

T.C.
MARMARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İŞLETME ANABİLİM DALI
SAYISAL YÖNTEMLER BİLİM DALI

**KURAL BAZLI VE SOSYOEKONOMİK YAKLAŞIMLA
MAĞAZA SEGMENTASYONU:
BİR SÜPERMARKET ZİNCİRİNDE UYGULAMA**

Doktora Tezi

EMRAH BİLGİÇ
762810701

İstanbul, 2016

T.C.

MARMARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İŞLETME ANABİLİM DALI
SAYISAL YÖNTEMLER BİLİM DALI

**KURAL BAZLI VE SOSYOEKONOMİK YAKLAŞIMLA
MAĞAZA SEGMENTASYONU:
BİR SÜPERMARKET ZİNCİRİNDE UYGULAMA**

Doktora Tezi

EMRAH BİLGİÇ

762810701

Danışman: YRD. DOÇ. DR. ÖZGÜR ÇAKIR

İstanbul, 2016

MARMARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜ

TEZ ONAY BELGESİ

İŞLETME Anabilim Dalı SAYISAL YÖNTEMLER Bilim Dalı DOKTORA öğrencisi
EMRAH BİLGİÇ'nin KURAL BAZLI VE SOSYOEKONOMİK YAKLAŞIMLA
MAĞAZA SEGMENTASYONU: BİR SÜPERMARKET ZİNCİRİNDE UYGULAMA adlı
tez çalışması, Enstitümüz Yönetim Kurulunun 1.09.2016 tarih ve 2016-34/12 sayılı kararıyla
oluşturulan jüri tarafından oy birliği / oy çokluğu ile Doktora Tezi olarak kabul edilmiştir.

Tez Savunma Tarihi 21.09.2016

Öğretim Üyesi Adı Soyadı

1. Tez Danışmanı	Yrd. Doç. Dr. ÖZGÜR ÇAKIR	İmzası
2. Jüri Üyesi	Prof. Dr. İSMAİL HAKKI ARMUTLULU	
3. Jüri Üyesi	Doç. Dr. FATMA NOYAN TEKELİ	
4. Jüri Üyesi	Prof. Dr. HAKAN YILDIRIM	
5. Jüri Üyesi	Doç. Dr. SEDA TOLUN	

KURAL BAZLI VE SOSYOEKONOMİK YAKLAŞIMLA MAĞAZA SEGMENTASYONU: BİR SÜPERMARKET ZİNCİRİNDE UYGULAMA

Doktora Tezi Özeti:

Bu tez çerçevesinde, Veri Madenciliğinin temel yöntemlerinden olan Kümeleme Analizi ve Birliktelik Kuralları Analizi kullanılarak bir süpermarket zincirine ait mağazalar segmente edilmektedir.

Tez çalışması, farklı bölgelerde bulunan mağazaları ve dolayısıyla da milyonlarca farklı müşterilere ulaşabilen perakende zincirlerinin veri tabanlarındaki verileri özellikle pazarlama faaliyetlerinde etkin bir şekilde nasıl kullanabilecekleri konusunda bir katkı sunmayı amaçlamaktadır. Verinin çok hızlı oluştuğu ve çok kolay bir şekilde ulaşılabilirdiği bu çağda perakende firmaları için gerek mağazalarına ait verilere gerekse de var olan veya var olabilecek müşterilerine ait verilere, Veri Madenciliği sürecini uygulayarak faydalı çıkarsamalarda bulunmak bu çalışmanın temel hedefidir.

Bu amaçla bir süpermarketler zinciri firmasına ait mağazalar üzerine araştırma yapılmış olup, Veri Madenciliği yöntemlerinden olan Kümeleme Analizi ile firmanın mağazaları segmente edilmiş, elde edilen segmentler için de pazarlama faaliyetlerinde kullanılabilecek sık satılan ve birlikte satılan ürünlerin tespiti için Birliktelik Kuralları Analizi uygulanmıştır.

Anahtar Kelimeler: Veri Madenciliği, Kümeleme Analizi, Birliktelik Kuralları Analizi, Mağaza Segmentasyonu

RETAIL STORE SEGMENTATION WITH RULE BASED AND SOCIOECONOMIC APPROACHES: AN APPLICATION ON A CHAIN COMPANY

Abstract Of PhD Dissertation:

Within the frame of this dissertation a company's supermarket chain stores will be segmented by Cluster Analysis and Association Rules Analysis which are basic methods of Data Mining.

This study aims to give some idea to retail chain companies which have different stores and different customers in different regions, about how to use the data in their databases effectively especially for the marketing efforts. Also the main objective of this thesis is to apply Data Mining process and infer useful information from the data about stores, customers of the stores or potential customers.

In this study a research is done on a supermarket company which has chain stores in Istanbul. Cluster Analysis is used for segmenting the stores and then Association Rules Analysis is applied for each segments for marketing efforts.

Keywords: Data Mining, Cluster Analysis, Association Rules Analysis, Store Segmentation

İÇİNDEKİLER

Sayfa

GİRİŞ.....	1
ÇALIŞMANIN AMACI VE ÖNEMİ	1
ÇALIŞMANIN ÇERÇEVE VE SINIRLARI.....	2
1. VERİ MADENCİLİĞİNDEN VERİ BİLİMİNE	4
1.1 VERİ MADENCİLİĞİ.....	4
1.1.1 Veri Madenciliği Süreci.....	4
1.1.2 Veri Madenciliği Yöntemlerinin Sınıflandırılması.....	5
1.1.3 Veri Madenciliği ve Veri Tabanlarında Bilgi Keşfi	6
1.1.4 Veri Madenciliği, İş Çözümlemesi ve Müşteri İlişkileri Yönetimi	7
1.1.4.1 Veri Madenciliği ve İş Çözümlemesi.....	7
1.1.4.2 Veri Madenciliği ve Müşteri İlişkileri Yönetimi	7
1.2 VERİ MADENCİLİĞİ, BÜYÜK VERİ VE VERİ BİLİMİ İLİŞKİSİ..	9
1.3 VERİ BİLİMİNİN ZAMAN AKIŞI.....	13
2. PAZAR SEGMENTASYONU VE MAĞAZA SEGMENTASYONU	19
2.1 PAZAR SEGMENTASYONU.....	19
2.1.1 Segmentasyon Çalışmalarında Kullanılacak Olan Değişkenler	23
2.1.1.1 Coğrafi ve Demografik Değişkenler	23
2.1.1.2 Psikografik Değişkenler.....	24
2.1.1.3 Davranışsal Değişkenler	25
2.1.2 Segmentasyon Analizinde Kullanılan İstatistiksel Teknikler.....	26
2.1.2.1 Ki-kare Otomatik Etkileşim Detektörü (CHAID).....	27
2.1.2.2 Konjoint Analizi.....	28
2.1.2.3 Regresyon Ağacı.....	29
2.1.2.4 Diskriminant Analizi.....	31

3.1.2.5 Kümeleme Analizi	32
2.2 MAĞAZA SEGMENTASYONU	32
3. KÜMELEME ANALİZİ VE BİRLİKTELİK KURALLARI ANALİZİ....	37
3.1 KÜMELEME ANALİZİNE GENEL BAKIŞ	37
3.2 KÜMELEME ANALİZİ SÜRECİ.....	39
3.2.1 Değişkenlerin Seçimi.....	39
3.2.2 Uzaklık Fonksiyonunun Seçimi.....	40
3.2.2.1 Nümerik Veriler için Uzaklık Ölçüleri: L_p formlarının özel halleri.....	41
3.2.2.2 Nümerik Olmayan Veriler için Benzerlik Ölçüleri.....	42
3.2.2.3 Karma Tipteki Veriler için Uzaklık Ölçüsü	44
3.2.3 Kümeleme Algoritmasının Seçimi	49
3.2.3.1 Hiyerarşik Olmayan Kümeleme Algoritmaları.....	49
3.2.3.1.1 k-Ortalamalar Algoritması.....	50
3.2.3.1.2 k-Medyanlar Algoritması.....	53
3.2.3.1.3 k-Medoids Algoritması	54
3.2.3.2 Hiyerarşik Kümeleme Algoritmaları	55
3.2.3.2.1 Birleştirici Yöntemler (Altan Üste)	56
3.2.3.2.1.1 Tek Bağlantı.....	56
3.2.3.2.1.2 Tam Bağlantı.....	58
3.2.3.2.1.3 Grup Ortalaması ile Bağlantı	60
3.2.3.2.1.4 Minimum Varyans: Ward'ın Yöntemi	61
3.2.3.2.2 Bölücü (Üstten Alta) Yöntemler	64
3.2.3.3 Uzaklık Matrisinin Güncellenmesi	65
3.2.4 Kümelemenin Geçerliliği	66
3.2.4.1 Dışsal Ölçütler	67
3.2.4.2 İçsel Ölçütler.....	68
3.2.4.3 Göreceli Ölçütler.....	69
3.2.5 İki Kümeleme Sonuçlarının Karşılaştırılması	70
3.3 SÜPERMARKETLERDE KÜMELEME ÇALIŞMALARI	75
3.4 BİRLİKTELİK KURALLARI ANALİZLERİ	77

3.4.1 Birliklik Kuralları Analizinin Süpermarket Uygulamaları	81
3.4.2 Kuralların Kümelenmesi Çalışmaları	84
4. ZİNCİR PERAKENDE MAĞAZALARI İÇİN MAĞAZA SEGMENTASYONU VE HEDEF PAZARLAMA UYGULAMASI	86
4.1 AMAÇ VE KAPSAM.....	86
4.2 VERİ KÜMESİ VE VERİ ÖN İŞLEME SÜRECİ.....	86
4.3 MAĞAZA SEGMENTASYONU UYGULAMASI	91
4.3.1 Sosyoekonomik Bazlı Segmentasyon Yaklaşımı Uygulaması	91
4.3.2 Kural Bazlı Segmentasyon Yaklaşımı Uygulaması.....	101
4.3.3 İki Yaklaşımın Karşılaştırılması.....	107
4.4 PAZARLAMA STRATEJİLERİ İÇİN SIK ÖĞE KÜME MADENCİLİĞİ VE BİRLİKTELİK KURALLARI MADENCİLİĞİ	110
4.4.1 Sosyoekonomik Bazlı Segmentlere Ait Analizler	110
4.4.2 Kural Bazlı Segmentlere Ait Analizler.....	117
4.5 Tartışma ve Öneriler	122
SONUÇ	124
KAYNAKÇA.....	126
EKLER	141

TABLolar

Sayfa No

Tablo 1 Büyük Veri ile Küçük Verinin Karşılaştırılması	11
Tablo 2 Büyük Veriye Kadar Kullanılan Teknikler	13
Tablo 3 Bazı Uzaklık Ölçüleri ve Formülleri	42
Tablo 4 Örnek Kontenjans Tablosu (2×2 boyutlu).....	43
Tablo 5 İkili Tipteki Değişkenler için Uzaklık Ölçüleri	44
Tablo 6 Asimetrik İkili Tipteki Değişkenler için Uzaklık Ölçüleri	45
Tablo 7 Lance ve Williams'ın Hiyerarşik Kümeleme Algoritmaları için Formülleri	67
Tablo 8 İkili Karşılaştırma için Kontenjans Tablosu Sembolleri	73
Tablo 9 Uygulamada Kullanılacak Olan Değişkenler ve Beş Mağazaya Ait Gözlem Değerleri	87
Tablo 10 Uygulamada Kullanılacak Olan Değişkenlerin Tipleri	89
Tablo 11 Ürünlerin Sınıflandırılması.....	91
Tablo 12 Değişkenlerin Segmentler Bazındaki Ortalama Değerleri	95
Tablo 13 Segmentlerin Özellikleri.....	96
Tablo 14 Segmentasyon için Kullanılacak Kural Örnekleri	103
Tablo 15 İki Yaklaşımla Elde Edilen Kümelere Ait Eşleştirme Tablosu	107
Tablo 16 Sosyoekonomik Bazlı Segmentlere Özel Sık Öğe Kümeleri Tablosu	111
Tablo 17 Sosyoekonomik Segmentlere Özel İlişkiler Tablosu (Güven Değeri Bazlı). 113	
Tablo 18 Sosyoekonomik Segmentlere Özel İlişkiler Tablosu (Kaldıraç Değeri Bazlı)	115
Tablo 19 Kural Bazlı Segmentlere Özel Sık Öğe Kümeleri Tablosu	117
Tablo 20 Kural Bazlı Segmentlere Özel İlişkiler Tablosu (Güven değerine göre)	118
Tablo 21 Kural Bazlı Segmentlere Özel İlişkiler Tablosu (Kaldıraç değerlerine göre) 120	

ŞEKİLLER

Sayfa No

Şekil 1: Veri Madenciliği ve Diğer Disiplinlerin Bütünleşmesi.....	5
Şekil 2: Genel Veri Madenciliği Süreci.....	6
Şekil 3: Kümeleme Analizinde Kullanılan Değişkenler.....	33
Şekil 4: Perakende Firmalarının Segmentasyondan Ne Gibi Faydalar Sağladıkları	34
Şekil 5: Kümeleme Prosedürü	40
Şekil 6: Bazı Uzaklıkların İki Boyutta Gösterimi.....	43
Şekil 7: Dendrogram Örneği.....	56
Şekil 8: Tek bağlantı	57
Şekil 9: Tam bağlantı.....	59
Şekil 10: Grup Ortalaması Bağlantı.....	60
Şekil 11: Sosyoekonomik Bazlı Segmentasyon Yaklaşımıyla Elde Edilen Dendrogram	94
Şekil 12: İstanbul Genelindeki 175 Mağazanın Ait Oldukları Segment Sembolleriyle Gösterilmiş Konumları	97
Şekil 13: Asya Yakasında Mağazaların Sembolleriyle Dağılımı	98
Şekil 14: Avrupa Yakasında Yer Alan Mağazaların Sembolleriyle Dağılımı.....	99
Şekil 15: Farklı Kaynaklardan Elde Edilmiş Verilere Dayalı Segmentasyon Yaklaşımı için Yol Haritası.....	100
Şekil 16: Kural Bazlı Segmentasyon Yaklaşımı İçin Yol Haritası	105
Şekil 17: İkinci Yaklaşımla Elde Edilen Kümeleme Analizine Ait Dendrogram	106

GİRİŞ

Tez çalışmamızda yapılacak olan uygulamanın amacı, çerçevesi ve sınırları bu bölümde yer alacaktır.

ÇALIŞMANIN AMACI VE ÖNEMİ

Çalışmamız, çok sayıda farklı tipteki müşterilere ulaşabilen perakende işletmelerinin veri tabanlarından ve dış kaynaklardan elde edebilecekleri verileri, Veri Madenciliği tekniklerinin de yardımıyla özellikle pazarlama süreçlerinde nasıl kullanabilecekleri hakkında bir fikir sunmayı amaçlamaktadır.

Zincir mağazalara sahip olan bir perakende firmasının farklı yerleşim yerlerinde yer alan birçok satış mağazası ve milyonlarca ifade edilebilecek müşterisi olabileceğinden, hem bu müşterilerin hem de mağazaların bulunduğu bölgelerin farklı özelliklere sahip olabileceği de kaçınılmaz bir unsurdur. Bu bakımdan firmanın pazarlama süreçleri için kararlar verirken genel olarak mağazaların bütününe değil, aksine bölgelere veya belirli mağaza gruplarına yönelik stratejiler geliştirmesi gerekebilmektedir.

Örneğin, İstanbul'un çeşitli semtlerinde mağazaları olan bir perakende işletmecisi pazarlama faaliyetleri geliştirirken şehrin kozmopolit yapısından ötürü farklı semtlerdeki müşterilerinin veya potansiyel müşterilerinin birbirlerinden farklı satın alma davranışları sergileyebileceğini göz önünde bulundurmalıdır. Bu bakımdan firmanın, veri tabanlarında tuttuğu satış verilerini, müşteri ile ilgili demografik özellikleri içeren verileri, mağaza ile ilgili olan satış alanı, ciro, kâr gibi verileri, ayrıca mağazaların bulunduğu bölge ile ilgili sosyoekonomik verileri doğru tekniklerle ve verimli bir şekilde nasıl analiz edebileceğine, sonrasında bu analizlerden nasıl faydalı çıkarımlarda bulunabileceğine önem vermesi gerekmektedir.

Firma için bu yaklaşım sayesinde, tespit edilen farklı müşteri profillerine farklı hizmet anlayışı geliştirilerek müşteri memnuniyetinin yükseltilmesini sağlamak ve dolaylı olarak hem mevcut müşterilerden daha fazla gelir elde etmek hem de potansiyel müşteri grubunu gerçek müşteri haline getirerek pazar payını ve kârlılığını arttırmak mümkün olacaktır.

Daha önce, perakende firmaları ile ilgili yürütülen pazarlama çalışmalarının çoğunda, firmaların sahip olduğu müşterilere ait bazı veriler (demografik veriler, satış verileri vb.) yardımıyla müşteri segmentasyonu uygulamaları yapılmıştır. Hem müşteri ile ilgili hem de mağaza ve mağaza bölgesi ile ilgili verilerin birlikte kullanımı yoluyla yapılmış olan segmentasyon çalışmaları nadirdir. Farklı kaynaklardan elde edilebilecek bu gibi verilerin kullanımı, zincir mağazalara sahip firmalar için yapılmış az sayıdaki “mağaza segmentasyonu” çalışmalarında karşımıza çıkmaktadır.

Tez çalışmamızın ana fikrini, zincir mağazalara sahip perakende firmalarının mağazalarını gerek sosyoekonomik faktörleri kullanarak gerekse de müşterilerin satın alma (tüketim) davranışlarını kullanarak segmentlere ayırabileceği oluşturmaktadır. Bu iki yaklaşımla elde edilen sonuçların uyumluluk gösterip göstermediği de araştırılacaktır. Neticede firmanın müşterilerine kitle pazarlama stratejilerini uygulamak yerine, günümüz rekabet şartlarında artık zorunlu bir hale gelen “hedef pazarlama” stratejilerini uygulayabilmesi hedefi gerçekleştirilebilir.

ÇALIŞMANIN ÇERÇEVE VE SINIRLARI

Çalışma, ulusal bir süpermarketler zincirinin, İstanbul ilindeki her bir mağazasına ait satışlarının yer aldığı veri tabanından rastlantısal olarak elde edilen bir tam günlük örnek veri kümesi üzerinde uygulanacak ve Veri Madenciliğinin Kümeleme ve Birliktelik Kuralları Analizi teknikleriyle sınırlandırılacaktır. Bu kapsamda, kümeleme analizi algoritmalarından olan Ward’ın tekniği, birliktelik kuralları analizi algoritmalarından da Apriori algoritması uygulanmaya çalışılacaktır.

Tez çalışmasında kullanılan satış verileri Türkiye’nin birçok ilinde faaliyet gösteren süpermarketler zincirine sahip bir firmaya aittir. Firma, gizlilik ilkeleri sebebiyle sadece İstanbul ilinde faaliyet gösteren 175 farklı mağazasına ait rassal olarak seçilmiş 1 günlük satış verilerinin kullanımına izin vermiştir. Dolayısıyla analizlerde aynı firmanın İstanbul ilinin farklı semtlerinde faaliyet gösteren 175 mağazasından alışveriş yapmış toplam 149 bin 402 müşteriye ait 598 bin 410 kayıt kullanılmıştır. Bu rakamlardan hareketle her bir alışveriş sepetinde ortalama dört ürün olduğu söylenebilir. Bahsi geçen bir günlük veri kümesinde, 23 bin 910 farklı ürün satılmıştır. Fakat, ürünler gruplara sınıflandırılarak tekrar kodlanmış, bu kodlama sonucunda, 103 farklı ürün grubu elde

edilmiş ve analizlerde bu 103 farklı ürün grubu ile çalışılmıştır. Ayrıca, her bir mağaza bölgesine ait bazı demografik ve sosyoekonomik değişkenlerden elde edilen verilere de çalışmada yer verilmiştir.

Uygulamada söz konusu veri kümesinin bilgisayar yardımıyla analizleri için, *Microsoft Excel 2010* ve *R Programlama Dili*'nden faydalanılmıştır.



1. VERİ MADENCİLİĞİNDEN VERİ BİLİMİNE

Geleneksel veri analizi gelişen veri saklama teknolojisi, bilgisayar teknolojisi ve ayrıca bilim ve teknikteki diğer gelişmelerle birlikte zamanla Veri Madenciliği, İş Çözümlemeleri, Büyük Veri, Veri Bilimi gibi isimler de almaya başladı. Günümüzde her ne kadar bu terimler aynı şeyi ifade ediyormuş gibi düşünülüyorsa da, bazı noktalarda kesişmelerine rağmen birbirinden farklıdırlar. Bu bölümde bu konulara sırasıyla değinilecek ve aralarındaki farklar ile sahip oldukları ortak özelliklerden kısaca bahsedilecektir.

1.1 VERİ MADENCİLİĞİ

Veri Madenciliği klasik istatistiksel çözümlemeden farklı olarak, daha fazla değişken ve daha fazla gözleme sahip veri kümelerini analiz etmek için kullanılmaktadır. Veri Madenciliği sahip olduğu algoritmaları kullanarak karmaşık veri kümelerini analiz edip, faydalı bilgiler ve modeller keşfetmeye yarayan bir süreçten oluşur. Bu süreçte Veri Tabanı Sistemleri, Yüksek Performanslı Programlama, Veri Görselleştirme, Görüntü ve Sinyal İşleme, İstatistik, Yapay Öğrenme ve Örüntü Tanıma gibi disiplinler arası kavramlar kullanılır (Ledolter, 2013, s.2; Zaki ve Meira Jr, 2014, s.25).

1.1.1 Veri Madenciliği Süreci

Veri madenciliği daha önce vurgulandığı gibi bir süreçtir ve bu süreç birbiri ile ilişkili birçok adımdan oluşmaktadır. Süreç doğru uygulandığı zaman istenilen sonuçların kalitesi de artar. Veri Madenciliği süreci genel olarak aşağıdaki adımlardan oluşur (Ahlemeyer-Stubbe ve Coleman, 2014, s.19; Ledolter, 2013, s.3; Olson ve Delen, 2008, s.9).

1. İş anlama aşamasında problemin ne olduğu tam olarak anlaşılır.
2. Veri kümesinin ne olduğuna karar verilir. Veri kümesi anlaşılmaya, keşfedilmeye çalışılır.
3. Veri kümesi analizler için hazır hale getirilir. Bu adımda veri ön işleme tabi tutulur.



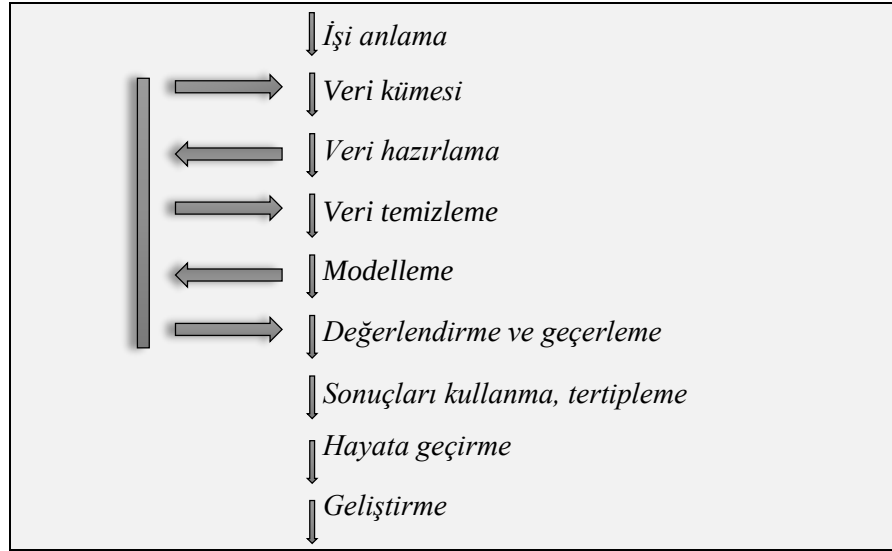
Şekil 1: Veri Madenciliği ve Diğer Disiplinlerin Bütünleşmesi

Kaynak: (Han, Kamber ve Pei, 2011, s.29)

4. Modelleme aşamasında, kullanılacak modelin test edilmesi amacıyla veri kümesi öğrenme ve sınama adı altında iki parçaya bölünür.
5. Değerlendirme aşamasında modelin sonuçları hedefler doğrultusunda değerlendirilir.
6. Son olarak elde edilen sonuçlar uygulamaya geçirilmelidir.

1.1.2 Veri Madenciliği Yöntemlerinin Sınıflandırılması

Bahsedildiği gibi veriden çeşitli analizlerle keşif yapmaya çalışan Veri Madenciliği süreci genel olarak iki veri analizi yönteminden faydalanır. Bunlardan ilki denetimli yöntemlerdir ki bu yöntemler bir veya daha fazla girdi (değişken) ile etki ettiği bir çıktıyı tahmin etmek veya kestirmek için bir model kurmaktadır. Yöntemin isminden de anlaşılacağı üzere, çıktı değişken daha önceden bilindiği için bu değişken adeta süreci bir gözetmen gibi yönlendirir. Denetimsiz yöntemlerde ise, herhangi bir gözetmen değişken olmaksızın, bir hedef değişkeni ölçülmez ve girdi olarak sadece veri kümesi sisteme tanıtılır. Bu yöntemde yapılan genel olarak, veri kümesinin özelliklerinin incelenmesidir (Ahlemeyer-Stubbe ve Coleman, 2014, s.17; Alpaydın, s.15; Hastie, Tibshirani ve Friedman, 2009, s.9).



Şekil 2: Genel Veri Madenciliği Süreci

Kaynak: (Ahlemeyer-Stubbe ve Coleman, 2014, s.19)

Denetimli ve denetimsiz tekniklere, yaygın kullanılan bazı örnekler şu şekilde verilebilir.

- ✓ **Denetimli Teknikler:** Regresyon, Lineer Diskriminant Analizi, Karar Ağaçları, Destek Destek Makineleri, Yapay Sinir Ağları
- ✓ **Denetimsiz Teknikler:** Kümeleme, Temel Bileşenler Analizi, İlişki Analizleri, Rassal Ormanlar, Kolektif Öğrenme

Denetimli ve denetimsiz birçok tekniği kullanan Veri Madenciliğinin dolayısıyla uygulama alanı da çok geniştir. Veri Madenciliği, pazarlama faaliyetlerinden her alandaki sahtekârlık tespitine, üretim kontrolünden sağlık bilimleri, astrofizik ve ekonomi alanlarına kadar birçok yerde uygulanmaktadır (Han ve diğerleri, 2011, s.1; James ve diğerleri, 2013, s.1). Bu çalışmada daha önce belirtildiği gibi, Veri Madenciliği tekniklerinin pazarlama uygulamalarında kullanılmasına temas edilecektir.

1.1.3 Veri Madenciliği ve Veri Tabanlarında Bilgi Keşfi

Veriden daha önce bilinmeyen ve potansiyel olarak faydalı olabilecek bilginin çekilmesi olarak tanımlanan Veri Tabanlarından Bilgi Keşfi (KDD) süreci genellikle yedi adımdan oluşmaktadır. Bu keşifte Veri Madenciliği, sürecin yedi aşamasından sadece bir tanesidir. Dolayısıyla KDD Veri Madenciliğine göre çok daha geniş bir alandır. Fakat

günümüzde Veri Madenciliği terimi KDD'den daha popüler hale gelmiş ve KDD'nin adımları artık Veri Madenciliği tarafından da kapsanmaktadır (Fayyad, Piatetsky-Shapiro ve Smyth, 1996; Han ve diğerleri, 2011, s.7).

1.1.4 Veri Madenciliği, İş Çözümlemesi ve Müşteri İlişkileri Yönetimi

Bu başlık altında birbirleri ile bağlantılı olan ve bazen birbirleri ile karıştırılan Veri Madenciliği, İş Çözümlemesi ve Müşteri İlişkileri Yönetimi arasındaki benzerlikler ve farklılıklar incelenecektir.

1.1.4.1 Veri Madenciliği ve İş Çözümlemesi

İş Çözümlemesi ile Veri Madenciliği arasındaki ilişkinin, İş Zekâsı kavramından tespit edilebileceğini söylemek mümkündür. Kısaca iş zekâsı kavramı; veri ve bilgi kaynaklarını kullanarak değerli bilginin ortaya çıkarılmasını sağlayan süreç ve yöntemlere verilen isim olduğu gibi, bu işi yapan araçlara da (yazılım vb.) atfedilen bir isimdir. İş Çözümlemesi ise İş Zekâsının temel bileşenlerinden sadece bir tanesidir. Veri Madenciliği teknikleri ise İş Çözümlemelerinde kullanılmakta, fakat ayrıca Metin Madenciliği, Ağ Madenciliği ve diğer karmaşık matematiksel/istatistiksel yöntemler de yine iş çözümlemesi kavramının içerisinde yer almaktadır (Demiray s.62). Dolayısıyla Veri Madenciliği teknikleri, İş Çözümlemelerinde kullanılan tekniklerden bir kaçını oluşturmaktadır.

1.1.4.2 Veri Madenciliği ve Müşteri İlişkileri Yönetimi

Hakkında birçok kitap ve makale yazılmış, hemen hemen bütün kurumsal firmaların halen uygulamaya çalıştığı Müşteri İlişkileri Yönetimi (MİY), müşteriler ve pazarlarla uzun dönemli ve kârlı bir ilişki kurmak ve devam ettirmek için ihtiyaç duyulan, insan ve teknolojiyle alakalı süreçlerin bütününe verilen addır (Chen ve Popovich, 2003; Kotler ve Armstrong, 2004, s.16). Dolayısıyla pazarlama alanı ile doğrudan bağlantılı olan MİY bir şekilde firmanın sahip olduğu müşteri veri tabanından karar verme süreçlerinde kendisine destek olabilecek faydalı bilgileri elde etmelidir. İşte bu noktada Veri Madenciliği devreye girer.

