(1)
                yil.append(df['makale_Yil'][i])
                dil.append(df['makale_Dil'][i])
                konu.append(df['makale_Konu'][i])
                kkategori.append(df['makale_konu_Kategorisi'][i])
keywords_df = pd.DataFrame()
keywords_df['anahtar_kelime'] = anahtar
keywords_df['sayi'] = say
keywords_df['yil'] = yil
keywords_df['dil'] = dil
keywords_df['konu'] = konu
keywords_df['kkategori'] = kkategori
keywords_df['anahtar_kelime'] = keywords_df['anahtar_kelime'].str.replace('İ', 'i').str.lower()
```

Kod 120: “keywords_df” anahtar kelime alt veri setinin oluşturulması

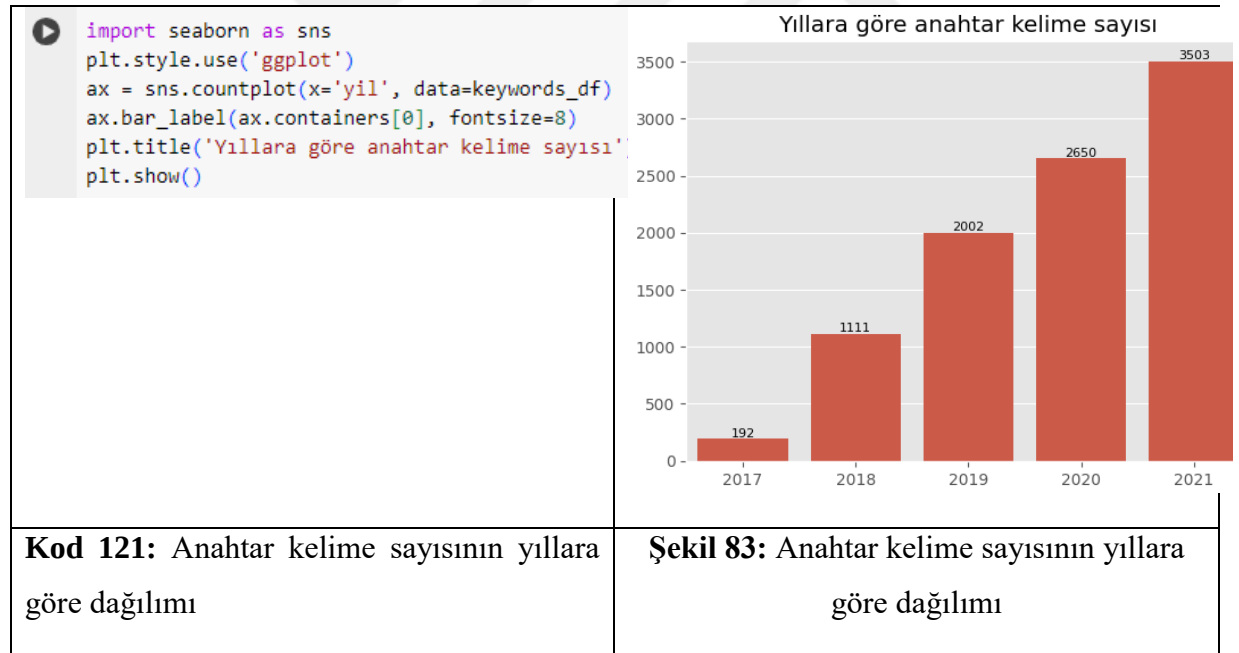
Anahtar kelime alanı dolu olan 2.161 kayıt için oluşturulan “keywords_df” alt veri setinde toplam 9.458 anahtar kelime olduğu tespit edilmiştir. “keywords_df” alt veri setinin kaç

kayıttan oluştuğunu görmek için “keywords_df.head()” kodu uygulanarak ilk beş kaydı Tablo 27’de verilmiştir.

Tablo 27: “keywords_df” alt veri setinin ilk beş satırı

	anahtar_kelime	sayi	yil	dil	konu	kkategori
0	Temporomandibular eklem disfonksiyonu	1	2021	Türkçe	Radyoloji ve Organ Görüntüleme	Fen
1	osteoartrit	1	2021	Türkçe	Radyoloji ve Organ Görüntüleme	Fen
2	konik-ışınli bilgisayarlı tomografi	1	2021	Türkçe	Radyoloji ve Organ Görüntüleme	Fen
3	Modern Dünya	1	2017	Türkçe	NaN	Sosyal
4	sekülerleşme	1	2017	Türkçe	NaN	Sosyal

“keywords_df” veri setindeki değişkenler dair teknik bir inceleme için bazı görselleştirmeler oluşturulmuştur. Kod 121 ile Şekil 83’te seaborn kütüphanesi kullanılarak oluşturulan çubuk grafikte hangi yıllarda kaç adet anahtar kelime kullanıldığı görselleştirilmiştir.



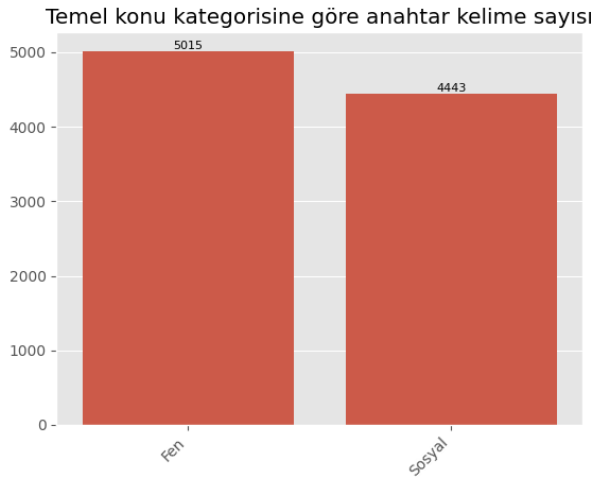
Şekil 83’e göre DergiPark üzerinden toplanan anahtar kelimelerin en çok 2021 yılına (3.503), en az ise 2017 (192) yılına ait olduğu, bu sayının yıllara göre düzenli bir şekilde arttığı görülmektedir. Bu durumda, TR Dizin makale detay sayfasında kaydedilmiş DergiPark adresli DOI linkleri sayısında yıllara göre artış olduğu söylenebilir.

Kod 122 ile Şekil 84’te seaborn kütüphanesi kullanılarak oluşturulan çubuk grafikte hangi dilde kaç adet anahtar kelime kullanıldığı görselleştirilmiştir.

<pre> ax = sns.countplot(x='dil', data=keywords_df) ax.bar_label(ax.containers[0], fontsize=8) plt.title('Dillere göre anahtar kelime sayısı') plt.show() </pre>	
Kod 122: Anahtar kelime sayısının dillere göre dağılımı	Şekil 84: Anahtar kelime sayısının dillere göre dağılımı

Şekil 84’e göre DergiPark üzerinden toplanan anahtar kelimelerin en çok Türkçe makalelerde olduğu (5.905), İngilizce makaleler için ise 3.511 anahtar kelimenin kullanıldığı tespit edilmiştir.

Kod 123 ile Şekil 85’te seaborn kütüphanesi kullanılarak oluşturulan çubuk grafikte hangi temel konu kategorisinde kaç adet anahtar kelime kullanıldığı görselleştirilmiştir.

<pre> ax = sns.countplot(x='kkategori', data=keywords_df) ax.bar_label(ax.containers[0], fontsize=8) plt.title('Temel konu kategorisine göre anahtar kelime sayısı') plt.xticks(rotation=45, ha='right') plt.show() </pre>	
Kod 123: Anahtar kelime sayısının konu kategorilerine göre dağılımı	Şekil 85: Anahtar kelime sayısının konu kategorilerine göre dağılımı

Şekil 85’e göre DergiPark üzerinden toplanan anahtar kelimelerin en çok “Fen” temel konu kategorisinde olduğu (5.015), “Sosyal” kategorisinde ise kullanılan anahtar kelime sayısının 4.443 olduğu tespit edilmiştir.

“anahtar_kelime” değişkeni farklı dillerdeki aynı anlama gelen kayıtlar içerebilmektedir. Örneğin Türkçe bir makaledeki “eğitim” konu başlığı İngilizce bir makalede “education” olarak geçebilmektedir. Bu nedenle öncelikle anahtar kelimeler alt veri setinde makale diline göre ayrıştırılmış ve “anahtar_kelime” değişkenine dair izleyen satırlardaki görselleştirmeler bu dillere göre yapılmıştır.

Öncelikle Türkçe makalelerin anahtar kelimeleri için görselleştirmeler oluşturulmuştur. “keywords_df” veri setindeki “anahtar_kelime” değişkeninde bulunan anahtar kelimelerin sıklığını görselleştirmek için Kod 124 ile “wordcloud” (kelime bulutu) kod kütüphanesi aktifleştirilmiştir. Aynı kod ile her yıl için Türkçe anahtar kelimeler wordcloud olarak Şekil 86’da görselleştirilmiştir. Aşağıdaki kod bloğu ile Türkçe yayınlara ait anahtar kelimelerin yıllara göre kelime bulutu şeklinde dağılımı verilmiştir. Kodda yer alan “keywords_df.yıl” alanına istenilen yıl değeri eklenerek filtreleme yapılabilir. İlgili kod beş yıl için ayrı ayrı uygulanarak görseller oluşturulmuştur.

```
from wordcloud import WordCloud
word_freq = dict(zip(keywords_df.loc[(keywords_df.dil == "Türkçe") & (keywords_df.yıl == 2021)].anahtar_kelime.value_counts().index,
                    keywords_df.loc[(keywords_df.dil == "Türkçe") & (keywords_df.yıl == 2021)].anahtar_kelime.value_counts().values))
wordcloud = WordCloud(width=1000, height=500, background_color='#0f0606').generate_from_frequencies(word_freq)
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.title("2021 yılı Türkçe anahtar kelimeler", fontsize=20)
plt.axis('off')
plt.show()
```

Kod 124: Türkçe anahtar kelimelerin yıllara göre kelime bulutu şeklinde dağılımı



Şekil 86: Türkçe anahtar kelimelerin yıllara göre kelime bulutu şeklinde dağılımı

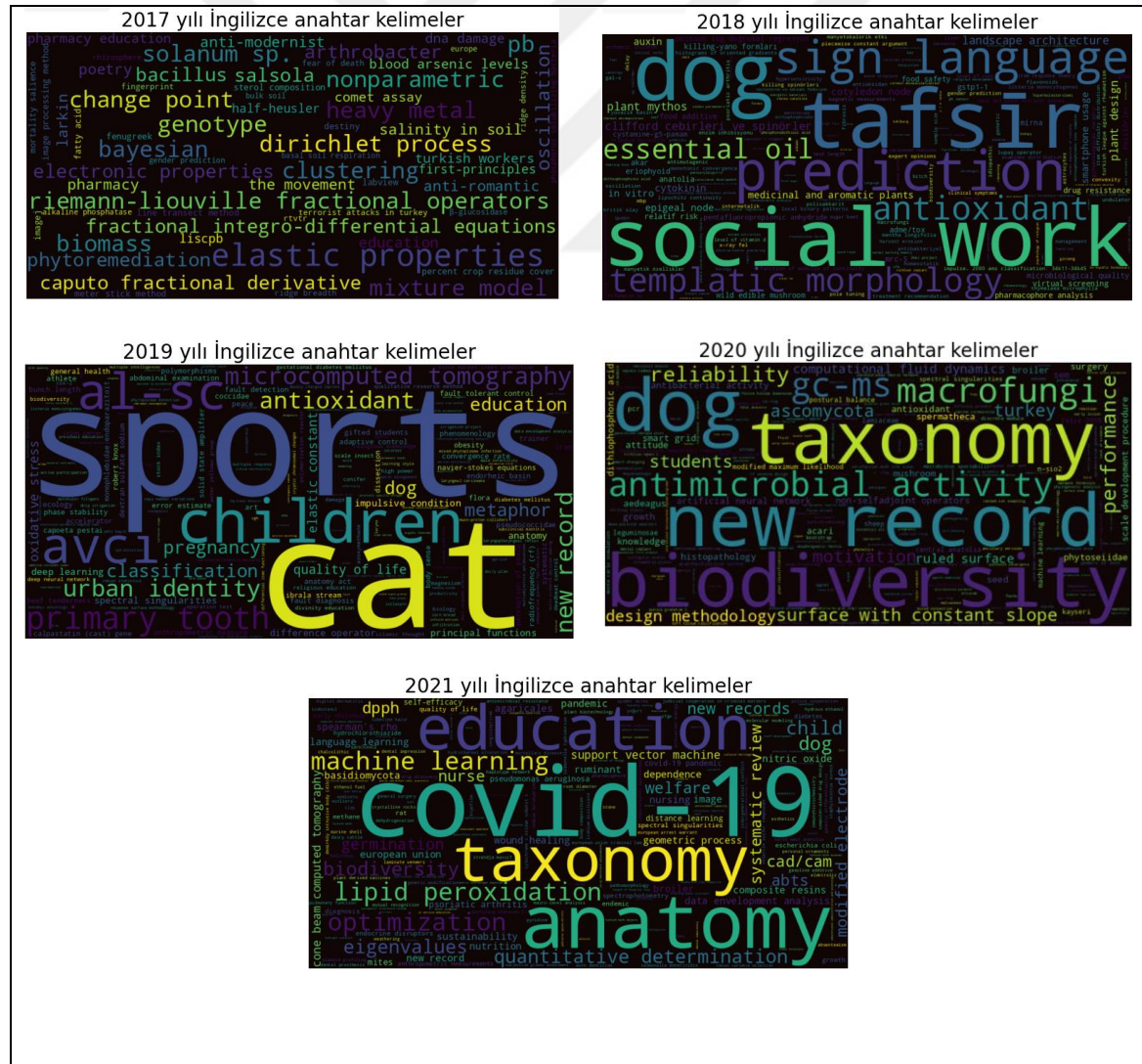
Şekil 86'ya bakıldığında; 2017 yılında Türkçe anahtar kelimelerinden “internet, modern dünya, okul güvenliği, ADH, Rusça” kelimelerinde sıklık vardır. 2018 yılında Türkçe anahtar kelimelerinden “metafor, eleştirel düşünme, Hint kültürü, sosyal hizmet”, 2019 yılında “Spor, çocuk, öğrenci, futbol, öğretmen adayları, performans, antrenör, hemşirelik”, 2020 yılında “COVID-19, beslenme, dinler tarihi, ölçek geliştirme, güvenilirlik” ve son olarak 2021 yılında Türkçe anahtar kelimelerinden “COVID-19, eğitim, pandemi, gerçeklik, spor, sosyal medya, fiziksel aktivite” anahtar kelimeleri sıklıkla kullanılmıştır.

4.1.7.4 İngilizce anahtar kelimelerin yıllara göre dağılımı nedir?

“keywords_df” veri setindeki “anahtar_kelime” değişkeninde bulunan İngilizce anahtar kelimelerin sıklığını görselleştirmek için Kod 125 ile "wordcloud" (kelime bulutu) kod kütüphanesi aktifleştirilmiştir. Aynı kod ile her yıl için anahtar kelimeler wordcloud olarak Şekil 87’de görselleştirilmiştir. Kodda yer alan “keywords_df.yil” alanına istenilen yıl değeri eklenerek filtreleme yapılabilir. İlgili kod beş yıl için ayrı ayrı uygulanarak görseller oluşturulmuştur.

```
from wordcloud import WordCloud
word_freq = dict(zip(keywords_df.loc[keywords_df.dil == "İngilizce"] & (keywords_df.yil == 2021)].anahtar_kelime.value_counts().index,
                    keywords_df.loc[keywords_df.dil == "İngilizce"] & (keywords_df.yil == 2021)].anahtar_kelime.value_counts().values)
wordcloud = WordCloud(width=1000, height=500, background_color='#0f0606').generate_from_frequencies(word_freq)
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.title("2021 yılı İngilizce anahtar kelimeler", fontsize=20)
plt.axis('off')
plt.show()
```

Kod 125: İngilizce anahtar kelimelerin yıllara göre kelime bulutu şeklinde dağılımı



Şekil 87: İngilizce anahtar kelimelerin yıllara göre kelime bulutu şeklinde dağılımı

2017 yılında İngilizce anahtar kelimelerine bakıldığında; “elastic properties, dirichlet process, clustering” kelimelerinde sıklık vardır. 2018 yılında İngilizce anahtar kelimelerinden “social work, tafsir, prediction, sign language, dog”, 2019 yılında “sports, children, cat, al-sc, urban identity, antioxidant”, 2020 yılında “biodiversity, taxonomy, dog, new record, antimicrobial activity, macrofungi, gc-ms” ve son olarak 2021 yılında; “COVID-19, taxonomy, anatomy, machine learning, lipid preoxidation” anahtar kelimelerinin sıklıkla kullanıldığı görülmektedir.

4.1.8 Soru 8: Makalelerin yayınlandığı dergi sayılarının yıllara göre dağılımı nedir?

AÜ adresli makalelerin yayınlandığı dergi sayısındaki değişimi görmek için ana veri setindeki (df) “dergi_adi” ve “makale_Yil” değişkenleri kullanılmıştır. Kod 126 ile dergiler birim/tekil dergiye indirgenmiştir.

```
len(df.dergi_adi.unique())  
938
```

Kod 126: 2017-2021 arası makalelerin yayımlandığı tekil dergi sayısı

TR Dizin’de indekslenen beş yıllık AÜ adresli makaleler 938 birim/tekil dergide yayınlanmıştır. Yıllara göre dergi sayılarını incelemek için Kod 127 ile “dergiSayisi_df” adlı bir alt veri seti oluşturularak “dergi_ad” değişkeni “makale_Yil” değişkenine göre gruplanmıştır.

```
dergiSayisi_df = df.groupby('makale_Yil')['dergi_adi'].nunique().reset_index(name='dergi_Sayisi')  
index = df.makale_Yil.value_counts().sort_index().index  
dergiSayisi_df
```

Kod 127: “dergiSayisi_df” alt veri setinin oluşturulması

Kod 127 ile oluşturulan, “makale_Yil ve dergi_Sayisi” değişkenlerinden oluşan “dergiSayisi_df” alt veri setinin görünümü Tablo 28’de verilmiştir.

Tablo 28: “dergiSayisi_df” alt veri seti görünümü

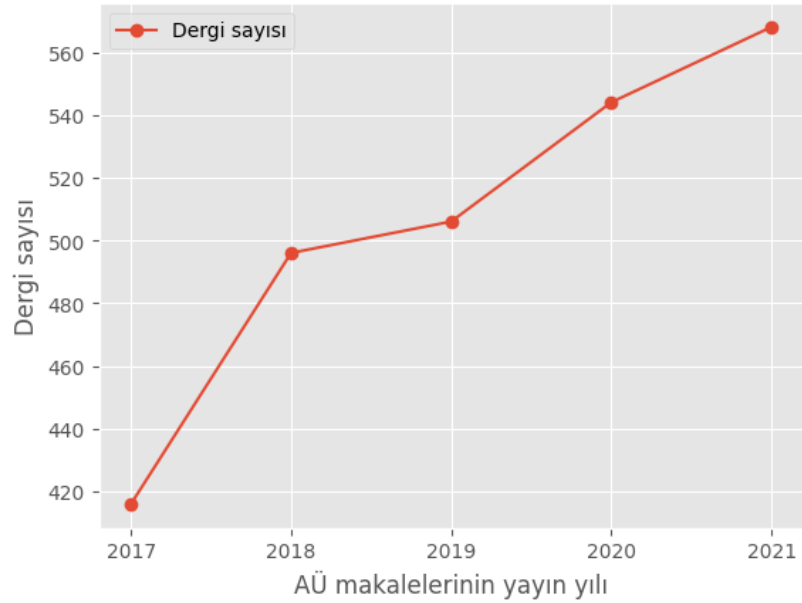
	makale_Yil	dergi_Sayisi
0	2017	416
1	2018	496
2	2019	506
3	2020	544
4	2021	568

“dergiSayisi” alt veri seti Kod 128 ile "matplotlib" kütüphanesi kullanılarak Şekil 88’deki line chart (çizgi grafik) görseli oluşturulmuştur.

```
unique_years = dergiSayisi_df['makale_Yil'].unique()
plt.plot(dergiSayisi_df['makale_Yil'], dergiSayisi_df['dergi_Sayisi'],
        marker='o', label='Dergi sayısı')
plt.xlabel('AÜ makalelerinin yayın yılı')
plt.ylabel('Dergi sayısı')
plt.xticks(unique_years)
plt.grid(True)
plt.legend()
```

Kod 128: AÜ makalelerinin yıllara göre yayınlandığı dergi sayısı dağılımı

```
2017    416
2018    496
2019    506
2020    544
2021    568
Name: dergi_adi,
```



Şekil 88: AÜ makalelerinin yıllara göre yayınlandığı dergi sayısı dağılımı

Şekil 88’e göre AÜ’nün beş yıllık makalelerinin yayınlandığı dergi çeşitliliğinde düzenli bir artışın olduğu görülmektedir. 2017 yılında makaleler 416, 2018’de 496, 2019’de 506,

2020’de 544 ve 2021 yılında toplam 568 çeşit dergide yayımlanmıştır. Dergi sayısındaki artış birçok açıdan yorumlanabilir. Örneğin bilimsel üretkenliğin artışına yanıt veren bir durum olarak görülebilir. Ancak; bu durum TR Dizin’de indekslenen dergilerin bütünü üzerinde bir inceleme gerektirdiği ve bu durumun tezin sınırlılıkları dışında olmasından dolayı ayrıca detaylı bir analiz yapılmamıştır.

4.1.8.1 Soru 8.1: Dergi sayısı artış oranı ile yayın sayısı artış oranı arasında bir korelasyon var mıdır?

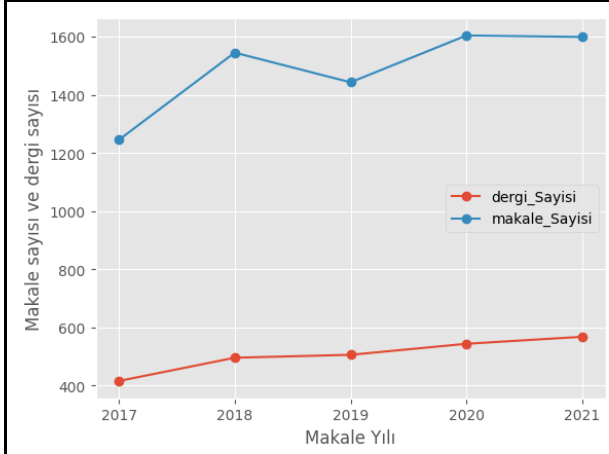
AÜ’nün beş yıllık süre zarfında yayımlandığı dergi çeşit sayısında artış olduğu Şekil 88’de tespit edilmişti. Dergi çeşidindeki bu artışın makale sayısı üzerinde de bir etkisinin olup olmadığını tespit etmek için Kod 129 ile “dergiSayisi_df” alt veri setine her yıl için toplam makale sayısını içeren “makale_Sayisi” değişkeni eklenmiştir.

```
dergiSayisi_df['makale_Sayisi'] = df.makale_Yil.value_counts().sort_index().values  
dergiSayisi_df
```

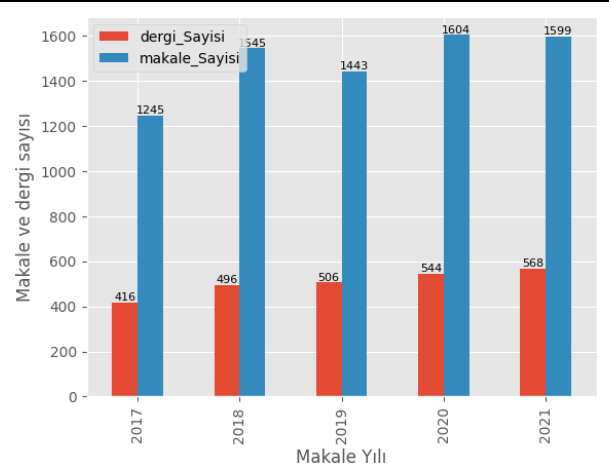
Kod 129: “dergiSayisi_df” veri setine “makale_Sayisi” değişkeninin eklenmesi

Yıllık dergi çeşit sayısı ile yıllık makale sayısının birlikte incelenmesi için Kod 130 ve Kod 131 ile Şekil 89’daki çizgi grafik ve Şekil 90’daki çubuk grafik görselleri oluşturulmuştur.

<pre>unique_years = dergiSayisi_df['makale_Yil'].unique() fig,ax= plt.subplots() plt.style.use('ggplot') dergiSayisi_df.plot(x="makale_Yil", y=["dergi_Sayisi", "makale_Sayisi"], ax=ax, kind="line", marker='o') for container in ax.containers: ax.bar_label(container, fontsize=8) plt.xticks(unique_years) plt.xlabel("Makale Yılı") plt.ylabel("Makale sayısı ve dergi sayısı")</pre> Kod 130: Makale yılına göre toplam makale ve dergi sayısı dağılımı (1)	<pre>fig,ax= plt.subplots() plt.style.use('ggplot') dergiSayisi_df.reset_index().plot(x="makale_Yil", y=["dergi_Sayisi", "makale_Sayisi"], ax=ax, kind="bar") for container in ax.containers: ax.bar_label(container, fontsize=8) plt.xlabel("Makale Yılı") plt.ylabel("Makale ve dergi sayısı")</pre> Kod 131: Makale yılına göre toplam makale ve dergi sayısı dağılımı (2)
---	---



Şekil 89: Makale yılına göre toplam makale ve dergi sayısı dağılımı (1) (çizgi)



Şekil 90: Makale yılına göre toplam makale ve dergi sayısı dağılımı (2) (çubuk)

Şekil 89 ve 90, her yıl için toplam makale sayısını ve ilgili yıldaki makalelerin yayınlandığı toplam dergi sayısını göstermektedir. 2017 yılından 2021 yılına kadar, makalelerin yayınlandığı dergi sayısında düzenli bir artış olduğu görülmektedir. Ancak makale sayısında artış olmadığı, yıllara göre değişkenlik gösterdiği gözlemlenmiştir. Bu bulguyu güçlendirmek için “Soru 8.2” sorulmuş ve dergi başına düşen makale sayısı belirlenmeye çalışılmıştır.

4.1.8.2 Soru 8.2: Yıllara göre dergi başına düşen makale sayıları nedir?

AÜ’nün beş yıllık süre zarfında her yıl için dergi başına düşen makale sayısını bulmak için Kod 132 ile “makale_Sayisi” değişkeni “dergi_Sayisi” değişkenine oranlanarak “dergiSayisi_df” alt veri setine değişken (sayi) olarak eklenmiştir.

```
dergiSayisi_df['sayi'] = dergiSayisi_df['makale_Sayisi'] / dergiSayisi_df['dergi_Sayisi']
dergiSayisi_df
```

Kod 132: “sayi” değişkeninin “dergiSayisi_df” alt veri setine eklenmesi

“sayi” değişkeni eklendikten sonra “dergiSayisi_df” alt veri setinin güncel görünümü Tablo 29’da verilmiştir.

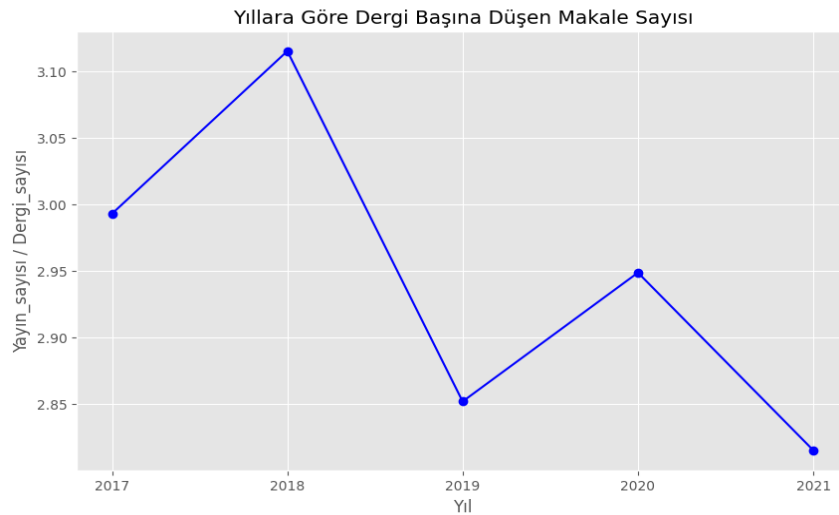
Tablo 29: “dergiSayisi_df” alt veri seti görünümü

makale_Yil	dergi_Sayisi	makale_Sayisi	sayi
2017	416	1245	2.992788
2018	496	1545	3.114919
2019	506	1443	2.851779
2020	544	1604	2.948529
2021	568	1599	2.815141

Her yıl için tekil dergi başına düşen makale sayısının (sayı değişkeni) görselleştirilmesi için Kod 133 ile matplotlib kütüphanesi kullanılarak Şekil 91’deki çizgi grafik oluşturulmuştur.

```
unique_years = dergiSayisi_df['makale_Yil'].unique()
fig,ax= plt.subplots()
plt.style.use('ggplot')
dergiSayisi_df.plot(x="makale_Yil", y=["sayi"],
                    ax=ax, kind="line", marker='o')
for container in ax.containers: ax.bar_label(container, fontsize=8)
plt.xticks(unique_years)
plt.xlabel("Makale Yılı")
plt.ylabel("Makale sayısı")
```

Kod 133: Yıllara göre dergi başına düşen makale sayısı



Şekil 91: Yıllara göre dergi başına düşen makale sayısı

Grafiğe bakıldığında, dergi başına düşen makale sayısının en yüksek olduğu yılın 2018, makale sayısının en az olduğu yılın ise 2021 yılı olduğu görülmektedir. 2020 yılından sonraki düşüşün COVID-19 salgınından kaynaklı olduğu düşünülmektedir.

4.1.8.3 Soru 8.3: AÜ yazarları hangi dergileri tercih etmektedir?

AÜ makalelerinin yayınlandığı dergilerle ilgili inceleme yapmadan önce veri setinde yer alan dergi adı değişkeni verileri Kod 134 ile başlık formatına dönüştürülerek standartlaştırılmıştır.

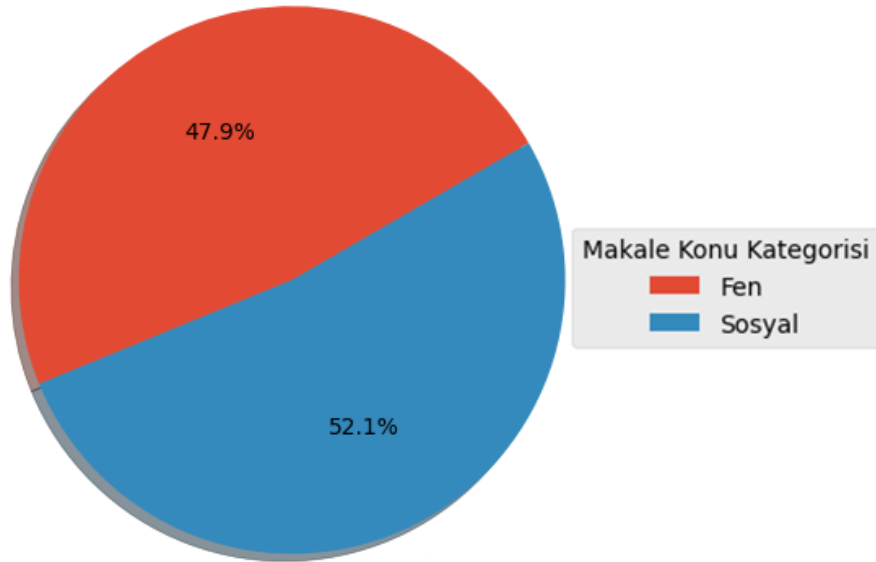
```
df.dergi_adi=df.dergi_adi.str.replace('İ','i').str.replace('İ','i').str.lower().str.title()
```

Kod 134: Dergi adlarının başlık formatına dönüştürülmesi

Daha sonra Kod 135 ile hangi temel konu kategorisinde ne kadar dergi olduğunu tespit etmek için dergi adları “makale_konu_Kategorisi”ne göre gruplanarak Şekil 92’de pasta grafik olarak gösterilmiştir.

```
plt.style.use('ggplot')
fig, ax = plt.subplots()
ax.pie(df.groupby("makale_konu_Kategorisi")["dergi_adi"].nunique(),
      labels=df.groupby("makale_konu_Kategorisi")["dergi_adi"].nunique().index,
      autopct='%1.1f%%',
      shadow=True,
      startangle=30)
ax.axis('equal')
plt.legend()
```

Kod 135: Tekil dergi sayısının “makale_konu_Kategorisi”ne göre dağılımı



Şekil 92: Tekil dergi sayısının “makale_konu_Kategorisi”ne göre dağılımı

AÜ yazarlarının tercih ettiği sosyal konu kategorisindeki dergileri tespit etmek için Kod 136 ile Şekil 93’teki “wordcloud” (kelime bulutu) görseli oluşturulmuştur.

Kod 136: Yazarların tercih ettiği Sosyal konu kategorisindeki dergiler



³⁴ Dergi sayısı TR Dizin dergi arayüzünden alınmıştır (TR Dizin, 2023).