Firmaların veri tabanlarından bilgi keşfetmeye yarayan uygun Veri Madenciliği teknikleri MİY kararları için adeta biçilmiş bir kaftan konumundadır (Alex, Stephen ve

Kurt, 2000, s.10). Bunun sebebi ise Veri Madenciliği araçlarının, bir taraftan MİY sürecinde kullanımı zorunlu olan veri tabanları ile diğer taraftan yine MİY kapsamına giren kampanya yönetimine kadar birbirleriyle bağlantılı olmalarındandır (Rygielski, Wang ve Yen, 2002, s.494).

Örneğin MİY için çok kritik konumda olan ve sonuçları direkt olarak MİY’i etkileyecek olan pazar segmentasyonu, Veri Madenciliği’nin temel araçlarından biri olan kümeleme analizi ile başarılı bir şekilde uygulanabilmektedir (Li, 2005, s.1048; Rygielski ve diğerleri, s.488). Kümeleme algoritmaları ile yapılmış pazar segmentasyonları hakkında sayısız çalışma mevcuttur. Bu çalışmaların bazılarına ilerleyen bölümlerde yer verilecektir.

Bir süpermarketin veya perakende firmasının, veri tabanında tuttuğu satış verileri ve müşteri ile ilgili diğer veriler, müşteri davranışlarının tespit edilmesi açısından ve dolayısıyla müşterilere daha iyi hizmet açısından büyük önem taşımaktadır. Firma, bu kayıtların Veri Madenciliği araçları ile analiz edilmesi sonucunda;

- ❖ Hangi müşterisinin firmaya daha bağlı,
- ❖ Hangisinin firmanın müşteriliğini terk etmek üzere olduğu,
- ❖ Hangi ürünlerin pazarlama faaliyetlerinde kullanılması gerektiği,
- ❖ Yeni şubenin nereye açılması gerektiği,
- ❖ Özel müşterilerin bir sonraki alışverişlerinde hangi ürünleri satın almak isteyecekleri,

gibi önemli sorulara yanıt bulabilecektir (Aiguo, Lanling ve Ping, 2005, s.1083; Berry ve Linoff, 1997, s.5).

Ngai ve diğerleri (2009) yapmış oldukları çalışmada, MİY’de Veri Madenciliği tekniklerinin kullanımını ayrıntılı bir şekilde, daha önce yapılan birçok çalışmayı ele alarak incelemişlerdir. Aşağıda Veri Madenciliği tekniklerinin hangi amaçla kullanıldığına dair bazı örnekler verilmiştir.

- **Sınıflandırma:** Genellikle evet-hayır gibi kategorik çıktıların sınıflandırılması için kullanılır. Örneğin, krediye başvuranların düşük,

orta, yüksek riskli şekilde sınıflandırılması, müşteri davranışlarının sınıflandırılması (Berry ve Linoff, 1997, s.9; Ngai ve diğerleri, 2009).

- **Kestirim (Estimation):** Gelir, boy, yaş gibi sürekli değer alan çıktılarının tahmininde kullanılabilir. Örneğin, bir firmanın müşterisinin kaç çocuklu olduğunun veya gelirinin kestirimi, satış kesirimi (Berry ve Linoff, 1997, s.9; Rygielski ve diğerleri).
- **Öngörü (Prediction):** Bir işin ilerisini kestirme olarak tanımlanan öngörü gelecekteki bir davranışa göre sınıflandırma işi için kullanılabilir. Hangi müşterinin alışveriş yapmayı bırakacağını öngörüsü örnek olarak verilebilir (Berry ve Linoff, 1997, s.9)
- **Görselleştirme:** Karmaşık yapıya sahip olan veri kümeleri görselleştirilerek daha kolay anlaşılır bir hale getirilir. 3 Boyutlu Grafikler, Hiper grafikler örnek olarak verilebilir (Ngai ve diğerleri, 2009).
- **Birliktelik Kuralları Analizi:** Hangi ürünlerin/hizmetlerin birlikte satıldığını tespit eden bir analizdir. En bilinen örneği ise pazar sepet analizidir (Giudici, 2005, s.175).
- **Kümeleme:** Heterojen bir yapıyı homojen alt gruplara bölme işidir. Sınıflandırmadan farklı olarak veri kümesi daha önceden belirlenen sınıflara ayrılmazlar. En yaygın uygulaması pazar veya müşteri segmentasyonudur (Berry ve Linoff, 1997, s.10)

1.2 VERİ MADENCİLİĞİ, BÜYÜK VERİ VE VERİ BİLİMİ İLİŞKİSİ

Bu bölümde Büyük Verinin tanımı, geçmişteki veri kümelerinden farkı, Veri Madenciliği ile ilişkisi ve günümüzdeki ulaştığı nokta ve Veri Bilimi ile ilgili kısaca bilgiler verilecektir.

Son yıllarda teknolojinin hızla gelişmesi ile birlikte adeta veri patlaması yaşanmakta, her geçen saniye, farklı alanlarda oluşan verinin Hacmi (Volume), Hızı (Velocity) ve Çeşitliliği (Variety) giderek artmaktadır. Gartner Şirketinin 2001 yılında Büyük Veri için tanımladığı bu 3V kuralında, hacim her saniyede üretilen veri miktarının büyüklüğüne, değişim hızı yeni verinin oluşturulma hızına ve verinin dolanma hızına, çeşitlilik ise yeni veri formatlarına ve tiplerine işaret eder (Laney, 2001). Bu tanımdan sonra 3V'ye ek olarak ilk önce Verinin Belirsizliği (Veracity) ve daha sonra da Değer de (Value) eklenmiştir ki bu kavramlar veri kümelerindeki belirsizliğin de giderek arttığına ve büyük veriden değer de elde edilebileceğine vurgu yapmaktadır (De Mauro, Greco ve Grimaldi, 2015, s.101).

Gerek internet ortamındaki milyarlarca taranıp eklenmiş veya direk oluşturulmuş belgelerden olsun, gerek internet sayfalarını ziyaret eden kullanıcıların bıraktığı bilgilerden olsun, gerek perakende mağazaları ve bankalar yoluyla oluşmuş işlemlerden olsun, gerekse de telekomünikasyon yoluyla veya akıllı teknolojilerle ortaya çıkan kayıtlardan olsun, hızla bir araya gelen veriler günümüzde petabaytlar, zettabaytlar hatta exabaytlarla ifade edilmektedir.

Örneğin ABD'deki Walmart süpermarketler zinciri firması her saatte bir milyonun üzerinde satış işlemi yapmakta ve yaklaşık 2,5 petabaytlık bir veriyle uğraşmakta, sosyal medya firması Facebook ise her gün yüklenen 250 milyon fotoğrafı işlemekte ve Google firması günlük karşılaştığı on Petabaytlık veri oluşumunu idare etmektedir (Camm ve Cochran, 2016, s.8; Chen ve diğerleri, 2014, s.2; Kitchen ve Lauriault, 2015, s.465). Fakat bu gibi şirketler artık Büyük Veri alanına yatırımlar yaparak, verileri analiz edip, müşterileri ile aralarındaki ilişkileri kuvvetlendirmeye, iş süreçlerini iyileştirmeye ve yeni iş fırsatlarını keşfetmeye çalışmaktadırlar (Dean, 2014, s.3-4; Tien, 2013, s.130).

Büyük Veri, “yeni veri işleme tekniklerini, teknolojilerini gerektiren büyük hacimli, çok hızlı, çok çeşitli ve değerli bilgilerden oluşmaktadır” (Gartner, 2016) tanımına dikkat edildiğinde, verinin nasıl saklanacağı ve nasıl etkin yönetileceğine dair vurgular yapılmıştır. Bunun sebebi geçmişteki geleneksel veri işleme uygulamalarıyla ve eski tip veri tabanı yönetimleriyle günümüzdeki yapısal olmayan verinin üstesinden

gelinememesidir (Camm ve Cochran, 2016, s.8; Chen ve diğerleri, 2014, s.1). Kitchin’in (2015) hazırlamış olduğu büyük verinin geçmişteki veri kümelerinden farklı olan özellikleri tablosu (Tablo 1), bu durumu açıkça ortaya koymaktadır.

Tablo 1 Büyük Veri ile Küçük Verinin Karşılaştırılması

ÖZELLİK	KÜÇÜK VERİ	BÜYÜK VERİ
<i>Hacim</i>	<i>Sınırlı genişlik</i>	<i>Çok geniş</i>
<i>Kapsam</i>	<i>Örnekler</i>	<i>Ana kütlelerin tamamı</i>
<i>Çözünürlük ve gösterim yeteneği</i>	<i>Kaba; zayıftan güçlüye doğru bir tutum</i>	<i>Sıkı ve güçlü</i>
<i>İlişkisellik</i>	<i>Zayıftan güçlüye doğru bir tutum</i>	<i>Güçlü</i>
<i>Hız</i>	<i>Yavaş</i>	<i>Hızlı</i>
<i>Çeşitlilik</i>	<i>Sınırlı genişlik</i>	<i>Çok geniş</i>
<i>Esneklik ve ölçeklenebilirlik</i>	<i>Düşük veya orta</i>	<i>Yüksek</i>

Kaynak: (Kitchin ve Lauriault, 2015, s.464)

Hem akademik çalışmalarda hem de bilgi teknolojileri ile ilgilenen şirketlerin raporlarında büyük verinin tanımı yapılırken bazı farklılıklar göze çarpmaktadır. Kimi tanımlar büyük verinin özelliklerini 4V veya 5V kuralı ile anlatmakta, kimi tanımlar ise büyük verinin özelliklerinin anlatılmasına ek olarak büyük verinin ayrıca “veriyi yönetme işi” de olduğunu vurgulamaktadır.

De Mauro ve diğerleri (2015) büyük veri ile ilgili 1500’den fazla akademik çalışmanın yanı sıra, bu iş ile ilgilenen şirketlerin de raporlarını ve web sitelerini incelemiş ve Büyük Veri tanımında genel olarak, “Bilgi”, “Teknoloji”, “Yöntem” ve “Etki” gibi 4 temanın ön plana çıktığını iddia etmişlerdir. Çalışmada, verinin günümüzde her yerden kolaylıkla elde edilebildiği, kitapların bile artık dijitalleştiği, teknolojinin veriyi toplayıp saklamak için çok müsait olduğu ve veriyi analiz eden yöntemlerin ise zaten eskiden beri var olduğu vurgulanmıştır. Ayrıca, farklı yöntemler kullanılarak büyük verinin analizi yapıldığında elde edilebilecek çıkarımların yapacağı ‘etki’nin, günlük yaşamımızdaki sorunları bile çözebileceği iddia edilmektedir.

Bir veri kümesinden gizli kalmış çeşitli modeller, veri ile ilgili anlamlı özet bilgiler ve ilgi çekici faydalı çıkarsamalar keşfetme süreci olan ve istatistiğin çok daha geniş bir boyutu olarak tanımlanan Veri Madenciliği en genel manada; veri toplama, değişken ortaya çıkarma ve veri temizleme, analitik süreç ve algoritmaların kullanımından oluşan üç ana aşamayı kapsamaktadır (Han ve diğerleri, 2011, s.7; Kantardzic, 2011; Ledolter, 2013, s.1). İlk aşamada, çeşitli veri kaynaklarından veriler bir araya getirilirken, ikinci aşamada veriler veri madenciliği algoritmaları için işlem yapılabilir hale getirilir, son aşamada ise hazır veri için etkili analitik teknikler tasarlanmaktadır (Aggarwal, 2015, s.3).

Büyük Veri ve Veri Madenciliği ilişkisine ilk olarak, Büyük Verinin tanımındaki 4 kavram (bilgi, teknoloji yöntem ve etki) ile yaklaşıldığında Veri Madenciliği de bu kavramlarla örtüşmektedir. Fakat Büyük Veri vurgulandığı gibi verinin elde edilmesi ve yönetimi ile alakalı iken, Veri Madenciliği sahip olduğu teknikleri kullanarak, yönetilmiş veriden faydalı bilgiler elde etme işidir. İlişkiye ikinci olarak, daha sonra anlatılacak olan Veri Madenciliği'nin aşamaları yoluyla bakılırsa, Veri Madenciliği sürecinin veri kümesinin elde edilmesi, işlenmesi vb. aşamaları Büyük Verinin alanı ile örtüşmektedir.

Karmaşık sistemlerin yapısını ortaya çıkarabilecek modeller yaratmaya çalışan ve bu modellerle uygulama yapabilmek için süreci bir sisteme bağlayan bir bilim dalı olarak karşımıza çıkan *Veri Bilimi*'nin ise şimdiye kadar tanımladığımız konuların hepsini; Büyük Veriyi, Veri Madenciliğini hatta Yapay Öğrenmeyi kapsadığı söylenebilir. Dolayısıyla Veri Bilimi demek Büyük Veri demek değildir, çünkü en basit şekliyle düşünüldüğünde; küçük veri de Veri Bilimi'nin bir parçasıdır (Aalst ve Damiani, 2015, s.813).

Davenport (2014), Büyük Veri teriminin kullanımına hangi terimlerin kullanılarak ulaşıldığını Tablo 2'de kısaca açıklamıştır. Daha ayrıntılı bir şekilde, büyük verinin geçmişten günümüze nasıl geliştiği ise bir sonraki bölümde ele alınacaktır.

Tablo 2 Büyük Veriye Kadar Kullanılan Teknikler

TERİM	TARİH	NE İŞE YARAR?
KARAR DESTEK	1970 – 1985	<i>Veri analizinin karar verme süreçlerini desteklemek için kullanılması</i>
YÜRÜTME DESTEK	1980 – 1990	<i>Yöneticiler tarafından verilecek kararları desteklemek amacıyla veri analizinin kullanılması</i>
ÇEVİRİM İÇİ ÇÖZÜMLEME SÜRECİ (OLAP)	1990 – 2000	<i>Çok boyutlu veri tablolarının analiz edilmesi için oluşturulmuş yazılım</i>
İŞ ZEKÂSI ÇÖZÜMLEMELERİ	1989 – 2005	<i>Veriden hareketle verilecek kararları destekleyen ve raporlamayı öne çıkaran araçlar</i>
VERİ MADENCİLİĞİ	2005 – 2010	<i>Karar verebilmek için istatistiksel ve matematiksel analizlere odaklanır</i>
BÜYÜK VERİ	2010 - şimdi	<i>Çok büyük, yapılandırılmamış ve çok hızlı hareket eden veriye odaklanır</i>

Kaynak: (H. Davenport, 2014, s.46)

1.3 VERİ BİLİMİNİN ZAMAN AKIŞI

Büyük Veri ile Veri Madenciliğinin tarihsel gelişimi aşağıda genel olarak verilmiştir (Chen ve diğerleri, 2014, s.6; Dean, 2014, s.5; James ve diğerleri, 2013, s.5; Marr, 2016; Press, 2016).

1958: IBM araştırmacısı Hans Peter Luhn, ilk defa İş Zekası'nı tanımladı.

1962: IBM mühendisi William Dersch, Shoebox isimli bir makineyi 1962 Dünya Fuarında tanıttı. Makine, sayıları ve İngilizce 16 kelimeyi algılayabiliyordu. Bu makine ses tanıma sistemlerinin öncüsüydü.

John W. Tukey “Veri Analizinin Geleceği” isimli bir çalışma yayınladı, bu çalışmasında verinin ve veri analizinin ne kadar önemli olduğuna vurgu yaptı.

1965: ABD Hükümeti, dünyanın ilk veri merkezi için harekete geçti. Amaç 742 milyon vergi kaydını ve 175 milyon parmak izini saklamaktı.

1970: IBM matematikçisi Edgar F Codd, ilişkisel veri tabanını tanıttı. Günümüzde de kullanılan bu modelde bilgiler hiyerarşik düzende saklanıyor ve herkes tarafından istenilen bilgiye kolayca ulaşılabilir.

Veri tabanı makineleri geliştirildi.

1974: Peter Naur “Bilgisayar Yöntemleri Üzerine Bir Araştırma” isimli kitabı yayınladı. Kitap veri işleme yöntemlerini geniş bir şekilde ele alıyordu. Ayrıca kitapta “Veri Bilimi” terimi kullanılmıştı.

1977: Tukey Açıklayıcı Veri Analizi kitabını yayınlayarak, hipotezlerin önerilmesi için daha fazla veriye gereksinim olduğuna vurgu yaptı.

Uluslararası İstatistiksel Programlama Birliği (IASC) kuruldu. Birliğin amacı geleneksel istatistiksel yöntemleri modern bilgisayar bilimi ile ilişkilendirmek ve veriyi bilgiye çevirmektir.

1984: Breiman, Friedman, Olshen and Stone Sınıflandırma ve Regresyon Ağaçları isimli çalışmayı tamamladılar.

1986: Teradata firması, ABD’de, 1 terabaytlık ilk paralel veri tabanı sistemini Kmart süpermarketler zinciri firmasına sattı.

1989: Gregory Piatetsky-Shapiro tarafından ilk Veri Tabanlarında Bilgi Keşfi (KDD) konferansı düzenlendi.

İş Zekâsı modeli kullanılmaya başlandı.

1991: İnternetin doğum günüdür.

1994: Business Week dergisi, Veri Tabanı Pazarlaması isimli bir çalışma yayınladı. Çalışmada firmaların nasıl müşteri bilgilerini topladığı ve bu bilgileri nasıl kullandığı anlatıldı.

1995: Sun firması Java platformunu çıkardı. 1991 yılında keşfedilen Java, C'den sonra en popüler olan dil oldu. Artık Web trafiğini kaydetme ve saklamak için kaynak vardı.

Global Pozisyon Belirleme Sistemi (GPS) tamamen devreye girdi.

1996: Uluslararası Sınıflandırma Birliği Federasyonu (IFCS) Japonya'da buluştu ve "Veri Bilimi" terimi ilk defa konferans isminde yer aldı.

Fayyad, Shapiro ve Smyth tarafından "Veri Madenciliğinden, Veri Tabanlarında Bilgi Keşfine" isimli kitap piyasaya sürüldü. Kitapta Veri Tabanlarında Bilgi Keşfi teriminin Veri Madenciliği, Bilgi Keşfi, Bilgi Çıkarsama, Veri Arkeolojisi ve Veri Örüntüsü İşleme gibi isimlerin hepsini kapsadığı vurgulandı.

1997: Michigan Üniversitesi'ndeki konuşmasında, Profesör Doktor C.F. Jeff Wu istatistik isminin Veri Bilimi olarak, istatistikçilerin isminin ise Veri Bilimcisi olarak değiştirilmesi çağrısında bulundu.

Veri Madenciliği ve Bilgi Keşfi akademik dergisi yayınlandı.

NASA araştırmacıları Michael Cox ve David Ellsworth çalışmalarında Büyük Veri terimini kullandı.

1998: Carlo Strozzi, erişime açık NoSQL isimli bir veri tabanı geliştirdi.

Google, Larry Page ve Sergey Brin tarafından kuruldu.

1999: Jacob Zahavi, Wharton Üniversitesi'ndeki konuşmasında, boyutu gittikçe büyüyen veri için özel veri madenciliği teknikleri geliştirilmesi gerektiğini vurguladı.

Kevin Ashton, nesnelerin interneti (Internet of things) terimini ortaya attı. İnternete bağlı olan cihazların sayısının giderek arttığını ve bu cihazların, arada insan olmadan, birbirleri ile bilgi paylaşılabileceğini; akıllı bir ağ oluşturabilecekleri fikrini savundu.

2001: Gartner firmasından Daug Laney, Büyük Verinin sahip olduđu 3V özelliđini yazdıđı makalede açıkladı.

Leo Breiman “İstatistiksel Modelleme: İki Kültür” isimli çalışmasında, eđer veri analizcilerin işi veriyi kullanarak problemleri çözmek ise, analizcilerin veri modellerinden çok daha farklı yöntemlere ihtiyacı olduğunu belirtti.

Wikipedia kuruldu.

2002: Veri Bilimi akademik dergisi yayına başladı.

Bluetooth piyasaya sunuldu. Bluetooth bilindiđi gibi, veriyi kısa mesafeler içerisinde kablosuz olarak transfer etmeye yarar.

2003: 2003 yılında yapılan bir araştırmaya göre, sadece 2003 yılında oluşan veri, bu yıla kadar oluşan verilerden daha fazlaydı.

İş yaşamı profesyonellerinin kullandığı LinkedIn piyasaya sunuldu. Şu anda 300 milyona yakın kullanıcısı vardır.

Veri Bilimi adlı başka bir akademik dergi yayına başladı.

2004: Wikipedia bir milyon makaleye ulaştı.

Facebook kuruldu, günümüzde 1,5 milyara yakın kullanıcısı vardır.

2005: Apache Hadoop projesi başlatıldı. Bu sayede, bir yazılım kütüphanesinde, büyük veri kümeleri saklanabiliyor ve analiz edilebiliyordu.

ABD’deki Ulusal Bilim Kurulu (NFS) “veri bilimcisi” aranıyor ilanı verdi.

Davenport, Cohen ve Jacobson “Analitik üzerine yarış” isimli bir çalışma yayınladılar. Çalışmada firmaların artık veri analizi konusunda birbirleri ile yarıştıkları belirtiliyordu. Çalışma Harvard Business Review’de de yayınlandı ve daha sonra kapsamı genişletilerek kitaplaştırıldı.

Ulusal Bilim Kurulu (NSB) yayınladığı çalışmada veri bilimcisinin tanımını yaptı.

2007: Bir grup IBM mühendisi Watson adında bir teknoloji piyasaya sürdüler. Amaç, internetteki mevcut çok geniş miktardaki veri, daha zekice kararlar verme konusunda nasıl işe yararlar?

Çin'deki Fudan Üniversitesi'nde Dataoloji ve Veri Bilimi isimli araştırma merkezi kuruldu. Bu merkezde çalışan iki araştırmacı "Dataoloji ve Veri Bilimi" isimli makale yayınladılar. Çalışmada Veri Bilimi'nin, Sosyal Bilimler gibi artık bir bilim dalı olduğunu iddia ettiler.

İlk akıllı telefon Iphone, Apple tarafından piyasaya sürüldü.

2008: İnternete bağlı cihazların sayısı, dünyadaki toplam insan sayısını geçti. Artık dünyadaki sunucular 9.57 zettabaytlık veriyi işlemekteydi.

JISC veri bilimcisinin kariyer gelişimi hakkında bir rapor sundu.

2009: "Dijital verinin gücünün toplum ve bilim için kullanımı" isimli rapor yayınlandı.

Nathan Yau "Veri Bilimcilerinin Yükselişi" çalışmasında, Google'ın yöneticilerinden Hal Varian'ın gelecek on yıl içinde istatistikçilerin en önemli meslek gruplarından biri olacağını söylediğini aktardı.

LinkedIn'de veri bilimcisi grubu kuruldu.

1000 çalışana sahip, ortalama bir ABD şirketi 200 terabaytlık veri stokluyordu.

2010: Kenneth Cukier çalışmasında yeni bir profesyonel meslek dalının ortaya çıktığını, bunun da "Veri Bilimcisi" olduğunu söyledi.

Mike Loukides "Veri Bilimi nedir" isimli çalışmasını yayınladı.

Drew Conway "Veri Biliminin Venn Diyagramı" isimli çalışmasında veri bilimcisinin çok farklı alanlardan bir çok konuya hakim olması gerektiğini yazdı.

2011: IBM'in Watson bilgisayarı, dört terabaytlık veriyi saniyeler içerisinde analiz edebiliyordu.

2012: ABD Devlet Başkanı Obama, Büyük Veri Araştırma ve Geliştirme İnisiyatifi'ni, altı departmanda 84 programla başlattı.

IDC ve EMC, 2013 yılında 2.8 ZB'lık verinin oluşacağını fakat bu verinin sadece %3'ünü büyük veri için kullanılabileceğini ve bundan daha azının analiz edilebileceğini açıkladı.

2013: Verinin demokratikleşmesi başladı. Akıllı telefonlar, tabletler ve Wi-Fi ile, herkes veri oluşturmaktadır. Ayrıca herkes büyük hacimli halka açık verilere ulaşabiliyor ve veriden anlamlı çıkarsamalar yapabilmektedir.

2014: İnsanlar artık, bilgiye ulaşmak için mobil cihazları ev ve işyerlerindeki bilgisayarlardan daha fazla kullanıyordu.

Gartner'in tahminine göre 2015 yılının sonuna kadar 4.9 milyar, 2020 sonuna kadar ise 25 milyar nesnenin internete bağlanacağını bildirdi.

2015: Araştırmalar günde 2,5 kantilyon bayt verinin oluştuğunu iddia etti.

Dünyanın en büyük Büyük Veri firması olan Google'ın günde on milyar gigabayt veriyi sakladığı bildirildi.

2. PAZAR SEGMENTASYONU VE MAĞAZA SEGMENTASYONU

Bu bölümde, segmentasyon, pazar segmentasyonu ve mağaza segmentasyonu konularına değinilip, segmentasyon işi için kullanılacak olan tekniklerden bahsedilecek ve mağaza segmentasyonu alanında yapılmış olan çalışmalardan örnekler verilecektir.

2.1 PAZAR SEGMENTASYONU

Segment kelimesi, büyük bir topluluk içindeki farklı ve tanımlanabilir karakteristik davranışlar sergileyen küçük seçkin bir grubu ifade etmektedir (Doyle, 2011, s.340). Pazar segmentleri ise; bir pazarlama stratejisine gruplar halinde benzer tepki veren insanlardan veya organizasyonlardan oluşmaktadır (Myers, 1996, s.16).

Segmentasyon, Wendell Smith'in (1956, s.5) çığır açan çalışmasında, ekonomistlerin diliyle şöyle tanımlanmıştır: "Segmentasyon, yaptığı etkiyle bütünü parçalarına ayırma yeteneğine sahiptir ve daha önce tek bir olarak algılanan talep planları (demand schedule) ortaya çıkarmaya çalışmaktadır.

William Robinson'un (1938) Kusurlu Rekabet'ine (imperfect competition) dayalı olarak, ürünlerin ve hizmetlerin talebinde heterojenlik olduğunu fark eden Smith, segmentlerin heterojen müşteri isteklerinden direkt olarak elde edilebildiğini söylemektedir. Ayrıca yaptığı pazar segmentasyonu tanımı günümüzün halen en iyi tanımı olarak görülmektedir. Smith'e göre pazar segmentasyonu; müşterilerin daha çok tatmin olmak için tercihlerinde gösterdikleri farklılıklar sebebiyle, heterojen olan pazarları birkaç küçük homojen pazar olarak görmektir (Wedel ve Kamakura, 2012, s.6).

Smith'in çalışmasından sonra pazar segmentasyonu, pazarlama alanında, hem araştırmacılar hem de uygulayıcılar için en önemli ve en çok araştırılan konulardan biri haline gelmiştir. Çünkü Smith artık "Kitle Pazarlama" çağının sonunun geldiğine işaret etmiş, basit bir şekilde bir grup ürün çeşidi sunan firmalarla (hedef homojen pazarlar), özenle seçilmiş ürün çeşitleri sunan firmaların (hedef çoklu heterojen pazar segmentleri) ayrı tutulması gerektiğine kanaat getirmiştir (Doyle, 2011, s.341).

Pazar segmentasyonu için Smith'den sonra da birçok tanımlar yapılmış, ayrıca segmentasyon işinin nasıl yapılacağı hususunda da farklı teknikler ve yöntemler geliştirilmiştir.

Bu bağlamda, Peter Bennett'in (1995, s.165) yaptığı basit ve anlaşılır tanımla ayrıca segmentasyon sürecinin nasıl işleyeceğinin ayrıntılı açıklanması önem arz etmektedir. Bennett pazar segmentasyonunu, bir pazarı, aynı şekilde davranan veya benzer ihtiyaçlar sergileyen müşteri alt gruplarına bölme süreci olarak açıklayıp, muhtemelen her bir alt grubun birer hedef pazar olarak seçilebileceğini ve bu gruplara farklı pazarlama stratejileri uygulanacağını vurgulamaktadır. Segmentasyon sürecini ise şu şekilde sıralayabiliriz (Bennett, 1995, s.165).

- Segmentasyon işinin hangi temelli yapılacağının belirlenmesi; ürün temelli bir uygulama ile müşteri farklılıklarını ortaya çıkarma veya pazarlama değişkenlerine tepki verme derecesini ölçme (satın alma davranışı, ürün/hizmet kullanım şekli, üründen/hizmetten ne tip fayda aradıkları, niyetler, tercihler veya bağımlılık)
- Segmentasyon işini yapabilecek segment belirleyicilerini seçme (demografik özellikler, coğrafik, psikografik özellikler vb.)

Myers ise (1996, s.18), segmentasyon sürecini şu şekilde anlatır.

- Segmentasyon değişkenlerine karar verilir (temel değişkenler adını alır)
- Veri analizi yöntemine karar verilir.
- Segmentleri elde etmek için metodoloji uygulanır.
- Temel değişkenleri ve diğer değişkenleri kullanarak segmentler tanımlanır.
- Hedef segmentler seçilir ve
- Her bir hedef segment için bir pazarlama karması geliştirilir.