4.1.8.4 Soru 8.4: AÜ makalelerinin yayınlandığı dergilerin yayın hayatına başlama tarihine göre dağılımı nedir?

Yayın hayatına başlama tarihine göre dergi sayılarını bulmak için öncelikle Kod 138 ile temel veri setindeki (df) dergi verilerini içeren sütunlar süzölmüş ve dergi adına göre tekilleştirilerek yeni oluşturulan “dergi_df” alt veri setine aktarılmıştır.

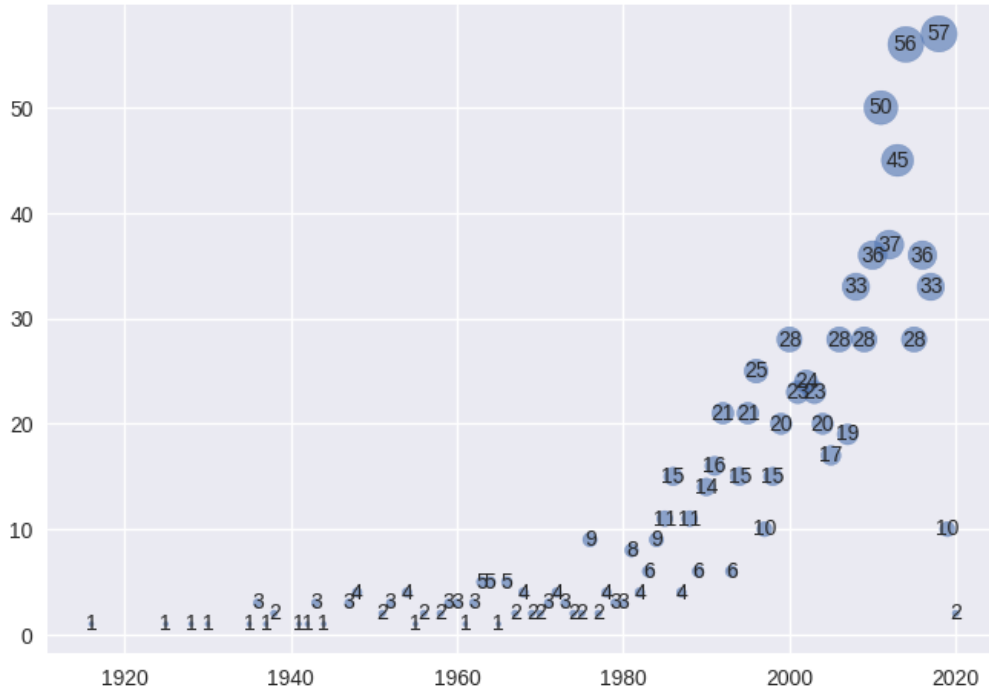
```
dergi_df=df.iloc[:, -15:]  
dergi_df=dergi_df.drop_duplicates(ignore_index=True)
```

Kod 138: “dergi_df” alt veri setinin oluşturulması

Daha sonra alt veri seti, Kod 139 ile matplotlib kütüphanesinin “scatter plot” (dağılım grafiği) türü kullanılarak yayın hayatına başlama tarihine göre Şekil 95’te görselleştirilmiştir.

```
import matplotlib.pyplot as plt  
plt.style.use("seaborn")  
x = dergi_df.der_yayin_yili.value_counts().index  
y = dergi_df.der_yayin_yili.value_counts().values  
plt.scatter(x,y, alpha=0.6, s=dergi_df.der_yayin_yili.value_counts().values * 5)  
for i, txt in enumerate(dergi_df.der_yayin_yili.value_counts().values):  
    plt.text(x[i], y[i], str(txt), ha='center', va='center')  
plt.show()
```

Kod 139: Dergi sayılarının yayın hayatına başlama tarihine göre dağılımı



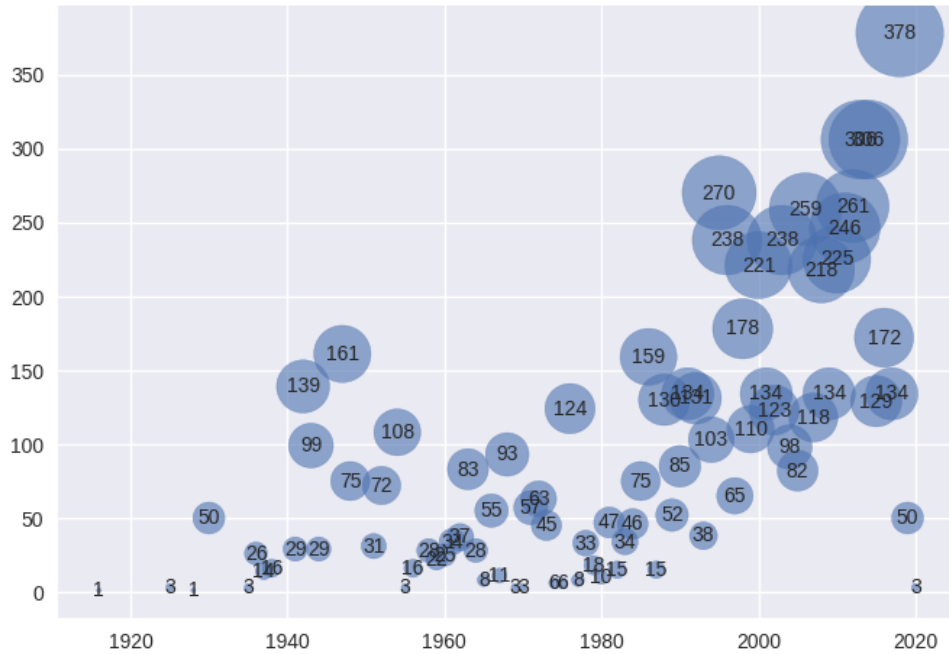
Şekil 95: Dergi sayılarının yayın hayatına başlama tarihine göre dağılımı

Beş yıllık süre zarfında AÜ yayınlarının yer aldığı dergilerin yayın hayatına başladığı tarihler incelendiğinde, en eski tarihli derginin (İstanbul Tıp Fakültesi Dergisi) 1916’da yayın hayatına başladığı görülmektedir. Sonraki eski tarihli dergiler ise (Türkiyat Mecmuası) 1925’te, (Türk İdare Dergisi) 1928’de ve (Veteriner Hekimler Derneği Dergisi) 1930’da yayın hayatına başlamışlardır. En son yayın hayatına başlayan dergiler ise; 2020 yılında yayın hayatına başlayan İslam Tetkikleri Dergisi, Türk Uzaktan Algılama ve CBS Dergisi ile İslam Tetkikleri Dergisi’dir.

Dergilerin yayın hayatına başlama tarihi ile bu dergilerde yayınlanan makale sayısı değişkenlerinin beraber incelenmesi için Kod 140 ile Şekil 96’daki “scatter plot” (dağılım grafiği) oluşturulmuştur.

```
import matplotlib.pyplot as plt
plt.style.use("seaborn")
x = df.der_yayin_yili.value_counts().index
y = df.der_yayin_yili.value_counts().values
plt.scatter(x,y, alpha=0.6, s=df.der_yayin_yili.value_counts().values * 5)
for i, txt in enumerate(df.der_yayin_yili.value_counts().values):
    plt.text(x[i], y[i], str(txt), ha='center', va='center')
plt.show()
```

Kod 140: Dergilerin yayına başlama tarihine göre makale sayıları



Şekil 96: Dergilerin yayına başlama tarihine göre makale sayıları

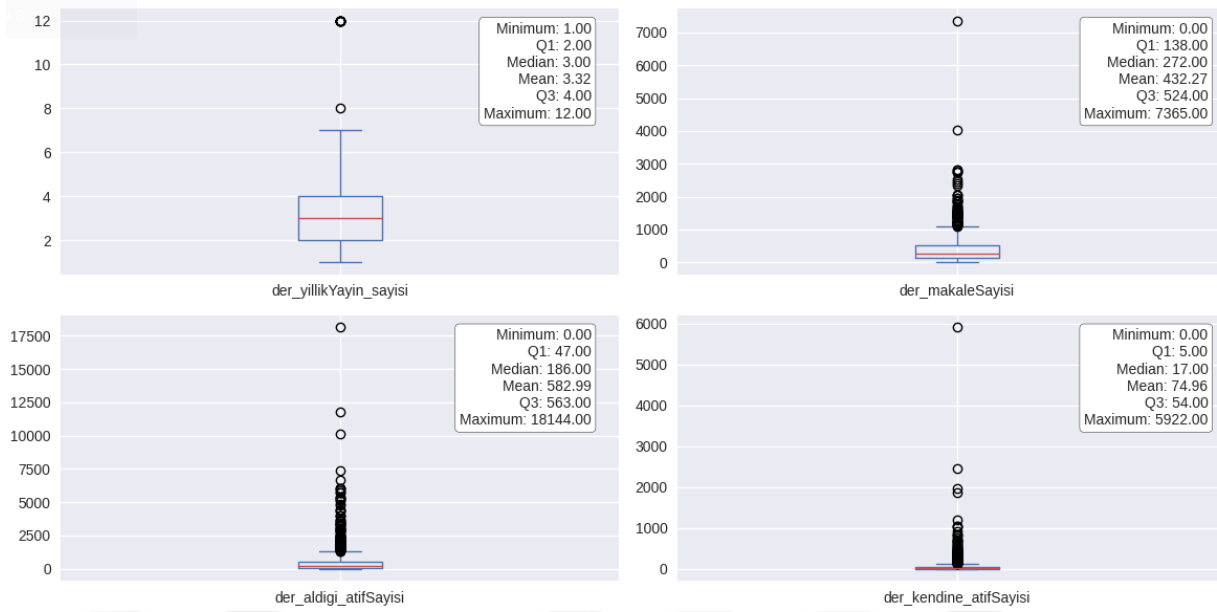
Dağılım grafiğine göre AÜ'nün beş yıllık süre zarfındaki makalelerinin büyük bir bölümünün 1990 yılı ve sonrasında yayın hayatına başlayan dergilerde yayınlandığı görülmektedir.

4.1.8.5 Soru 8.5: AÜ makalelerinin yayınlandığı dergilerin yıllık yayın periyodları, makale sayıları, aldığı atıf sayıları ve kendine atıf sayıları nedir?

Dergilere dair “dergi_df” alt veri setinde bulunan “yıllık yayın periyodları, makale sayıları, aldığı atıf sayıları ve kendine atıf sayıları” verilerinin genel bir incelemesini yapmak için Kod 141 ile her değişken için Şekil 97’de oluşturulan “boxplot” (kutu grafik) görselleri oluşturulmuştur. Boxplot, bir değişkendeki verilerin dağılımını görselleştirmek için kullanılan bir grafik türüdür. Boxplot’ta çeşitli istatistiksel parametreler kullanılarak veri setinin merkezi eğilimi, dağılımı ve aykırı değerleri hakkında bilgi edinilebilmektedir.

```
boxplot_cols = ["der_yillikYayin_sayisi", "der_makaleSayisi", "der_aldigi_atifSayisi", "der_kendine_atifSayisi"]
fig, axes = plt.subplots(nrows=2, ncols=2, figsize=(12, 6))
dergi_df[boxplot_cols].plot(kind="box", subplots=True, layout=(4, 1), ax=axes)
for ax, col in zip(axes.flatten(), boxplot_cols):
    statistics = {
        'Minimum': dergi_df[col].min(),
        'Q1': dergi_df[col].quantile(0.25),
        'Median': dergi_df[col].median(),
        'Mean': dergi_df[col].mean(),
        'Q3': dergi_df[col].quantile(0.75),
        'Maximum': dergi_df[col].max()
    }
    annotation = '\n'.join([f'{key}: {value:.2f}' for key, value in statistics.items()])
    ax.annotate(annotation, xy=(0.95, 0.95), xycoords='axes fraction', ha='right', va='top',
                bbox=dict(boxstyle='round,pad=0.3', edgecolor='black', facecolor='white'))
plt.tight_layout()
```

Kod 141: “der_yillikYayin_sayisi, der_makaleSayisi, der_aldigi_atifSayisi, der_kendine_atifSayisi” değişkenlerinin boxplot ile incelenmesi



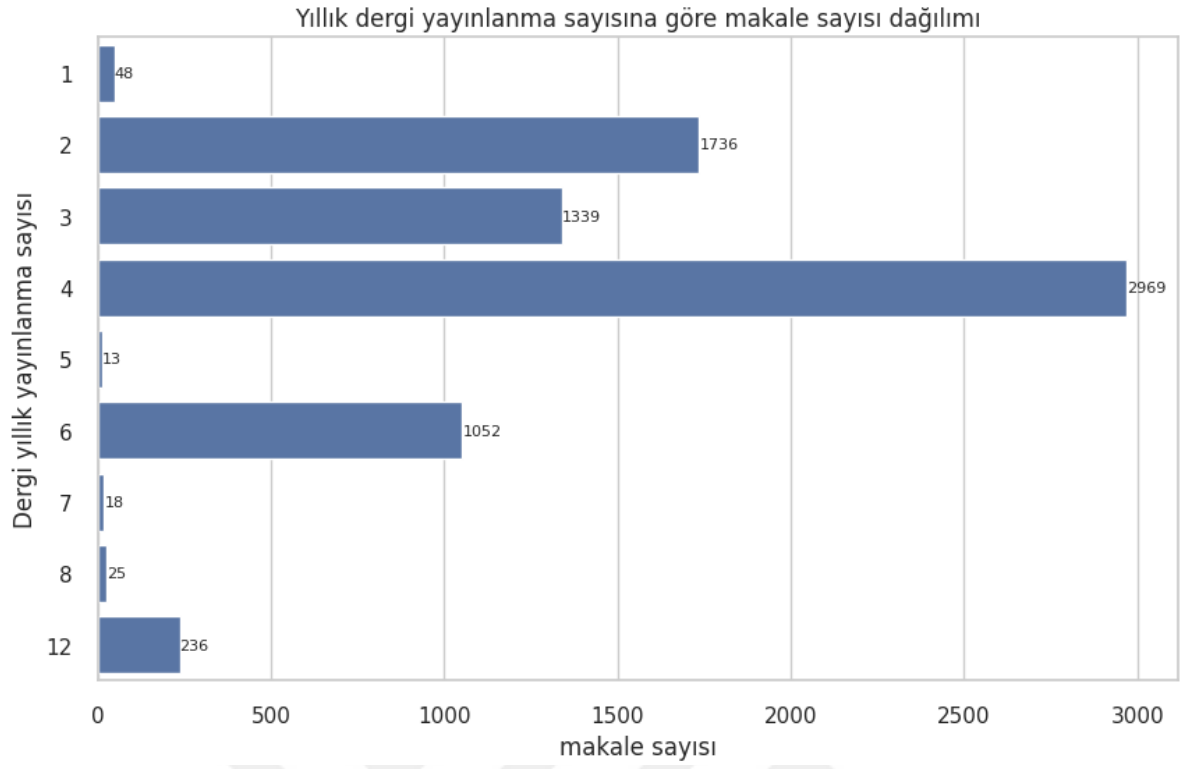
Şekil 97: “der_yillikYayin_sayisi, der_makaleSayisi, der_aldigi_atifSayisi, der_kendine_atifSayisi” değişkenlerinin boxplot ile incelenmesi

Şekil 97 incelendiğinde; AÜ yayınlarının çoğunlukla yılda 4 defa yayın yapan dergilerde yayınlandığı, yılda 8 ve 12 defa yayın yapan dergilerde de az da olsa yayınlandığı görülmektedir. Tercih edilen dergilerin ortalama makale sayıları 432’dir. Toplam makale sayısı 1.000 ile 7.365 arasında değişen dergiler de bulunmaktadır. Ortalama atıf sayıları 582 olup 18.144 atıf alan bir dergi de bulunmaktadır. Bu dergilerin kendine atıf sayısı ortalaması 74.96’dır. Ayrıca kendine atıf sayısı 5.922 olan istisna bir dergi de bulunmaktadır.

AÜ yazarlarının en çok hangi periyodlarla çıkan dergileri tercih ettiği durumu Kod 142 ile incelenmiştir. Bu kod ile “der_yillikYayin_Sayisi” değişkeni ile makale sayıları Şekil 98’de yatay çubuk grafik olarak görselleştirilmiştir.

```
sns.set(style="whitegrid")
plt.figure(figsize=(10, 6))
ax = sns.countplot(y='der_yillikYayin_sayisi', data=df)
ax.bar_label(ax.containers[0], fontsize=8)
plt.xlabel("Dergi sayısı")
plt.ylabel("Yıllık yayınlanma sayısı")
plt.title("Yıllık yayınlanma sayısına göre dergi sayısı")
```

Kod 142: Dergi yıllık yayınlanma sayısına göre makale sayısı



Şekil 98: Dergi yıllık yayınlanma sayısına göre makale sayısı dağılımı

Grafik incelendiğinde AÜ yazarlarının çoğunlukla yıllık dört defa yayın çıkaran dergileri tercih ettiği görülmektedir.

5. BÖLÜM: SONUÇ VE ÖNERİLER

Bilim ve teknolojiadaki hızlı ilerleme, verilerin hacminde ve çeşitliliğinde benzeri görülmemiş bir artışa yol açmıştır. Verilerdeki bu artış, bu kadar büyük miktardaki verinin verimli bir şekilde işlenmesi zorluğunu da beraberinde getirmiştir. Bu durum, veri selinin nasıl yönetileceği sorusunun önemini ortaya koymaktadır.

Büyük ve çeşitli türdeki verinin bütünleşik olarak işlenmesi ve analiz edilmesi ihtiyacına yanıt olarak veri görselleştirme kritik bir önem kazanmaktadır. Veri görselleştirme, anlamayı ve yorumlamayı kolaylaştırmak amacıyla verilerin görsel temsilini ifade etmektedir. Veri görselleştirme; veri setlerinde bulunan sayısal biçimdeki veriler içerisinde, ayırt edilmesi zor olabilecek örüntülerin, eğilimlerin ve ilişkilerin keşfedilmesine olanak tanıyan ve içgörülerin çeşitli hedef kitlelere daha net şekilde iletilmesini kolaylaştıran bir yöntemdir.

Etkili veri analizi, veri kümeleri içindeki ilişkileri ve örüntüleri aydınlatmak için uygun görselleştirme türlerinin kullanılmasıyla sağlanabilir. Bilgisayar teknolojisi ve veri görselleştirme yazılımlarındaki ilerlemeler sayesinde çok sayıda görselleştirme seçeneği mevcut olduğundan, amaçlanan mesajın iletilmesinde uygun görselleştirme seçimi büyük bir öneme sahiptir. Örneğin; kelime bulutları, çubuk grafikler, dağılım grafikleri, ısı haritaları, kutu grafikler, histogramlar vb. görselleştirme türlerinden hangisinin veri seti için uygun olduğu ve görselleştirmenin ne amaçla kullanılacağı hedef kitleye göre değişiklik gösterebilmektedir. Yararlanıcı grubun durumu ve seviyesi, seçim sürecini etkiler. Özetlere ihtiyaç duyan paydaşlar için çubuk grafikler veya pasta grafikleri gibi anlaşılması daha kolay görselleştirmeler kullanılabilir. Daha ayrıntılı içgörülere gereksinim duyan veri analistleri veya uzmanlar için ise ısı haritaları veya dağılım grafikleri gibi karmaşık görselleştirmeler kullanılabilir.

Veri görselleştirme için gerek ticari gerekse ücretsiz ve açık kaynak kodlu olarak kullanılabilecek birçok yazılım ve programlama dili kütüphanesi mevcuttur. Bunlardan en yaygın olarak kullanılan programlar; Tableau, PowerBI, Google Charts ve Gephi'dir. En yaygın kullanılan programlama dilleri ise; Rstudio, Python ve D3.js'dir.

İncelenen yabancı literatürde veri görselleştirmenin ve araçlarının, kütüphanecilik ve bilgi bilimi alanları bağlamında yoğun olarak ele alındığı tespit edilmiştir. Veri görselleştirme; işlem günlükleri, kaynak kullanımı, koleksiyon yapıları, kullanıcı davranışları ve dijital hizmet kullanım verilerinin analiz edilmesi için kullanılması gereken bir yöntem olarak değerlendirilmiştir. Ayrıca veri görselleştirme, araştırmacılara ve öğrencilere araştırma destek

hizmetleri kapsamında sunulan hizmetlerde ve bibliyometrik analizler bağlamında performans belirlemek için kullanılmaktadır.

TR Dizin'in indirmeye izin verdiği dosya verileri içinde, çalışmada yanıt aranan bazı soruların gerektirdiği bazı verilerin bulunmadığı ya da standart formatta verilmediği görülmüştür. Örneğin konu analizi için gerekli olan anahtar kelime ve makale konu alanlarına ilişkin verilerin bulunmaması ve kurum verilerinin standart olmayışı gibi. Bu nedenle bu veriler kullanılamamıştır.

TR Dizin'de yayın arama arayüzünde “kurum adı” parametresi ile arama seçeneğinin olmaması ve önceki paragrafta bahsedilen indirilebilir arama sonuçları dosyası ile ilgili sorunlar nedeniyle veri toplama yöntemi olarak web kazıma (web scraping) kullanılmıştır. Web kazıma işlemi Python programlama dili kullanılarak uygulanmıştır. Çalışmanın kapsamını Ankara Üniversitesi adresli 2017-2021 yılları arasındaki makale verileri oluşturmaktadır. Web kazıma yönteminin uygulanması sırasında karşılaşılan sorunlar şu şekildedir:

- “Yazar Kurum” bilgileri standart bir formatta gelmemektedir. Kurum adı, fakülte adı, birim adı, disiplin adı, yabancı kurum adı vb. alanlara ait veriler standart bir formatta kaydedilmemiştir ve ayrıştırılamamaktadır. Örneğin; bazı kayıtlarda, üniversite adı + fakülte adı + disiplin adı şeklinde kaydedilmiş, bazı kayıtlar sadece üniversite adı ya da üniversite + fakülte adı şeklinde verilmiştir. Bu alandaki verilerin standart olmaması nedeniyle, anlamlı görselleştirmeler oluşturulamayacağı düşünülerek, disiplin ve fakülte bazlı analiz ve görselleştirme yapılamamıştır.
- TR Dizin'de yer alan “makale konusu” ve “anahtar kelimeler” üst veri alanlarındaki verilerin eksik girildiği görülmüştür. Bu nedenle TR Dizinde DOI linki üst veri alanı dolu olan ve hedefi DergiPark'a giden makalelerin ilgili üst veri alanları DergiPark'tan toplanmıştır. Anahtar kelimeleri toplanan makale oranı, veri setindeki toplam makalelerin %26'sını, makale konusu toplanan makale oranı ise veri setindeki toplam makalelerin %21'ini oluşturmaktadır. Toplanan bu veriler ile görselleştirmeler yapılmış olsa da düşük orandaki bu verilerle yapılan konu analizlerinin yeterli olmayacağı düşünülmektedir.
- Hakemli bilimsel makale türü olarak nitelendirilebilecek türler “araştırma makalesi, derleme ve olgu sunumu” olmasına rağmen TR Dizin'de “kitap tanıtımı, editoryal, editöre mektup, çeviri, kısa bildiri, bildiri, mektup, düzeltme, toplantı özetleri” gibi türler de “makale türü” alanında kaydedilmiştir. Bu durum yayın türü olarak geniş bir içerik kapsamına işaret etse de hakemli araştırma makalelerinin kolaylıkla saptanması

mükün olamamaktadır. Oysa araştırma makalelerinin hakemli olduğunu anlamamızı sağlayacak ayrı bir üst veri alanının bulunması yayın türlerinin analizi için uygun olacaktır. Ayrıca “diğer” makale türündeki kategori adlarının açık bir şekilde yazılması ile tür çeşit sayısının belirlenmesi ve ayrıntılı analiz için daha fazla veri sağlanacaktır.

- TR Dizin sistemin ürettiği ve yazar benzersizliğini sağlayan bir tanımlayıcı bulunmadığı ve araştırmacı arayüzünde birçok yazarın ORCID verisi eksik olduğu için yazar verileri tekilleştirilememiştir. Bu nedenle yazarlara dair analiz ve görselleştirme yapılamamıştır.

Bu araştırmada; ulusal akademik yayınların TR Dizin sisteminden üst verilerin çekilmesi ve veri setine dönüştürülmesi, verilerin ön işleme tabi tutulması, görselleştirme destekli keşifsel analiz yapılması ile ilgili uygulama bizzat araştırmacı tarafından geliştirilmiştir. Uygulama kapsamında veri setine çeşitli sorular sorulmuş (Tablo 16) ve elde edilen sonuçlar görselleştirilmiştir. Sorular; makalelere dair genel niteliklerin belirlenmesi, ulaşılabilirlik, iş birliğinin (yazar/disiplin/kurum) ne düzeyde olduğu ve etki (yazar/makale/dergi) durumlarının tespit edilmesine yöneliktir. Bu çalışma kapsamında toplanan verilerden elde edilen bulgulara yönelik değerlendirmeler aşağıda verilmiştir.

TR Dizin veri tabanındaki makalelere ait bibliyografik kimlikler sadece “Fen Bilimleri” ve “Sosyal Bilimler” temel konu kategorilerinde gruplanmaktadır. 2017 yılı hariç diğer yıllarda Fen Bilimleri temel konu kategorisindeki makale sayısı daha fazladır. Bu durumun nedeni, üniversitenin Tıp Fakültesi, Mühendislik Fakültesi gibi birimlerinin varlığı ve Fen Bilimleri alanında yayın üretimi sıklığının daha fazla olmasıdır. Konuların sadece iki temel kategoride sınıflandırılması, bölüm adlarına göre kategorilerin olmaması, TR Dizin ve DergiPark’ın anahtar sözcüklerin ve makale konu alanlarının kavramsal dizin veya konu başlıklarına dair ayrıntılı konu sınıflamasının yetersizliği, disiplinlere (bölüm) dair derinlemesine analiz yapılmasını engellemektedir.

Türkçe haricindeki makaleler; İngilizce, Arapça, Almanca, Rusça, Fransızca, İspanyolca, diğer şeklindedir. “Diğer” kategorisindeki makale sayısı Almanca, Rusça, Fransızca, İspanyolca dillerindeki makale sayılarından fazla olmasına rağmen, bu kategoride hangi dillerin olduğu verisine ulaşamamıştır. 2017’den 2021’e kadar İngilizce makalelerde, özellikle Fen Bilimleri alanındakilerde, önemli bir artış olması, AÜ’nün Fen Bilimleri alanındaki araştırma çıktılarının giderek uluslararasılaştığını göstermektedir.

TR Dizin’de yer alan, “TR Dizin Başvuru Koşulları ve Dergi Değerlendirme Kriterleri” 10. maddesinde “*Dergide yer alan makalelerde, makale dilinde öz ve anahtar kelimeler olmalıdır. Öz yazımı için kelime sınırı verilmiş olmalı ve bu kurala uyulmalıdır. Öz ve Anahtar kelimeler uluslararası standartlara uygun olmalıdır*” ifadesi yer almaktadır (bkz. <https://trdizin.gov.tr/kriterler/>). Bu kriter kapsamında dizinlenen makale PDF dosyalarında anahtar kelimeler bulunmasına rağmen, DergiPark platformundaki belli sayıdaki anahtar kelime hariç, bu verilerin TR Dizin makale üst veri alanlarına kaydedilmediği tespit edilmiştir. Bir veri kaynağı olarak başvuru TR Dizin’de anahtar kelimelere ilişkin üst veri alanında eksik verilerin olması, güvenilir görselleştirmelerin üretilmesini güçleştirmektedir. Bu nedenle anahtar kelime verileri, DOI link hedefi DergiPark olan makalelerden toplanmıştır. Bu durum sınırlı sayıda anahtar kelime ile analiz yapılmasına neden olmuştur.

DergiPark’ta makaleler için kaydedilen anahtar kelimelerin girişinin, bir denetimli bir kavramsal dizin kullanılmadan, yazarlar tarafından yapıldığı düşünülmektedir. Bu anahtar kelimelerin makale konusu verileriyle de ilişkilendirilmemiş olması konu analizlerini zorlaştırmaktadır. Ayrıca DergiPark’ta makale konusu verilerinin girişinde de bir sınıflandırma uygulanmamaktadır. Örneğin; bazı makale konuları “engineering” şeklinde genel bir terim ile kaydedilmişken, bazılarının daha spesifik olarak “civil engineering, mechanical engineering” şeklinde kaydedildiği gözlemlenmiştir. Bu durum makale konusu verilerinin tekilleştirilmesini ve anahtar kelimeler ile ilişkilendirilmesini zorlaştırmaktadır. Bu nedenle DergiPark’tan toplanan anahtar kelimeler, makale konusu verileri ile ilişkilendirilemeden bağımsız olarak görselleştirilmiştir. Bu makalelerde kullanılan mevcut anahtar kelimeler yıllara göre görselleştirilmiştir. Veri setindeki son yıl (2021) anahtar kelime verilerine bakıldığında, sıklıkla kullanılan anahtar kelimelerin, Türkçe makalelerde; “COVID-19, eğitim, pandemi, gerçeklik, spor, sosyal medya, fiziksel aktivite”, İngilizce makalelerde ise “COVID-19, education, taxonomy, anatomy, machine learning, lipid prexidation” kelimeleri olduğu tespit edilmiştir. 2020 yılında dünya çapında, yoğun olarak etkisini gösteren COVID-19 pandemisinin, çeşitli disiplinlerdeki akademik araştırmalarda yoğun olarak ele alındığı düşünülmektedir. Sadece DergiPark’ta bulunan makale anahtar kelime ve makale konusu verilerinin toplanabiliyor olması, makale konusu verileri için bir sınıflandırma yapılmamış olması gibi nedenlerden dolayı AÜ’nün ilgili dönemdeki araştırma eğilimleri tespit edilememiştir.

TR Dizin’deki AÜ makalelerinin erişilebilirliği değerlendirildiğinde, makalelere tam metin olarak yüksek düzeyde (%95,9) bir erişilebilirliğin olduğu görülmektedir. Yalnızca bibliyografik bilgilerine erişilebilen makalelerin oranı ise %4,1’dir. “Sosyal” temel konu

kategorisindeki tam metin olarak erişilemeyen yayınların sayısı “Fen” alanındaki yayınlardan daha fazladır. Bu durum, araştırmacıların veri paylaşım tutumları bağlamında ayrıca araştırılması gereken bir bulgu olarak değerlendirilmektedir.

Makale türü “araştırma, derleme ve olgu sunumu” olan ortak yazarlı makaleler, tek yazarlı makalelere oranla yaklaşık üç kat fazladır. AÜ adresli makalelerde yazar iş birliğine yönelik güçlü bir eğilim olduğu söylenebilir. Tek yazarlı makale sayısı, toplam makale sayısının yaklaşık dörtte biri, ortak yazarlı makale sayısı ise toplam makale sayısının dörtte üçü oranındadır. Makalelerin almış oldukları atıflar incelendiğinde; tek yazarlı makaleler, toplam atıf sayısının yaklaşık dörtte biri, ortak yazarlı makaleler ise dörtte üçü oranında atıf almıştır. Bu durumda ortak yazarlılık veya tek yazarlılık durumunun atıf sayısı üzerinde bir etkisinin olup olmadığı konusunda bir çıkarım yapılamamıştır.

Ortak yazarlı makalelerde diğer kuruluşlarla yapılan iş birliğinin oranı, sadece AÜ içerisindeki akademisyenlerle yapılan iş birliğinden fazladır. Ortak yazarlı makalelerde en çok iş birliği yapılan kurumlar sırasıyla; T.C. Sağlık Bakanlığı, Gazi Üniversitesi, Hacettepe Üniversitesi, Millî Eğitim Bakanlığı, Yabancı ülke kurumları, Sağlık Bilimleri Üniversitesi ve Yıldırım Beyazıt Üniversitesi’dir. En çok iş birliği yapılan bu kurumların çoğunlukla aynı şehirden olması, bakanlıklar ve yabancı ülke kurumlarındaki yüksek oran dikkat çeken durumlar olarak değerlendirilmektedir.

TR Dizin’de yazar üst veri alanında AÜ’nün iş birliği yaptığı yabancı ülke kurumları “yabancı kurum” şeklinde kaydedilmiştir. Kurum verilerinin toplandığı TR Dizin üst veri alanında ülke ve kurum isimleri olmadığı için yabancı kurumlara dair bir analiz yapılamamıştır. Bu durum, TR Dizin’deki veriler ile üniversitenin uluslararası iş birliği ağının çıkarılamayacağını göstermektedir.

AÜ’nün diğer kuruluşlarla iş birliği sonucu ürettiği makalelerde, diğer kuruluşlardan daha fazla sayıda ilk kurum statüsündedir. İlk yazarlık aynı zamanda araştırmanın liderliğini ve yönlendirilmesini ifade eden bir gösterge niteliğindedir.

Yıllık toplam yayın sayısı ve toplam atıf sayıları karşılaştırıldığında, en çok atıf 2018 yılında, en az atıf ise 2021 yılındadır. En yüksek atıf alan yayın araştırma makalesi türünde, ikinci en yüksek atıf alan yayın ise “derleme makalesi” türündedir.

Makale yılına göre görüntüleme ve indirme sayılarındaki artış ve azalışların doğru orantılı olduğu görülmektedir. Her iki değişkende de 2018’den 2020’ye kadar bir düşüş olmuş ve 2021’de ise yükseliş eğilimi gözlemlenmiştir.