İlerleyen bölümlerde, tez çalışmasının uygulama kısmında yapılacak olan segmentasyon çalışmasında kullanılacak olan değişkenlerden, yöntemlerden, elde edilecek segmentlerin özelliklerinden ve yürütülebilecek pazarlama faaliyetlerinden ayrıntılı bir şekilde bahsedilecektir.

Segmentasyon analizi için, yukarıdaki adımları uyguladıktan sonra ortaya çıkacak olan pazar segmentlerinin bazı şartları taşıması gerekli görülmüştür. Sebebi ise, firmanın ancak bu şartları taşıyan segmentlerdeki müşterilerine pazarlama aktivitelerini uygulayabilecek olmasındandır. Şartlara göre; segmentler ölçülebilir olmalı ve segmentlerdeki müşterilerin detaylı karakteristikleri düşük maliyetle tespit edilebilir olmalıdır. Segmentler ayrıca farklı pazarlama aktiviteleri uygulayabilmek için mantıklı olmalı, yani farklı tercihleri olan yeterli sayıdaki müşteri pazarlama kriterleri açısından belirlenebilir olmalıdır (Doyle, 2011, s.341).

Kotler daha bilimsel olarak bir modeldeki pazar segmentlerinin; homojenlik, tutumluluk (parsimony), ve erişilebilirlik gibi üç özelliği taşıması gerektiğini vurgulamaktadır. Bu üç özellik kısaca şöyle açıklanabilir. Homojenlik bir segmentteki potansiyel müşterilerin, bazı değişkenlere vermiş oldukları benzer tepkilerin derecesinin ölçümü iken tutumluluk, segmentasyon işinin her bir potansiyel müşteriye eşsiz bir hedef yapma derecesi ve erişilebilirlik ise, gözlemlenebilen açıklayıcı değişkenlerin segmentleri karakterize edilebilme derecesidir (Lilien ve Kotler, 1983, s.291).

Amacı kısaca pazarları analiz etmek, niş fırsatlar yakalamak ve üst düzey rekabet edebilme pozisyonunu alabilmek olan segmentasyonun, firmalara ne denli önemli avantajlar sağladığı, günümüzde bu iş için yapılan büyük yatırımlardan da anlaşılmaktadır (Weinstein, 2004, s.3). Segmentasyon sayesinde artık, pazarlamacılar üründen ziyade müşteriye odaklanmaya başlamış, müşteri ile diyalogun artması sonucunda da müşterilerin ihtiyaçlarının tespiti kolaylaşmış, böylece daha önce kitle pazarlama sebebiyle bilinmeyen yeni ve kârlı müşteri grupları keşfedilmiştir. Segmentasyon ayrıca, pazarın bundan sonra sadece büyük üreticiler tarafından domine edilemeyeceği müjdesini de beraberinde getirmiştir, çünkü küçük üreticilerin en azından bir tanesi, bazı segmentlerde baskın olabilmektedir (Doyle, 2011, s.342).

Pazar segmentasyonu analizinin şirketlere olan faydalarını şu şekilde dört ana başlıkta toplamak da mümkündür (Weinstein, 2004, s.16).

1. Pazarın ihtiyaçlarını karşılamak için firma yeni ürünler tasarlar.
2. Firma, etkili ve düşük maliyetli promosyon stratejileri gerçekleştirir.

3. Firma, pazar rekabetini, özellikle pazardaki pozisyonunu değerlendirir.
4. Güncel pazarlama stratejilerinin daha detaylı olarak incelenmesi sağlanır.

Segmentasyon, anlaşılacağı üzere firmayı, müşterilerle daha çok ilgilenmeye dolayısıyla pazarı sürekli incelemeye zorlamakta ve bu sayede küresel rekabet ortamında daha canlı olmalarına vesile olmaktadır. Segmentasyon analizinin önemi arttıkça günümüzdeki kapsamı daha da geliştirilmiş ve farklı segmentasyon yaklaşımları ortaya çıkmıştır. Bu farklı tekniklere coğrafik segmentasyon, demografik segmentasyon, marka bağımlılığına göre segmentasyon, bir ürünün kullanım düzeylerine göre segmentasyon, ürünlere göre segmentasyon, fayda segmentasyonu, yaşam tipi segmentasyonu, niş pazar segmentasyonu ve psikografik segmentasyon örnek olarak verilebilir.

Fakat segmentasyon çalışmalarının birçoğu şu iki sınıflandırma yaklaşımını kullanmaktadır (Myers, 1996, s.17).

- Müşteri temelli sınıflandırmaya karşı ürün/hizmet temelli sınıflandırma
- Öncüle karşı ardıl segmentasyon

Müşteri temelli sınıflandırma yaklaşımında, müşterilerin bazı özelliklerine bakılır ve müşterileri birbirinden ayırt edebilecek ve pazarlama planları yapılabilecek özellikler seçilir (demografik özellikler, değerler, ihtiyaçlar vb.). Tersine, ürün temelli yaklaşımda ise bir ürünün veya hizmetin belirli bir fiziksel özelliği ile ilgilenilir (müşterilerin bunlardan ne gibi faydalar istediği, kullanım oranları ve örüntüleri vb.) (Myers, 1996, s.17).

Diğer sınıflandırma planı olan öncül (a priori) segmentasyonda, firma daha önceden bilinen, ürün/hizmet ile ilişkilendirilebilecek bazı faktörleri (demografik özellikler, coğrafik bölge, satın alma yoğunluğu vb.) benzer tüketici grupları için inceler. Çünkü öncül yaklaşımda, müşterilerin karakteristikleri ve ihtiyaçları arasında bir ilişkinin olduğu düşünülür. Tüketici araştırmaları sonucunda segmentasyon için kullanılacak olan değişkenler analiz *öncesinde* belirlenebilir hale gelmiştir, segmentasyon bu değişkenlerle yapılır (Machauer ve Morgner, 2001; Myers, 1996, s.17; Neal, 2016, s.1).

Ardıl segmentasyonda ise segmentler ilk etapta bilinmemektedir, ancak bir grup değişken (ürün özelliği tercihleri, ürün satın alma örüntüleri, marka bağımlılığı, elde

edilen fayda vb.) ile araştırma bittikten sonra ortaya çıkan tepkiler temel alınarak segmentler oluşurlar. Modern yaklaşımlar ardıl segmentasyonu kullanmaktadır. Heterojen bir topluluk araştırılır ve segmentler bu toplulukta homojen tepkiler veren gruplardan oluşur. Bu teknikte genel olarak faktör analizi ve kümeleme analizi kullanılır (Machauer ve Morgner, 2001; Myers, 1996, s.17; Neal, 2016, s.3).

Çalışmamızın uygulama kısmında yapılacak olan segmentasyon analizinde de ardıl segmentasyon (post hoc) yaklaşımı kullanılacaktır.

2.1.1 Segmentasyon Çalışmalarında Kullanılacak Olan Değişkenler

Segmentasyon çalışmalarında, kullanılacak olan değişkenlerin belirlenmesi zor bir süreçtir. Daha önce bahsedilen iki tip sınıflandırma yaklaşımından biri tercih edildikten sonra, segmentasyon değişkenlerinin belirlenmesi gereklidir. Basit segmentasyon çalışmalarında genellikle sadece kullanım oranları ve/veya demografik özellikler temel değişken olmakta fakat bu durum her ne kadar başlangıç için iyi olsa da, yanlış pazarlama kararları vermeye ve kaynakların israfına sebebiyet verebilmektedir (Myers, 1996, s.18; Wind, 1978, s.319). Bu bakımdan değişkenlerin doğru belirlenmesi segmentasyon çalışmasından faydalı geri dönüşler alınabilmesi açısından çok önemlidir.

Lilien ve Kotler (1983, s.291) segmentasyon çalışmalarında kullanılacak olan değişkenleri üç ana başlık altında toplamış ve kısaca her biri için açıklayıcı tanımlamalar yapmıştır.

2.1.1.1 Coğrafik ve Demografik Değişkenler

Coğrafik segmentasyon pazarı farklı bölgelere; milletler, eyaletler, ülkeler, şehirler veya mahallelere bölmekte, demografik ve sosyoekonomik değişkenler ise potansiyel müşterileri sınıflandırmada en çok kullanılan değişkenler olarak karşımıza çıkmaktadır. Örneğin; yaş, cinsiyet, medeni durum, aile büyüklüğü, gelir, eğitim, meslek, din, ırk, milliyet ve sosyal sınıf. Coğrafik, demografik ve sosyoekonomik değişkenler uzun zamandan beri segmentasyon için kullanılan en popüler değişkenler olarak bilinmektedir (Lilien ve Kotler, 1983, s.291).

Bu değişkenleri kullanarak, perakende sektöründe yapılmış olan bazı pazar segmentasyonu çalışmaları aşağıdaki gibi özetlenmiştir.

Çadırcı (2012), moda pazarını segmente etmek amacıyla, İstanbul'daki bazı alışveriş mağazalarından alışveriş yapan 369 kişiye yönelik bir anket çalışması yapmış ve tüketicileri sosyopsikolojik ve demografik özelliklerine göre altı segmente ayırmıştır.

Gonzalaez ve Benito (2005), jeo-demografik segmentasyonun perakendecilerin mağaza yeri belirlemesindeki önemini vurgulamak açısından, İspanya'da faaliyet gösteren bütün süpermarketleri jeo-demografik değişkenleri kullanarak segmentlere ayırmaya çalışmış, bu bağlamda 12769 süpermarketi analiz etmiştir. Süpermarketle ilgili veriler hangi firmaya ait oldukları, adresleri, satış alanları, formatları gibi değişkenleri içermektedir. Ayrıca İspanyol istatistik kurumundan elde ettikleri bölgesel demografik değişkenleri; turizm, doğa, ticaret, işsizlik, aileler, şirketler vb. ile ilgili bölgedeki verileri de kullanmışlardır.

Bayus ve Mehta (1995), dayanıklı tüketim mallarının pazar segmentasyonu hakkında yapmış olduğu çalışmada, beş ürünle alakalı, tüketicilere anket uygulamıştır. Anket sorularında tüketicilerin birçok demografik özellikleri de sorgulanmış ve sonuçta 389 tüketici her ürün için iki segmente ayrılmış ve segmentler pazarlama faaliyetleri için detaylı olarak incelenmiştir.

Segal ve Giacobbe (1994), ABD'deki bir süpermarkete ait on bin müşterinin demografik bilgilerini içeren veri kümesini kümeleme yöntemini kullanarak dört segmente ayırmıştır.

Bai ve diğerleri (2009) çalışmalarında, Çin'deki bir süpermarket müşterilerini, pazarlama faaliyetleri için segmente etmek amacıyla, müşterilerden elde ettikleri demografik verilere göre 20 gruba ayırmıştır.

2.1.1.2 Psikografik Değişkenler

Tüketiciler, yaşam tarzlarına göre veya kişilik farklılıklarına göre farklı gruplara ayrılırlar. Aynı demografik özellikli gruptaki insanlar çoğu zaman psikografik profillerine göre ayrılırlar. Çok hızlı yaşayan kişiler, hippiler, jet sosyete farklı hayat tarzlarına birer örnek olarak verilebilir (Lilien ve Kotler, 1983, s.291). Bu değişkenleri kullanarak yapılmış olan bazı çalışmaların özetleri aşağıdaki gibidir.

Boote (1982), “Avrupa’da Psikografik Segmentasyon” isimli çalışmasında, üç ülkeden toplam 500 bayan katılımcıya yapmış olduğu anket çalışmasında genel olarak katılımcıların kişilik özellikleri sorgulanmış ve katılımcılar iki, üç ve dört segmente ayrılıp incelemiştir.

Ziff (1971), ABD’deki 1600 ev hanımına, yaşam tarzları ile ilgili uygulamış olduğu anket çalışması ile segmentasyon analizi yapmıştır.

Lesakova (2014), emeklilerin yaşam tarzlarını tespit edip, yaşam tarzlarının mağaza seçiminde ne denli etkili olduğunu öğrenmek için, 65 yaşından büyük kişilere yaşam tarzlarına ait soruların da yer aldığı bir anket uygulamış ve 265 kişi, yaşam tarzlarına göre dört segmente ayrılmıştır.

Mehmedovic ve diğerleri (2015), Batı Balkan ülkelerinde, yaşam tarzına göre pazar segmentasyonu ile alakalı bir çalışma yapmak üzere, dört ülkeden 762 katılımcıya yaşam tarzları ile ilgili soruların yer aldığı bir anket çalışması düzenlemiştir. Katılımcılar 3 segmente ayrılmışlardır.

2.1.1.3 Davranışsal Değişkenler

Tüketiciler tavırlarına, farkındalıklarına, durumlarına ve ürüne/ürünün özelliklerine verdikleri tepkilere göre de segmente edilebilirler. Birçok pazarlamacı davranışla ilgili değişkenlerin segmentasyon çalışmasının başlangıç noktası olduğuna inanmaktadır (Lilien ve Kotler, 1983, s.292). Bu değişkenleri kullanarak yapılmış olan çalışmaların özetleri ise aşağıda sunulmuştur.

Cooil ve diğerleri (2008) Türkiye’nin birçok şehrinde faaliyet gösteren Migros süpermarketler zincirinin segmentasyon çalışmalarını durum çalışması olarak değerlendirmiştir. Çalışmada, firmanın yapmış olduğu değer temelli segmentasyon, davranış temelli segmentasyon, yaşam tarzı segmentasyonu ve aktivite seviyesi temelli segmentasyon çalışmalarından örnekler verilmiştir.

Kımıloğlu ve diğerleri (2010) cep telefonu pazarı için, farklı davranış sergileyen müşterileri segmente etmek amacıyla, Türkiye’deki 302 tüketiciye anket çalışması uygulamış ve farklı davranışlar sergileyen dört farklı segment elde etmiştir.

Hjort ve diğerleri (2013) müşterilerin satın alma ve iade davranışlarındaki farklılıkları tespit etmek ve müşterileri segmente etmek amacıyla, Danimarka, Finlandiya, Norveç ve İsveç’te faaliyet gösteren bir e-ticaret firmasından elde ettiği 502 bin 429 sipariş verisini analiz etmiştir.

Bacila ve diğerleri (2013) cep telefonları kullanıcılarını segmente etmek amacıyla, müşterilerin telefon firması ile olan ilişkilerini cep telefonu faturaları ile değerlendirmiş ve telefon operatörünü kullanımlarına göre 9993 abone dokuz gruba ayrılmıştır.

Azizi ve Hosseinabadi, (2015) medikal sektörünün bankalardan beklentilerini belirlemek amacıyla, sağlık sektöründe çalışan 456 kişilik bir anket uygulamış ve analiz sonucu üç segment elde etmiştir.

2.1.2 Segmentasyon Analizinde Kullanılan İstatistiksel Teknikler

Segmentasyon analizini yapabilmek için bazı tekniklere ihtiyaç vardır. Çalışmalarda genel olarak aşağıda sıralanan istatistiksel teknikler, ya direk segmentasyon işini yapmada ya da segmentasyon sürecine destek olarak kullanılmaktadırlar (Lilien ve Kotler, 1983, s.298; Myers, 1996, s.31).

- Ki-kare Otomatik Etkileşim Detektörü (CHAID)
- Konjoint Analizi
- Sınıflandırma ve Regresyon Ağacı (C&RT)
- Diskriminant Analizi
- Kümeleme Analizi

2.1.2.1 Ki-kare Otomatik Etkileşim Dedektörü (CHAID)

CHAID, bir karar ağacı tekniğidir. Literatürde Otomatik Etkileşim Dedektörü (AID), Sınıflandırma ve Regresyon Analizi, Genetik Algoritma gibi farklı tekniklerle benzerlik gösteren bu teknik Gordon Kass (1980) tarafından geliştirilmiştir (Díaz-Pérez ve Bethencourt-Cejas, 2016). CHAID sınıflandırma ve tahmin amacıyla, ayrıca değişkenler arasındaki etkileşimi keşfetmek amacıyla da kullanılmaktadır. Pazarlama alanındaki en sık kullanım amacı ise, müşteri gruplarının belirlenmesi ve müşterilerin bazı değişkenlere verdikleri tepkilerin diğer değişkenleri nasıl etkileyeceğinin tahminidir. CHAID uygulanmadan önce bağımlı değişken ve bağımsız değişkenlerin belirlenmesi gerekir, çünkü segmentasyon işi bu değişkenlerin etkileşimine göre yapılır (Bijak ve Thomas, 2012, s.2436). CHAID kullanılarak yapılmış olan bazı segmentasyon çalışmalarını şu şekilde özetleyebiliriz.

Perzz ve Cejas (2016), turizm pazarını segmente etmek amacıyla, İspanya'daki La Palma adasına gelen 324 turiste ait verileri CHAID tekniğiyle analiz etmiştir. Harcama değişkeni bağımlı değişken olarak kabul edilmiş, CHAID tekniğinin çıktısı olan ağaç diyagramında turistlerin milliyeti segmentasyonu yapabilecek ilk anlamlı değişken olarak belirlenmiştir. Alman ve Hollandalı turistler ortalama harcama miktarının altında harcama yapan bir grup olarak, İspanyol turistler ise ortalamanın üstünde harcama yapan diğer grup olarak belirlenmiştir.

Park ve diğerleri (2013), Kore'de insanların neden internetten alışveriş yapmadıklarını ve internetten alışveriş yapanlarla yapmayanlar arasındaki farklılıkları anlayabilmek amacıyla 4298 kişiyle yapılan bir anket çalışmasını analiz etmiştir. Çalışmanın, katılımcıları gruplara ayırma bölümünde CHAID tekniğini uygulamıştır. “İnternette alışveriş tecrübesi” değişkeni bağımlı değişken kabul edildiğinde “ürün/hizmet ile ilgili bilgilerin öğrenilmesi” değişkeninin en önemli bağımsız değişken olduğu bulunmuştur. İnternette alışveriş yapan katılımcıların % 62,3'ünün bu değişkene olumlu cevap verdiği (internette ürün/hizmet bilgisi öğrenebilme) dolayısıyla internetten alışveriş yapma sebebinin ürün/hizmet hakkında bilgi alabiliyor olma olduğunu tespit etmiştir. İkinci önemli değişkenin ise “kredi kartı kullanımı” olduğu ortaya çıkmıştır. İnternette alışveriş yapmayan katılımcıların %60,2'sinin bu değişkene olumsuz cevap

verdiği (internette kredi kartı kullanmaya sıcak bakmama) tespit edilmiş ve internetten alışveriş yapmama sebebinin “kredi kartı kullanımı” olduğu ortaya çıkmıştır.

Bijak ve Tomas (2012), İngiltere’deki iki bankanın müşteri verilerinden yola çıkarak, kredi skorlaması için CHAID tekniği ile segmentasyon yapmışlardır. Segmentasyon sonucu her bir grup için lojistik regresyon kullanarak skor kartları oluşturmuşlardır. Müşteri ile ilgili birçok değişkenin olduğu veri kümesinde, “bankanın ne kadar yıldır müşterisi oldukları” değişkeni, müşterileri üç yıldan az ve üç yıl ile üç yıldan fazla şeklinde iki gruba ayırmıştır. Üç yıl ve üç yıldan fazla müşteri grubu tekrar “sigortaya sahip olup olmama” değişkeni ile iki gruba ayrılmıştır. Aynı şekilde, sigortası olan müşteriler de aynı adreste 5,2 senden az ve 5,2 seneden fazla yaşıyor olmak şeklinde iki gruba bölünmüştür. Segmentasyon sonucunda araştırmacılar, farklı kümelerde yer alan müşteriler için kredi skorları tahminlemesi yapmışlardır.

2.1.2.2 Konjoint Analizi

Pazarlama Araştırmasında sıkça kullanılan konjoint analizi, bir ürün/hizmetin özelliklerine tüketicilerin ne kadar değer atfettiğini ölçebilen bir tekniktir (Nakip, 2006, s.553). Green ve Rao (1971) pazarlama çalışmalarında konjoint analizini kullanmaya başlamışlardır. Konjoint analizi kullanılarak yapılmış olan bazı çalışmalara aşağıda yer verilmiştir.

Djokic ve diğerleri (2013) Sırbistan’daki bir şehirde öğrencilerin yoğurt markası tercihlerine yönelik bir araştırma için, 123 öğrenciyle yaptığı çalışmada, konjoint analizi ve kümeleme analizi birlikte kullanılarak öğrenciler iki gruba ayrılmıştır. Yoğurta ait ilgili değişkenler marka, süt tipi, yağ oranı, meyve tipi vb. olarak belirlenmiştir. Cevaplayıcılar bu değişkenlere 1’den 10’ a kadar tercih puanları vermişlerdir. Anket cevaplayıcılarına ait demografik özellikler de elde edilmiştir. Konjoint analizine göre yoğurt tercihini belirleyen en etkili değişken kullanılan süt tipidir. Kümeleme analizi sonucu iki gruba ayrılmış olan müşteriler için, kullandıkları “süt tipi” değişkeni gruplara göre araştırılmıştır. Ek olarak daha az öneme sahip değişkenler de müşteri grupları için araştırılmış ve gruplar bu analizlere göre karşılaştırılmıştır.

Guo ve diğerkleri (2015) Spa pazarını müşterilerin spa tercihlerine göre segmente etmek amacıyla Hong Kong’da Çin’den gelen 372 turiste bir anket çalışması düzenlemiş ve konjoint analizi ile kümeleme analizini birlikte kullanarak segmentasyon analizini gerçekleştirmişlerdir. Anket çalışmasında, birçok demografik sorunun yanı sıra, spa tercihleri ile ilgili sorular da yer almış, konjoint analizi sonrasında “fiyat” değişkeni ve ardından “rezervasyon kolaylığı” en önemli tercih sebepleri olarak tespit edilmiştir. Hiyerarşik kümeleme ve K-ortalamlar analizi sonucunda ise, hizmet odaklı ziyaretçiler, garanti rezervasyona hassas ziyaretçileri, fiyata hassa ziyaretçiler ve birkaç gün önce rezervasyona hassas ziyaretçiler gibi dört küme elde edilmiştir. Bu dört küme demografik özelliklerine göre ve tercih puanlarına göre tek tek incelenmiştir.

Ceylan (2013), fayda temelli pazar segmentasyonu amacıyla, Uşak ilinde öğrenim gören 427 üniversite öğrencisinin süpermarketlerden bekledikleri faydaların önem düzeyleri konjoint analizi ile belirlenmiş ve bu faydalar kümeleme analizinin girdileri olarak seçilmiştir. Konjoint analizi sonucu en önemli faydalar şunlardır; fiyat, erişim, müşteriye yaklaşım ve ürün çeşidi. Bu faydalara göre yapılan k-ortalamlar kümeleme analizi sonucunda dört küme elde edilmiştir. Yine her küme için en önemli faktörler belirlenmiş; sırasıyla müşteriye yaklaşım, ürün çeşidi ve erişim, fiyat, indirim ve kapanış saatleri faktörleri kümeler için önem arz etmektedir.

2.1.2.3 Regresyon Ağacı

Regresyon Analizi, Karar Ağaçları ile birlikte kullanılarak müşterileri farklı özelliklerine göre gruplara ayırma işine destek olur. Sınıflandırma ve Regresyon tekniği (C&RT) çok yaygın bir şekilde pazarlama alanındaki uygulamalarda kullanılmaktadır. Bazı örnek çalışmalara aşağıda yer verilmiştir.

Timor ve Şimşek (2008), Türkiye’de faaliyet gösteren bir süpermarketler zincirinin 999 müşterisine ait alışveriş kayıtları ve bazı demografik özellikleri analiz etmiştir. Çalışmada müşterilerin en fazla satın aldığı ürünler hedef değişken olarak tanımlanmış ve karar ağaçları kullanılarak müşterilerin satın alma davranışını en çok etkileyen değişkenler belirlenmiştir. Şekerleme ürünleri, unlu mamuller ve meşrubat-kuruyemiş ürünlerinin her üçünü birden alan müşterilerin özelliklerini belirleyebilmek için bu üç ürün hedef değişken olarak belirlenmiş ve C&RT tekniği kullanarak müşteriler

gruplara ayrılmıştır. Alışverişin toplam tutarı ve müşterinin evlilik yılı değişkenleri müşterileri gruplara ayırma konusunda en nitelikli değişkenler olarak bulunmuş ve müşteriler ilk önce toplam alışveriş tutarı 17,77 TL'den düşük olanlar ve yüksek olanlar olarak iki gruba ayrılmıştır. 17,77 TL'den büyük alışveriş tutarına sahip müşteriler de evlilik yılı 22,5 seneden düşük ve büyük olanlar olarak iki gruba ayrılmıştır.

Mehrotra ve Agarwal (2009), Hindistan'daki iki şehirde 400 müşteriye ait anket çalışmasını Sınıflandırma ve Regresyon Ağacı Tekniği (C&RT) ile analiz etmiştir. Amaç, tele pazarlamaya olan olumlu veya olumsuz yaklaşımlara göre, müşterileri iki gruba ayırma ve olumlu yaklaşan müşterilere hedef pazarlama yapılmasını sağlamaktır. Değişkenler önce faktör analizi ile beş faktöre indirgenmiş, bu beş faktöre C&RT tekniği uygulanmıştır. Kullanılan bu teknik ile tele pazarlamaya olumlu ve olumsuz yaklaşanları ayırt edebilen yedi kural elde edilmiştir.

De Ona ve De Ona (2015), İspanya'nın Granada kentindeki toplu taşıma araçlarının hizmet kalitesini analiz etmek amacıyla, müşterilerin demografik özelliklerinin her biri için (farklı segmentler) ve ayrıca bütün pazar için C&RT tekniğini kullanarak müşterileri sınıflandırmıştır. Bu amaçla, toplu taşıma araçlarının temizliği, onlara ulaşımın kolaylığı, ücret, onlar hakkında bilgiye ulaşmanın kolaylığı, araçların geçiş sıklığı vb. kriterler anket yoluyla 3664 müşteri tarafından 11 puan üzerinden puanlanmıştır. Bütün pazar için yapılan C&RT analizinde “bilgi” değişkeninin puan derecesi 6'dan yüksek olduğu zaman, genel hizmet kalitesi müşteriler tarafından “iyi” olarak algılanmaktadır. Bilgi değişkenine verilen puan altı ve altıdan az olduğunda C&RT ağacında ikinci önemli değişken olarak “frekans” değişkeni karşımıza çıkmaktadır ki, bu değişken müşterileri iki gruba (genel kalite iyi ve kötü olarak düşünenler) ayırmaktadır.

Ekinci ve diğerleri (2014), müşteri değerini (customer value) baz alarak, bir Türk bankasındaki müşterileri gruplara ayırmıştır. C&RT tekniğini kullanılan çalışmada, her biri 2500 müşteriden oluşan dört grup örnek müşteriye veri tabanından ulaşılmış ve her birine altı ay boyunca farklı bir promosyon teklif edilmiştir. İlk aşamada, müşteri tabanında yer alan müşterilerin hepsi, üç değişken baz alınarak, C&RT tekniği ile 29 gruba ayrılmıştır. C&RT ağacındaki ilk altı düğüm incelendiğinde, müşterilerin sahip oldukları üç finans ürününün parasal değerini gösteren V1, V2, V3 (sırasıyla, bankanın

müşterilerinin ürünlerinden elde ettiği kâr, müşterilerin kullandıkları hizmetlerden elde edilen kâr ve müşterilerin nakitlerinden elde edilen kâr) değişkenleri düğümleri birbirinden ayırmaktadır. İlk düğüm müşterilerinin ürünlerinden elde ettiği kâr değişkenine göre, 23105 dolardan düşük ve 23105 dolardan yüksek değere sahip olan müşterilere göre iki gruba ayrılmıştır. Çalışmada daha sonra, rastsal olarak, her bir müşteriye promosyonlar uygulanmıştır. Uygulanan promosyonların her bir gruptaki etkisi analiz edilerek, her bir grup için optimal promosyon bulunmaya çalışılmıştır.

3.1.2.4 Diskriminant Analizi

Çok değişkenli istatistik tekniklerinden biri olan Diskriminant Analizi, önceden sınıflandırılmış iki ya da daha fazla grubu birbirinden ayıran faktörleri tespit eden ve grup dışından alınan bir gözlemin hangi grup ile ilişkilendirilebileceğini ortaya çıkaran bir tekniktir (Nakip, 2006, s.477). Diskriminant analizi ile sınıflandırma işi, farklı gruplardaki nesnelerin birbirleri arasındaki varyansın maksimum, aynı sınıftaki nesnelerin birbirleri arasındaki varyansı da minimum yapan bir anlayışa dayanmaktadır (Badea, 2014, s.239). Bu tekniği kullanarak yapılmış olan bazı segmentasyon çalışmalarına aşağıda yer verilmiştir.

Zhiyu ve Congdong (2009), bir gıda firmasının 210 müşterisine ait dokuz değişkenli (satın alınan ürün tipi, ödeme yöntemi, satın alınan ürün miktarı vb.) verileri kullanarak diskriminant analizi ile sınıflandırma yapmıştır. İki değişken (satış sonrası müşteri memnuniyeti ve ürün kalitesine müşterinin tavrı) müşterileri iki gruba sınıflandırması için yeterli ve önemli görülmüştür. Çalışma sonunda, firmanın pazarlama çalışmalarında, satış hizmetleri ve ürün kalitesi üzerinde durması gerektiği vurgulanmıştır.