AÜ'nün beş yıllık süre zarfındaki yayınlarının büyük bir bölümü 2000 yılı ve sonrasında yayın hayatına başlayan dergilerde yayınlanmıştır. Bu durumun nedeni, 2000 yılından sonra yayın hayatına başlayan dergi sayısında ciddi bir artışın olmasıdır (Şekil 95). TR Dizin'de 2017-2021 yıl aralığında dergi başına düşen AÜ makale sayılarında en yüksek oran 2018 yılında, en düşük oran ise 2021 yılındadır (Şekil 91). Bu durum AÜ akademisyenlerinin çok sayıda farklı dergide makale yayınladıklarını göstermektedir. Dergi çeşitliliğinde düzenli bir artış olmasına rağmen, makale sayısında aynı oranda bir artış olmadığı tespit edilmiştir. TR Dizin'de indekslenen beş yıllık AÜ adresli makaleler toplam 938 birim/tekil dergide yayınlanmıştır. Bu sayı TR Dizin'deki toplam dergi sayısının %57,70'ini oluşturmaktadır.

Veriler, AÜ araştırmacılarının yeni dergilere doğru bir yöneliminin olduğunu göstermektedir. Ancak bu yönelime rağmen yayınlanan makale sayısında önemli bir artışın olmaması ve dergi başına makale oranının yıllar içinde dalgalanması, yeni dergilere ilgi artsa da genel araştırma çıktısı hacminin nispeten sabit kaldığını göstermektedir. Ayrıca AÜ bünyesinde, tercih edilen dergilerin performans değerlendirmelerine yönelik kalite kontrolünün yapıp yapılmadığı verisine ulaşılamamıştır. Bu, ayrıca araştırılması gereken bir durum olarak değerlendirilmektedir.

Keşifsel veri analizi ve görselleştirmeler uygulayarak ortaya konan bulgular doğrultusunda; veri setinde yer alan makale sayısı, makale temel konu kategorisi, makale dili, erişim türü, yazar iş birliği, kurumsal iş birliği ile tercih edilen dergi bilgileri gibi veri alanlarına keşifsel veri analizi uygulayarak görselleştirmeler oluşturulabilmektedir. Ancak örüntüler tespit etmek, iç görü elde etmek için bu veri alanları ile makale konusu, makale anahtar kelime, yabancı kurum bilgisi, atıf sayısı gibi değişkenlerin beraber görselleştirilmesi gerekmektedir. Makale konusu, makale anahtar kelime, yabancı kurum bilgisi gibi alanlardaki verilerin bir sınıflamaya göre girişlerinin yapılmış olması ve veri alanlarının dolu olması gerekmektedir. Ayrıca atıf sayısı verilerinin sadece TR Dizin'de indekslenen makaleler ile sınırlı olması da bu atıf verilerinin diğer veri alanları ile beraber analiz edilmesini ve görselleştirilmesini engellemektedir. Bu bağlamda; “Bilimsel akademik yayınlara ait üst verilerden oluşan veri setlerine, veri görselleştirme destekli keşifsel veri analizi teknikleri uygulayarak veri alanlarının görselleştirmeye elverişli olup olmadığı ve içgörü sağlayıp sağlayamayacağı tespit edilebilir.” şeklindeki araştırma hipotezinin doğrulandığını söylemek mümkündür.

Araştırma kapsamında elde edilen bu sonuçlar bağlamında aşağıdaki öneriler sıralanabilir:

Veri sağlayıcısı (TR Dizin) hem üniversiteler hem de bağımsız araştırmacılar için, yayınlara ait bütün üst veri alanlarının olduğu bir API hizmeti ya da veri seti oluşturma hizmeti verebilir. Sistem üzerinde, yayınlara ait daha kapsamlı üst veri alanlarına yer verilmesi ve bu üst veri alanlarının belli bir standartta eksiksiz veri barındırması, yapılacak veri analizi ve veri görselleştirmelerin güvenilir sonuçlar vermesi için önemlidir.

Yabancı kurumlarla yapılan makale iş birliklerinde yabancı kurum verilerinin, örneğin; ülke, kurum, disiplin vb. verilerin standart bir formatta işlenmesi durumunda yabancı kurumlarla yapılan iş birliği ağına dair istatistiksel analizler ve görselleştirmeler yapılabilecektir.

DergiPark platformu, akademik süreli yayıncılığın, Türkiye’de kaliteli ve standartlara uygun olarak gelişmesini sağlamak amacıyla oluşturulan bir platform olup, Türkiye adresli, akademik ve hakemli olduğunu beyan eden tüm dergileri elektronik ortamda bulunduran ve editoryal süreç yönetimi sağlayan bir alt yapı hizmetidir. DergiPark sistemindeki makale anahtar kelimeleri yazarlar tarafından işlenmektedir. Sistem belli bir standardın sağlanması için anahtar kelimelerin oluşturulacak bir terimler listesinden seçilmesi sağlanabilir. Makale konu alanları DergiPark’ın sunduğu dizinden seçilmektedir. Disiplinleri temsil eden bu konu alanlarında da bir standardın olmadığı gözlemlenmiştir. Örneğin bir makale için genel olarak “mühendislik” konu alanı tanımlanırken, başka bir konu için “jeoloji mühendisliği” ya da “inşaat mühendisliği” gibi konu alanları tanımlanmıştır. Makale konu alanları (disiplinler) için hiyerarşik bir standardın oluşturulması ve yazarlara bu konu alanlarının bir dizinden seçtirilmesi disiplinlerle ilgili yapılacak analizlerin sağlıklı yapılabilmesi için önemlidir.

TR Dizin’de makale konusu (disiplin) ve anahtar kelime üst veri alanlarındaki veri eksikliği, var olan verilerin ise belli bir taksonomik sınıflamaya göre düzenlenmemiş olması, keşifsel veri analizi yapmayı ve anlamlı görselleştirmeler oluşturmayı engelleyen faktörler olmuştur. Bu alanlardaki verilerin içgörü sağlamak için veri görselleştirmeye için elverişli olmadığı düşünülmektedir.

Ulusal akademik dergilerdeki makale üst verileri ve tam metin erişimlerinin DergiPark sistemi üzerinden takip edilmesi, TR Dizin tarafından harici sistemlerden çekilen üst veri alanlarında da standardı sağlayabilir.

TR Dizin makale dili veri alanında bulunan “diğer” kategorisinde bulunan dillerin açık bir şekilde yazılması, makale dili ile ilgili yapılacak analizleri çeşitlendirecektir. Ayrıca TR

Dizin’de yazarları niteleyen bir tekil tanımlayıcı oluşturulması, yazarlara dair yayın performans analizlerinin yapılabilmesini sağlayacaktır.

TR Dizin ve DergiPark’taki veri alanlarının bir üst veri standardına (örn: Dublin Core) göre girişlerinin yapılması ve her veri alanının ne tür veri barındırdığına ilişkin kapsamlı bir kılavuzun oluşturulması, sistemler arası veri transferi ve kullanıcılar için veri toplama, veri seti oluşturma gibi süreçleri kolaylaştıracaktır. TR Dizin’de, bilimsel yayın verilerinin yönetimine dair bir politika oluşturulabilir. Bilimsel yayın verilerinin yorumlanması ve sunumunda görselleştirme konusunda kurumlara hangi olanakların sunulduğu paylaşılabilir. Bu bağlamda kurumlar için ayrıntılı keşifsel veri analizi (exploratory data analysis) ve görselleştirmelerin yer aldığı ya da bunların yapılabileceği araçlar sunulabilir.

Çalışma kapsamında yapılan literatür incelemesi, kütüphanecilik alanındaki veri görselleştirme çalışmalarında üniversite kütüphanelerinin ve kütüphanecilerin önemli rolünü vurgulamaktadır. Bilginin organizasyonu ve bilgi yönetimi konusundaki uzmanlıklarından yararlanan kütüphaneciler, görselleştirme sürecini destekleyen; akademik yayınlarla ilişkili üst verilerin toplanması, düzenlenmesi ve yönetimi için iyi bir konuma sahiptirler. Ayrıca kütüphanecilerin veri kütüphaneciliği alanında kendilerini geliştirmeleri, üniversite kütüphanelerinde araştırma veri hizmetlerinin kapsamının genişlemesine ve etkili hizmetler tasarlanmasına katkıda bulunabilecektir (Karcı Kandemir, 2023, s.267).

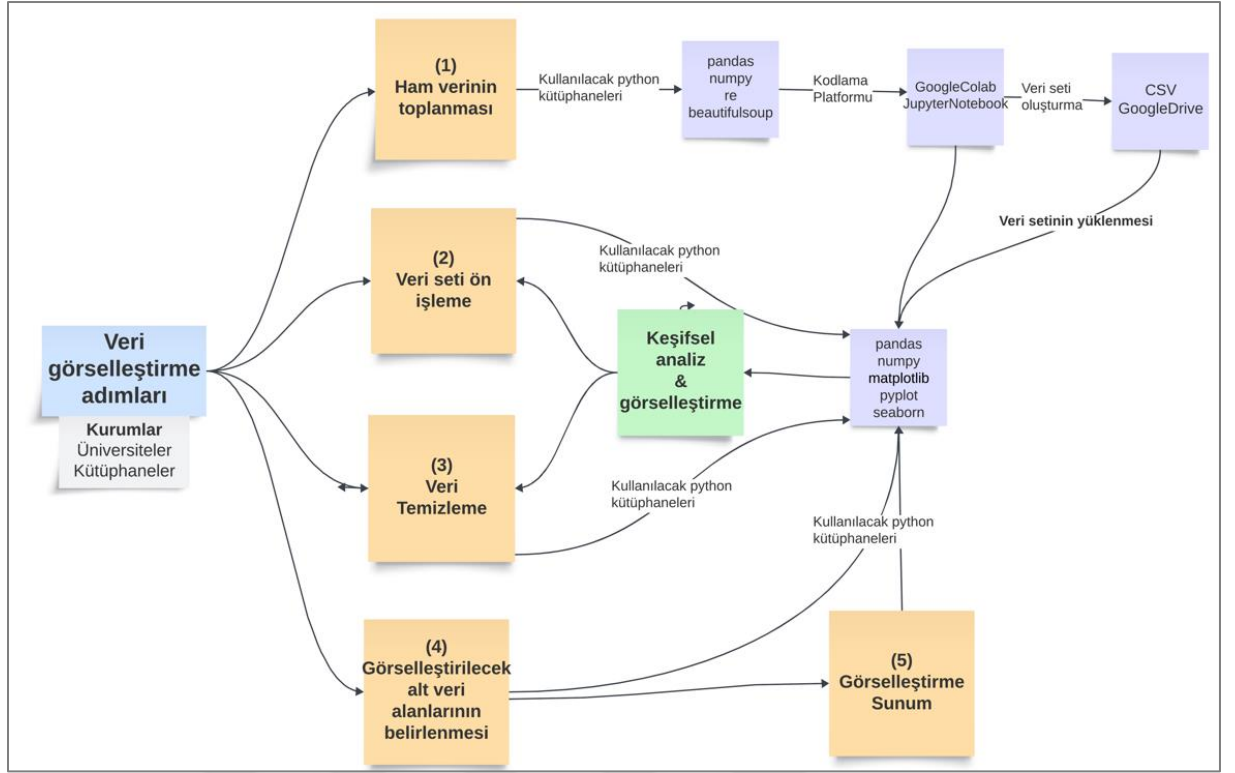
Veri görselleştirme araçlarını kullanmadaki yeterlilik, kütüphanecilere ham verileri eyleme geçirilebilir iç görümlere dönüştürme gücü verir ve veri kümeleri içindeki eğilimlerin ve korelasyonların tanımlanmasını kolaylaştırır. Dolayısıyla veri görselleştirme araçlarına ilişkin eğitim ve rehberliğin sağlanması; kütüphanecilerin veri görselleştirme yeteneklerinin geliştirilmesi açısından önem taşımaktadır. Bu bağlamda; kütüphane ve yönetim iş birliği ile bu uygulamaların benimsenmesi ve gerekli planlamanın yapılması önemlidir.

Veri görselleştirme araçları, önemli bilgiler sağlayarak, karar verme sürecine yardımcı olarak ve keşfedici öğrenme fırsatları sunarak kütüphanecilerin çalışmalarına değer katabilir. Örneğin, bir veri görselleştirme aracı ile belirli bir dönemde bir kütüphaneden ödünç alınan kitapların sayısının, erişilen çevrimiçi kaynak sayısının, kullanıcı profillerinin, akademik yayınlara dair sayısal değerlerin ve hizmetlere dair yapılan harcamaların etkili görselleştirmeler yaparak analiz edilebilmesi mümkündür. Kütüphaneciler, veri görselleştirme araçlarını kullanarak kütüphaneleri hakkında bilgi edinebilir, teknik becerilerini geliştirebilir ve görselleştirmelerden elde edilen içgörülere dayanarak bilinçli kararlar alabilirler.

Üniversite kütüphaneleri, web sayfalarında, açık arşiv ve kütüphane otomasyon sistemleri içerisinde yürütülen işlemlere dair önemli miktarda veriye sahiptir. Bu verilerin görselleştirmelerle desteklenmiş analizi, kütüphanedeki hizmet geliştirmeyle ilgili karar verme süreçlerine önemli ölçüde destek sağlayacaktır. Bu analizler; kütüphane kaynaklarının ve kullanımının değerlendirilmesi ve iyileştirme stratejilerinin formüle edilmesi gibi faaliyetler için de destekleyici bir özelliğe de sahiptir. Ayrıca, kütüphane sistemlerinde barındırılan akademik yayınlarla ilişkili üst veriler de çeşitli analizlerde kullanılabilecek verilerdir.

“Ankara Üniversitesi 2024-2028 Stratejik Planı”nda belirtilen “*Veri yönetimi ve öz değerlendirme süreçlerini etkin kılarak kurumsal yapıyı güçlendirmek*” şeklindeki A.5 nolu amaç ve bu amaç altındaki “*Kurum genelinde öz değerlendirme çalışmalarının geliştirilmesi (H5.1)*” ve “*Veri yönetimi konusunda farkındalık oluşturma (H5.2)*” hedefleri 2024-2028 yılları arasında kurumun veri yönetimi ve öz değerlendirme çalışmalarına yönelik faaliyetler yürütmeyi planladığı anlamına gelmektedir. Yine aynı dokümanda 2022 yılında kurulan Araştırma Dekanlığı bünyesindeki farklı birimler aracılığıyla, araştırma faaliyetlerinin izlenmesi ve iyileştirilmesi çalışmalarının veriye dayalı olarak sürdürüleceği belirtilmiştir (Ankara Üniversitesi, s.79). Ayrıca veri yönetiminin tek bir birim altında toplanmasıyla veri doğruluğunun kontrol edileceği vurgulanmıştır (Ankara Üniversitesi, ss.98,100-101). Kurumun bu amaç ve hedefler doğrultusunda yürütmeyi planladığı veri yönetimi faaliyetlerinde, mevcut çalışmada oluşturulan uygulama adımları ile kullanılan keşifsel veri analizi ve görselleştirme adım ve teknikleri örnek teşkil edebilecek niteliktedir.

Çalışma kapsamında uygulanan veri görselleştirme adımlarının ve geliştirilen uygulama kodlarının üniversiteler veya kütüphaneler tarafından nasıl kullanılabileceğini gösteren bir model tasarlanmıştır (Şekil 99). Öncelikle üniversiteler, bu modeli TR Dizin üzerindeki yayınlarına ait üst verileri toplamak, veri setine dönüştürmek ve görselleştirmek için kullanabilirler.



Şekil 99: Veri görselleştirme modeli

Ham verilerin toplanması (adım 1) için, bölüm 3.1.3'te bulunan, Kod 4'teki “facet-authorInstitution” değişkenine istenilen üniversite ismi yazılarak web kazıma işlemi için sorgu oluşturulabilir. Daha sonra uygulamanın Kod 4 ile Kod 35 arasındaki kod blokları sırasıyla uygulanarak makalelere ait üst veri alanları toplanıp bir dataframe'e dönüştürülebilir. Veri toplama için Python programlama dilinin “pandas, numpy, re ve beautifulsoup” kütüphaneleri kullanılacaktır. Kodlama platformu olarak da Google Colab üzerinde bulunan Jupyter Notebook kodlama ortamı kullanılacaktır. Toplanan veriler ise Kod 36 ile CSV formatına dönüştürülerek bilgisayarların hard diskine ya da Google Drive depolama alanına bir veri seti olarak kaydedilecektir. TR Dizin'in yapacağı herhangi bir arayüz değişikliği, ya da üst veri alanlarının bulunduğu HTML etiketlerindeki herhangi bir değişiklik durumunda ilgili kodlarda gerekli düzenlemelerin yapılması gerekmektedir.

Veri seti ön işleme (adım 2) için, çalışmanın 3.2.1'de yer alan Kod 37 ile Kod 46 arasındaki kod bloklarını uygulanarak veri setlerine dair ön izleme işlemleri yapılacaktır. Bu durumda toplanan veri alanlarına ait değişken isimlerinin Tablo 6 ve 7 ile birebir aynı olması gerekmektedir. Veri temizleme için (adım 3) ise Kod 47 ile Kod 55 arasındaki kod bloklarının uygulanması gerekmektedir. Adım 3 ve adım 4 aşamalarında görselleştirme destekli keşifsel veri analizi yapılabilir. Bu süreçlerde “pandas, numpy, matplotlib, pyplot ve seaborn”

kütüphaneleri kullanılabilir. Çalışmada olduğu gibi, görselleştirilecek alt veri alanlarının belirlenmesi (adım 4) ve görselleştirmeler (adım 5) için, Kod 56 ile Kod 142 arasındaki kod blokları uygulanarak çalışmada oluşturulan bütün alt veri setleri ve görselleştirmeler uygulanabilir. Adım 4 ve 5'te veri setinde filtrelemeler yapılabilir, var olan değişkenlerden alt veri setleri oluşturulabilir veya analizlerde kullanılmak üzere veri setinde olmayan yeni değişkenler oluşturularak analiz ve görselleştirmelere dâhil edilebilir. Bu model doğrultusunda oluşturulan anonim görselleştirmeler, bir gösterge paneli (dashboard) şeklinde tasarlanarak kütüphane web sayfasında kullanıcıların hizmetine sunulabilir.

Şekil 99'daki veri görselleştirme modelinin kütüphanelerde nasıl kullanılabileceğini açıklamadan önce, kütüphanelerin veri görselleştirmeyi hangi amaçlar için kullanabileceklerini anlatmakta yarar vardır. Kütüphaneler veri görselleştirmeyi aşağıdaki amaçlarla kullanarak hizmetlerini iyileştirebilir ve kullanıcılarına daha etkili bir şekilde ulaşabilirler:

- **Kullanıcı Davranışlarını Analiz Etme:** Kütüphaneler; kitap ödünç alma, e-kaynakların kullanımı gibi kullanıcı aktivitelerini görselleştirerek hangi kaynakların ne kadar kullanıldığını, hangi konulardaki kaynakların talep edildiğini tespit edebilirler. Bu, popüler yayınlar ve mevcut kaynakların kullanımı hakkında detaylı bilgiler sağlar. Ayrıca kullanım istatistiklerin kütüphane web sayfasında görsel olarak sunumu hizmet kalitesini de arttıracaktır.
- **Koleksiyon Yönetimi:** Yayınların konu başlıklarına, kategorilere, türlere veya yayın yıllarına göre dağılımını görselleştirerek koleksiyonun dengesini analiz edebilirler. Hangi türlerin veya konuların eksik olduğunu belirleyerek koleksiyonlarını zenginleştirebilirler.
- **Etkinlik Katılımı:** Kütüphane etkinliklerine katılım oranlarını ve etkinlik türlerine göre katılımı görselleştirerek, hangi etkinliklerin daha popüler olduğunu ve hangi zaman dilimlerinde daha fazla katılım sağlandığını ve etkinliklerin ödünç verme sayısı üzerindeki etkilerini belirleyebilirler.
- **Kütüphane Kullanımı:** Ziyaretçi sayıları, kütüphane içindeki hareketlilik ve kaynak kullanımı gibi verileri görselleştirerek, kütüphanenin fiziksel ve dijital olarak kullanım verilerini analiz edebilirler.
- **Finansal Analiz:** Yayın satın alma, personel ve bina giderleri ile gelirler, bağışlar ve yayın bağışları gibi finansal verileri görselleştirerek mali durumlarını daha iyi yönetebilirler.

- **Kullanıcı Geri Bildirimleri:** Kullanıcı anketleri ve geri bildirimlerini görselleştirerek, hizmet kalitesini ve kullanıcı memnuniyetini değerlendirebilirler.
- **Tarama Verileri:** Kullanıcıların en çok aradığı terimler ve hangi kaynakların en sık erişildiğini görselleştirerek, kullanıcı ihtiyaçlarını daha iyi anlayabilir ve otomasyon sistemlerini buna göre optimize edebilirler.
- **Personel ve Hizmet Performansı:** Personel performansı, hizmet süresi ve kullanıcı memnuniyeti gibi verileri görselleştirerek, hizmet kalitesini artırmak için gerekli düzenlemeleri yapabilirler.

Kütüphaneler, yukarıdaki amaçlar doğrultusunda güncel veri setleri oluşturarak, veri görselleştirme modelinin ilk adımı olan “ham verilerin toplanması” (adım 1) adımını tamamlamış olacaklardır.

Kütüphane sunucu bilgisayarına veya kurumsal bir bulut depolama alanına CSV formatında kaydedilmiş olan bu veri setlerine modelin “veri seti ön işleme” (adım 2), “veri temizleme” (adım 3), “görselleştirilecek alt veri alanlarının belirlenmesi” (adım 4) ve “görselleştirme / sunum” (adım 5) adımları sırasıyla uygulanarak ihtiyaç duyulan görselleştirmeler oluşturulabilecektir. Kütüphane hizmetleri için oluşturulacak veri setleri ve bu veri setlerindeki değişken adları çalışma kapsamında oluşturulan veri seti değişkenlerinden farklı olacağı için, bu çalışmada oluşturulan kod blokları olduğu gibi kullanılamayacaktır. Ancak bu kod blokları kütüphane veri setlerinin; ön işleme (adım 2), temizleme (adım 3), keşifsel veri analizi (adım 2-3), alt veri seti oluşturma (adım 4) ve görselleştirme (adım 5) süreçlerinde rehber olarak kullanılabilir.

Veri görselleştirme, kütüphanelerin bu tür verilerini daha anlaşılır ve eyleme geçirilebilir hale getirmesini sağlar, böylece stratejik kararlar almak için güçlü bir araç sunar. Bu çalışma kapsamında gerçekleştirilen uygulama ve tasarlanan model hem üniversitelere hem de kütüphanelere, öncelikle akademik yayınlar için veri görselleştirmede yararlı bilgiler sağlayabilecek niteliktedir. Ayrıca diğer kütüphane hizmet verilerinin analizi ve görselleştirilmesi için de faydalı olacağı düşünülmektedir.

KAYNAKÇA

- Al, U., Sezen, U., ve Soydal, İ. (2012). *Türkiye'nin bilimsel yayınlarının sosyal ağ analizi yöntemiyle değerlendirilmesi*. <https://bby.hacettepe.edu.tr/bs/projeler/SOBAG-110K044.pdf>
- Ali, S. M., Gupta, N., Nayak, G. K. ve Lenka, R. K. (2016). Big data visualization: Tools and challenges. *2nd International Conference on Contemporary Computing and Informatics (IC3I)* içinde (ss. 656-660). IEEE.
- Altın, N. C. (2018). Veri görselleştirme ve infografiklerin tasarım eğitimi içerisindeki yeri. *İdil Sanat ve Dil Dergisi*, 7(45), 575-588. doi: 10.7816 /idil-07-45-10
- Amadebai, E., (t.y.). Data visualization vs information visualization explained. *Analytics For Decisions*. 25 Ocak 2024 tarihinde https://www.analyticsfordecisions.com/data-visualization-vs-information-visualization/?expand_article=1 adresinden erişildi.
- American Psychological Association. (2013). *Publication manual of the American Psychological Association (6th ed.)*. Washington, DC.
- Andrews, S., Segonds-Pichon, A., Virk, B. ve Montgomery, J. (2023). *Scientific Figure Design*. Bahraham Bioinformatics. https://www.bioinformatics.babraham.ac.uk/training/Figure_Design_Course/Figure%20Design%20Slides.pdf
- Ankara Üniversitesi. (2024). *Ankara Üniversitesi 2024-2028 Stratejik Planı*. Strateji Geliştirme Daire Başkanlığı, Stratejik Yönetim ve Planlama Şube Müdürlüğü http://sgdb.ankara.edu.tr/wp-content/uploads/sites/172/2024/01/AU_2024-2028_Stratejik_Planı.pdf
- Artut, H. S. (2017). Bilim kurgu sineması bağlamında veri görselleştirme ile geleceği yorumlamak. *The Turkish Online Journal of Design, Art and Communication*, 7(4), 626-633.
- Aslan, A. (2019). TR Dizin. *Acta Medica Alanya*, 3(1), 1-2. <https://doi.org/10.30565/medalanya.557393>

- Asan, M.A. (2017, 8-10 Şubat). *Güncel imge doku yumuşatma yöntemlerinin bir incelemesi*. [Konferans sunumu]. AB'17 19. Akademik Bilişim Konferansı, Aksaray Üniversitesi, Aksaray, Türkiye. <https://ab.org.tr/ab17/bildiri/73.pdf>
- Atalay, İ. E. ve Neslihanoglu, S. (2021). Türkiye'deki illerin partikül madde (PM10) miktarının değerlendirilmesi ve R programlama dili ile görselleştirilmesi. *Doğal Afetler ve Çevre Dergisi*, 7(2), 354-361. <https://doi.org/10.21324/dacd.882682>
- Barnard, F. R. (1921). One look is worth a thousand words. *Printers Ink*, 8, 96-97.
- Bayter, M. (2008). *Web sitelerinin kimliklenmesi: Üst veri (Metadata)*. (1.Baskı). İstanbul: Hiperlink Yayınları.
- BeautifulSoup. (2023). *Beautifulsoup Documentation*. <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>
- Beautifulsoup. (2024, 14 Ocak). *Beautifulsoup project description*. <https://pypi.org/project/beautifulsoup4/>
- Börner, K., Chen, C., ve Boyack, K. W. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology*, 37(1), 179–255. <https://doi.org/10.1002/aris.1440370106>
- Brigham, T. J. (2016). Feast for the eyes: an introduction to data visualization. *Medical Reference Services Quarterly*, 35(2), 215-223. doi: 10.1080/02763869.2016.1152146
- Brillinger, D. R. (2002). John W. Tukey: His life and professional contributions. *Annals of Statistics*, 30(6), 1535-1575. <https://www.stat.berkeley.edu/~brill/Papers/life.pdf>
- Brown, J. M., ve Stowers, E. D. (2013). Use of data in collections work: An exploratory survey. *Collection Management*, 38(2), 143-162. <https://doi.org/10.1080/01462679.2013.763742>
- Bucur, D. (2022). The network signature of constellation line figures. *PLoS ONE*, 17(7), e0272270. <https://doi.org/10.1371/journal.pone.0272270>
- Budiu, R. (2019, 10 Kasım). Information foraging: A theory of how people navigate on the Web. *Nielsen Norman Group*. <https://www.nngroup.com/articles/information-foraging/>

- Buhler, J., Lewellen, R., ve Murphy, S. (2016). Tableau unleashed: Visualizing library data. *Research Library Issues: A Report from ARL, CNI, and SPARC*, 288, 21-36. <https://publications.arl.org/rli288/21>
- Bytescout. (t.y). CSV format: History, advantages and why it is still popular. <https://bytescout.com/blog/csv-format-history-advantages.html>
- Cairo, A. (2012). *The Functional Art: An introduction to information graphics and visualization*. New Riders.
- California Digital Library. (t.y.). *Data management general guidance*. DMPTool. https://dmptool.org/general_guidance#types-of-data
- Cansız, S. (2019). Parçalı Dairesel veri görselleştirme tekniğinin R “Ggplot2” paketi ile noktasal tabanlı oluşturulması. *Tasarım Enformatiği*, 1(1), 62-66.
- Card, S. K., Mackinlay, J. D. ve Shneiderman, B. (1999). *Readings in information visualization: Using vision to think*. Morgan Kauffmann.
- Carranza, E.J.M. (2021). Exploratory Data Analysis. Daya Sagar, B., Cheng, Q., McKinley, J., Agterberg, F. (editörler) *Encyclopedia of Mathematical Geosciences. Encyclopedia of Earth Sciences Series*. Springer, Cham. https://doi.org/10.1007/978-3-030-26050-7_105-1
- Categorical data. (t.y). https://pandas.pydata.org/docs/user_guide/categorical.html
- Chen, C. (2004). Searching for intellectual turning points: Progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences*, 101(suppl 1), 5303-5310.
- Chen, C. (2005). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for information Science and Technology*, 57(3), 359-377. <https://doi.org/10.1002/asi.20317>
- Chen, C. (2010). Information visualization. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 387–403. <https://doi.org/10.1002/wics.89>
- Chen, H. M. (2017). Real-world uses for information visualization in libraries. *Library Technology Reports*, 53(3), 21–25. <https://journals.ala.org/index.php/ltr/article/view/6291>

- Cheng, B. ve Wood, J. (t.y.). *Chernoff Faces*, CS 465: *Information Visualization*.
https://www.cs.middlebury.edu/~candrews/showcase/infovis_techniques_s16/chernoff/chernoff.html
- Chernoff, H. (1973). The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association*, 68(342), 361-368.
- CiteSpace. (2021). *CiteSpace visualizing patterns and trends in scientific literature*.
<http://cluster.cis.drexel.edu/~cchen/citespace/>
- Classics of data visualization. (2021).
<https://www.arcgis.com/apps/MapJournal/index.html?appid=537fe15974c04df24f50ffe3d519de#>
- Conner-Simons, A. (2015, 5 Kasım). *How to make better visualizations*. Massachusetts Institute of Technology. <https://news.mit.edu/2015/how-make-better-infographic-visualizations-1105>
- Çalışkan, C., ve Kahraman, M. E. (2020). Viyana Toplum ve Ekonomi Müzesi'nin sergileme tasarımında verinin görselleştirilmesi. *İstanbul Aydın Üniversitesi Güzel Sanatlar Fakültesi Dergisi*, 6(11), 1-9. doi: 10.17932/IAU.SANAT.2015.015/sanat_v06i11001
- Çelik, S., ve Akdamar, E. (2018). Büyük veri ve veri görselleştirme. *Akademik Bakış Uluslararası Hakemli Sosyal Bilimler Dergisi*, (65), 253-264.
- Çuhadar, M., Ongun, U., ve Topsakal, Y. (2022). Kültürel miras ve dijitalleşme konusunda yayınlanmış çalışmaların bibliyometrik analizi (Bibliometric analysis of published studies on cultural heritage and digitalization). *Journal of Tourism & Gastronomy Studies*, 10(4), 3418-3443. <https://doi.org/10.21325/jotags.2022.1148>
- Datacarpentry. (2023, 6 Mayıs). *Data types and formats*. <https://datacarpentry.org/python-ecology-lesson/04-data-types-and-format.html>
- Datawrapper. (t.y.). *Choropleth map*. <https://www.datawrapper.de/maps/choropleth-map>
- Dataquest. (2022, 9 Şubat). Tutorial: An Introduction to Python Requests Library. <https://www.dataquest.io/blog/tutorial-an-introduction-to-python-requests-library/>

Deardorff, A. (2016). Tableau (version. 9.1). *Journal of the Medical Library Association*, 104(2), 182-183. [doi: 10.3163/1536-5050.104.2.022](https://doi.org/10.3163/1536-5050.104.2.022)

DergiPark. (2022). *DergiPark hakkında*. <https://dergipark.org.tr/tr/pub/page/about>

Dilke, O. A. (1987). Maps in the service of the state: Roman cartography to the end of the Augustan Era. *The history of cartography*, 1, 201-211. https://press.uchicago.edu/books/HOC/HOC_V1/HOC_VOLUME1_chapter12.pdf

DOAJ. (t.y). *About DOAJ*. <https://doaj.org/about/>

Doğan, G. ve Taşkın, Z. (2019). Başlangıcından bugüne entelektüel birikimi ve bilime katkılarıyla Hacettepe Üniversitesi. Ankara: Hacettepe Üniversitesi Yayınları (ISBN: 978-975-491-490-0) <https://library.hacettepe.edu.tr/books/fulltext/gdzt2019.pdf>

Doğanay, A., Ataizi, M., Şimşek, A., Balaban Salı, J., ve Akbulut, Y. (2018). *Sosyal bilimlerde araştırma yöntemleri*. Anadolu Üniversitesi.

Doğanay, H. A., Orman, A. ve Dener, M. (2022). Siber güvenlik amaçlı büyük veri görselleştirme: BETH veri seti. *El-Cezerî Fen ve Mühendislik Dergisi*, 9(4); 1572- 1582.