Nakahara ve Yada (2012), Japonya'daki bir süpermarketten 7000 müşteriye ait verileri, müşterilerin satın alma davranışlarını tespit edebilmek amacıyla analiz etmiştir. Veri kümesi müşterilere ait bazı değişkenleri ve satış verilerini içermektedir. İlk önce geçmiş alışverişlerine göre müşteriler üç gruba ayrılmıştır. Bu gruplarda satın alma açısından yüksek değerli müşteriler ve düşük değerli müşteriler olarak 2'şer sınıf elde edebilmek amacıyla diskriminant analizi uygulanmış ve ziyaret edilen ürün kategorisi

(taze ürünler, donmuş ürünler, süt ürünler vb.) ile mağazada kalma süresi değişkenleri, müşterileri sınıflandırmada en önemli faktörler olarak tespit edilmiştir.

3.1.2.5 Kümeleme Analizi

Çalışmamızda segmentasyon analizi ile ilgili uygulama yapılırken, kümeleme analizi kullanılacağı için bu teknik, 3. Bölümde ayrıntılı olarak incelenecektir.

Çalışmamızın amacı, sayısız defa uygulanmış olan müşteri segmentasyonu çalışmalarının aksine, yine müşteri ile ilgili gerek demografik, gerek sosyoekonomik gerekse de müşterilerin satın alma davranışlarından olsun, birçok değişkenden faydalanarak perakende mağazalarını segmente etmek olduğundan bir sonraki bölümde bu alanda yapılmış olan çalışmalara değinilecektir.

2.2 MAĞAZA SEGMENTASYONU

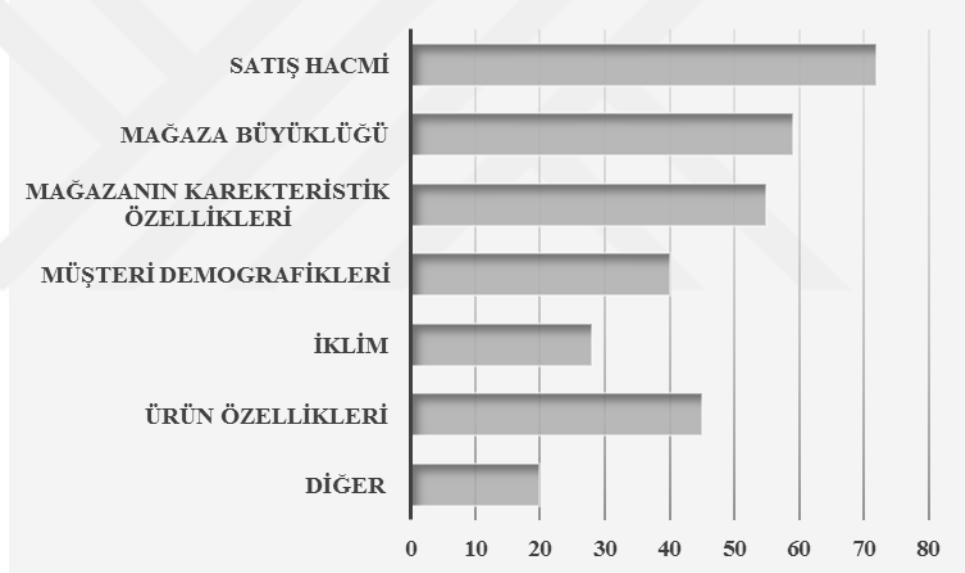
Mağaza segmentasyonu ile ilgili akademik çalışmaların sayısının, yaptığımız araştırmalara göre çok az olduğu saptanmıştır. Yapılan araştırmaya göre erişilebilen çalışmalara bu bölümde kısaca yer verilecektir. Mağaza segmentasyonu Bermingham ve diğerlerinin (2013, s.68) tanımıyla, birbiri ile bağlantılı olan mağazaları anlamlı gruplara ayırmaktır. Örneğin, 350 mağazaya sahip olan bir perakendeci, daha önce belirlenen bazı değişkenler ışığında, bütün mağazalarından sadece altı farklı mağaza grubu oluşturabilir, her bir grupta bazı özelliklere göre birbirine benzeyen mağazalar yer almaktadır.

Birçok mağazası ve milyonlarca müşterisi olan bir perakende firmasının, bu kadar yüksek sayıdaki müşterilerini segmentlere ayırmak yerine, sahip olduğu mağazalarını segmentlere ayırması daha kolay bir iş olarak görünmektedir (Vohra, 2011). Örneğin, müşterilerinin satın alma davranışlarına göre mağazalarını gruplara ayıracak olan bir firma bu sayede sadece müşterileri aynı davranışı sergileyen mağazalarını tespit etmekle kalmayacak aynı zamanda aynı grupta yer alıp da grubundaki mağazalara göre daha düşük veya daha yüksek performans sergileyen mağazalarını da tespit etmiş olacaktır (Bornac, 2015).

Ayrıca kümeleme yoluyla belki yüzlerce mağaza belirli özelliklerdeki birkaç gruba ayrılacağı için firmaların her türlü planlama, satış, hatta satın alma süreçleri daha da kolaylaşacaktır. Mağazaların satış performanslarına göre, büyüklüklerine veya

konseptlerine göre, bulundukları bölgelere veya hedef müşterilere göre (emekliler, öğrenciler vb.) kümelere ayrılması ile beraber firma mağazalarını mikro pazarlama için hazırlamış olacaktır (Donofrio, 2009).

Boston Retail Partners'in (2012), sayısı 500'ün üzerindeki önde gelen ABD'li perakende firmaları ile yapmış olduğu bir araştırmada, mağazalarını kümeleme yoluyla segmentlere ayıran firmalara hangi değişkenleri kümeleme analizi için kullandıkları sorulmuştur. Şekil 3'de görüldüğü gibi satış hacmi değişkeni en sık verilen yanıt olduğundan (katılımcıların %75'e yakını) birinci sıradaki değişkendir. Mağaza büyüklüğü değişkeninin de katılımcıların yaklaşık % 60'ı tarafından kullanıldığı tespit edilmiştir.



Şekil 3: Kümeleme Analizinde Kullanılan Değişkenler

Kaynak: Boston Retail Partners 2012 Raporu, s.12

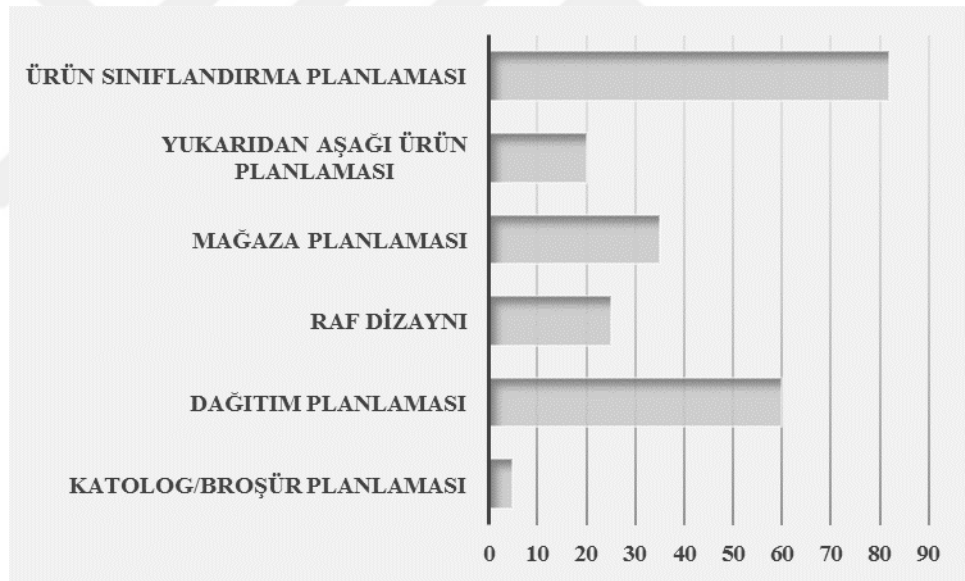
Raporda ayrıca mağazaların kümelenmesinin hangi süreçlere destek sağlayacağı ile ilgili bir soru da sorulmuş, Şekil 4'te görüldüğü gibi, ürün sınıflandırma planlaması sürecine destek en sık yanıt alan süreç iken, dağıtım planlaması süreci ise en sık yanıt alan ikinci sıradaki süreçtir.

Perakende sektöründe faaliyet gösteren şirketlere danışmanlık hizmeti veren birçok kurum, mağaza kümeleme fikrini danışmanı olduğu şirketlere tavsiye etmektedirler. Infoys, Oracle, Category Management Knowledge Group, The Parker

Avery Group, Wilson Perumal & Company, Symphony EYC, Revionics, Pitney Bowes, Data Ventures vb. kurumlar mağaza kümeleme konusunda şirketlere danışmanlık yapmaktadır.

Mağazaların segmentasyonunda kullanılacak olan değişkenler en genel anlamda şu şekilde sınıflandırılmıştır (Wilson Perumal Company, 2013;(Bermingham ve diğerleri, 2013, s.69-70).

- Mağaza ile ilgili: Ciro, kâr, satış alanı, konsept vb.
- Müşteri ile ilgili: Demografik özellikler, satın alma verileri vb.
- Bölge ile ilgili: Bölgede yaşayan insanların demografik özellikleri, iklim koşulları, bölgenin sosyoekonomik özellikleri, rakip mağazaların sayısı vb.



Şekil 4: Perakende Firmalarının Segmentasyondan Ne Gibi Faydalar Sağladıkları
Kaynak: Boston Retail Partners 2012 Raporu, s.12

Mağaza segmentasyonu ile ilgili yapılmış olan bazı akademik çalışmalar aşağıda kısaca özetlenmiştir.

Hawkes ve McLaughlin (1994) çalışmalarında mağazaların geo-demografik kümelenmesi ismini verdiği yöntem ile 78 adet süpermarketi, satış performanslarına göre kümelemiştir. Bu çalışmada amaç ürün bazlı (margarin) satışların artırılmasıdır.

Koehn (2001), Starbucks firmasının mağazalarını kümeleyerek, markasının farkındalığını arttırdığını vurgulamış, fakat firmanın bu uygulamayı nasıl yaptığı bilgisine çalışmasında yer verilmemiştir. Yine Lippmann da (2003) çalışmasında birçok farklı bölgede mağazası bulunan market perakendecilerine mağaza segmentasyonunu çok önemli bir strateji olarak tavsiye etmiş fakat makalede herhangi bir uygulamaya yer verilmemiştir.

Clarke ve diğerleri (2003), ayrıca Bermingham ve diğerleri (2013) mağaza kümeleme tekniğini, perakende firmalarının yer tespiti için karar verme süreçlerinde kullanmışlardır. Her iki çalışmada da MIRSA ismi verilen, perakende firmaları için hazırlanmış erişime kapalı özel bir paket program ile mağazalara ait değişkenlere ait veriler programa girildikten sonra, mağazalar kümelere ayrılmaktadır. Bu programın temelinde bilişsel haritalar (cognitive maps) kullanılmaktadır. Program, şirketlere satılan özel bir program olduğu için içeriği hakkında fazla bilgi verilmemiş, ayrıca çalışmanın uygulama bölümünde analiz edilen firma ve kullanılan değişkenler gizli tutulmuştur. Fakat sadece değişkenler genel olarak; ticaret yapılan bölgenin demografik özellikleri, rekabet, ekonomik durum gibi, açıklanmışlardır. Bu gibi değişkenler bizim çalışmamızda da kullanılacak olan değişkenlerdir. MIRSA programı segmentasyon için kümeleme tekniklerinden K-ortalamlar tekniği kullanılmaktadır.

Mendes ve Cardoso (2006) da outlet mağazalarını kümeleme yöntemiyle, performans değerlendirmesi yapmak ve yer tespit çalışmalarını desteklemek amacıyla kümelere ayırmışlardır. Bu çalışmada ise kullanılan değişkenler açıkça anlatılmıştır. Değişkenler mekân ve mağaza ile ilgili değişkenler (örneğin satış alanı, demografik değişkenler), müşteri tercihleri, tutum ve davranışları ile ilgili değişkenler olarak sınıflara ayrılmıştır. Bu çalışmada kümeleme tekniklerinden olan Ward'ın tekniği kullanılmıştır.

Çalışmada az sayıda mağaza (25 mağaza) çok sayıdaki değişkenlere göre (250 değişken) kümelendiği için, sonuçlar uzman görüşlerine başvurularak değerlendirilmiştir. Çalışmada mağaza müşterilerine anket yapılarak veri toplanmıştır.

Kargari ve Sepehri (2012), otomotiv parçaları üreten bir firmanın sahip olduğu mağazalarını, ulaştırma maliyetlerini azaltmak amacıyla, kümelere ayırmıştır. Bu çalışmada kümeleme tekniklerinden k-ortalamlar tekniği kullanılmıştır. 75 ürün firma için çok kritik kabul edilmiş ve bu ürünlerin satışına göre mağazalar gruplara ayrılmıştır. Son üç yılın verisi 815 mağaza için incelenmiştir.

Bilgiç, Kantardzic ve Çakır (2015), İstanbul ve çevre illerde faaliyet gösteren bir süpermarket zinciri firmasına ait 73 mağazayı hedef pazarlama amaçları için 5 segmente ayırmıştır. Segmentasyon için kullanılan değişkenler mağazaya, mağaza bölgesine ve mağaza bölgesinde yaşayan insanların demografisine ait bazı değişkenlerdir.

Yapılan bu ve benzeri çalışmaların genelinde segmentasyon için hangi değişkenlerin kullanıldığına açıkça yer verilmemiş, uygulama ve sonuç kısımları detaylandırılmamıştır.

3. KÜMELEME ANALİZİ VE BİRLİKTELİK KURALLARI ANALİZİ

Veri Madenciliğinin en çok kullanılan tekniklerinden olan Kümeleme Analizi ve Birliktelik Kuralları Analizi, çalışmamızın uygulama kısmında kullanılacak olan tekniklerden olduğu için, her iki konu da bu bölümde ayrıntılı olarak incelenecektir. Ayrıca, çalışmamızın uygulamasının yapıldığı alan olan süpermarketlerde daha önce yapılmış olan kümeleme analizi uygulamalarına ve yine çalışmamızda kullanılacak segmentasyon yaklaşımlarından biri olan “birliktelik kurallarını kümeleme” yolu ile segmentasyon uygulaması ile ilgili geçmişte yapılmış olan bazı çalışmalara da yer verilecektir.

3.1 KÜMELEME ANALİZİNE GENEL BAKIŞ

Kümelemenin tanımını yapmadan önce sınıflandırma konusuna değinmek yerinde olacaktır. Çünkü kümeleme, sınıflandırmanın özel bir çeşididir (Jain ve Dubes, 1988, s.56). Sınıflandırma aslında, insanoğlunun varoluşundan beri yapageldiği bir iştir. İnsanoğlu en başından beri canlılara (hayvanlar, bitkiler vb.), yiyeceklere (et, süt, ot vb.), hayvanlara (kedi, köpek, at vb.) farklı isimler vererek belki de farkında olmadan dili sınıflandırma işini yapmışlardır. En geniş anlamda, sınıflandırma, büyük bir veri kümesini faydalı bilgileri daha etkin bir şekilde elde edebilmek amacıyla organize etmek için en basit yollardan biridir (Everitt ve Hothorn, 2011, s.1).

Kümeleme, kimi zaman yanlış bir şekilde sınıflandırma ile aynı anlamda kullanılmaktadır. Ancak bu iki yöntem temelde alt veri kümelerinin verilerin kendisinden türetilmesi ya da önceden tanımlanmış olması açısından farklılık göstermektedir (Çakır, 2008, s.32). Daha açık bir ifade ile sınıflandırmada sınıflar daha önceden belirlidir ve yeni gözlemler bu sınıflara atanmaya çalışılırlar. Fakat kümeleme analizinde grup sayısı ile veya grup yapısı ile ilgili bir varsayım yokken, gruplandırma işi gözlemlerin birbirleri ile olan benzerlik (similarity) veya benzeşmezliklerine (dissimilarity) göre yapılır (Johnson ve Wichern, 1992, s.671).

Everitt (1974) bir kümenin tanımını basitçe şöyle yapar: Bir küme, birbirine benzeyen bir takım birimden oluşur. Kümelemenin ise birçok tanımı mevcuttur, yapılan tanımlar genel olarak birbirine benzemektedir. Bazı tanımlamaları şöyle sıralayabiliriz.

Kümeleme veri noktalarını, küme adı verilen doğal gruplara ayırma işidir, öyle ki bir grup içerisindeki veri noktaları birbirine çok benzemekte, farklı kümelerdeki veri noktaları ise mümkün olduğunca birbirine benzememektedir (Zaki ve Meira Jr, 2014, s.28).

Daha teknik bir tanım ise şu şekildedir. Kümeleme, sonlu bir grup nesneye yapılan bir sınıflandırma çeşididir. Nesneler arasındaki ilişki bir yakınlık matrisinde temsil edilirler. Eğer bahsi geçen nesneler d-boyutlu ölçü uzayındaki örnekler veya noktalar ise, yakınlıklar nokta çiftleri arasındaki uzaklıklardır (örneğin Öklidyen uzaklık gibi). Eğer, nokta çiftleri arasında bir uzaklık veya benzerlik olmazsa kümeleme analizi de olmaz. Bunun sebebi benzerlik (yakınlık) matrisinin kümeleme analizini yapabilmek için tek girdi olmasıdır (Jain ve Dubes, 1988, s.55).

Benzerlik matrisi konusu ilerleyen bölümlerde daha ayrıntılı bir biçimde incelenecektir. Bu tanımdan anlaşılacağı üzere, kümeleme analizi ham veri kümesi ile değil, kümelemesi yapılacak gözlem çiftlerinin benzerlik (yakınlık) veya benzeşmezlik (uzaklık) değerlerinin oluşturduğu matris ile yapılmaktadır.

Kümeleme analizi sonucu heterojen olan veri kümesi artık homojen gruplar oluşturmuştur. Bu homojen gruplar oluşturulduğu zaman artık resmin bütününe bakmak yerine, bu küçük gruplara odaklanılabilir. Ayrıca, içerisinde birçok değişken değeri barındıran büyük nesneler yerine, aynı nesnelerin daha az sayıdaki değişken değerleriyle ilgilenilebilir. Bu bakımdan, kümeleme analizi, veri analizinde her zaman veri kümesini açıklayıcı bir yöntem olarak bilinmektedir ve hipotezleri test etmekten ziyade, hipotez üretmeyle ilgilenmektedir (Wedel ve Kamakura, 2012, s.39).

Kümeleme analizi, birçok istatistiksel yöntemin aksine, verinin dağılımı ile ilgili herhangi bir varsayım yapmaya ihtiyaç duymamaktadır (Norusis, 2010, s.361). Analiz sonucunda, çok sayıdaki veri noktasının düşük sayıdaki gruplarda toplanmış olması, verinin daha iyi anlaşılması ve özetlenmesi açısından veri madenciliği uygulamalarına önemli bir şekilde destek olmaktadır. Aggarwal (2015, s.153) kümelemenin bazı önemli uygulamalarını şu şekilde anlatmaktadır.

- **Verinin özetlemesi:** Daha önce belirtildiği gibi, kümeleme problemi, bir çeşit veri özetlemesi olarak düşünülebilir. Zaten birçok veri madenciliği uygulamasında da verinin daha iyi anlaşılması açısından kümeleme analizi sürecin ilk aşamasıdır. Örneğin kümeleme, sınıflandırma modelleri için, birliktelik analizleri için ve aykırı değer analizleri için ilk aşamadır.
- **Müşteri Segmentasyonu:** Benzer müşterilerinin ortak satın alma davranışlarını inceleyebilmek, firmalar tarafından daima arzu edilen bir uygulamadır. Bu uygulama müşteri segmentasyonu ile yapılır. Kümeleme analizinin belki de en çok kullanıldığı alan budur.
- **Sosyal ağlar analizi:** Ağ verisinde, bağlantı ilişkileri ile kümelenen düğüm noktaları, çoğu zaman benzer arkadaş gruplarını veya cemiyetleri işaret etmektedir. Sosyal ağlar analizinde bu tip cemiyetleri veya arkadaş gruplarını saptamak en çok çalışılan konulardan biridir.

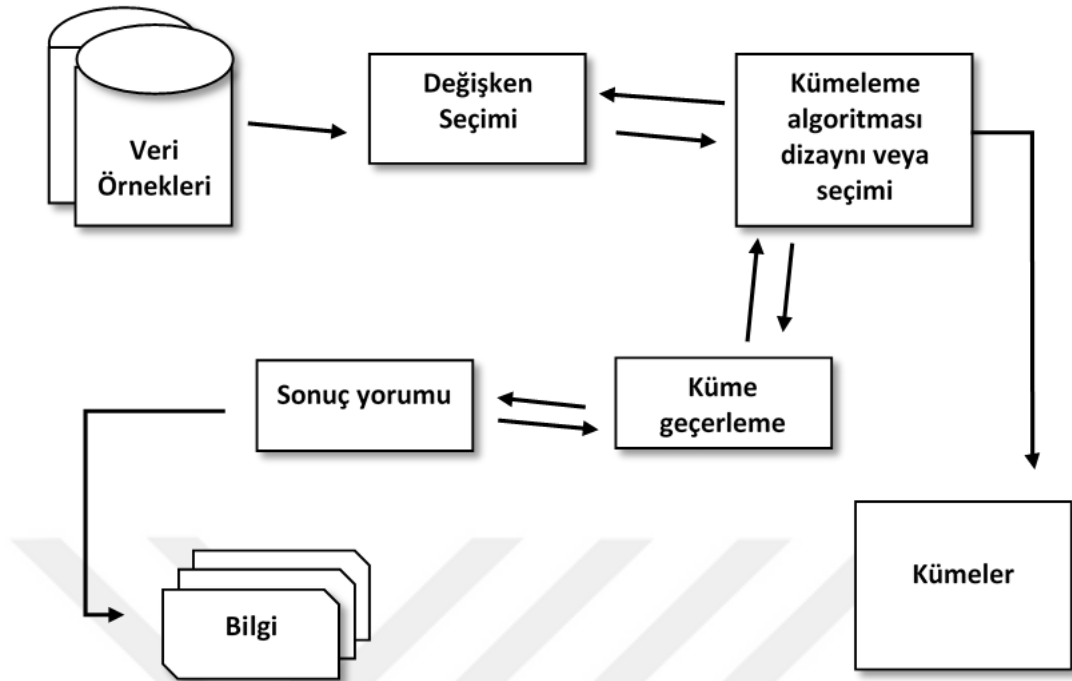
3.2 KÜMELEME ANALİZİ SÜRECİ

Kümeleme analizinin nasıl yapılacağı en genel anlamda; değişkenlerin seçimi, uzaklık fonksiyonu ile kümeleme algoritmasının seçimi, kümelemenin geçerlemesi ve sonuçların yorumlanması şeklinde sıralanabilir. Dört adımdan oluşan aşamalarının her biri birbiriyle bağlantılı olup birbirlerinden beslenirler ve elde edilen kümeleri belirlerler (Xu ve Wunsch, 2009).

3.2.1 Değişkenlerin Seçimi

Bu adım kümeleme analizinde kullanılacak olan, nesneleri (örnekleri) birbirinden ayırt edici özelliği bulunan değişkenlerin seçimidir. Ayırt edici özelliği bulunmayan değişkenlerin seçilmesi, analiz sonucunda herhangi bir kümenin bulunamamasına sebebiyet verebilir. Bu bakımdan değişken seçimi çok önemli bir adımdır.

Uzaklık matrisi hesaplamadan önce, her bir değişken arasındaki korelasyonun değerinin tespit edilmesi gerekmektedir. Bunun sebebi, örneğin yüksek korelasyon değerine sahip iki değişken mevcutsa bu iki değişkenin her ikisinin birden uzaklık matrisi hesabında yer almasının gereksiz olduğudur.



Şekil 5: Kümeleme Prosedürü

Kaynak: (Xu ve Wunsch, 2009, s.6)

Çalışmamızın uygulama bölümünde, kümeleme analizinde kullanılacak olan değişkenlerin seçimi hususunda iki farklı yaklaşım sergilenecektir. Dolayısıyla segmentasyon işi iki farklı yaklaşımla uygulanacaktır. Yaklaşımların ilkinde, kullanılacak olan değişkenler Türkiye İstatistik Kurumu, Google Haritalar ve bazı emlak şirketlerinden, yani farklı kaynaklardan bir araya getirilmiştir. İkinci yaklaşımda kullanılacak olan değişkenler ise, çalışmamızın uygulama bölümü boyunca analizi yapılacak olan bir süpermarketler zincirinin satış verilerinden elde edilecek olan Birliktelik Kurallarıdır.

3.2.2 Uzaklık Fonksiyonunun Seçimi

Uzaklık fonksiyonunu belirleme veri madenciliği uygulamalarında oldukça önem taşıyan bir kavramdır. Çünkü çok yaygın olarak kullanılan kümeleme yöntemleri bir yakınlık (proximity), benzerlik veya ilişki göstergesine ihtiyaç duyarlar. Bu gösterge ham veriden elde edilmektedir (Aggarwal, 2015, s.88).

Yakınlıktan kastedilen literatürde sıkça karşılaşılan benzeşmezlik (dissimilarity) veya benzerliğe (similarity) karşılık gelmektedir (Jain ve Dubes, 1988, s.11). Benzerlik,

iki nesnenin birbirine ne kadar yakın olduğunun ölçüsü iken, benzeşmezlik ise aksine ne kadar uzak olduğunun ölçüsüdür. Benzeşmezlik genel olarak bir uzaklık ölçüsü ile hesaplanmaktadır, Öklidyen uzaklık en bilinen uzaklık ölçüsüdür.

Uzaklık ölçüsü için literatürde Öklidyenden başka farklı uzaklık fonksiyonları da mevcuttur ve uzaklık fonksiyonunun seçiminin kritik olma sebebi, elde edilecek sonuçların kalitesini direkt olarak etkilemesindendir. Uzaklık fonksiyonları için birçok yöntemin karşımıza çıkma nedeni ise, fonksiyonun veri tipine, veri boyutuna ve veri dağılımına oldukça hassas olmasındandır (Aggarwal, 2015, s.88).

Uzaklık ölçülerinin aldığı değerler negatif olamazlar. Eğer birbirine yakın iki nokta ile ilgileniliyorsa bunlar arasındaki uzaklık sıfıra yaklaşırken birbirine uzak iki nokta söz konusu ise uzaklıkları bire yaklaşırlar.

Farklı tipteki değişkenler için farklı yöntemler kullanarak, gözlem değerleri (veri noktaları) arasındaki uzaklıkları bulabiliriz.

3.2.2.1 Nümerik Veriler için Uzaklık Ölçüleri: L_p formlarının özel halleri

Sayısal veriler için bilinen en yaygın uzaklık fonksiyonu L_p - normlarıdır. Sayısal değer almış iki veri noktası $x_i = (x_1 \dots x_n)$ ve $y_i = (y_1 \dots y_n)$ arasındaki L_p -normu aşağıdaki gibi ifade edilir. Bu ifade sayesinde p yerine değerler konularak ilgilenilen iki noktanın arasındaki uzaklık (d) hesaplanır.

$$d(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}} \quad (1)$$

L_p - normlarının en çok bilinen ve kullanılanları Manhattan ($p=1$) ve Öklidyen ($p=2$) ölçüleridir. Öklidyen uzaklık iki veri noktası arasındaki düz çizgi uzaklığı kadardır. Manhattan uzaklık ise, ismini New York'taki Manhattan adasından almıştır. Manhattan'daki iki nokta arasındaki uzaklık yürüme mesafesi ise; Kuzey-Güney ve Doğu-Batı yönlerindeki binaların sayısı bu iki noktanın arasındaki uzaklığa eşittir.

Tablo 3 Bazı Uzaklık Ölçüleri ve Formülleri

<i>Uzaklık ölçüsü</i>	<i>Formül</i>
<i>Manhattan Uzaklığı</i>	$d(x, y) = \sum_{i=1}^n x_i - y_i $
<i>Öklit Uzaklığı</i>	$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$
<i>Tchebyshev Uzaklığı</i>	$d(x, y) = \max_{i=1,2,\dots,n} x_i - y_i $
<i>Minkowski Uzaklığı</i>	$d(x, y) = \sqrt[p]{\sum_{i=1}^n (x_i - y_i)^p}, p > 0$
<i>Canberra Uzaklığı</i>	$d(x, y) = \sum_{i=1}^n \frac{ x_i - y_i }{x_i + y_i}$

Kaynak: (Cios, Pedrycz ve Swiniarski, 2012, s.259)

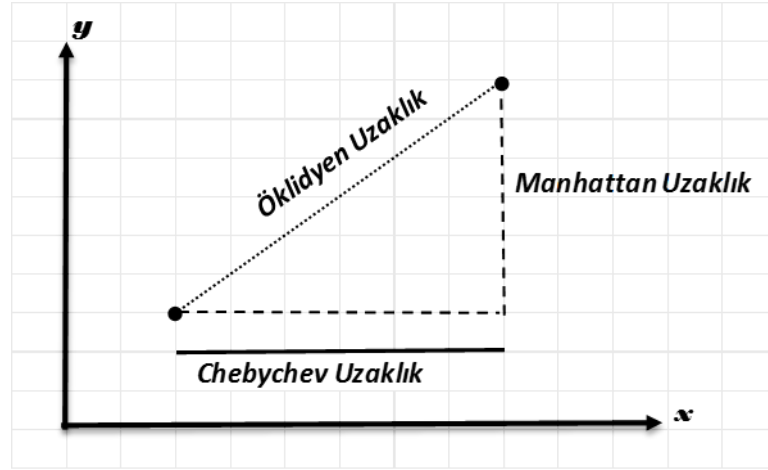
L_p normlarının veri analistleri tarafından en çok kullanılan uzaklık fonksiyonları olmasının sebebi bu ölçülerin kolay anlaşılabilir ve yorumlanabilir olmasıdır (Aggarwal, 2015, s.64).

3.2.2.2 Nümerik Olmayan Veriler için Benzerlik Ölçüleri

Kategorik verilerin kesikli değerlerinde, kategoriler arasındaki uzaklıklar belirsizdir. Dolayısıyla uzaklıkların hesaplanabilmesi için kategoriler ikili değerler halinde nümerik veriye dönüştürülebilir. Fakat oluşacak ikili vektörde birçok sıfır değeri olacağı için, uzaklık fonksiyonları yerine, benzerlik fonksiyonları ile işlem yapmak daha yaygın olan bir iştir.

Daha önce belirtildiği gibi, benzerlik ölçüsü belirli ortak değişkenler için iki nesne arasındaki ilişkinin gücünü göstermektedir. Örneğin elimizde iki kayıt $\bar{X} = (x_1 \dots x_d)$ ve $\bar{Y} = (y_1 \dots y_d)$ olsun. $S(x_i, y_i)$ x_i ve y_i değişken değerlerinin arasındaki benzerlik ise toplam benzerlik;

$$Sim(\bar{X}, \bar{Y}) = \sum_{i=1}^d S(x_i, y_i) \text{ olur.} \quad (2)$$



Şekil 6: Bazı Uzaklıkların İki Boyutta Gösterimi

En basit yol $S(x_i, y_i)$ değerini, $x_i = y_i$ olduğunda 1'e eşitlemek, diğer halde ise 0'a eşitlemektir. Buna örtüşme (overlap) ölçüsü denir (Aggarwal, 2015, s.74). Benzerlik ölçüsü aşağıda ifade edilen aralıkta değerler alabilir.

$$s_{ij} = f(x_i, x_j); -1 \leq s_{ij} \leq 1. \quad (3)$$

Benzeşmezlik ölçüleri benzerlik ölçülerinin tamamlayıcısıdır ve benzerlik katsayısı ile benzeşmezlik katsayısı arasında aşağıdaki şu ilişki mevcuttur.

$$d_{ij} = 1 - s_{ij} \quad (4)$$

Örneğin ikil (binary) değişkenle ölçülmüş gözlem değerleri, sadece iki değer almaktadırlar ($x_{if} = 0$ veya $x_{if} = 1$, i:gözlem, f:değişken) ve bu tip verilerin birbirlerine olan uzaklıklarını hesaplamak için tasarlanmış özel yöntemler vardır (Kaufman ve Roussew, 1990, s.23).

Tablo 4 Örnek Kontenjans Tablosu (2×2 boyutlu)

		nesne j		
		1	0	
nesne i	1	a	b	$a+b$
	0	c	d	$c+d$
		$a+c$	$b+d$	p

Tablo 4’te yer alan a değeri, her iki nesnenin (gözlem değeri) 1 değerini aldığı değişken sayısını göstermektedir, b değeri ise $x_{if} = 1$ ve $x_{jf} = 0$ durumunun olduğu, c değeri ise $x_{if} = 0$ ve $x_{jf} = 1$ durumunun olduğu değişken sayısını göstermektedir (f toplam değişken sayısıdır). d değeri de anlaşılacağı üzere her iki nesnenin 0 değerini aldığı ilgilenilen değişkenlerin sayısıdır. Tablo 5’te bu değişken tipi için kullanılan bazı uzaklık fonksiyonları gösterilmiştir.

Tablo 5 İkili Tipteki Değişkenler için Uzaklık Ölçüleri

Yöntem ismi	$d(i,j)$
<i>Basit Eşleştirme Katsayısı</i>	$\frac{b + c}{a + b + c + d}$
<i>Rogers and Tanimoto’nun Katsayısı (1960)</i>	$\frac{2(b + c)}{(a + d) + 2(b + c)}$
<i>Sokal ve Sneath’in Katsayısı (1963)</i>	$\frac{b + c}{2(a + d) + (b + c)}$

Kaynak: (Kaufman ve Roussew, 1990, s.24)

Bazı durumlarda ikili değişkenlerin her iki kategorisi de aynı öneme sahip olmayabilirler. Örneğin cinsiyet değişkeni için erkek kadın gözlem değerlerinden herhangi birini 0 ile kodlamak veya 1 ile kodlamak önemsizdir. Fakat değeri “var”, “yok” olan bir değişkende “var” değeri araştırmacı için önem arz ediyorsa bu değer 1 olarak kodlanır. Bu özellikteki ikili değişkenler asitmerik olarak adlandırılır ve Jaccard’ın (1908) katsayısı bu tip değişkenler için idealdir. Tablo 6’da bu tip veri için bazı diğer uzaklık fonksiyonları gösterilmiştir (Kaufman ve Roussew, 1990, s.26).

3.2.2.3 Karma Tipteki Veriler için Uzaklık Ölçüsü

Gerçek hayat uygulamalarında yaygın olarak farklı tipteki değişkenlerin aynı veri kümesi içerisinde yer aldığını görürüz. Bu tip verilerle kümeleme analizi yapmanın farklı yolları mevcuttur. Örneğin her bir farklı değişkenle ölçülmüş verilere ayrı ayrı kümeleme yapılabilir. Fakat kümelemelerin sonuçları farklı çıkarsa, bu sonuçları bağdaştırmak zor olabilir. Bu sebeple en iyi yol veri kümesini bir bütün olarak alıp analizi yapmaktır. Bütün değişken tiplerini tek bir yakınlık/uzaklık matrisinde birleştirmek en ideal yol olarak görülmektedir. Log-Olabilirlik Uzaklığı ve Gower’in uzaklığı akla ilk

gelen kullanılabilecek ölçülerdir. Çalışmamızda Gower'in önerdiği uzaklık ölçüsü temel olarak alınacaktır.

Tablo 6 Asimetrik İkili Tipteki Değişkenler için Uzaklık Ölçüleri

Yöntem ismi	$d(i,j)$
<i>Jaccard'ın katsayısı (1908)</i>	$\frac{b + c}{a + b + c}$
<i>Dice (1945) ve Sorenson (1948)</i>	$\frac{b + c}{2a + b + c}$
<i>Sokal ve Sneath (1963)</i>	$\frac{2(b + c)}{a + 2(b + c)}$

Kaynak: (Kaufman ve Rousseeuw, 1990, s.26)

Gower (1971) farklı tipteki değişkenleri barındıran veri kümeleri için (aralıklı, nominal ve ikili değişkenler) uzaklık ölçüsü tasarlanmıştır. Daha sonraki 1990 yılında ise Gower'in formülünü genelleştirilerek sıra ve oran ölçekli değişkenler de hesaplamaya dâhil edilmiştir (Kaufman ve Rousseeuw, 1990, s.32).

Her ne kadar Gower'in tanımı benzerlik katsayısı hesaplasa da daha önce belirtildiği gibi bu katsayı benzeşmezlik katsayısına $d(i,j) = 1 - s(i,j)$ yardımıyla dönüştürülebilir.

Kauffman ve Rousseeuw, (1990, s.35) benzeşmezlik matrisini hesaplamak için karma tipteki değişkenleri nasıl kullanacaklarını şöyle anlatmışlardır: Diyelim ki veri kümemiz p sayısı kadar farklı ölçekle ölçülmüş değişkeni barındırsın. i ve j nesneleri arasındaki benzeşmezlik şöyle bulunur.

$$d(i,j) = \frac{\sum_{f=1}^p \delta_{ij}^{(f)} d_{ij}^{(f)}}{\sum_{f=1}^p \delta_{ij}^{(f)}} \quad (5)$$

Buradaki $\delta_{ij}^{(f)}$, f inci değişken için eğer x_{if} ve x_{jf} kayıp değilse, yani veri kümemizde yer alıyorsa 1'e; herhangi biri kayıpsa 0'a eşittir. Ayrıca f değişkeni asimetrik ikili değişkense ve i ile j nesneleri 0-0 eşleşmesine sahipse, $\delta_{ij}^{(f)}$ yine 0'a eşittir.

Diğer durumlarda ise 1'e eşittir.

Formüldeki $d_{ij}^{(f)}$ ise, f değişkeninin, i ve j arasındaki benzeşmezliğe katkısıdır.

Örneğin, f değişkeni ikili veya nominal değişken ise o halde $d_{ij}^{(f)}$ şu şekilde tanımlanır.

$$d_{ij}^{(f)} = \begin{cases} 1 & \text{eğer } x_{if} \neq x_{jf} \\ 0 & \text{eğer } x_{if} = x_{jf} \end{cases} \quad (6)$$

Diyelim ki, f değişkeni aralıklı ölçekle ölçülmüşse o zaman $d_{ij}^{(f)}$

$$d_{ij}^{(f)} = \frac{|x_{if} - x_{jf}|}{R_f} \quad (7)$$

Buradaki R, f değişkeni için yayılma bandıdır (range).

Eğer f değişkeni asimetrik ikili (binary) ile ölçülmüşse, Jaccard'ın Katsayısı (1908) benzeşmezlik matrisini hesaplamak için uygun olur.

Eğer bütün değişkenler aralıklı ölçülmüşse, Gower'in formülü Manhattan uzaklık formülü halini almaktadır, ilk olarak değişkenler yayılma bantlarına bölünür.

Gower'in bu formülü $d_{ij}^{(f)}$ değerlerini 0 - 1 arasına sınırlamaktadır, yani her bir değişken benzeşmezliğe 0 ile 1 arasında katkı yapmaktadır ve $d(i,j)$ sonucu da 0 ile 1 arasında olmaktadır.

Gower'in uzaklık ölçüsü, birçok alanda kümeleme çalışmaları için kullanılan bir ölçüdür (Pavoine ve diğerleri, 2009). Gower'in ölçüsünü kullanarak yapılmış bazı çalışmalardan kısaca şu şekilde söz edebiliriz.

Everitt ve Merette (1990), karma tipteki verilerin kümelenmesi için kullanılabilecek uzaklık ölçülerini ve kümeleme tekniklerini karşılaştırmış, Gower'in uzaklık ölçüsü de karşılaştırılan dört ölçüden biri olmuştur. Karşılaştırmada dört benzetim verisi kullanılmıştır ve Gower'in ölçüsü, bir veri kümesinde çok başarılı olmuştur. Başarılı olduğu veri kümesindeki nümerik değişkenlerin varyansları çok yüksektir,

Gower'in formülü bu sayısal değişkenlerin hesaplamalarda baskın gelmesine engel olmuştur. Ayrıca Ward'ın hiyerarşik kümeleme yöntemi de diğer hiyerarşik yöntemlere bazı durumlarda üstün gelmiştir. Her ne kadar çıkan sonuçları genelleştirmek mümkün olmasa da, Gower'in uzaklık ölçüsü ile Ward'ın yönteminin birlikte kullanımı genel olarak olumlu sonuçlar vermiş, fakat çalışmada en tatmin edici sonuçlar sonlu karma modelden (finite mixture model) elde edilmiştir. Yine de sonlu karma modelin bu çalışmada sadece bir veya iki kategorik değişkeni içeren veri kümesi için yapıldığı vurgulanmıştır.

Wishart (2003), karma tipteki değişkenleri barındıran veri kümeleri ve eksik gözleme sahip veri kümeleri için, Gower'in uzaklık ölçüsünün kullanılabileceğini vurguladıktan sonra, k-ortalamalar tekniğini tekrar düzenleyerek, tekniğin karma tipteki değişkenler için de uygulanabileceğini iddia etmiştir. Teori ile sınırlandırılan çalışmada, herhangi bir uygulamaya yer verilmemiştir.

Morlini ve Zani (2010), karma tipteki değişkenleri içeren bir veri kümesinin farklı yöntemlerle analiz edilmesi ve sonuçların karşılaştırılması ile ilgili çalışmasında, Gower'in uzaklık ölçüsünü de kullanmıştır. Bir ürüne ait farklı markaların piyasaya sunmuş oldukları modelleri inceleyen çalışmada, kategorik ve nümerik değişkenlerden elde edilmiş gözlem değerleri ile çalışılmış, Gower'in formülünün oluşturduğu uzaklık matrisi k-ortalamalar tekniği ile üç kümeye ayrılmıştır. Gower'in uzaklık ölçüsünün nümerik tipteki değişkenlerin veri kümesindeki ağırlığını azaltabileceğini iddia eden ve farklı uzaklık ölçülerinin ve kümeleme tekniklerinin de analiz edildiği çalışmada, son olarak fazi mantık kullanılarak kümelemenin yapılması önerilmiştir.

Berg (2011) öğrencilerin sportif faaliyetlere karşı tutumlarını belirlemek için, sigara/alkol pazar segmentasyonu analizinden faydalanmış, öğrenciler için anket tasarlanmış, sonuçta Ward'ın kümeleme tekniği ile, dört segment tespit edilmiştir. Çalışmada hem nümerik hem kategorik değişkenler olduğu için Gower'in uzaklık ölçüsü kullanılmıştır.

Cao ve diğerleri (2012) kategorik tipteki değişkenleri içeren veri kümeleri için, k-modlar kümeleme tekniğinin ideal olduğunu önerdikten ve uzaklık matrisinin

hesaplanmasında Gower'in uzaklık ölçüsünün de kullanılabileceğine değindikten sonra, kendilerine ait, kaba kümeler teoremine dayalı yeni bir uzaklık ölçüsü bulmuşlardır. Simülasyonla üretilen veri kümelerine uyguladıkları bu yeni uzaklık ölçüsü başarılı sonuçlar vermiş fakat diğer uzaklık ölçüleri ile karşılaştırılmamıştır.

Ben Ali ve Massmoudi (2013), Gower'in uzaklık ölçüsünü ve diğer uzaklık ölçülerini K-algoritmalar kümeleme tekniğini kullanarak kıyaslamıştır. Birkaç veri kümesi için yapılan bu çalışmada, Gower'in uzaklık ölçüsü ile yapılan kümeleme daha iyi sonuç vermiştir.

Hendrickson (2014), Gower'in formülünü farklı tipteki veriler için genişletmiş ve meteoroloji ile ilgili bir veri kümesi için uygulamasını yapmıştır.

Cebeci ve Kayaalp, gerçek hayattaki birçok veri kümesinin karma tipteki değişkenleri içermesine karşın, kümeleme çalışmalarında bu değişkenlerin birbirlerine dönüştürülerek kullanıldığını vurgulamıştır. Fakat yapılması gerekenin bu olmadığı, değişkenlerin benzerliklerinin tek bir ölçü ile hesaplanması gerektiği iddia edildikten sonra, Gower'in uzaklık ölçüsü incelenmiştir. Örnek uygulamada, R programlama dili ile, farklı tipteki değişken tiplerinden oluşan çok küçük bir veri kümesine hiyerarşik kümeleme analizi uygulanmıştır.

Hennig (2015) yapmış olduğu çalışmada Gower'in yaklaşımını kayıp veriler olduğu durumlar için de geliştirmiştir.

Swenson ve diğerleri (2016), sağlık pazarını segmente etmek amacıyla, ABD'deki bir hastanede hizmet almış 700 hastaya ait demografik veriler ve uygulanan anket yoluyla elde edilen verileri kullanarak kümeleme analizi yapmıştır. Çalışmada uzaklık ölçüsü olarak Gower'in formülü ve kümeleme analizi olarak da k-ortalamlar tekniği kullanılmıştır.

Karma tipteki verileri içeren veri kümelerini kümeleme analizi için girdi haline getirmek, yani uygun bir uzaklık ölçüsü ile uzaklık matrisini elde etmek işi halen araştırılmakta olan bir konudur. Örneğin Cheung ve Jia (2013), He ve Xu (2005), Jourdan

ve diğerleri (2004), Milenova ve Campos (1997), Gower'in ölçüsü dışında farklı uzaklık ölçüleri geliştirerek kümeleme analizini yapmayı denemişlerdir.

3.2.3 Kümeleme Algoritmasının Seçimi

Kümeleme algoritmasının seçimi de uzaklık fonksiyonu ile bağlantılı bir durumdur. Çünkü algoritmaların çoğu bir benzerlik ölçüsüne bağlı olarak çalışmaktadır. Bu adımda benzerlik ölçüsü seçilir ve bir amaç fonksiyonu ışığında kriterler belirlenir. Benzerlik ölçüsü kümeleme işini doğrudan etkileyecektir.

Daha önce belirtildiği gibi, çok boyutlu veri kümelerinin yapısındaki çeşitliliği kümeleyebilecek veya bu veri kümesinden doğru veya en iyi kümeleri ortaya çıkarabilecek araştırmacılar tarafından ortak kabul görmüş bir kümeleme tekniği yoktur (Hennig ve Liao, 2013, s.8; Xu ve Wunsch, 2009, s.7).

Literatürde birçok kümeleme tekniği olduğu gibi, bu teknikler gün geçtikçe geliştirilmekte ve hatta yeni teknikler de bu teknikleri uygulayabilecek yeni yazılımlar ve paket programlarla birlikte ortaya çıkmaktadır. Genel olarak kümeleme teknikleri iki yaklaşım altında toplanmaktadır.

- ✓ **Hiyerarşik olmayan kümeleme**
- ✓ **Hiyerarşik kümeleme**

3.2.3.1 Hiyerarşik Olmayan Kümeleme Algoritmaları

Hiyerarşik olmayan kümeleme tekniklerinde, küme sayısı (k) daha önce araştırmacı tarafından bilinmektedir. Bu tekniklerin en bilinenleri, temsilciye dayalı algoritmalar başlığı altında incelenebilir.

Temsilciye dayalı algoritmaların kümeleme algoritmaları arasında en basit ve yaygın olmasının temel sebebi, bu algoritmaların veri noktalarını kümelemek için direkt olarak uzaklıkları (veya benzerlikleri) temel almalarıdır. Bu algoritmalarda kümeler arasında herhangi bir hiyerarşi söz konusu olmayıp, kümeler tek hamlede oluşturulmaktadır. Bu oluşum, bir grup temsilcinin birleşmesiyle ortaya çıktığı için bu algoritmalar temsilciye dayalı algoritmalar ismini almıştır. Temsilciler belirlendiğinde,

veri noktalarını en yakın temsilcilere atamak için, bir uzaklık fonksiyonu da kullanılabilir (Aggarwal, 2015, s.159).

D veri kümesinin, d boyutlu uzayda, n kadar veri noktası $\bar{X}_1 \dots \bar{X}_n$ içerdiğini düşünelim. Amaç, aşağıdaki amaç fonksiyonu O 'yu minimize edecek, k kadar temsilci $\bar{Y}_1 \dots \bar{Y}_k$ belirlemektir.

$$O = \sum_{i=0}^n [\min_j \text{Dist}(\bar{X}_i, \bar{Y}_j)] \quad (8)$$

Bu şu anlama gelmektedir: Farklı veri noktalarının kendilerine en yakın temsilcilere olan uzaklıkları toplamı minimize edilmelidir. Veri noktalarının temsilcilere atanması, temsilcilerin $\bar{Y}_1 \dots \bar{Y}_k$ seçimine bağlıdır. Temsilciye dayalı algoritmaların genel olarak şu adımları takip ettiği söylenebilir (Aggarwal, 2015, s.160).

- (Atama adımı) Her bir veri noktasını, uzaklık fonksiyonunu $\text{Dist}(.,.)$ kullanarak, kendilerine en yakın olan temsilciye ata ve kümeleri $C_1 \dots C_k$ olarak adlandır.
- (Optimizasyon adımı) Her bir küme C_j için, yerel amaç fonksiyonunu $\sum_{\bar{X}_i \in C_j} [\text{Dist}(\bar{X}_i, \bar{Y}_j)]$ minimize edecek en uygun temsilciyi $\bar{Y}_j$ belirle.

İlerleyen bölümlerde, temsilciye dayalı algoritmalarından olan k-ortalamlar, k-medyanlar ve k-medoids teknikleri incelenecektir.

3.2.3.1.1 k-Ortalamlar Algoritması

Kümeleme analizinde belki de en çok kullanılan tekniktir. Bu teknikte genellikle Öklidyen uzaklık girdi olarak kullanılmaktadır.

k-ortalamlar algoritmasında; veri noktalarının, kendilerine en yakın temsilcilere olan Öklidyen uzaklıklarının karesi, kümelemenin amaç fonksiyonunun değerini belirler. Bu yüzden aşağıdaki ifadeyi yazmak mümkündür (Aggarwal, 2015, s.162)

$$\text{Dist}(\bar{X}_i, \bar{Y}_j) = \|\bar{X}_i - \bar{Y}_j\|_2^2 \quad (9)$$

Burada $\|\cdot\|_p$, L_p normunu temsil etmektedir. $Dist(\bar{X}_i, \bar{Y}_j)$ ifadesi, bir veri noktasının kendisine en yakın olan temsilciye yaklaştırılmasıdır. Böylece, farklı noktalar üzerinden hesaplanan Hata Kareleri Toplamı (HKT) minimize edilir.

Daha geniş bir anlatımla, k-ortalamlar ilk etapta, veri uzayında rastgele k kadar nokta oluşturularak bu noktaları küme ortalamaları olarak alır. Her bir yineleme iki adımdan oluşmaktadır. Birinci adımda küme ataması yapılır, ikinci adımda ise merkez güncellenir. Küme atama aşamasında; örneğin elimizde k küme ortalaması olsun; ilk etapta her bir nokta (gözlem) kendisine en yakın olan ortalama atanır ve böylece kümeler oluşmaya başlar. Her bir kümeyi (C_i) oluşturan noktalar, μ_i 'ye diğer küme ortalamalarına göre daha yakın olan noktalardır. Merkez güncelleme aşamasında ise, her bir kümedeki (C_i ; ($i=1,2,\dots,k$)) noktalar ile yeni ortalamalar hesaplanır. Küme atama ve merkez güncelleme aşamaları, sabit bir noktaya ulaşınca veya yerel minimuma ulaşınca kadar yinelenir durur.

$$HKT(C) = \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - \mu_i\|^2 \quad (10)$$

K-ortalamlar tekniğini kullanarak yapılmış olan bazı segmentasyon çalışmalarına da bu bölümde yer verilmiştir.

Zakrzewska ve Murlewski (2005), bir bankanın müşterilerini segmente etmek amacıyla, bir bankanın müşterilerine ait veri kümesini analiz etmiş ve kümeleme sürecinde K-ortalamlar tekniğini de kullanmıştır.

El Adly (2007), BAE'deki alışveriş merkezlerini ziyaret eden müşterilere, alışveriş merkezlerinin çekiciliği ile ilgili bir anket çalışması uygulamış, 404 müşteri K-ortalamlar tekniği ile, üç segmente ayırmıştır.

Bai ve diğerleri (2009) Çin'deki bir süpermarket müşterilerini segmente etmek amacıyla iki tip segmentasyon yaklaşımı kullanmışlar ve bu yaklaşımları birleştirmişlerdir. İlkinde, demografik verilere göre müşteriler k-ortalamlar tekniği ile 20 gruba ayrılmıştır. İkinci yaklaşımda 13025 satış fişinden oluşan veri kümesi yine K

ortalamlar tekniđi ile 30 gruba ayrılmıř, iki yaklařımın birleřtirilmesi sonucu oluřan veri kümesine uygulanan segmentasyon analizi sonucu 27 segment elde edilmiřtir.

Namvar ve diđerleri (2010), iki ařamalı segmentasyon ismini verdikleri yöntemle, İran'daki bir bankanın müřterilerini bankadan kullandıkları hizmetlere göre, k-ortalamlar tekniđini kullanarak dokuz segmente ayırmıř, ortaya çıkan segmentler demografik özelliklere göre tekrar kümelenmiřtir.

Lefait ve Kechadi (2010) k-ortalamlar tekniđini kullanarak müřterileri tüketim derecelerine göre beř segmente ayırmıřtır. Veri kümesi, 10 bin müřterinin 62 hafta boyunca iki süpermarketten yapmıř oldukları üç markaya ait alıřveriřleri içermektedir.

Ying ve Yuanyuan (2010) řangay'daki bir bankanın müřterilerini k-ortalamlar tekniđini kullanarak, üç segmente ayırmıřtır. 1000 müřterinin demografik bilgilerini ve finansal bilgilerini içeren veri kümesinin oluřturduđu üç segment pazarlama amaçları için ayrıntılı bir řekilde incelenmiřtir.

Li ve arkadaşları (2010) Çin'deki bir bankanın kredi kartı müřterilerini, pazarlama faaliyetleri için k-ortalamlar tekniđi ile dört segmente ayırmıřtır. Kullanılan veri kümesi 65012 kredi kartına ait, müřterilerin demografik ve finansal verilerini de içeren 27 deđiřkenden de oluřmaktadır.

Kımılođlu ve diđerleri (2010) cep telefonu pazarı için, müřteri davranıřlarını segmente etmek amacıyla 302 tüketiciye anket çalıřması uygulamıř, tüketiciler k-ortalamlar yöntemi ile davranıř bakımında dört farklı segmente ayrılmıřlardır.

Memery ve diđerleri (2012) süpermarketten alıř veriř yapan etik müřterilerin tespit etmek amacıyla, İngiltere'deki 220 müřteriye anket uygulamıř, sonuçlar K-ortalamlar ve hiyerarřik kümeleme yöntemleri ile analiz edilmiřtir. k-ortalamlar tekniđini müřterileri farklı dört gruba ayırmıřtır.

Garver ve diđerleri (2014) müřterilerin indirim kuponu tercihleri hakkında yapmıř oldukları çalıřmada, 250 kiřiye anket yapılmıř ve müřteriler k-ortalamlar tekniđi kullanılarak kupon tipi tercihlerine göre üç segmente ayrılmıřlardır.

Chengduan (2015), müşteri segmentasyonu sonuçlarının kalitesinin artması, değerli müşterilerin daha etkin ortaya çıkarılması için bir kümeleme algoritması geliştirmiş ve geliştirdiği algoritmayı k-ortalamlar tekniği ile karşılaştırmıştır. Çalışmada Foodmart veri kümesinin kullanıldığı çalışmada, yazılan algoritma k-ortalamlar tekniğinde farklı açılardan üstün gelmiştir.

Seegebarth ve diğerleri (2016) organik gıda ve müşterilerin değer algısı üzerine yapmış oldukları çalışmada, Almanya ve ABD'deki toplam 446 tüketiciye anket uygulamış, ilk etapta k-ortalamlar tekniği ile daha sonra Ward'ın tekniği kullanılarak tüketiciler dört gruba ayrılmıştır.

3.2.3.1.2 k-Medyanlar Algoritması

k-Medyanlar algoritmasında, amaç fonksiyonu için Manhattan uzaklığı kullanılmaktadır. Bu yüzden, uzaklık fonksiyonu $Dist(\bar{X}_i, \bar{Y}_j)$ şöyledir (Aggarwal, 2015, s.164);

$$Dist(\bar{X}_i, \bar{Y}_j) = \|X_i - Y_j\|_1. \quad (11)$$

Bu durumda, en ideal temsilci $\bar{Y}_j$, C_j 'deki veri noktalarının medyanıdır. Çünkü bir veri kümesinin medyanı bu veri kümesinin bir noktaya uzaklıkları toplamını (L_t) en küçük yapan değerdir.

K-medyanlar, k-ortalamlar tekniğine göre temsilci seçiminde daha tutucu davranmaktadır. Sebebi medyanın kümedeki aykırı değerlere k-ortalamlara göre duyarlı olmamasıdır. K-medyanlar algoritmasının pazarlama alanı ile ilgili uygulamaları hemen hemen yok gibidir, literatürde var olan çalışmalar görüntü işleme ile ilgili çalışmalardır.

Wei ve diğerleri (2004) k-medyanlar tekniğini videoları segmente etmek amacıyla kullanmışlardır. Bir videoyu parça parça ele alan çalışmada, videodaki bir görüntünün homojen bölgelere ayrılması sürecinde k-medyanlar tekniği kullanılmıştır.

Zabih ve Kolmogorov (2004) da k-medyanlar tekniğini, görüntü işleme (image processing) konusuyla ilgili bir uygulama amacıyla bir resme segmentasyon yapmışlardır.

3.2.3.1.3 k-Medoids Algoritması

k-medoids algoritmasında temsilci her zaman veri tabanından seçilmektedir. K-medoids algoritmasının diğer algoritmalara tercih edilmesinin nedenleri şöyle sıralanabilir. Birincisi k-ortalamlar algoritmasındaki temsilcinin o kümede yer alan aykırı bir değer sebebiyle sapmış olabileceği ihtimali ve ikinci sebep ise, karma veri tipinde yer alan veri noktalarının, ortalama veya medyan ile temsilci seçiminin bazen zor olmasıdır. Bu yüzden, k-medoids algoritmasının en önemli özelliği, veri kümesine uygun benzerlik veya uzaklık fonksiyonu tanımlanabildiği sürece, bu algoritmanın her veri tipi için kullanılabiliyor olmasıdır (Aggarwal, 2015, s.165; Zaki ve Meira Jr, 2014, s.333).

k-medyanlar algoritması, ile k-medoids algoritması birbirine karıştırılmamalıdır. k-medoids temsilciyi orijinal veri tabanından seçerken, her bir boyutta bağımsız olarak seçilecek medyanların, d boyuttaki temsilcisi orijinal veri kümesine ait olmayabilir.

k-medoids algoritmasını kullanarak yapılmış olan bazı çalışmalar şöyle sıralanabilir.

Reynolds ve diğerleri (2004) ileriki bölümlerde ayrıntılı anlatılacağı üzere, k-medoids tekniği ile birliktelik kurallarını kümelere ayırmıştır.