Doğru, S. (2021). 2013-2017 yılları arasında sınıf yönetimi alanında, SSCI dergilerinde yayınlanmış olan çalışmaların bibliyometrik analizi. *PEARSON JOURNAL*, 6(16), 321-331. <https://doi.org/10.46872/pj.434>

DOI Foundation. (2022). *What Is The Digital Object Identifier (DOI)?*. <https://www.doi.org/>

D'ORGEIX, É. (2020). Cartographic representation of Europe during the Renaissance. *Encyclopédie d'histoire numérique de l'Europe [online]*, ISSN 2677-6588. <https://ehne.fr/en/node/12465>

Dos Santos, S. ve Brodlie, K. (2004). Gaining understanding of multivariate and multidimensional data through visualization. *Computers & Graphics*, 28(3), 311-325. <https://doi.org/10.1016/j.cag.2004.03.013>

Dragicevic, P. ve Jansen, Y. (2012). *List of Physical Visualizations*. <http://dataphys.org/list/>

- Driscoll, M. (t.y). *Jupyter notebook: An introduction*. Real Python. <https://realpython.com/jupyter-notebook-introduction/>
- DundasBI. (2019, Eylül). *A brief history of data visualization*. <https://www.dundas.com/resources/blogs/introduction-to-business-intelligence/brief-history-data-visualization>
- Eaton, M. (2017). Seeing library data: A prototype data visualization application for librarians. *Journal of Web Librarianship*, 11(1), 69-78. <https://doi.org/10.1080/19322909.2016.1239236>
- Eken, S. (2020). Büyük verinin interaktif görselleştirilmesi: Tableau üzerine öğrenci deneyimleri. *Avrupa Bilim ve Teknoloji Dergisi*, (18), 262-271. <https://doi.org/10.31590/ejosat.659823>
- Enago Academy. (2019, 19 Temmuz). *Report research data using box plots and bar charts*. <https://www.enago.com/academy/report-research-data-using-box-plots-and-bar-charts/>
- Er, S. ve Akman, G. (2022). Kurumsal kaynak planlama ve veri görselleştirme teknolojisi: Müşteri ilişki analizi. Luigi Pio Leonardo Cavaliere, Ali Ahmad ve Ruhul Amin Sardar (editörler), *Communication Economics Organization (20-22 May 2022), India, 4th Proceedings Book* içinde (ss. 645-657).
- Erkurt, E. (2020). *Büyük veri görselleştirme ve Türkiye’de konut sektörüne ilişkin infografikler* [Yayımlanmamış doktora tezi]. Marmara Üniversitesi.
- Etiket bulutu. (2023, 3 Kasım). *Wikipedia* içinde. https://tr.wikipedia.org/wiki/Etiket_bulut
- EU Open Data Portal. (2021). Visualisations. https://web.archive.org/web/20191107031704/http://data.europa.eu/euodp/en/knowledge-center/?ff0=field_type:118
- Falkowitz, R. (2019, 21 Kasım). Information visualization or data visualization. *Concentric Circle Consulting*. <https://www.3cs.ch/information-visualization-data-visualization/>

- Few, S. ve Edge, P. (2007). Data visualization: past, present, and future. *IBM Cognos Innovation Center*, 1-12. https://mail.perceptualedge.com/articles/Whitepapers/Data_Visualization.pdf
- Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., ve Berners-Lee, T. (1999). *Hypertext transfer protocol--HTTP/1.1* (No. rfc2616).
- Finch, J. L. ve Flenner, A. R. (2016). Using data visualization to examine an academic library collection. *College & Research Libraries*, 77(6), 765. <https://doi.org/10.5860/crl.77.6.765>.
- Flowingdata. (2021, 20 Nisan). *Pandemic migrations*. <https://flowingdata.com/2021/04/20/pandemic-migrations/>
- Friendly, M. (2000). Re-visions of Minard. *Statistical Computing & Statistical Graphics Newsletter*. 11(1), 1.
- Friendly, M. (2006). A brief history of data visualization. C. Chen, W. Härdle ve A. Unwin (editörler), *Handbook of Computational Statistics: Data Visualization (Vol. III)* içinde. Heidelberg: Springer-Verlag. 04 Mayıs 2022 tarihinde <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=7f5eaf018578aa5f6ca829b42cc2a8fcc103e763> adresinden erişildi.
- FusionCharts. (2024a). *Sunburst chart*. <https://www.fusioncharts.com/resources/chart-primers/sunburst-chart>
- FusionCharts. (2024b). *Treemap chart*. <https://www.fusioncharts.com/resources/chart-primers/treemap-chart>
- Gleicher, M., Albers, D., Walker, R., Jusufi, I., Hansen, C. D., ve Roberts, J. C. (2011). Visual comparison for information visualization. *Information Visualization*, 10(4), 289-309. <https://doi.org/10.1177/1473871611416549>
- Google (t.y.). *Colaboratory*. <https://research.google.com/colaboratory/faq.html>
- Google Charts (t.y.). *Charts Gallery*. <https://developers-dot-devsite-v2-prod.appspot.com/chart/interactive/docs/gallery>

- Hardin, M., Hom, D., Perez, R. ve Williams, L. (2012). Which chart or graph is right for you? *Tell Impactful Stories with Data. Tableau Software.* http://www.e330.org/s/CCSS-Math-which_chart_v51-fbkd.pdf
- Heat Map. (2024, 19 Haziran). *Wikipedia* içinde. https://en.wikipedia.org/wiki/Heat_map
- Highnam, K., Arulkumaran, K., Hanif, Z. ve Jennings, N. R. (2021). BETH Dataset: Real Cybersecurity Data for Unsupervised Anomaly Detection Research. *CEUR Workshop Proc* içinde (Vol. 3095, pp. 1-12).
- Hillery, A. (2020, 19 Ekim). The evolution of data visualization. *ChartIO.* <https://chartio.com/blog/the-evolution-of-data-visualization/>
- Hoelscher, J. ve Mortimer, A. (2018). Using Tableau to visualize data and drive decision-making. *Journal of Accounting Education*, 44, 49-59.
- Horizon Telescope Collaboration, T. E. (2019). First M87 Event Horizon Telescope Results. IV. Imaging the Central Supermassive Black Hole. *ArXiv.* <https://doi.org/10.3847/2041-8213/ab0e85>
- HTML. (2024, 31 Mayıs). *Wikipedia* içinde. <https://tr.wikipedia.org/wiki/HTML>
- IBM. (2023). *What is exploratory data analysis? (EDA).* Erişim adresi: <https://www.ibm.com/topics/exploratory-data-analysis>.
- İnterpolasyon. (2024, 1 Mayıs). *Wikipedia* içinde. <https://tr.wikipedia.org/wiki/%C4%B0nterpolasyon>
- Jaadi, Z. (2021). A step-by-step explanation of principal component analysis (PCA). *Built In.* <https://builtin.com/data-science/step-step-explanation-principal-component-analysis>
- Jupyter. (2022). *Jupyter Notebook.* <https://jupyter-notebook.readthedocs.io/en/latest/>
- Karcı Kandemir, P. (2023). *Türkiye’de üniversite kütüphanelerinde veri kütüphaneciliği yeterlikleri ve araştırma veri hizmetlerinin kapsamı: Bir karma yöntem çalışması* [Yayımlanmamış doktora tezi]. Ankara Üniversitesi.

- Keim, D. A. (2002). Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, 8(1), 1-8. [doi: 10.1109/2945.981847](https://doi.org/10.1109/2945.981847)
- Kent Harita Eğitim. (2016, 11 Şubat). *Kontur Gabari nedir?*. <https://www.kentharita.com/kontur-gabari-nedir/>
- Kirk, A. (2019). *Data visualisation: A handbook for data driven design* (2. basım.). Sage.
- Kolko, J., Badger, E. ve Bui, Q. (2021, 19 Nisan). How the pandemic did, and didn't, change where Americans move. *The New York Times*. <https://www.nytimes.com/interactive/2021/04/19/upshot/how-the-pandemic-did-and-didnt-change-moves.html>
- Kowieski, J. (2022, 2 Aralık). Data cleaning: What it is, examples, and how to keep your data clean in 7 steps. *ThoughtSpot*. <https://www.thoughtspot.com/data-trends/data-science/what-is-data-cleaning-and-how-to-keep-your-data-clean-in-7-steps>
- Krum, R. (2014). *Cool infographics: Effective communication with data visualization and design*. John Wiley & Sons.
- LaPolla, F. W. Z. (2020). Excel for data visualization in academic health sciences libraries: a qualitative case study. *Journal of the Medical Library Association: JMLA*, 108(1), 67-75. [doi: 10.5195/jmla.2020.749](https://doi.org/10.5195/jmla.2020.749).
- Laurels. (2009, 10 Eylül). *Historical milestones. Escaping the 2D plane: The stereogram*. <http://euclid.psych.yorku.ca/SCS/Gallery/historical.html>
- Library Carpentry (2016, Haziran). *Data Types and Formats*. <http://LibraryCarpentry.github.io/library-data-intro/>.
- LinkedIn. (2024, 01 Haziran). *Data visualization tools and software market size and analysis research report 2031*. <https://www.linkedin.com/pulse/data-visualization-tools-software-market-size-analysis-znbye/>
- LinkedIn. (2024, 5 Haziran). *How do you use data type conversion to improve data visualization and interpretation?*. <https://www.linkedin.com/advice/1/how-do-you-use-data-type-conversion-improve-visualization>

Liu, J., Tang, T., Wang, W., Xu, B., Kong, X., ve Xia, F. (2018) A survey of scholarly data visualization, *IEEE Access*, 6, 19205-19221. [doi: 10.1109/ACCESS.2018.2815030](https://doi.org/10.1109/ACCESS.2018.2815030).

matplotlib. (2023). *matplotlib.pyplot.boxplot*.
https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.boxplot.html

Mazza, R. (2009). *Introduction to information visualization*. Springer Science & Business Media.

Mcleod, S. (2023, 31 Temmuz). Box plot explained: Interpretation, examples, & comparison. *Simplypsychology*. <https://www.simplypsychology.org/boxplots.html>

Merriam-Webster. (t.y.). Data. *Merriam-Webster.com sözlük* içinde. 15 Şubat 2022 tarihinde <https://www.merriam-webster.com/dictionary/data> adresinden erişildi.

Merriam-Webster. (t.y.). Histogram. *Merriam-Webster.com sözlük* içinde. 16 Mayıs 2021 tarihinde <https://www.merriam-webster.com/dictionary/histogram> adresinden erişildi.

Merriam-Webster. (t.y.). Nomogram. *Merriam-Webster.com sözlük* içinde. 16 Mayıs 2021 tarihinde <https://www.merriam-webster.com/dictionary/nomogram> adresinden erişildi.

Merriam-Webster. (t.y.). Visualization. *Merriam-Webster.com sözlük* içinde 10 Şubat 2022 tarihinde <https://www.merriam-webster.com/dictionary/visualization> adresinden erişildi.

Miller, J. D. (2017). *Big data visualization*. Packt Publishing Ltd.

Mixon, E. ve Wigmore, I. (2023, Nisan). What is Google Drive?. *TechTarget*.
<https://www.techtarget.com/searchmobilecomputing/definition/Google-Drive>

Moffitt, C. (2018, 26 Mart). Overview of pandas data types. *Practical Business Python*.
https://pbpython.com/pandas_dtypes.html

Murphy, S. A. (2013). Data visualization and rapid analytics: Applying tableau desktop to support library decision-making. *Journal of Web Librarianship*, 7(4), 465-476.
<https://doi.org/10.1080/19322909.2013.825148>

NumPy. (2023). *About Us*. <https://numpy.org/about/>

NumPy. (t.y). *Datetimes and Timedeltas*.
<https://numpy.org/doc/stable/reference/arrays.datetime.html>

Office for National Statistics. (2015, 18 Aralık). *How popular is your birthday?*
<https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/articles/howpopularisyourbirthday/2015-12-18>

Ogier, A. L. ve Stamper, M. J. (2018). Data visualization as a library service: Embedding visualization services in the library research lifecycle. *Journal of eScience Librarianship*, 7(1): e1126. <https://doi.org/10.7191/jeslib.2018.1126>

Oxford Learners Dictionaries. (t.y). Dataset. *Oxford Learners Dictionaries* içinde.
<https://www.oxfordlearnersdictionaries.com/definition/english/data-set?q=data+set>

Öznur, E. ve Bayram, Ö. (2021). Görsel imgelemin canlılığı ölçeğinin Türkçe'ye uyarlama, geçerlik ve güvenirlik çalışması. *Ekev Akademi Dergisi*, (87), 525-536.
<https://dergipark.org.tr/en/pub/sosekev/issue/71551/1151426>

Pandas. (t.y.). *Intro to data structures*. https://pandas.pydata.org/docs/user_guide/dsintro.html

Pandas. (t.y.). *Working with missing data*. https://pandas.pydata.org/pandas-docs/stable/user_guide/missing_data.html#working-with-missing-data

Patnaik, R. (2019). *Data analytics and visualization in libraries*. INFLIBNET Centre, Gandhinagar. 5 Aralık 2022 tarihinde
<https://ir.inflibnet.ac.in/bitstream/1944/2331/1/3.pdf> adresinden erişildi.

Pavlopoulos, G. A., Paez-Espino, D., Kyripides, N. C. ve Iliopoulos, I. (2017). Empirical comparison of visualization tools for larger-scale network analysis. *Advances in Bioinformatics*, 2017(1), 1278932. <https://doi.org/10.1155/2017/1278932>

Picture Superiority Effect. (2024, 13 Mayıs). *Wikipedia* içinde.
https://en.wikipedia.org/wiki/Picture_superiority_effect

Pirolli, P. ve Card, S. (1999). Information foraging. *Psychological Review*, 106(4), 643-675. <https://doi.org/10.1037/0033-295X.106.4.643>

Python. (t.y.). *Python 3.12.4 documentation*. <https://docs.python.org/3/>

Python. (t.y.). *re-Regular expression operations*. <https://docs.python.org/3/library/re.html>

Raciborski, R. (2009). Graphical representation of multivariate data using Chernoff faces. *The Stata Journal*, 9(3), 374-387. <https://doi.org/10.1177/1536867X0900900302>

Reitz, J. M. (2014a). Application programming interface (API). *ODLIS: Online dictionary for library and information science* içinde 15 Şubat 2022 tarihinde https://odlis.abc-clio.com/odlis_a.html#api adresinden erişildi.

Reitz, J. M. (2014b). Data visualization. *ODLIS: Online dictionary for library and information science* içinde 15 Şubat 2022 https://odlis.abc-clio.com/odlis_v.html adresinden erişildi.

Reitz, J. M. (2014c). Metadata. *ODLIS: Online dictionary for library and information science* içinde 15 Şubat 2022 tarihinde https://odlis.abc-clio.com/odlis_m.html#metadata adresinden erişildi.

Rodgers, T. (t.y.). Which type of chart or graph is right for you?. *Tableau*. <https://www.tableau.com/learn/whitepapers/which-chart-or-graph-is-right-for-you>

Roldós, I. (2021, 18 Ekim). 8 effective data cleaning techniques for better data. *MonkeyLearn*. <https://monkeylearn.com/blog/data-cleaning-techniques/>

Ruhland, K., Bioletti, S. ve Sedlmair, M. (2009). LibViz: Data visualisation of the old library. *International Journal of Architectural Computing*, 7(1):177-192. <https://doi.org/10.1260/147807709788549402>

Seattle Bubble. (2010, 24 Ağustos). “Entry Level” home sales fell furthest in july. <https://seattlebubble.com/blog/2010/08/24/entry-level-home-sales-fell-furthest-in-july>

Statcounter Global Stats. (2024a, Mayıs). Social media stats worldwide. *Statcounter*. <https://gs.statcounter.com/social-media-stats#monthly-202101-202405-bar>

Statcounter Global Stats. (2024b, Mayıs). Browser market share worldwide. *Statcounter*. <https://gs.statcounter.com/browser-market-share#monthly-201001-202405>

- Stedman, C. (t.y.). Data cleansing (data cleaning, data scrubbing). *TechTarget*.
<https://www.techtarget.com/searchdatamanagement/definition/data-scrubbing>
- Sterman, L. ve Borda, S. (2017). Making visualization work for institutional repositories: Information visualization as a means to browse electronic theses and dissertations. *Journal of Librarianship and Scholarly Communication*, 5(1), eP2140.
<https://doi.org/10.7710/2162-3309.2140>
- Stojiljković, M. (2024). The pandas DataFrame: Make working with data delightful. *Real Python*. <https://realpython.com/pandas-dataframe/>
- Su, X., Li, X. ve Kang, Y. (2019). A bibliometric analysis of research on intangible cultural heritage using CiteSpace. *SAGE Open*, 9(2), 1-18.