Mao ve diğerleri (2005) resimleri segmente etmek amacıyla, Sreedharan ve diğerleri (2014), Khatamai ve diğerleri de (2015) yine görüntü çözünürlüğü ile ilgili bir çalışmada K-medoids yöntemini kullanmışlardır.

Pardeshi ve Toshniwal (2010) k-medoids tekniğinin bazı dezavantajlarını sıraladıktan sonra, oluşturmuş oldukları k-medoids benzeri bir algoritmayı çok bilinen bazı veri kümelerinde uygulamış ve kümeleme açısından başarılı sonuçlar elde etmelerine rağmen, işlem zamanı açısından olumlu sonuç elde edememişlerdir.

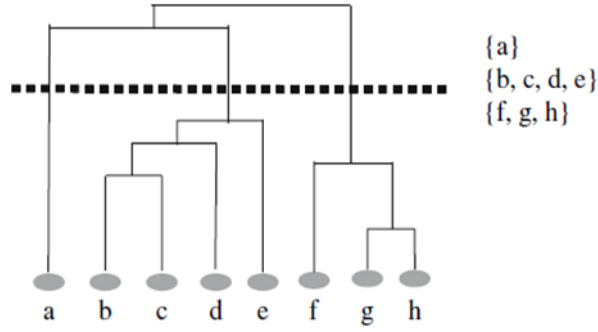
Han ve diğerleri (2014) Çin'deki bir telekomünikasyon firmasının 1000 müşterisine ait 28 günlük web hizmetlerini kullanma verilerini analiz etmiş, müşteriler kullandıkları web hizmetlerine göre, k-medoids tekniği kullanılarak beş segmente ayrılmıştır.

3.2.3.2 Hiyerarşik Kümeleme Algoritmaları

Hiyerarşik algoritmalar veri kümesine uzaklıkların oluşturduğu matrisi kullanarak kümeleme analizi uygular. Hiyerarşik kümeleme yönteminin en önemli faydası, oluşacak olan hiyerarşinin farklı seviyelerinin uygulama açısından farklı bakış açıları sunmasıdır (Aggarwal, 2015, s.166; Everitt ve Hothorn, 2011, s.55).

Hiyerarşik teknikler birleştirici (alttan üste) ve bölücü (üstten alta) yöntemler olarak ikiye ayrılırlar. Birleştirici yaklaşım alttan üste doğru mantığını kullanır, ilk başta n veri noktasının her biri ayrı ayrı küme olarak düşünülür ve yineleme esasına dayanarak, birbirine en benzer parçalar, tek bir küme oluşuncaya kadar birleşir. Bölücü yaklaşım ise birleştirici yaklaşımın aksine üstten alta doğru mantığını kullanır. İlk etapta bütün veri noktalarının oluşturduğu tek bir küme yineleme esasına dayanarak, her bir nokta başlı başına birer küme oluncaya kadar bölünür.

Hiyerarşik kümelemenin amacı, d boyutlu bir uzayda n veri noktası için, bir dizi bölümler oluşturmaktır ve bu bölümler bir ağaç yardımı ile veya dendrogram denilen kümelerin hiyerarşisini gösteren kullanışlı bir şekilde görselleştirilir. Ağacın tabanında her bir nokta ayrı ayrı birer küme oluşturmakta, en üst noktasında ise bütün noktalar birleşerek tek bir küme oluşturmaktadır. Dolayısıyla dendrogram kümeleme analizinde, her bir adımda gerçekleşen birleşmeleri veya bölünmeleri gözler önüne sermektedir (Zaki ve Meira Jr, 2014, s.364). Aşağıdaki şekilde (Şekil 7) a, b, c, d, e, f, g ve h gözlemlerinin nasıl birleşip kümeler oluşturduğu dendrogramla gösterilmiştir.



Şekil 7: Dendrogram Örneği

Kaynak: (Cios ve diğerleri, 2012, s.260)

3.2.3.2.1 Birleştirici Yöntemler (Altan Üste)

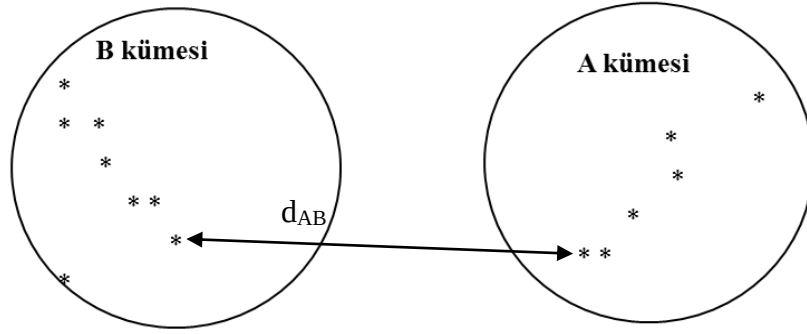
Birleştirici yöntemleri şu şekilde bir örnekle açıklayabiliriz. $C=\{C_1, C_2, \dots, C_m\}$ kümelerinde C_i ve C_j kümeleri birbirlerine en yakın kümeler olsun. Dolayısıyla daha önce belirtildiği gibi, bunlar yeni bir küme $C_{ij} = C_i \cup C_j$ oluştururlar. Sonraki aşamada ise güncelleme yapılması gerektiğinden, C_i ve C_j kümeleri birleştiği için küme setinden kaldırılıp onların yerine C_{ij} eklenmesi gerekir.

Yukarıda anlatılan süreç sadece tek bir küme kalıncaya kadar; C tek bir küme içerinceye kadar devam eder. Eğer daha önceden kullanıcı tarafından küme sayısı k belirlenmişse, birleşme süreci k adet küme oluşunca durur (Zaki ve Meira Jr, 2014, s.366).

Kümelerin birleşebilmesi için, içerisinde birden çok veri noktası barındıran kümeler arasındaki yakınlığı/uzaklığı ölçebilecek bir yöntem ihtiyacı vardır. İşte bu uzaklığı hesaplamak için farklı teknikler geliştirilmiştir. Bunlardan en bilinenleri tek bağlantı, tam bağlantı, ortalama bağlantı ve Ward'ın yöntemi gibi tekniklerle elde edilecek ölçüler ile iki küme arasındaki uzaklık hesaplanabilir.

3.2.3.2.1.1 Tek Bağlantı

Tek bağlantı yöntemi, birleştirici yöntemler arasında en eski ve en basit olanıdır. İlk olarak Florek ve diğerleri (1951) tarafından, daha sonra Sneath (1957b) ve sonraları Johnson (1967) tarafından tanımlanmışlardır. Tek bağlantı yöntemine “en yakın komşu yöntemine” de denir (Everitt ve Hothorn, 2011, s.57).



Şekil 8: Tek bağlantı

Örneğin, birleşecek olan, C_i ve C_j kümeleri arasındaki uzaklık $\delta(C_i, C_j)$ ise, tek bağlantı için tanımlamay; C_i ve C_j kümelerinde birbirine uzaklıkları minimum olan nokta çiftinin uzaklığı şeklinde yapabiliriz.

$$\delta(C_i, C_j) = \min\{\delta(x,y) \mid x \in C_i, y \in C_j\} \quad (12)$$

İki kümede bulunan noktalar arasındaki uzaklıkların minimum olanını seçtiğimizde ve bu noktaları birbirine bağladığımızda, diğer veri noktası çiftleri uzakta kalacağı için, bu kümelerde sadece tek bir bağlantı oluşacaktır, bu yüzden bu bağlantıya tek bağlantı denilmiştir.

İki kümenin birleşmesi gerçekleştikten sonra, uzaklık çiftlerinin oluşturduğu matris güncellenmelidir. i . ve j . satır ve sütunlar matristen silinir ve bunların yerine bu iki kümenin birleşerek oluşturmuş olduğu kümeyi temsil eden yeni bir satır ve sütun yazılır. Yeni satır (sütun) tek bağlantının senaryosunda, matriste bir adım önce silinen satır (sütun) çiftindeki değerlerin en küçüklerini kullanarak hesaplanır. Çünkü belirtildiği gibi, diğer kümelerin, birleşmiş kümeye olan uzaklıkları, diğer kümelerin birleşmiş kümeye minimum olan uzaklıklarıdır (Aggarwal, 2015, s.169).

Bu yöntemin ana noksanlığı şudur. Açıkça birbirinden farklı iki küme herhangi bir noktada birbirlerine yakınlarsa, yöntem bunların ikisini her ne kadar farklı kümeler olsalar da, ayrıymış gibi düşünemez. İki kümenin birleşmesi için sadece bir bağlantı

yeterli gelir ve kümeler birleşmemesi gerekirken birleşirler (Kaufman ve Roussew, 1990, s.226). Bu yöntemle yapılmış olan bazı çalışmalara da bu bölümde yer verilmiştir.

Doyle ve Saunders (1985) endüstriyel pazarların segmentasyonu ile ilgili çalışmalarında tek bağlantı yöntemi kümelerdeki aykırı değerleri tespit etmek için kullanmış, Ward'ın yöntemi ile de segmentasyon yapmıştır.

Mangiameli ve diğerleri (1996) SOM kümeleme tekniğinin hiyerarşik kümeleme tekniklerine üstün olduğunu ispatlamak için, 252 farklı veri kümesi ile çalışmış, çalışmada tek bağlantı yöntemi de kullanılmış ve Som bu tekniğe üstün gelmiştir.

Vakharia ve Mahajan (2000), ortaya çıkardıkları yeni bir kümeleme algoritmasında, değişkenlerin ve gözlemlerin kümelemesi için tek bağlantı, ortalama bağlantı ve k-ortalamalar tekniğini kullanmışlardır.

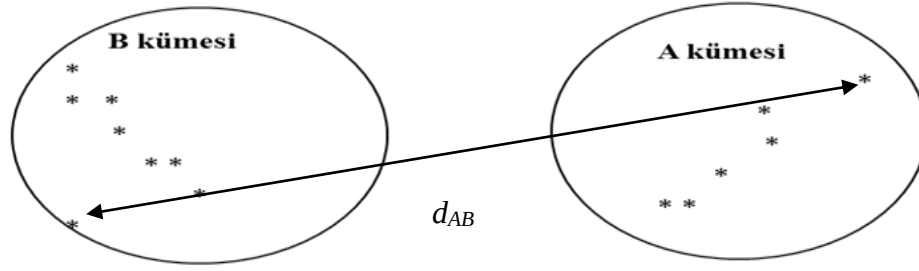
Hung ve Tsai (2008), 227 müşteriye yapmış oldukları anket çalışması ile, kümeleme analizini kullanarak müşterileri segmente etmek istemiş, çalışmada tek bağlantı yöntemi de kullanılmış, yöntemin yetersizliğinden dolayı SOM tekniği kullanılarak segmentasyon yapılmıştır.

Bittman ve Gelbard (2009), yeni bir arabanın fiyatlandırılması konusundaki karar verme sürecine destek olması amacıyla, birçok aracın ve özelliklerinin yer aldığı bir veri kümesinden faydalanmıştır. Araçları kümeleme analizi aşamasında tek bağlantı yöntemi de kullanılmış fakat Ward'ın yöntemi daha iyi sonuç vermiştir.

Devi ve Sharma (2013) Hindistan'da bir bankanın çalışanları üzerinde yapmış oldukları çalışmada, 501 çalışan ile görüşmüş ve çalışanlar iş stresine neden olan faktörlerine göre üç segmente ayrılmış, tek bağlantı yöntemi, aykırı gözlem değerlerinin tespiti amacıyla kullanılmıştır. Segmentasyon ise Ward'ın yöntemi ile yapılmıştır.

3.2.3.2.1.2 Tam Bağlantı

Bu ölçüde, iki küme arasındaki uzaklık; C_i ve C_j kümelerinde birbirine uzaklıkları maksimum olan noktaların uzaklığı olarak tanımlanır. Bu durum daha açık bir ifade ile iki grup arasında birbirine en uzak olan nokta çiftine karşılık gelmektedir.



Şekil 9: Tam bağlantı

$$\delta(C_i, C_j) = \max\{\delta(x,y) \mid x \in C_i, y \in C_j\} \quad (13)$$

İki kümenin birleşmesi gerçekleştikten sonra, uzaklık matrisi, satırların (sütun) en yüksek değerini kullanarak güncellenir.

Mingoti ve Lima (2006) benzetimle oluşturdukları 2530 veri kümesine farklı kümeleme teknikleri uygulamıştır. Çalışmada tek, tam ve ortalama bağlantı ile fazi kümeleme tekniği uygulanmış, fazi tekniği hiyerarşik tekniklere üstün gelmiştir.

Panwar (2008), Hindistan’da bir tür sigara tüketicilerine yaptığı fayda segmentasyonu çalışmasında 1200 kişiye anket uygulamış ve tam bağlantı yöntemini kullanarak, kullanıcıları üç segmente ayırmıştır.

Fonseca (2013) farklı tipteki değişkenlerin oluşturduğu üç veri kümesi ile yaptığı çalışmada, kümeleme yöntemleri ile Gizli Sınıf Modellerini karşılaştırmıştır. Tam bağlantı yönteminin de kullanıldığı çalışmada, kümeleme yöntemlerinin sadece nümerik değişkenleri içeren veri kümelerinde, Gizli Sınıf Modellerine göre daha başarılı olduğu ortaya çıkmıştır.

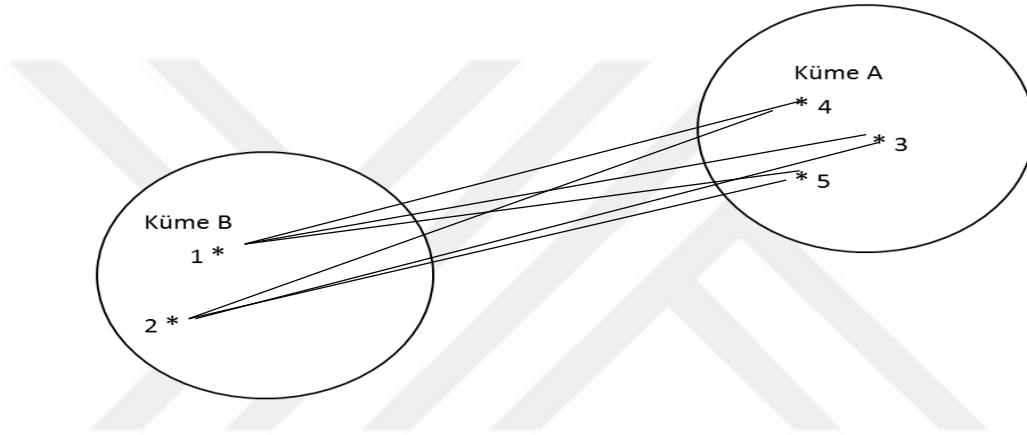
Güçdemir ve Selim (2015) bir yedek parça üreticisi firmanın satışlarını incelemiş, çalışmada, 317 müşteri firmaya ait 28840 kayıt analiz edilmiştir. Tam bağlantı yönteminin yanında başka kümeleme yöntemlerinin de denendiği çalışmada sonuç olarak beş segment oluşturulmuştur.

3.2.3.2.1.3 Grup Ortalaması ile Bağlantı

Bu ölçüde, iki küme arasındaki uzaklık; C_i ve C_j kümelerindeki nokta çiftlerinin ortalama uzaklıkları olarak tanımlanır (Zaki ve Meira Jr, 2014, s.368).

$$\delta(C_i, C_j) = \frac{\sum_{x \in C_i} \sum_{y \in C_j} \delta(x, y)}{n_i \cdot n_j} \quad (14)$$

$n_i = |C_i|$ C_i 'deki noktalardır.



Şekil 10: Grup Ortalaması Bağlantı

Uzaklık matrisinde, birleşmiş kümenin uzaklık matrisindeki satırını (sütun) hesaplamak için, matristeki i . ve j . satırların (sütun) ağırlıklı ortalaması kullanılır (Aggarwal, 2015, s.170).

Haddad ve diğerleri (2007) tüketicilerin labne ürününü satın alma niyetini etkileyen faktörler konusunda yaptığı bir araştırmada, 300 öğrenciye deney uygulamıştır. Labne çeşitlerini tadan öğrencilere anket uygulanmıştır. Faktörlerin, konjoint analizi ile tespit edildiği çalışmada, öğrenciler ilk etapta ortalama bağlantı tekniği ile dört segmente ayrılmıştır.

Bigne ve diğerleri (2010), yaşlı pazarını segmente etmek amacıyla, 85015 İspanyol yaşlının harcamalarına ait bir veri kümesini İspanyol istatistik kurumundan elde etmiş ve analiz etmiştir. Segmentasyon çalışması için birçok tekniğin karşılaştırıldığı

çalışmada ortalama bağlantı yöntemi de kullanılmış, fakat hiyerarşik olmayan yöntemlerin hiyerarşik yöntemlere göre daha iyi segmentler verdiği gözlemlenmiştir.

Godoy ve diğerleri (2013) Brezilya’da, çay tüketicilerinin algısı ve davranışları üzerine yapmış oldukları bir çalışmada, deney yaptıkları 100 tüketiciye anket de uygulamış ve tüketiciler algılarına göre üç segmente ayrılmışlardır. Kümeleme analizinde hiyerarşik kümeleme tekniklerinden biri olan ortalama bağlantı tekniği kullanılmıştır.

Haverila ve diğerleri (2013) Finlandiya, BAE, Çin, Kanada ve Yeni Zelanda’da eğitim gören 403 lise ve 892 üniversite öğrencilerine, kullandıkları cep telefonlarının özellikleri ile ilgili yaptıkları bir pazar araştırmasında, öğrencilere bir anket uygulanmış, sonuçların analizinde ortalama bağlantı yöntemi de kullanılmıştır.

Mehmedovic ve diğerleri (2015) Batı Balkan ülkelerinde bir yaşam tarzına göre pazar segmentasyonu çalışması yapmak üzere, dört ülkeden 762 katılımcıya yaşam tarzlarıyla ilgili bir anket çalışması uygulamıştır. Çalışmada ortalama bağlantı yöntemi ile katılımcılar üç segmente ayrılmışlardır.

3.2.3.2.1.4 Minimum Varyans: Ward’ın Yöntemi

Bu ölçüt birleşme sonucunda daha önceden belirlenmiş bir amaç fonksiyonunda oluşacak olan değişmeyi minimum yapmayı hedefler. Birleşmeler belirlenen amaç fonksiyonunda kötüleşmeye neden olmaktadır. Bu sebeple amaç fonksiyonunda en az değişim olmasını sağlayacak birleşme seçilir (Aggarwal, 2015, s.170).

Varyanstaki değişimi kullanmak yerine, birleşme kriteri olarak Hata Kareleri Toplamı da kullanılabilir. Bu yöntem aslında merkezi (centroid) yöntemin diğer bir çeşididir. Ward her bir adımda birleşebilecek her bir küme çiftini göz önüne almış ve birleşmeleri halinde “malumat kaybı” artışının minimum olacağı iki küme seçmiştir. Ward bu kaybı, Hata Kareleri Toplamı olarak tanımlamıştır (Aggarwal, 2015, s.171).

Dolayısıyla bu ölçüde iki küme arasındaki uzaklık iki küme birleştiğinde Hata Kareleri Toplamındaki artış olarak tanımlanacaktır. C_i kümesinin HKT şu formülle hesaplanır (Zaki ve Meira Jr, 2014, s.368).

$$HKT_i = \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (15)$$

$C = \{C_1, C_2, \dots, C_m\}$ kümelemesindeki HKT şöyle hesaplanır.

$$HKT = \sum_{i=1}^m HKT_i = \sum_{i=1}^m \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (16)$$

Belirtildiği gibi örneğin, C_i ve C_j arasındaki uzaklık, C_i ve C_j kümeleri birleşip C_{ij} kümesini oluşturduğu zaman, HKT'deki net değişim değeri olarak açıklanır (Zaki ve Meira Jr, 2014, s.369).

$$\delta(C_i, C_j) = \Delta HKT_{ij} = HKT_{ij} - HKT_i - HKT_j \quad (17)$$

Ward'ın yöntemini kullanarak yapılmış olan bazı çalışmaları şu şekilde kısaca özetleyebiliriz.

Punj ve Steward (1983, s.138), daha önce yapılmış olan kümeleme algoritmalarının karşılaştırılması çalışmalarının birçoğunu ele almış ve bu çalışmalar ışığında Ward'ın yönteminin, ortalama bağlantı yönteminin ve K-ortalamalar yönteminin diğer yöntemlere üstün geldiğini tespit etmiştir. Ward'ın yönteminin ortalama bağlantı yöntemine gözlem değerleri arasında aykırı değerler olmadığı zaman üstün geldiğini, K ortalamalar tekniğinin ise hem Ward'ın yöntemine hem de ortalama bağlantı yöntemine rassal olmayan bir başlangıç noktası belirlendiğinde üstün geldiğini tespit etmiştir.

Behe, Prince ve Tayama (1992) süpermarketten çiçek alan müşterileri segmentlere ayırmak amacıyla, ABD'de 1389 müşteriye anket çalışması düzenlemiş, ankette demografik bilgilerin yanı sıra müşterilerin neden çiçek aldığını sorgulamıştır. Ward'ın kümeleme algoritmasının ve k-ortalamalar algoritmasının kullanıldığı çalışmada müşteriler beş segmente ayrılmıştır.

Rangan, Moriarty ve Swartz (1992), bazı sanayi malzemeleri üreten bir fabrikanın ürünlerini sattığı özel 174 müşterisini satın alma davranışlarına göre segmentlere ayırmak istemiştir. Binlerce müşterisi olan bu fabrikanın bahsedilen 174 müşterisi, fabrikanın satışlarının %40'ını oluşturmaktadır. Bu amaçla Ward'ın tekniği kullanılarak, müşteriler satın alma davranışlarına göre dört segmente ayrılmıştır.

Segal ve Giacobbe (1994) ABD’deki bir süpermarkete ait 10 bin müşterinin demografik bilgilerini içeren veri kümesini Ward’ın yöntemini kullanarak dört segmente ayırmıştır.

Dibb ve Stern (1995), kümeleme analizinin ne kadar etkili olduğunu tespit etmek amacıyla, 270 kişiye bir anket çalışması düzenlemiş, cevaplayıcılar Ward’ın yöntemi ile altı kümeye bölünmüştür ve bu sonuç, tek bağlantı, tam bağlantı ve ortalama bağlantı yöntemi ile karşılaştırılmıştır.

Machauer ve Morgner (2001), Almanya’daki 285 banka müşterilerine yapmış oldukları anket çalışması ile, müşterilerin bankadan bekledikleri faydaları anket soruları ile ölçmüştür. Müşterilerin verdikleri puanlarla, Ward’ın yöntemini kullanarak, dört segment elde edilmiştir.

Durkin (2004), 480 İngiliz banka müşterisine yapmış oldukları anket çalışması ile, müşterilerin İnternet Bankacılığı kullanımını sorgulamış ve müşteriler dört segmente ayrılmışlardır. Çalışmada Ward’ın kümeleme tekniği ve k-ortalamlar tekniği kullanılmıştır.

Mendes ve Cardoso (2006), outlet mağazalarını kümeleme yöntemiyle, performans değerlendirmesi yapmak ve yer tespit çalışmalarını desteklemek amacıyla gruplara ayırmışlardır. Mağaza müşterilerine anket yoluyla veri toplanan çalışmada, kümeleme tekniklerinden olan Ward’ın tekniği kullanılmıştır.

Füller ve Matzler (2008), Avusturya İsviçre ve İtalya’da yer alan Alp Dağlarındaki otelleri ziyaret eden müşterilere, müşteri tatmini ile ilgili bir anket çalışması uygulamıştır. 6172 katılımcının olduğu bu çalışmada müşteriler Ward’ın kümeleme yöntemi kullanılarak beş segmente ayrılmışlardır.

Kennedy ve diğerleri (2008) ABD’de, tüketicilerin güvenli yiyeceklere olan algıları ile ilgili yaptıkları bir araştırma için 1014 tüketiciye anket yapılmış, tüketiciler Ward’ın kümeleme yöntemi ile beş segmente ayrılmışlardır.

Shahrabi ve diğerleri (2009) metal pazarını segmente etmek amacıyla, 210 kişiye anket uygulamış ve Ward'ın kümeleme analizi sonucunda pazar, beş segmente ayrılmıştır.

Murtagh ve Legendre (2011), Ward'ın yönteminin bazı çalışmalarda yanlış uygulandığının altını çizerek, farklı uygulamalar ile gerçekte olması gereken yöntemi karşılaştırmıştır. R programlama dilindeki Ward'ın yöntemini uygulayan işlevler ile analizler yapılan çalışma sonucunda, R programlama Dilindeki “ward.D2” işlevinin Ward'ın yöntemini tam olarak uyguladığını vurgulamıştır.

Eusebia ve Carneiro (2015) genç turizm pazarını incelemek amacıyla, Portekiz’de eğitim gören 399 üniversite öğrencisi genç turizmciyi bir anket çalışması ile segmente etmiş, çalışmada bütün hiyerarşik kümeleme yöntemleri denenmiş ve Ward'ın yöntemi en iyi sonucu vermiştir.

Brentari ve diğerleri (2016) İtalya’daki McDonald firmasına ait restoranlarla ilgili yaptıkları çalışmada, pazar segmentasyonunu, farklı restoranlardan da seçtikleri 2085 müşteriye uyguladıkları anket yoluyla yapmışlardır. Segmentasyon çalışmasında Ward'ın tekniği kullanılmıştır.

Şu ana kadar değinilen her bir yöntemin farklı avantajları ve dezavantajları vardır. Tek bağlantılı yöntemde, aykırı gözlem değerlerinin varlığı bu yöntemin davranışını bozabilmektedir. Tam bağlantı yöntemi bir kümedeki herhangi iki noktanın maksimum uzaklığını minimize etmek istemektedir ki bu yaklaşım bir kümenin çapını akla getirir. Yöntem bir taraftan çapı minimize etmeye çalışırken diğer taraftan aynı anda benzer çapa sahip kümeler de oluşturmaya çalışmaktadır. Dolayısıyla eğer veri kümesindeki doğal bazı kümeler diğer kümelerden daha büyükse kümeleme sonucunda bu doğal kümelerin elde edilmesi gerekirken, bu yaklaşım bu büyük kümelerin bozulmasına sebep olabilir. Bu gibi sebeplerden ötürü Grup Ortalaması ve Ward'ın yöntemi diğer yöntemlere nazaran daha elverişlidir. (Aggarwal, 2015, s.171).

3.2.3.2.2 Bölücü (Üstten Alta) Yöntemler

Bölücü hiyerarşik yöntemler basit bir tanımla şu şekilde açıklanabilir: Bölücü yöntemler, bütün veri kümesinin ilk etapta tek bir küme oluşturması, daha sonra ise daha

küçük alt kümelere yavaş yavaş bölünmesidir (Gan, Ma ve Wu, 2007, s.137). Bu yöntemler için kullanılan en yaygın algoritma ise Kaufman ve Roussew'in geliştirdiği DIANA algoritmasıdır (1990). Ayrıca Grafik kümeleme yöntemi olarak bilinen, bir grafiğin düğüm noktaları arasındaki benzerliği kullanarak kümeleme yapan yöntemin temelini oluşturur (Zaki ve Meira Jr, 2014, s.394).

Birleştirici yöntemlerin aksine fazla kullanılmayan bir tekniktir. Bu tekniğin birleştirici yöntemlere karşısındaki avantajı sadece, araştırmacı için kümelerin bölünmesinin önem arz ettiği zaman olduğu söylenebilir. Bölücü yöntemler herhangi bir algoritmayı bölünme için kullanabilmektedirler. Örneğin k-ortalamlar veya k-medoids her iterasyonda k=2 alınarak kümelerin bölünmesini sağlayabilir (Hastie ve diğerleri, 2009, s.526).

3.2.3.3 Uzaklık Matrisinin Güncellenmesi

Gerek tek bağlantı yönteminde, gerek tam bağlantı yönteminde gerekse de Ward'ın minimum varyans yönteminde, Lance ve Williams'ın formülü (1967) kullanılarak yineleme yapılır. Bu formül her bir adımda (küme birleşmeleri yapılırken) küme uzaklıklarını güncellemek için kullanılır.

C_i ve C_j sembolü iki küme C_{ij} kümesi olarak birleşeceği zaman, yeni oluşmuş olan C_{ij} kümesinin diğer bütün kümelere C_k ($k \neq i$ ve $k \neq j$) olan uzaklığının yeniden hesaplanması gerektiği için, uzaklık matrisi güncellenmelidir. Bu noktaya kadar bütün uzaklıklar daha önceden belirlenmiş ve dolayısıyla elimizde uzaklık matrisi mevcuttur (Zaki ve Meira Jr, 2014, s.370).

d_{ij} , d_{ik} ve d_{jk} , C_i , C_j ve C_k kümeleri arasındaki uzaklıklar olsun.

$d_{(ij)k}$ ise, $C_i \cup C_j$ (yeni küme) ve C_k kümesi ile arasındaki uzaklık olsun.

$$d_{(ij)k} = \alpha_i d_{ik} + \alpha_j d_{jk} + \beta d_{ij} + \gamma |d_{ik} - d_{jk}| \quad (18)$$

Burada α_i , α_j , β ve γ katsayıları, küme boyutuna bağlı olabilirler ve küme uzaklık fonksiyonu d_{ij} ile birlikte kümeleme algoritmasını belirlerler.

$\alpha_i = \alpha_j = 1/2$, $\beta = 0$ ve $\gamma = -1/2$ olduğu zaman formül aşağıdaki hali alır ve algoritma tek bağlantı yöntemine karşılık gelir;

$$d_{(ij)k} = \min(d_{ik}, d_{jk}) . \quad (19)$$

$\alpha_i = \alpha_j = \gamma = 1/2$ ve $\beta = 0$ olduğunda ise, formül tam bağlantı yöntemine karşılık gelir;

$$d_{ijk} = \max(d_{ik}, d_{jk}) \quad (20)$$

Tek bağlantı, çok bağlantı ve ortalama bağlantı yöntemleri, uzaklık hesaplanırken, bir çift kümedeki bütün noktaları göz önünde bulundurur ve bu yüzden bunlara grafik yöntemler de denir. Diğer yöntemler ise küme temsilcisi için geometrik merkezleri kullandıkları için, geometrik yöntemler olarak adlandırılırlar.

Formüldeki $\alpha_i, \alpha_j, \beta$ ve γ katsayıları her ölçü için Tablo 7’de gösterildiği gibi farklı değerler alır.

3.2.4 Kümelemenin Geçerliliği

Kümeleme analizi için birbirinden farklı birçok algoritma vardır ve her biri farklı farklı kümelerin dolayısıyla farklı sonuçların oluşmasını sağlamaktadır. Bu sebeple kümelemenin uygunluğunu, performansını ve ortaya çıkardığı sonuçlarını iyi değerlendirmek gerekmektedir. Kümeleme sonucunda ortaya çıkan gruplar bir uzman tarafından incelenebilir. Bunun dışında bir kümelemenin geçerliliğini ölçebilecek ölçütler de vardır ve genel olarak Dışsal Ölçütler, İçsel ölçütler ve Göreceli Ölçütler olarak üçe ayrılabilirler (Halkidi, Batistakis ve Vazirgiannis, 2001, s.109; Jain ve Dubes, 1988, s.161).

Tablo 7 Lance ve Williams’ın Hiyerarşik Kümeleme Algoritmaları için Formülleri

KÜMELEME ALGORİTMASI	α_i	α_j	β	γ
TEK BAĞLANTI	1/2	1/2	0	-1/2
TAM BAĞLANTI	1/2	1/2	0	1/2
GRUP ORTALAMASI BAĞLANTI	$\frac{n_i}{n_i + n_j}$	$\frac{n_j}{n_i + n_j}$	0	0
AĞIRLIKLİ ORTALAMA BAĞLANTI	1/2	1/2	0	0
MEDYAN BAĞLANTI	1/2	1/2	-1/4	0
MERKEZİ BAĞLANTI	$\frac{n_i}{n_i + n_j}$	$\frac{n_j}{n_i + n_j}$	$\frac{-n_i n_j}{(n_i + n_j)^2}$	0
WARD’IN YÖNTEMİ	$\frac{n_i + n_k}{n_i + n_j + n_k}$	$\frac{n_j + n_k}{n_i + n_j + n_k}$	$\frac{-n_k}{n_i + n_j + n_k}$	0

Kaynak: (Zaki ve Meira Jr, 2014, s.371)

3.2.4.1 Dışsal Ölçütler

Dışsal Ölçütlerde bir kümeleme algoritmasının sonuçları daha önceden belirli olan veya insanlar tarafından bir şekilde bilinen veya tahmin edilen sonuçlarla karşılaştırılır. Bu ölçüt için en yaygın kullanılan ölçüler Rand’ın Endeksi, Jaccard’ın Katsayısı, Fowlkes-Mallows’un Endeksi, Hubert’in İstatistiğidir (Gan ve diğerleri, 2007, s.302; Jain ve Dubes, 1988).

Kümelemenin geçerliliği için ortaya çıkarılmış ölçülere kümelemeler arasındaki benzerlik ölçüleri (similarity of clusterings) veya kümelemelerin görüş birliği ölçüleri (agreement of clustering methods) isimleri de verilmiştir. Kümelemelerin geçerliliğinde kullanılacak ölçüler temellerinde yatan farklı yaklaşımlar sebebiyle genel olarak ikiye ayrılırlar. Bunlardan ilki *Çiftleri Sayma* (counting pairs) yöntemi olarak bilinen ve her iki kümeleme yönteminde de aynı kümeler veya her iki yöntemde de ayrı kümeler düşmüş veri noktalarının sayılması esasına dayanır. Çiftleri sayma yöntemleri veri sayısına, küme sayısına ve kümedeki eleman sayılarına duyarlıdır (Rabbany ve Zaiane, 2015, s.2; Yeung ve Ruzzo, 2001, s.1). Bu kategorideki ölçülere ilerleyen bölümlerde ayrıntılı olarak değinilecektir.

İkinci yaklaşım ise *Malumata Dayalı* (information theoretic) ölçülerdir. Bu ölçülerin en çok bilineni olan Malumatın Varyansı (Variance of Information) ismindeki Meila'nın (2007) ölçüsüdür. Bu ölçünün mantığı bir kümelemeden (C) diğer bir kümelemeye (C') değişim yapılırken kaybedilen ve kazanılan malumatın miktarıdır (Meilă, 2007, s.873). Ayrıca Ortak Bilgi (Mutual Information), Düzeltilmiş Ortak Bilgi (Adjusted Mutual Information) ölçüleri de bu kategoriye dâhildir (Wagner ve Wagner, 2007, s.12).

3.2.4.2 İçsel Ölçütler

İçsel Ölçütlerde ise bir algoritma tarafından üretilen kümelemenin yapısı sadece veri kümesinden intikal eden bazı özellikler ve nicelikler kullanılarak değerlendirilir. Hiyerarşik bir kümelemede oluşan hiyerarşiler bu duruma örnek olarak verilebilir (Gan ve diğerleri, 2007, s.303). Bu ölçütte en çok kullanılan ölçüler ise Kophenetik Korelasyon Katsayısı (Sokal ve Rohlf, 1962) ve Kaufman ve Roussew'in (1990) önerdiği Birleştirici Katsayısıdır (Agglomerative Coefficient).

Kophenetik Korelasyon Katsayısı, kophenetik matrisinden hesaplanmaktadır. Bu matris hiyerarşik bir algoritmanın ortaya çıkardığı hiyerarşi diyagramını temsil edebilir niteliktedir. Matrisin herhangi bir elemanın değeri, o elemanla ilgili iki nesnenin ilk olarak hangi seviyede bir kümede yer aldığını temsil eder (Halkidi ve diğerleri, 2001, s.128). Çünkü bir hiyerarşik kümeleme ağacında, orijinal veri kümesindeki herhangi iki nesne belirli bir seviyede (yükseklikte) birbirleri ile ilişkilendirilerek kümeler oluştururlar ve bu ilişkilendirmenin yüksekliği bu iki nesneyi içeren iki küme arasındaki uzaklığı temsil eder. İşte bu yüksekliğe iki nesne arasındaki kophenetik uzaklık denir.

Daha önceden belirlenen bağlantı fonksiyonu (tek, tam, ortalama bağlantı veya Ward'ın yöntemi) ile oluşmuş olan kümeleme ağacının veri kümemizi ne kadar iyi temsil ettiğini, kophenetik uzaklık ile orijinal veri kümesinden oluşan uzaklık matrisinin karşılaştırılması ile anlayabiliriz. Karşılaştırma sonucu ortaya çıkan sonuç Kophenetik Korelasyon Katsayısıdır. Bu katsayı 1'e yaklaştıkça kümeleme ile ortaya çıkan sonuç veri kümesini o kadar iyi yansıtmaktadır. Başka bir deyişle yüksek bir korelasyon kümeleme

algoritmasının orijinal benzeşmezlik matrisinin sahip olduğu malumatı çarpıtmadığını gösterir (Legendre ve Legendre, 2012, s.375; Singh, 2008, s.6).

Kophenetik katsayısının anlamlılığının testi yapılamamaktadır. Bunun sebebi, kophenetik matrisinin orijinal benzerlik matrisinden bağımsız olmamasıdır. Anlamlılık ancak iki farklı veri kümesinin benzerliği araştırılıyorsa yapılabilir (Legendre ve Legendre, 2012, s.376)

İçsel ölçütler için diğer bir ölçüt ise Kaufman ve Roussew'in (1990) önerdiği Birleştirici Katsayısıdır (Agglomerative Coefficient). Bu ölçüt kümeleme algoritmasının ortaya çıkardığı küme yapısına dayalı bir ölçüttür. Genellikle kümelemenin gücünü ve kalitesini tespit etmek amacı ile kullanılır. Veri kümesindeki her bir nesne i ile gösterilsin ayrıca $d(i)$ ise bu nesnenin ilk defa birleştiği kümeyle olan uzaklığının algoritmanın son adımındaki birleştiği kümeyle olan uzaklığına bölünmesini gösterebilir. Birleştirici Katsayısı bütün $(1-d_i)$ 'lerin ortalamasına eşit olur (Struyf, Hubert ve Rousseeuw, 1997, s.19).

$$\text{Birleştirici Katsayı} = \text{ortalama} \left(1 - \left(\frac{\text{ilk birleşme}}{\text{son birleşme}} \right) \right) \quad (21)$$

Veri setinin bir algoritma tarafından üretilmiş kümeleme yapısını ölçen bu katsayı, 0 ile 1 arasında değer alır. Kullanılan algoritma, veri kümesinde doğal bir kümeleme yapısı bulamadığı durumda 0 değerini alırken, algoritmanın açık bir küme yapısını bulduğu durumda ise 1 değerini alır. Veri kümesindeki gözlem sayısı arttıkça ölçünün değeri de artacağından bu katsayı farklı büyüklükteki veri kümelerini karşılaştırmada kullanılamaz (Carvalho ve diğerleri, 2011, s.877; Frees, Derrig ve Meyers, 2014, s.308).

3.2.4.3 Göreceli Ölçütler

Göreceli Ölçütler üçüncü ve son olarak kümeleme sonuçlarının değerlendirilmesi konusunda karşımıza çıkmaktadır. Bu ölçütteki temel fikir bir kümeleme yapısının aynı algoritmayla fakat başka parametrelerle üretilmiş olan kümeleme yapısıyla karşılaştırılmasıdır (Halkidi ve diğerleri, 2001, s.123). Bu noktada

akla ilk gelen örnek, hiyerarşik kümelemede karşımıza çıkan Tek, Tam ve Ortalama gibi bağlantı yöntemleridir (Jain ve Dubes, 1988, s.161).

3.2.5 İki Kümeleme Sonuçlarının Karşılaştırılması

Bir kümeleme algoritmasının sonuçlarının değerlendirilmesi kümelemenin geçerliliği başlığı altında yapılmaktadır (Halkidi ve diğerleri, 2001, s.123). Literatürde kümelemenin geçerliliği başlığı altında yer alan dışsal ölçütler ayrıca kümelemelerin karşılaştırılması başlığı altında da incelenmektedir (Meilă, 2007; Romano ve diğerleri, 2014; Vinh, Epps ve Bailey, 2009; Wagner ve Wagner, 2007).

İki kümeleme sonuçlarının karşılaştırılabilmesi için her iki kümelemede elde edilen küme sayısının eşit olması zorunluluğu yoktur (Meilă, 2007, s.875; Rabbany ve Zaïane, 2015, s.2). Kümelemelerin karşılaştırılması konusunda yapılan çalışmalar incelendiğinde genel amaç aynı veri kümesine ait farklı algoritmaların performanslarının karşılaştırılması iken, bu karşılaştırmada kullanılan ölçülerin farklı veri kümelerinin kümelenebilirliğini de karşılaştırabiliyor olması beklenebilir (Xiang ve diğerleri, 2012, s.1)

Her ne kadar çalışmamızda iki farklı segmentasyon yaklaşımında kullanılacak iki kümelemenin girdi değişkenleri farklı olsa da, her iki yaklaşım için de kümelenecek olan veri noktaları aynı olan 175 mağazadır. “İki kümeleme yönteminde veri noktası çiftleri eğer her iki kümeleme yönteminde de aynı kümede yer almışsa veya her iki yöntemde de farklı kümelerde yer almışsa bu kümelemeler birbirine benzerdir” (Rand, 1971, s.847) ana fikrini savunan Çiftleri Sayma Yöntemlerine göre, çalışmamızda kullanılacak olan kümeleme yöntemlerinin benzerlik göstermesi beklenmektedir. Çünkü ilk segmentasyon yaklaşımında kümeleme girdisi olarak kullanılan sosyoekonomik faktörlerin ikinci segmentasyon yaklaşımında kümeleme girdisi olarak kullanılan satın alma davranışlarını etkilediği düşünülmektedir. Dolayısıyla mağazaların bir kısmının her iki yaklaşımda da aynı kümelerde veya her iki yaklaşımda da farklı kümelerde yer alması beklenmektedir. Bunun tespitinin yapılması da Çiftleri Sayma yöntemi ile mümkündür.

Çiftleri Sayma yönteminde yapılan iş, her iki kümelemede de aynı yolla sınıflanmış veri noktası çiftlerinin sayısının hesaplanmasıdır. Her iki kümelemede de veri çiftlerinin aynı kümede yer almaları veya farklı kümelerde yer almış olmaları hesaplanır

(Wagner ve Wagner, 2007, s.3). Vurgulandığı gibi, Rand Endeksi (Rand, 1971) bu hesaplamayı yapabilmek için önerilmiş bir endeks olup, Rand'ın bu çalışmasından sonra endeksin değerini şanstan arındırmak amacıyla araştırmacılar tarafından birçok farklı düzeltme faktörü önerilmiştir. Rand çalışmasında aşağıdaki tanımlamaları yaptıktan sonra kümelemelerin benzerliği formülünü sunmuştur (Rand, 1971, s.846).

(X, Y, m) ifadesinde, X kümelemesi yapılacak N kadar nesneyi (veya veri noktası) temsil etsin $(X = X_1, X_2, \dots, X_N)$.

Y ise bu nesnelerin K kadar özel bir kümeye bölündüğünü belirtsin. Bu bölünme “bir kümeleme” olarak adlandırılacak olup $(Y = Y_1, Y_2, \dots, Y_K)$ kümeleri söz konusu olacaktır. Burada, her bir küme $Y_K = X_{k_1}, X_{k_2}, \dots, X_{k_{n_k}}$ nesnelerini veya veri noktalarını içermektedir ($\sum n_k = N$ ve $k = 1, 2, \dots, K$ için $n_k \geq 1$).

N kadar nesnenin (veya veri noktasının) yer aldığı bir veri kümesinde, $K = 1, 2, \dots, N$ tamamı için bahsedilen bölünmelerin hepsinin oluşturduğu küme $\mathcal{Y}$ olsun.

m sembolü, verilen bir X veri kümesi için, özel bir Y seçmek amacıyla kullanılacak olan yöntemi temsil etmektedir.

Aynı veri kümesiyle kullanılan iki farklı kümeleme yöntemi için (Y ve Y') bir benzerlik ölçüsü hesaplanabileceğini öne süren Rand (1971, s.847), Bu ölçüyü $c(Y \text{ ve } Y')$ ile göstermiş ve bu ölçünün değerini ise veri noktası çiftlerinin benzer şekilde kümelere atanma sayısının, toplam veri çiftlerinin sayısına normalize edilmesi olarak hesaplamıştır. N noktası için $(X = X_1, X_2, \dots, X_N)$, ve bunların iki yöntemle kümeleneceği ile $(Y = Y_1, Y_2, \dots, Y_{K_1})$ ve $(Y' = (Y'_1, Y'_2, \dots, Y'_{K_2}))$;

$$\frac{\sum_{i < j} Y_{ij}}{\binom{N}{2}} \quad (22)$$

Burada Y_{ij} aşağıda gösterildiği gibi ya 0 ya da 1 değerini alır:

1; eğer k ve k' mevcutsa öyle ki hem X_i hem X_j , Y_k ve $Y'_{k'}$ da yer alıyorsa

1; eğer k ve k' mevcutsa öyle ki X_i hem Y_k ve $Y_{k'}$ da yer alıyorsa ve X_j ise ya Y_k ya da $Y_{k'}$ da yer alıyorsa

0; diğer durumlarda.

Rand çalışmasında c 'yi hesaplamanın daha kolay ve genelleştirilmiş bir yolunu da önermiştir. n_{ij} sayısı Y 'nin i . Y' 'nün ise j . kümesinde aynı anda yer alan noktaların sayısı olsun.

$$c(Y, Y') = \left[\binom{N}{2} - \left[\frac{1}{2\{\sum_i (\sum_j n_{ij})^2 + \sum_j (\sum_i n_{ij})^2\}} - \sum \sum n_{ij}^2 \right] \right] / \binom{N}{2} \quad (23)$$

Hesaplanan c değeri, benzerlik ölçüsü olduğu için sıfır ile bir arasında değer almaktadır ($0 \leq c \leq 1$). İki kümeleme arasında tam bir benzerlik konusu iken ölçü bire, hiçbir benzerlik söz konusu değilse sıfıra eşit olur.

Çiftleri sayma yöntemlerinden bir diğeri ise Hubert ve Arabie'nin (1985) Rand Endeksini genişletmek ve şanstı arındırmak amacıyla önerdiği bir düzeltme faktörüdür. Modelin rassallığı için genelleştirilmiş hipergeometrik dağılımı varsayım olarak kullanmışlardır. Bu varsayımda karşılaştırılması yapılacak iki kümeleme rassal olarak seçilmiş, başka bir anlatımla kümelerdeki nesne sayısı ve kümelerin sayısı sabit olmaktadır. Yeung ve Ruzzo (2001) çalışmalarında hem Rand'ın formülünü hem de Hubert ve Arabie'nin önerdiği Düzeltilmiş Rand Endeksini detaylı ve anlaşılır bir biçimde şu şekilde anlatmışlardır (Yeung ve Ruzzo, 2001, s.1-2).

Elimizde n nesneli bir veri kümesi olsun. Ayrıca $S = \{O_1, O_2, \dots, O_n\}$ iken U ile V ise S 'nin iki farklı kümelemesi olsun; $U = \{u_1, \dots, u_R\}$ ve $V = \{v_1, \dots, v_C\}$ ile gösterilsin.

$$U_{i=1}^R u_i = S = U_{j=1}^C v_j \text{ ve ayrıca } u_i \cap u_{i'} = \emptyset = v_j \cap v_{j'} \text{ olmak üzere;} \\ 1 \leq i \neq i' \leq R \text{ ve } j \neq j' \leq C.$$

U kümelemesi daha sonra kendisi ile değerlendirme yapılabilecek harici bir ölçüt olsun ve V kümelemesi de yapılan bir kümelemenin sonucu olsun. a değeri hem U 'da hem de V 'de aynı kümede yer almış nesne çiftlerinin sayısı olsun. b değeri ise U 'da aynı kümede yer almış fakat V 'de farklı kümede yer almış nesne çiftlerinin sayısını, aynı

şekilde c değeri de V 'de aynı kümede yer almış fakat U 'da farklı kümede yer almış küme çiftlerinin sayısını gösterebilir. Son olarak d değeri her iki kümelemelerde de farklı kümelerde yer almış nesne çiftlerinin sayısı olsun. Kısaca a ve d değerleri aynı fikirde olunanlar (agreement) olanlar b ve c ise aynı fikirde olunmayanlardır. Bu doğrultuda Rand Endeksi aşağıdaki formülle de hesaplanabilir.

$$\frac{a+d}{a+b+c+d} \quad (24)$$

Düzeltilmiş Rand Endeksinin formülü için ise; n_{ij} hem u_i 'de hem de v_j 'de yer almış nesne sayısı olarak tanımlansın. Ayrıca $n_{i.}$ ve $n_{.j}$ ise u_i ve v_j kümelerinde yer alan nesnelerin sayısını gösterebilir. Aşağıdaki tabloda semboller daha da anlaşılır hale gelmektedir.

Tablo 8 İkili Karşılaştırma için Kontenjans Tablosu Sembolleri

Küme	v_1	v_2	...	v_c	Toplam
u_1	n_{11}	n_{12}	...	n_{1c}	$n_{1.}$
u_2	n_{21}	n_{22}	...	n_{2c}	$n_{2.}$
...	...	...	...	...	...
u_R	n_{R1}	n_{R2}	...	n_{Rc}	$n_{R.}$
Toplam	$n_{.1}$	$n_{.2}$	...	$n_{.c}$	$n_{..}=n$

Bu bağlamda Düzeltilmiş Rand Endeksinin formülü ise aşağıdaki gibidir.

$$\frac{\sum_{ij} \binom{n_{ij}}{2} - \left[\sum_i \binom{n_{i.}}{2} \sum_j \binom{n_{.j}}{2} \right] / \binom{n}{2}}{\frac{1}{2} \left[\sum_i \binom{n_{i.}}{2} + \sum_j \binom{n_{.j}}{2} \right] - \left[\sum_i \binom{n_{i.}}{2} \sum_j \binom{n_{.j}}{2} \right] / \binom{n}{2}} \quad (25)$$

Çiftleri sayma yönteminde diğer bir endeks olan Fowlkes-Mallow Endeksi (Fowlkes ve Mallows, 1983) ise hiyerarşik kümelemelerin karşılaştırılması amacıyla ortaya çıkmıştır. Fakat hiyerarşik olmayan kümelemeler için de kullanılabilir. Çalışmalarında karşılaştırılması yapılacak olan iki kümelemedeki küme sayısının birbirine eşit olmasını önermişlerdir fakat Wagner ve Wagner (2007, s.6) farklı sayıda

kümeye sahip olan kümelemeler için, genelleştirilmiş Fowlkes-Mallow Endeksi formülünü şu şekilde açıklamışlardır.

X sonlu bir küme olsun, $|X| = n$. Bir kümeleme olan C , X 'in boş olmayan ayrık bir alt kümesidir ve $\{C_1, \dots, C_k\}$ alt kümelerinden oluşmaktadır öyle ki bunların birleşimi X 'i verir. X 'in mümkün bütün kümelemelerinin kümesi $P(X)$ ile gösterilsin. Herhangi bir $C = \{C_1, \dots, C_k\}$ kümelemesinde bütün $i = 1, \dots, k$ için $|C_i| > 0$ olduğu varsayılır. $C' = \{C'_1, \dots, C'_\ell\}$ kümelemesi ise X 'in başka bir kümelemesi olsun. C ve C' ikilisinin $M = (m_{ij})$ konfüzyon veya kontenjans matrisi $k * \ell$ boyutlu bir matristir. Bu matrisin ij . girişinin değeri C_i ve C'_j kümelerinde kesişmiş elemanların sayısı kadardır. Fowlkes-Mallow'un önermiş olduğu formül aşağıdaki gibidir.

$$\mathcal{FM}(C, C') = \frac{\sum_{i=1}^k \sum_{j=1}^{\ell} m_{ij}^2 - n}{\sqrt{(\sum_i |C_i|^2 - n)(\sum_j |C'_j|^2 - n)}} = \frac{n_{11}}{\sqrt{(n_{11} + n_{10})(n_{11} + n_{01})}} \quad (26)$$

$$m_{ij} = |C_i \cap C'_j|, \quad 1 \leq i \leq k, \quad 1 \leq j \leq \ell.$$

Fowlkes-Mallow Endeksi de Düzeltilmiş Rand Endeksi gibi güçlü varsayımlara bağlı olduğundan sonuçlarının yorumlanması zordur. Düzeltilmiş Rand Endeksinde olduğu gibi bu benzerlik ölçüsünün değeri de, sabit küme büyüklüğünde bağımsızlık ifade eden yokluk hipotezi ışığında, beklenen değerden olan sapmaya uymaktadır. Küme sayısının az olduğu durumlarda sonuç çok yüksek çıkmaktadır (Wagner ve Wagner, 2007, s.6).

Rand Endeksi, Düzeltilmiş Rand Endeksi ve Fowlkes-Mallows Endeksi dışında, çiftleri sayma yaklaşımıyla Gamma, Hubert, Pearson ve Jaccard'ın ölçüleri de kullanılmaktadır.

3.3 SÜPERMARKETLERDE KÜMELEME ÇALIŞMALARI

Bu bölümde, çalışmamıza konu olan süpermarketlerde, daha önce yapılmış olan bazı kümeleme çalışmalarına yer verilecektir.

Kolyshkina ve arkadaşları (2010) hedef pazarlama amacıyla, 106 mağazaya ait 151 değişken içeren (mağaza ile ilgili değişkenler, mağazalarda satılan ürünler, satılan ürünlerin satış performansları, mağazaların satış performansları vb.) veri setine birkaç kümeleme tekniği uygulamış ve mağazalar altı segmente ayrılmıştır. Daha sonra segmentlerde alkolsüz ürünlerin satışları incelenmiştir.

Şekerkaya ve Cengiz (2010), İstanbul'da 568 kişiye uygulamış oldukları anket çalışması ile, tüketicilerin perakende zincirlerini tercih sebeplerine göre kümelenmesini amaçlamıştır. Bu bağlamda A,B,C,D olarak isimlendirdikleri dört büyük perakende firması hakkında cevaplayıcılara, bu mağazalardan toplam 17 ürün grubu üzerinden (gıda, giyim, kozmetik, züccaciye vb.) yapmış oldukları satın almaların sıklıkları sorgulanmıştır. Cevaplayıcıların demografik özellikleri sunulduktan sonra, k-ortalamar tekniği ile elde edilen dört kümenin özellikleri şöyle bulunmuştur; Müşterilerinin çoğu lise mezunu olan 1. Kümedeki müşteriler en çok A marketini tercih etmektedir. A marketinden gıda, kozmetik, züccaciye, kişisel bakım gibi ürünleri satın almakta, ayrıca reklam çeşidi olarak dergi reklamlarını ve internet reklamlarını önemli bulmaktadırlar. Genellikle üniversite mezunu olan, 2. Kümedeki müşterilerin özellikle gittikleri bir market bulunmazken A marketinden yalnız gıda, B marketinden ise giyim, kozmetik gibi ürünleri satın almaktadırlar. Ağırlıklı olarak evli ve lise mezunu müşterilerin olduğu 3. Kümedeki cevaplayıcılar en çok C marketine gitmekte, D marketine hiç gitmemekte ayrıca C marketinden sıklıkla kişisel bakım ürünleri ve temizlik ürünleri satın almaktadırlar. 4. Kümedeki müşterilerin geneli lise mezunu olup, en çok A marketine gitmekte, A marketinden gıda, kozmetik, züccaciye, ev tekstili, kitap, kırtasiye ürünleri satın almaktadırlar.

Liao ve diğerleri (2011) Tayvan'daki Carrefour süpermarketlerinin sunmuş olduğu internetten alışveriş ve eve teslim hizmetlerini sorgulamak amacıyla bir anket çalışması hazırlamıştır. Taze ürün ve kolay bozulmayan ürün alan 352 müşteriye yönelik anket çalışması sonucunda, iki aşamalı algoritma kullanılarak müşteriler üç kümeye

ayrılmıştır. Ankette müşteri davranışlarının yanı sıra, taze ürün alışverişleri, taze olmayan ürün alışverişleri, müşterilerin adrese teslim hizmeti hakkındaki görüşleri ve internetten alışveriş hizmeti hakkındaki görüşleri de sorgulanmıştır. Birinci kümede yer alan müşteriler az ürün satın alan, ikinci grup orta sayıda ürün alan ve üçüncü grup ise çok ürün alan müşteriler olarak adlandırılmıştır. Her bir grubun demografik özellikleri incelendikten sonra, ankette yer alan “hangi ürünlerin evinize teslim edilmesini istersiniz?” sorusuna verilen cevaplara göre, birliktelik kuralları üretilmiştir. Örneğin birinci kümedeki müşteriler sos ve et ürünlerini adrese teslim isterken, ikinci ve üçüncü gruptaki müşteriler ise sırasıyla sebze ve ekmek ürünlerinin en çok eve teslim edilmesini istemektedirler. İnternette alışveriş tercihleri konusunda ise, üç kümede yer alan müşteriler sırasıyla, makyaj ürünleri, bilgisayar/iletişim ürünleri ve müzik ürünlerini tercih etmektedirler.

Migueis ve diğerleri (2012), Avrupalı bir perakende firmasının müşterilerini birlikte satın aldıkları ürünlerden yola çıkarak, “yaşam tarzlarına” göre segmente etmek amacıyla, 100 bin müşterinin, 1831 ürün üzerinden satın alma verilerini analiz etmiştir. Bu amaçla müşterileri kümelemek yerine bu çalışmada VARCLUS tekniği kullanılarak, değişkenler olan ürünler kümelenmiştir. Her bir kümedeki ürünler incelenmiş, bu ürünleri sıklıkla satın alan müşteriler kümelerdeki ürün çeşitlerine göre altı grupta temsil edilmiştir. Birinci kümede orta gelirli insanların alabileceği ürünler, ikinci kümede diyet yapan sağlıklı insanlar, üçüncüde et sever müşteriler, dördüncü kümede düşük gelirli, beşincide geliri yüksek ve bebeği olan müşteriler, altıncı kümede ise beşinci kümeye benzer fakay geliri biraz daha düşük müşteriler yer almaktadır.

Çil (2012), Türkiye’de faaliyet gösteren süpermarket zincirine ait bir mağazanın, altı aylık satış verilerini kullanarak, 35 ana kategoriye indirgelediği ürünlerin birlikte satışlarını, birliktelik analizi ile incelemiş ve ortaya çıkan 518 anlamlı kuralı (destek değeri 0,05 ve güven değeri 0,20) mağazanın dizilişi için kullanmayı amaçlamıştır. Bu bağlamda, ortaya çıkan kurallara göre ürün gruplarının mağazanın neresinde olacağı, yan yana hangi ürün gruplarının yer alması gerektiği, müşterilerin alışverişlerini kolaylaştırma ve satış hacmini arttırma açısından önerilmiştir.