<https://doi.org/10.1177/2158244019840119>
- Tableau (t.y.). *Guide to data cleaning: definition, benefits, components, and how to clean your data*. <https://www.tableau.com/learn/articles/what-is-data-cleaning>
- TechTarget (t.y.). *Datatype*. <https://www.techtarget.com/searchapparchitecture/definition/data-type>
- Terra, J. (2024, 16 Şubat). Why Python is essential for data analysis and data science. *simplilearn*. <https://www.simplilearn.com/why-python-is-essential-for-data-analysis-article>
- The Data Visualisation Catalogue. (t.y.). *Pictogram chart*.
<https://datavizcatalogue.com/methods/pictogram.html>
- Think Insights. (2017, 8 Ağustos). *History of data visualization*.
<https://thinkinsights.net/data/data-visualization-history>
- Thompson, C. (2016, Temmuz). The surprising history of the infographic. *Smithsonian Magazine*. <https://www.smithsonianmag.com/history/surprising-history-infographic-180959563/?no-ist>
- Tominski, C. (2006). *Event based visualization for user centered visual analysis* [Yayımlanmamış doktora tezi]. University of Rostock.
- TR Dizin. (t.y.). *Hakkında*. <https://trdizin.gov.tr/hakkinda/>

TR Dizin. (t.y). *TR Dizin Yardım*. <http://trdizin.gov.tr/yardim>

TRUECHART. (2021, 22 Haziran). *A quick history lesson in data visualization*.
<https://www.truechart.com/data-visualization-history/>

Tufte, E. R. (1997). *Visual explanations: Images and quantities, evidence and narrative*. Graphics Press.

Tufte, E. R. (2001). *The visual display of quantitative information*. Graphics Press Cheshire, CT. <https://faculty.salisbury.edu/~jtanderson/teaching/cosc311/fa21/files/tufte.pdf>

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley Series in Behavioral Science: Quantitative Methods. Reading, MA: Addison-Wesley.

TÜBİTAK ULAKBİM TV. (2019, 30 Eylül). *TR Dizin tanıtım ve eğitim etkinliği*. [Video]. YouTube.
https://www.youtube.com/watch?v=_zKkeFyXNM&list=PLRofGbhRbm7gZHFGAb3g9a4uS4-15_PKI

w3resource. (2024, 4 Ocak). *Pandas exercises, practice, solution*.
<https://www.w3resource.com/python-exercises/pandas/index.php>

Ward, M.O., Grinstein, G. ve Keim, D. (2015). *Interactive data visualization: Foundations, techniques, and applications*, Second Edition (2nd ed.). A K Peters/CRC Press.
<https://doi.org/10.1201/b18379>

Waskom, M. L., (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60), 3021, <https://doi.org/10.21105/joss.03021>.

Web kazıma. (2024, 21 Nisan). *Wikipedia içinde*
https://tr.wikipedia.org/wiki/Web_kaz%C4%B1ma

Wood, B. ve Bolch, C. (2018). Seeing and understanding data. *Statistics and Probability*. 2.
https://digitalcommons.ursinus.edu/cgi/viewcontent.cgi?article=1002&context=triumphs_statistics

- World Economic Forum. (2017). *The Global Competitiveness Report 2017 - 2018*. K. Schwab (editör). <https://www3.weforum.org/docs/GCR2017-2018/05FullReport/TheGlobalCompetitivenessReport2017%E2%80%932018.pdf>
- Wu, C., Kennedy, A., Qiu, L., ve Smyth, N. (2019). Library data visualisation transformations: global patterns for research evaluation metrics in China. *International Conference on Library and Information Science (LIS 2019)*, 19-21 January, 2019, Sapporo, Japan içinde. 11 Eylül 2021 tarihinde <https://eprints.nottingham.ac.uk/56031/1/Library%20Data%20Visualisation%20Transformations%20global%20patterns%20for%20research.pdf> adresinden erişildi.
- Veltman, N. (t.y.). *Effects of daylight saving time*. <https://noahveltman.com/daylight/>
- Yau, N. (2008, 15 Şubat). How to read and use a box-and-whisker plot. *Flowingdata*. <https://flowingdata.com/2008/02/15/how-to-read-and-use-a-box-and-whisker-plot/>
- Yi M. ve Sapountzis, M., (2024). Essential chart types for data visualization. *Atlassian*. <https://www.atlassian.com/data/charts/essential-chart-types-for-data-visualization>
- Zakaria, M. S. (2021). Data visualization as a research support service in academic libraries: An investigation of world-class universities. *The Journal of Academic Librarianship*, 47(5), 102397. <https://doi.org/10.1016/j.acalib.2021.102397>
- Zeeshan, A.A. (2022, 4 Kasım). Secure Python URL validation. *Snyk Blog*. <https://snyk.io/blog/secure-python-url-validation/>
- Zengin, Z. B., ve Cemelelioğlu, N. (2022). Tanzimat’tan günümüze en çok eser vermiş 10 edebiyatçı kadına yönelik veri görselleştirme projesi. Turan Sağer (Editör), *Yıldız Teknik Üniversitesi Yaratıcı Endüstriler Sempozyumu II (13-14 Ekim 2022)*, *Bildiriler Kitabı* içinde, (ss. 30-38).
- Zhang, L. ve Eichmann-Kalwara, N. (2019). Mapping the scholarly literature found in Scopus on “research data management”: A bibliometric and data visualization approach. *Journal of Librarianship and Scholarly Communication*, 7(1). Erişim adresi: <https://www.iastatedigitalpress.com/jlsc/article/12841/galley/12521/download/>
- Zodyak. (2024, 1 Mayıs). *Wikipedia* içinde. <https://tr.wikipedia.org/wiki/Zodyak>

EKLER

EK-1: Terimler Listesi

Terim	Türkçe	Tanım
API-Application Programming Interface	Uygulama programlama arayüzü	Kullanıcıya yazılım işlevlerinin bir uygulamadan diğerine aktarıldığına dair hiçbir belirti vermeden, Web tabanlı uygulama programları arasında iletişimin sorunsuz bir şekilde gerçekleşmesini sağlayan bir yazılımdan yazılıma veri transferi arayüzüdür (Reitz, 2014a).
Beautifulsoup		Beautiful Soup, HTML ve XML dosyalarından veri çekmeye yönelik bir Python kod kütüphanesidir (BeautifulSoup, 2023).
CSV (comma separated values)	Virgülle ayrılmış değerler	Verilerin tabloda yapılandırılmış biçimde düz metin olarak kaydedilmesine olanak tanıyan virgülle ayrılmış değerler dosyası formatıdır (Bytescout, t.y).
Data	Veri	Akıl yürütme, tartışma veya hesaplama için temel olarak kullanılan gerçek bilgiler (ölçümler veya istatistikler gibi) (Merriam-Webster, t.y.).
Data cleaning	Veri temizleme	Veri temizleme, bir veri kümesindeki yanlış, bozuk, yanlış biçimlendirilmiş, yinelenen veya eksik verileri düzeltme veya kaldırma işlemidir (Tableau, t.y).
DataFrame		DataFrame, satır ve sütunlardan oluşan, değişkenleri (sütun) ve buna karşılık gelen verileri (satır) içeren iki boyutlu bir veri yapısıdır (DataFrame, 2022).
Data set	Veri seti	Bilgisayar tarafından tek bir birim olarak işlenen veri topluluğudur (Oxford Learners Dictionaries, t.y).
Data type	Veri Türü	Bilgisayara bilimde veri türü, bir değişkenin sahip olduğu değer türünü ve ne tür matematiksel, ilişkisel veya mantıksal işlemlerin hataya neden olmadan uygulanabileceğini ifade eder (Techtarget, t.y.).
Data type-category	Kategorik veri tipi	İstatistiklerdeki kategorik değişkenlere karşılık gelen pandas veri türüdür. Kategorik bir değişken sınırlı ve genellikle sabit sayıda olası değer alır (Categorical data, t.y).
Data type-datetime64		datetime64(), tarih ve saat bilgilerini 64 bitlik bir tamsayı nesnesi olarak saklayan bir Python Numpy fonksiyonudur (NumPy, t.y).
Data type-float64		Ondalık sayı içeren sayısal karakterlerdir (Datacarpentry, 2023).
Data type-Int64		int64, veriyi 64 bitlik bir tamsayı olarak saklayan nümerik bir veri türüdür (Library Carpentry, 2016).

Data visualization	Veri görselleştirme	İçeriğin anlaşılmasını kolaylaştırmak için verileri grafikler, haritalar, etiket bulutları, animasyonlar veya herhangi bir grafiksel araç biçiminde temsil etmek için elektronik araçların (yazılım uygulamaları) kullanılmasıdır (Reitz, 2014b).
DergiPark		“TÜBİTAK ULAKBİM çatısı altında, Türkiye’de yayınlanan akademik hakemli dergiler için elektronik ortamda barındırma ve editoryal süreç yönetimi hizmeti sunan bir platformdur.” (DergiPark, 2022).
DOAJ (Directory of Open Access Journals)	Açık Erişim Dergiler Dizini	DOAJ, dünyanın dört bir yanından çeşitli açık erişimli dergilerin yer aldığı ve ücretsiz olarak çevrimiçi erişilebilen kapsamlı bir dizindir (DOAJ, t.y).
Exploratory Data Analysis (EDA)	Keşifsel Veri Analizi	Keşifsel veri analizi (EDA), veri bilimcileri tarafından veri kümelerini analiz etmek, araştırmak ve temel özelliklerini özetlemek için sıklıkla veri görselleştirme tekniklerinin kullanıldığı yöntemidir (IBM, 2023).
Google Colab		Colab, kullanmak için herhangi bir kurulum gerektirmeyen ve GPU’lar ve TPU’lar dâhil bilgi işlem kaynaklarına ücretsiz erişim sağlayan, barındırılan bir Jupyter Notebook hizmetidir. Colab özellikle makine öğrenimi, veri bilimi ve eğitime çok uygundur (Google, t.y).
Google Drive		Google Drive, kullanıcıların çevrimiçi olarak dosya depolamasına ve erişmesine olanak tanıyan bulut tabanlı bir depolama hizmetidir. Depolanan belgeleri, fotoğrafları ve daha fazlasını mobil cihazlar, tabletler ve PC’ler de dâhil olmak üzere kullanıcının tüm cihazlarında senkronize edilmesini sağlayan bir hizmettir (Mixon ve Wigmore, 2023)
HTML (Hypertext Markup Language)	Hipermetin İşaretleme Dili	HTML, web sayfalarını oluşturmak için kullanılan standart metin işaretleme dilidir (HTML, 2024).
HTTP (Hyper-Text Transfer Protocol)	Hiper-Metin Transfer Protokolü	İstek ve yanıt mesajlarının değişimi için TCP/IP bağlantılarına dayanan, istemci-sunucu mimarisine dayanan bir istek/yanıt protokolüdür. Dağıtık, hiper ortam bilgi sistemleri için uygulama düzeyinde bir protokoldür (Fielding vd., 1999, s.7; Dataquest, 2022).
Jupyter Notebook		Jupyter Notebook, canlı kod, denklemler, görselleştirmeler ve metin içeren belgeler oluşturmak ve paylaşmak için kullanabileceğiniz açık kaynaklı bir web uygulamasıdır (Driscoll, t.y).
metadata	Üst veri	Verilerle ilgili veriler. Çeşitli amaçlara yönelik bilgi kaynaklarını/nesnelerini açıklayan yapılandırılmış bilgi (Reitz, 2014c). Belge içeriğinin anlaşılmasına ve

		bunlara teknolojik olarak erişilebilmesine olanak sağlayan bir araç (Bayter, 2008. s.3)
Pandas		Pandas, Python programlama dili üzerine inşa edilmiş, hızlı, güçlü, esnek ve kullanımı kolay bir açık kaynak veri analizi ve manipülasyon aracıdır (Pandas, t.y. “Intro to data structures”).
Web scraping	Web kazıma	Web scraping yöntemi (web hasat veya web veri çekimi olarak da bilinmektedir) web sitelerinden bilgi toplamayı sağlayan bir yazılımdır. Bu yazılımların insanların web üzerindeki araştırma davranışlarını simüle edebilme özelliği de vardır (“Web kazıma”, 2024).

ÖZET

Veri analizi, ham verileri çeşitli projelerde stratejik karar alma için eyleme geçirilebilir içgörülere dönüştürmede hayati bir rol oynar. Veri kümeleri giderek daha karmaşık hale geldikçe, önemli ve ilgisiz bilgiler arasında ayırım yapmak daha zor hale gelir. Analiz tekniklerindeki ve sunum yöntemlerindeki ilerlemeler, veri kümeleri içindeki örüntülerin ortaya çıkarılmasına yardımcı olarak verilerin etkili bir şekilde yorumlanmasını ve anlaşılmasını kolaylaştırır. Veri kümelerinden kritik bilgilerin çözülmesine yardımcı olan önemli bir husus veri görselleştirme. Bireyler, veri görselleştirme araçlarını kullanarak, veri kümeleri içindeki temel bilgileri daha verimli bir şekilde ayırt edebilir ve karmaşık veri yapılarının genel anlayışını ve yorumunu geliştirebilir.

Veri görselleştirme, bilgiyi görsel formatlara dönüştürerek karmaşık veri kümelerinin anlaşılmasını ve özümsemesini basitleştirir. Verilerdeki görsel temsiller, örüntüler, eğilimler ve korelasyonlar daha erişilebilir hale gelir ve ham verilerde gizlenmiş olabilecek değerli içgörülerin ortaya çıkarılmasına yardımcı olur. Bu görsel yaklaşım yalnızca açık iletişimi teşvik etmekle kalmaz, aynı zamanda karmaşık bulguların, uzmanlık bilgisi olmayanlar da dâhil olmak üzere farklı hedef kitlelere etkili bir şekilde aktarılmasını sağlar.

Bu çalışma kapsamında öncelikle kavramsal çerçeveyi oluşturmak, sorunu sınırlamak ve yeni yaklaşımlar bulabilmek için alanyazın taraması yapılmıştır. Alanyazın taramasında veri görselleştirme kavramı başta olmak üzere veri görselleştirme ile ilgili kavramlar, veri görselleştirme yöntemleri, türleri ve araçları hem uluslararası hem de ulusal kaynaklardan detaylı olarak incelenmiştir.

Araştırmaya dair veriler TR Dizin’de yer alan 2017-2021 yılları arasındaki Ankara Üniversitesi adresli akademik yayınlara ilişkin üst veriler çekilerek bu veriler ile bir veri görselleştirme uygulaması oluşturulmuştur.

Görselleştirme uygulamasının geliştirilmesinde açık kaynak kodlu programlama dili olan Python’ın pandas, numpy, beautifulsoup, re, request, matplotlib, seaborn ve plotly kod kütüphaneleri kullanılmıştır. Görsellerin elde edilmesi için veri setine sorular sorulmuştur (bkz. Tablo 16). Her bir soru için uygun görselleştirme aracı belirlenmiş ve keşifsel veri analizi aracılığı ile verilerin doğru bilgi veren görseller üretilip üretilmediği test edilmiştir.

Anahtar kelimeler: Veri görselleştirme, Veri görselleştirmenin tarihi, Veri görselleştirme yöntemleri, Veri görselleştirme araçları, Veri seti oluşturma, Web kazıma, matplotlib, seaborn, plotly, TR Dizin, Ankara Üniversitesi



SUMMARY

Data analysis plays a vital role in transforming raw data into actionable insights for strategic decision-making in various projects. As data sets become increasingly complex, it becomes more difficult to distinguish between important and irrelevant information. Advances in analysis techniques and presentation methods help uncover patterns within data sets, making it easier to effectively interpret and understand data. An important aspect that helps unravel critical information from datasets is data visualization. By using data visualization tools, individuals can more efficiently discern key information within datasets and improve overall understanding and interpretation of complex data structures.

Data visualization simplifies the understanding and assimilation of complex data sets by transforming information into visual formats. Visual representations, patterns, trends, and correlations in data become more accessible and help uncover valuable insights that may be hidden in raw data. This visual approach not only encourages open communication, but also enables complex findings to be effectively conveyed to different audiences, including those without specialized knowledge.

Within the scope of this study, a literature review was first conducted to create the conceptual framework, limit the problem and find new approaches. In the literature review, concepts related to data visualization, especially the concept of data visualization, data visualization methods, types and tools were examined in detail from both international and national sources.

Data regarding the research was extracted from the metadata of academic publications from Ankara University between 2017 and 2021 in TR Dizin, and a data visualization application was created with these data.

Pandas, numpy, beautifulsoup, re, request, matplotlib, seaborn and plotly code libraries of Python, an open source programming language, were used in the development of the visualization application. Questions were asked to the data set to obtain visualizations (see Table: 16). The appropriate visualization type was determined for each question and it was tested whether the data produced visuals that provided accurate information through exploratory data analysis.

Keywords: data visualization, history of data visualization, data visualization methods, data visualization tools, creating data-set, web scraping, matplotlib, seaborn, plotly, TR Dizin, Ankara University