Chen ve diğerleri (2012) Xaffinity paket programında yer alan, 20245 üründen oluşan ve 98742 satış fişine sahip olan veri kümesini, ilk etapta birliktelik analizi kullanarak 152 farklı kural elde etmiş ve bu kuralların ne kadar etkili olup olmadığını tespit etmek için ise Veri Zarflama Analizi kullanmıştır. Daha sonra bu kurallar iki sınıfa; etkili olanlar ve etkili olmayanlar olarak ayrılmıştır. Sınıflama analizinde ise karar ağaçları kullanılmıştır.

Atkins ve diğerleri (2016), insanların alışveriş yaparken dikkatlice düşünüp mü, yoksa fazla düşünmeden mi alışveriş yaptıklarını öğrenmek için, market müşterilerini akıllıca alışveriş yapıp yapmadıklarına göre segmentlere ayırmaya çalışmıştır. Bu amaçla ABD’de son üç ayda herhangi bir markette alışveriş yapmış olan 751 tüketiciye anket çalışması yapılmış, ankette müşterilerin akıllıca yaptıkları alışveriş aktiviteleri; alışveriş değerini maksimize etmek, zamanı, harcanan parayı ve harcanan eforu minimize etmek gibi özellikler sorgulanarak ölçülmeye çalışılmıştır. Kümeleme analizinde k-ortalamlar tekniği ve Ward’ın hiyerarşik kümeleme analizi kullanılmış, sonuçta üç farklı segment ortaya çıkmıştır. Bu segmentler; kendiliğinden zeki müşteriler (satın alma öncesi ve araştırma ve plan konusunda zayıf müşteriler), duyarsız zeki müşteriler (araştırma konusunda iyi fakat diğer bütün zeki satın alma davranışlarında zayıf), karışık zeki müşteriler (zeki satın alma davranışları yüksek fakat bilgi araştırması zayıf müşteriler) olarak isimlendirilmiştir.

3.4 BİRLİKTELİK KURALLARI ANALİZLERİ

Veri Madenciliği tekniklerinin en önemlilerinden biri olan Birliktelik Kuralları Analizi, Bilgisayar Bilimleri alanında geliştirilmiş olan ve veri nesneleri arasındaki birliktelikleri (ilişkileri, bağımlılıkları) bulmaya yarayan bir tekniktir. Pazar sepet analizleri; bir müşteri tarafından satın alınmış ürünlerin (süpermarket veya hizmet ürünleri) birliktelik derecesini ölçen analizler, web sitelerini ziyaret etme ve sayfalarını tıklama analizleri vb. önemli Veri Madenciliği uygulamalarında sıkça kullanılmaktadır (Aggarwal, 2015, s.93; Cios ve diğerleri, 2012, s.289; Giudici, 2005, s.90).

Birliktelik Kuralları Madenciliği (Association Rules Mining) ismi ile de anılan Birliktelik Kuralları Analizlerinin en çok kullanıldığı alan olan pazar sepet analizinde, bir elimizde nesneler (items) diğer elimizde alışveriş sepetleri (baskets) bulunmaktadır.

Sepetler genellikle ticari işlem (transaction), örneğin süpermarketlerdeki her bir satış fişinde yer alan işlemler olarak karşımıza çıkmakta ve her bir sepette bir ürün kümesi (itemset) yer almaktadır (Leskovec, Rajaraman ve Ullman, 2014, s.202).

Bir süpermarketten veya perakende firmasından alışveriş yapmış müşterilere ait satış verilerinin oluşturduğu veri tabanının analizinde, firma en temel olarak sıkça birlikte satılan ürün kümelerini (Sık Ürün Küme Madenciliği) belirleyebilir. Ürün kümesinden kasıt, içerisinde ürün/ürünler barındıran bir küme, sık ürün kümelerinden (frequent itemsets) kasıt ise ilgili ürün kümesinin birçok sepette görülmesidir.

Analizde, k-ürün kümesi, kümenin içerisinde kaç ürünün var olduğunu, örneğin süt ve çikolatanın yer aldığı bir ürün kümesinde k'nın iki olduğunu, bu kümenin ikili ürün kümesi olduğu vurgulanır. Sık ürün kümeler tespit edildikten sonra ($k=2, 3, 4, \dots$), bu ürün kümelerinin kurallara dönüşmesi kolaydır, çünkü ileride ayrıntılı olarak anlatılacağı üzere, sık iki ürün kümesinin hangi olasılıkla birlikte oluştuğunu veya koşullu olarak oluştuğunu belirlemek zor değildir. Bu arada, bazı ürün kümeleri birden fazla kural oluşturabilirken bazıları hiçbir kural oluşturamayabilir (Hall, Witten ve Frank, 2011; Zaki ve Meira Jr, 2014, s.217).

Birliktelik Kuralları Analizleri ile ilginç ilişkilerin tespiti işletmelerin karar verme süreçleri için oldukça faydalı olabilir. Örneğin, pazar sepet analizleri sayesinde firma hem daha önceden müşteri kartlarıyla takibe almış olduğu müşterilerini göstermiş oldukları satın alma davranışları ile daha iyi tanıyabilecek, hem de pazarlama faaliyetlerini desteklemesi açısından, hangi ürünlerin birlikte satıldığını ve hangi ürünlerin promosyonlara (çapraz pazarlama, insert/katalog tasarımı) uygun olduğunu belirleyebilecektir. Ayrıca bu bilgi, mağazanın raf tasarımı, hususunda da önem arz etmektedir (Berry ve Linoff, 1997, s.287; Cios ve diğerleri, 2012, s.289).

Birliktelik Analizi Kurallarının ayrıntılı bir şekilde incelenmesine geçmeden önce bazı tanımlamaları yapmak gereklidir. D bir süpermarkete ait satış işlemlerinin veri tabanı olsun ve bu veri tabanındaki her bir müşteriye ait satış işlemleri (satış fişi) T olsun. $I = \{I_1, I_2, \dots, I_m\}$ bir ürün kümesi ise, $T \subseteq I$.

Bir birliktelik kuralı $A \Rightarrow B$ şeklinde gösterilmektedir. Kuralın solunda yer alan nesneye/ürüne sol taraf, sağındakine sağ taraf denir. Bu ifade basitçe *Sebze $\Rightarrow$ Meyve* olarak düşünüldüğünde, ifadenin sebze alan müşterilerin meyve de aldığı anlamı taşıdığı açıktır. Kuralın sol tarafında birden fazla ürün yer alabilir. $A \Rightarrow B$ kuralında, $A \subset I, B \subset I$ ve $A \cap B = \emptyset$ 'dir.

Bir kuralın minimum destek ve minimum güven değerini sağlaması zorunludur. Destek ve güven değerini kısaca özetlemek gerekirse, öncelikle her bir kuralın destek, güven ve kaldırmaç isimli üç değerinin mevcut olduğunu belirtmek gerekir. Bunlardan destek değeri (s), D veri tabanındaki satış fişleri arasında yer alan $A \cup B$ fişlerinin oranıdır. Bu da hem A 'yı hem B 'yi içeren fişlerin bütün fişlere oranına karşılık gelmektedir.

Destek değeri sadece kurallar için değil, ürün kümesi için de geçerlidir. O ürün kümesinin D veri tabanındaki kaç fişte yer aldığı, destek değeridir. Diğer bir değer olan güven (c), A 'yı içeren fişlerin ne kadarının B 'yi de içerdiğini gösterir. Bu değer bir koşullu olasılık değeridir. Bu iki değer dışında daha önce belirtildiği gibi bir de kaldırmaç değeri hesaplanmaktadır. Bu değerler hakkındaki yorumlara ilerleyen aşamalarda yer verilecektir.

$$\text{destek}(A \Rightarrow B) = P(A \cup B) \quad (27)$$

$$\text{güven}(A \Rightarrow B) = P(B|A) \quad (28)$$

$$\text{kaldırmaç}(A \Rightarrow B) = \frac{\text{güven}(A \Rightarrow B)}{\text{destek}(B)} = \frac{\text{destek}(A \Rightarrow B)}{\text{destek}(A)\text{destek}(B)} \quad (29)$$

Genel olarak, birliktelik kuralı analizi iki aşamadan oluşmaktadır (Aggarwal, 2015, s.98; Han ve diğerleri, 2011, s.231).

1. Kullanıcı tarafından daha önce belirlenen bir destek düzeyinin (minimum destek değeri) üzerindeki bütün sık ürün kümeleri tespit edilir.

2. Sık ürün kümelerinden güçlü birliktelik kuralları oluşturulur; kurallar kullanıcı tarafından belirlenen minimum bir destek ve güven değerinin üzerinde olmalıdır.

Destek güven ve kaldırıcı değerleri, bir kuralın kullanılabilirliği (gücü) ve kesinliği açısından o kuralın kalitesini ölçen değerlerdir. Daha önce belirtildiği gibi destek değeri, bir veri kümesinde, bir kuralın sol ve sağ tarafında yer alan ürünlerin birlikte kaç fişte (işlemede) yer aldığını tespit eder. Güven değeri ise, kuralın sol tarafındaki ürünü içeren fişlerin kaçının aynı zamanda sağ taraftaki ürünü de içerdiğini belirtir (Cios ve diğerleri, 2012, s.290).

Kaldırıcı değerinin 1'den büyük olması kuraldaki ürünler arasındaki birlikteliğin pozitif olmasını, yani A ve B ürünlerinin beklenenden daha fazla birlikte görüldüğünü, 1'den küçük olması ise aralarında negatif korelasyonun olmasını ifade eder. Değerin tam 1'e eşit olması durumunda hiçbir korelasyonun mevcut olmadığına, bağımsızlığa işaret eder. Kaldırıcı değeri, kuralın şans olarak mı ortaya çıktığını yoksa gerçekten beklenen ve iyi bir kural mı olduğunu belirlediği için, kuralın gücü olarak ifade edilir. Kaldırıcı ismini almasının sebebi ise, örneğin iki ürünlü bir kuralda kaldırıcı değeri 1'den büyükse bir ürünün satışı diğer ürünün satışını yükseltebilir yorumundan kaynaklanmaktadır (Berry ve Linoff, 1997, s.300).

Güven değeri bir kuralın, sağ tarafında hangi öğenin yer alacağını öngörmede ne kadar iyi olduğunun derecesini de ölçer. Fakat sağ taraftaki öğeler çok bilindik öğeler ise kural bizim için ilginç olmayabilir. Bu sebeple kuralın gücü olan kaldırıcı değeri farklı olarak, kuralın tamamını rassal olarak seçilmiş olan sağ taraf öğeleri ile karşılaştırır. Dolayısıyla, bir kuralın güven derecesi ile birlikte, kaldırıcı derecesinin de göz önünde bulundurulması gerekmektedir.

Birliktelik kurallarını bulmak için geliştirilen algoritmalar arasında en yaygın olanı Apriori Algoritmasıdır (Agrawal ve diğerleri, 1996). Algoritma, k -ürün kümesinin $k+1$ -ürün kümesi için kullanıldığı, en basit kuraldan başlayıp bu kurala tek tek ürünler ekleyen aşamalı bir tekniktir. Bu algoritma, verilen bir ürün kümesindeki, önceden belirlenmiş bir destek değerinin üzerinde değer alan ürün alt kümesini seçerek (ilk etapta

sık tekli ürünler seçilir) ve destek değerinin altında kalan diğer ürünleri göz ardı ederek işe başlar. İlk adımdan geçen ürünler, iki ögeli ürün kümeleri oluştururlar. Bu ürün kümelerinin hesaplanan destek değerleri de ilk başta belirlenen destek değeriyle karşılaştırılır, altında kalan ürün kümeleri göz ardı edilirler. Fakat göz ardı edilen bu ürün kümeleri ileride iki ürünlü kurallar için (kuralların sağ tarafları için) birer adaydırlar. Bu süreç, belirlenen destek değerine ulaşınca kadar sık ürünleri birleştirir, en sonunda ise artık hiçbir sık k -ürün küme bulunamaz. Sık ürün kümelerinin tespitinden sonra yine daha önceden belirlenmiş olan minimum bir destek değeri ve ayrıca bir güven değerinin üzerindeki birliktelik kuralları bulunur (Aggarwal, 2015, s.100; Giudici, 2005, s.93).

3.4.1 Birliktelik Kuralları Analizinin Süpermarket Uygulamaları

Birliktelik kuralları analizi ile ilgili yapılan çalışmalarda özellikle bilgisayar bilimciler, var olan algoritmaların etkinliklerinin geliştirilmesi ya da yeni algoritmaların keşfedilmesi hususunu ele almaktadır. Sosyal Bilimler alanındaki çalışmaların büyük bir kısmını ise perakende sektöründeki uygulamalar oluşturmaktadır. Genel olarak pazarlama stratejilerine destek olmak için yapılan birliktelik kuralları analizleri aslında sadece pazarlama stratejileri için firmaya karar verme süreçlerinde destek sağlamaz, aynı zamanda satış, satın alma, hizmet faaliyetlerinde de kullanılırlar (Chen Y. L. ve diğerleri, 2005). Süpermarketlerde yapılmış olan ve öne çıkan bazı birliktelik kuralları analizi çalışmaları şöyledir:

Brijs ve diğerleri (1999) birliktelik kurallarını, çapraz satış gibi faaliyetler için ürün seçimi konusunda kullanmış, bu teknikle firmanın kârlılığını arttırabileceklerini iddia etmişlerdir. Kullanılan veri kümesi bir mağazanın beş buçuk aylık, 27 bin 148 satış fişinden oluşan bir settir.

Chen, Tang ve diğerleri (2005) pazar sepet analizlerini, birbirinden farklı konseptteki birden çok mağazası bulunan firmalar için, Apriori algoritmasına benzer, fakat biraz daha farklı bir teknikle uygulamıştır. Farklı bir teknik uygulanmasının sebebi bir ürünün bütün mağazalarda yer almaması olasılığını göz önünde bulundurmalarındandır. Bu sebeple ortaya çıkarmış oldukları algoritma elde edilen kuralların hangi mağazaya ve hangi tarihe ait olduğunu da belirtmektedir. Kullanılan veri kümesi simülasyon ile elde edilmiştir.

Chen ve diğerleri (2005) birliktelik kuralları analizini kullanarak ilk etapta müşteri davranışlarını ortaya çıkarmış, daha sonra aynı müşteriler için farklı bir zamanda yine birliktelik kurallarını kullanarak davranışlar arasındaki farkları tespit etmiştir. Apriori algoritması ile analiz edilen veriler Microsoft SQL Serverda yer alan “Foodmart” veri kümesidir.

Rafların çok etkin tasarlanması gerektiğini vurgulayan Chen ve Lin (2007) birliktelik kuralları analizini hangi ürünlerin ön plana çıkarılacağını ve raflarda hangi ürünlere ne kadar boşluk ayrılacağını konusunda kullanmışlardır. Kullandıkları çok katmanlı birliktelik analizi ile, hem ürünler arasındaki hem de ürün kategorileri arasındaki ilişki tespit edilmiştir. Microsoft SQL Serverda yer alan “Foodmart” veri kümesini ve Apriori algoritması kullanarak yapılan uygulama sonucunda, ürünlerin yer aldığı farklı raf tasarımları şekillerle sunulmuştur. Ayrıca Özcan ve Esnaf da (2010), raf alanının etkin kullanımı ile ilgili oluşturdukları modellerde birliktelik analizlerini kullanmışlardır.

Timor ve Şimşek (2008), Türkiye’de faaliyet gösteren bir süpermarketler zincirinin 628 müşterisine ait 27 bin 924 alışveriş kayıtları ve bazı demografik özellikleri analiz etmiştir. Müşterilerin en fazla satın aldığı ürünler hedef değişken olarak tanımlanmış ve karar ağaçları kullanılarak müşterilerin satın alma davranışını en çok etkileyen değişkenler belirlenmiştir. Örneğin, şekerleme ve pasta ürünlerini alan müşteriler için “toplam değer” isimli alışveriş tutarını belirten değişken çok önemlidir. Çünkü toplam alışveriş tutarı 17 TL’nin altında olan müşterilerin %81’i bu ürünü almamıştır. Yine örneğin evlilik yılı değişkeni 22,5’den küçük olan müşterilerin %79’u şekerleme ve pasta ürünlerini satın almaktadır.

Kronberger ve Affenzeller (2011), çalışmalarında Apriori ve PRIM algoritmalarını Belçika’daki bir süpermarketin 88163 satış fişinden oluşmuş bir veri kümesi ışığında karşılaştırmalarını yapmıştır. Apriori algoritması sadece 0-1 (ikili) formatındaki veri kümelerini analiz edebilirken, PRIM algoritması nümerik değişkenli veri kümelerini de (örneğin bir müşteri bir üründen birkaç tane satın alabilir) analiz edebilme yeteneğine sahiptir. Sonuçta, her iki algoritmanın birliktelik analizleri için ideal algoritmalar olduğunu, Apriori algoritmasının büyük veri kümeleri için daha ideal çalıştığını, PRIM’in ise içerisinde nümerik değişken barındıran veri kümelerini de analiz

edebildiğini ve çok küçük destek değere sahip kuralları bile tespit edebildiğini vurgulamıştır.

Çil (2012), müşteri davranışlarını anlamada birliktelik kurallarının ne kadar önemli olduğuna vurgu yapıp, birliktelik kuralları yardımıyla mağaza planlamasını geliştirmeyi amaçlamıştır. Bu sayede mağaza planının müşteri odaklı bir plan olacağını iddia etmiştir. Çalışmada Migros süpermarketler zincirinden elde edilen bir mağazaya ait altı aylık veri kümesi kullanılmış, veri kümesi yaklaşık bir milyon satış fişinden oluşmuştur. Apriori algoritması kullanılarak elde edilen birliktelik kuralları ile, ürün kategorileri arasındaki ilişkiler bulunmuş ve çalışma sonunda ürün kategorilerinin mağaza planında (yerleşim) ne şekilde yer alması gerektiği bir şema ile gösterilmiştir.

Annie ve Kumar (2012), Hindistan'daki bir süpermarketin müşterilerini segmentlere ayırmış ve her bir segmente birliktelik kuralları analizi ile elde ettikleri sonuçlar ışığında pazarlama stratejileri uygulamışlardır. Çalışmada kullanılan veri kümesi 9620 satış fişi Apriori algoritması ile analiz edilmiştir.

Riaz ve diğerleri (2014) çapraz satış vb. pazarlama faaliyetleri için ürün seçme konusunda birliktelik kuralları analizini Apriori algoritmasını ile uygulamıştır. Çalışmada kullanılan veri kümesi ABD'deki bir süpermarketin web sitesi aracılığıyla yaptığı satışlardır. Alışveriş web sitesi aracılığıyla yapılmadan önce firma üyelik için müşterilerinden bazı demografik bilgileri istemektedir. Demografik özelliklere göre müşteri segmentasyonu da yapılan çalışmada ortaya çıkan segmentlerin satın alma davranışları birliktelik kuralları ile incelenmiş ve segmentlere özel ürünler önerilmiştir.

Güllüoğlu da (2015) yapmış olduğu çalışmada, müşteri profillerini Apriori algoritması kullanılarak ortaya çıkan birliktelik kuralları ile ilişkilendirmiş ve en sık tercih edilen ürün gruplarını cinsiyetlere göre ve yaş dağılımlarına göre belirlemiştir. Çalışmada İstanbul'daki bir süpermarketin bazı demografik özellikleri belirli 300 müşteriye ait satış verileri kullanılmıştır.

Tekin ve diğerleri (2015) Konya'daki bir süpermarketten elde ettikleri bir yıllık, 962 bin satış fişinden oluşan veri kümesini kullanmışlardır. Bu veri kümesi ile müşterilerin ürün kategorileri bazındaki satın alma davranışları Apriori algoritması ile

incelenmiş, sonuçlar firmanın pazarlama faaliyetlerini desteklemek için ayrıntılı bir şekilde incelenmiştir. Örneğin et ürünlerinin ve günlük gıdaların ilişkili olduğu ürün grubu taze sebzeler iken, hazır giyim grubunun ise bisküvi grubu olduğu tespit edilmiştir.

3.4.2 Kuralların Kümelenmesi Çalışmaları

Bu bölümde, uygulama bölümünde yapılacak olan segmentasyon çalışmasına konu olan, birliktelik kurallarının kümelenmesi ile ilgili yapılmış olan bazı çalışmalara yer verilecektir. Birliktelik kuralları ile elde edilen ilişki kurallarının kümelenmesi fikri ile ilgili çok az sayıda çalışma vardır.

Lent ve diğerleri (1997) süpermarket dışındaki verileri, birliktelik kuralları için analiz ettiğimizde yüzlerce veya binlerce kural bulabileceğimizden bahsettikten sonra, kuralları kümeleme fikrini ortaya atmıştır. İlk önce iki tane sol tarafta yer alacak ürün seçilmesini ve bu ürünlere sahip olan kuralların kümelenmesini tavsiye etmektedir. Buna iki değişkenli birliktelik kurallarının kümelenmesi ismini veren araştırmacılar, uygulamada süpermarket verisi kullanmamış ve kümeleme analizi için uygulanan algoritmayı ise kendileri geliştirmiştir.

Kosters ve diğerleri (1999) ilk etapta belirli bir destek değerinin üzerindeki kuralları tespit etmiştir. Daha sonra bu kurallar arasından en iyi kural; kuralın destek değerine ve güven değerine bakılarak tespit edilmiştir. Kuralın sol tarafındaki ürünü/ürünleri satın alan müşteriler bir kümede toplanmışlardır. Daha sonra bu kümenin elemanları veri kümesinden kaldırılmış, süreç başka bir kural için işletilmiştir. Bu model ilki 800 müşterinin, ikincisi 10 bin müşterinin satış verilerinin yer aldığı iki veri kümesinde denenmiştir. Elde edilen kümelerin anlamlı olduğu çalışmada, Apriori algoritması ve hiyerarşik kümeleme tekniği kullanılmıştır.

Strehl ve diğerleri de (1999) birliktelik kuralları analizinin çoğu zaman 10 binden daha fazla kural oluşturduğunu, bu kuralların bir şekilde azaltılması gerektiğini savunarak, kuralları kümeleme yoluna gitmişlerdir. Bu amaçla oluşacak kurallar arasındaki uzaklığı ölçecek yeni bir uzaklık ölçüsü bulan yazarlar ayrıca Çok Boyutlu Ölçekleme (MDS) ve Kendini Düzenleyen Haritalar (SOM) kümeleme tekniğinden faydalanarak yeni bir kümeleme yöntemi de elde etmişlerdir. 172 bin satış fişinin

oluşturduğu süpermarket verisi ile uygulama yapılan çalışmada, kurallar 19 kümede toplanmışlardır.

Reynolds ve diğerleri (2004) “adults” isimli çok bilinen bir veri kümesini birliktelik kurallarını bulmak amacıyla “all-rules” isimli algoritma ile analiz etmiş ve başka veri kümelerinde olduğu gibi, bu veri kümesinde de binlerce kural üretilebildiğini tespit etmiştir. Kuralların kümelenerek daha iyi anlaşılabilceğini iddia ederek, k-medoids tekniği ile kuralları kümelere ayırmıştır.

Hahsler ve Hornik (2007), bir süpermarketin satışlarına ait veri kümesini ilk önce kümeleme analizi ile sekiz kümeye ayırmış ve anlamlı bulduğu iki kümeyi incelemeye almıştır. Satışlara ait veri kümesini kümeleme fikri ile birbirine benzer satın alma davranışlarının aynı kümelerde yer alacağını iddia etmiştir. Gerçekten de kümelerin birinde müşteriler bir ürünü sıkça satın almışlardır. Kümeleme tekniği olarak PAM algoritmasının birliktelik kuralları analizi için de Apriori algoritmasının kullanıldığı çalışmadaki veri kümesi 30 gün boyunca bir süpermarkette yapılan 9835 satış fişini kapsamaktadır.

Plasse ve diğerleri de (2007) veri kümelerinde elde edilen birliktelik kuralları sayısının çok fazla olduğuna dikkat çekerek birliktelik kurallarının kümelenmesi fikrini savunmuşlardır. İlk önce veri kümesindeki belirli bir destek ve güven değerinin üzerindeki birliktelik kuralları bulunmuştur. Daha sonra değişken kümeleme yoluyla, birliktelik kuralları kümelenmiş, ortaya çıkan her bir küme içerisindeki birliktelik kuralları incelenmiştir. İki farklı veri kümesinin kullanıldığı çalışmada, süpermarket verilerine sahip ikinci veri kümesi için birliktelik kuralları üretilmiş, bu kurallar farklı kümeleme teknikleri kullanılarak kümelenmiştir. Kümeleme tekniklerinin ve ürettikleri sonuçların detaylı bir şekilde incelendiği çalışmada, her bir tekniğin oluşturmuş olduğu kümelerde tekrar birliktelik kuralları araştırılmıştır. Sonuçta, belirli bir destek ve güven değerinin altındaki en az sayıda kural veren tekniğin başarılı bir kümeleme tekniği olduğuna karar verilmiştir.

4. ZİNCİR PERAKENDE MAĞAZALARI İÇİN MAĞAZA SEGMENTASYONU VE HEDEF PAZARLAMA UYGULAMASI

4.1 AMAÇ VE KAPSAM

Türkiye’de faaliyet gösteren bir süpermarketler zinciri firmasına ait, İstanbul ili sınırları içerisinde bulunan 175 mağaza hedef pazarlama stratejileri amaçları doğrultusunda segmentlere ayrılacaktır. Segmentlerin tespitinden sonra segmentlerin satış verilerine uygulanacak olan Birliktelik Kuralları Analizleri ile segmentlere özel, sık satılan ürün grupları ve ürün ilişkileri tespit edilecektir.

Segmentasyon işi Kümeleme Analizi kullanılarak yapılacak olup, analizde iki farklı yaklaşıma başvurulacaktır. Yaklaşımların farklılıkları kümeleme analizi için girdi olacak olan değişkenlerin seçiminden kaynaklanmaktadır. İlk yaklaşımdaki değişkenler Türkiye İstatistik Kurumu, Google Haritalar ve bazı emlak şirketlerine ait web siteleri aracılığıyla farklı kaynaklardan bir araya getirilen sosyoekonomik değişkenlerdir. Diğer yaklaşımda kullanılacak olan değişkenler ise firmanın satış verilerinin birliktelik kuralları analizi yoluyla elde edilecek olan bazı kuralların destek, güven ve kaldırma değerleridir.

4.2 VERİ KÜMESİ VE VERİ ÖN İŞLEME SÜRECİ

Segmentasyon için başvurulacak olan ilk yaklaşımda kümeleme analizine girdi olacak sosyoekonomik değişkenler firmaya ait 175 mağazanın adresleri kullanılarak farklı kaynaklardan elde edilmiştir. İlk etapta bu mağazaların ticari çevresi; mağazanın yakınlarında bir veya birkaç fabrika, ticaret merkezi, üniversite ve turistik merkez bulunup bulunmadığı tespit edilerek incelenmiştir. Bu inceleme Google Haritalar kullanılarak yapılmış ve sonunda ikili değişkenler (0: Hayır, 1: Evet) oluşturulmuştur. Bu değişkenlerin aldığı değerler, örnek teşkil etmesi amacıyla beş mağaza için Tablo 9’da sunulmuştur.

Google haritalar yardımı ile ikinci etapta mağazaların yakın çevresinde faaliyet gösteren rakip mağazaların sayısı da tespit edilmiştir. Örneğin incelenen bölgede firmanın büyük ölçekli bir mağazası yer alıyorsa, rakip olarak yine büyük ölçekli mağazaların yakın çevrede yer alıp almadığı, varsa kaç tane yer aldığı tespit edilmeye çalışılmıştır. Bu

amaçla tek tek 175 mağazanın çevresinde yer alan rakip mağazalarının sayısı bulunmuş ve rakip mağaza sayısı değişkeni oluşturulmuştur. Yine Tablo 9'dan bu değişkenlerin aldığı değerler örnek beş mağaza için incelenebilir.

Tablo 9 Uygulamada Kullanılacak Olan Değişkenler ve Beş Mağazaya Ait Gözlem Değerleri

Değişkenler	Mağaza 1	Mağaza 2	Mağaza 3	Mağaza 4	Mağaza 5
<i>Ev Kirası</i>	1200	2200	900	3000	600
<i>Fabrika Bölgesi</i>	0	0	0	1	0
<i>Üniversite Bölgesi</i>	1	0	0	0	1
<i>Ticaret Merkezi</i>	0	0	0	0	0
<i>Turistik Bölge</i>	0	0	0	1	0
<i>Rakip Sayısı</i>	5	1	7	1	1
<i>Yaş 0-4</i>	0.0854	0.0546	0.0668	0.0376	0.0928
<i>Yaş 5-14</i>	0.1714	0.1502	0.1268	0.0749	0.1702
<i>Yaş 15-34</i>	0.4498	0.2675	0.3363	0.2632	0.3647
<i>Yaş 35-54</i>	0.2272	0.3631	0.3069	0.3246	0.2921
<i>Yaş 55+</i>	0.0663	0.1646	0.1631	0.2996	0.0802
<i>Bekâr</i>	0.5062	0.4475	0.4156	0.6010	0.3724
<i>Evli</i>	0.4938	0.5525	0.5844	0.3990	0.6276
<i>Eğitim Yok</i>	0.1826	0.0932	0.0908	0.0579	0.1356
<i>Eğitim Düşük</i>	0.3832	0.2049	0.4821	0.2252	0.5622
<i>Eğitim Orta</i>	0.3224	0.2246	0.2629	0.2954	0.1757
<i>Eğitim Yüksek</i>	0.1118	0.5021	0.1642	0.4481	0.1283

Ayrıca yine mağazaların adresleri kullanılarak Türkiye İstatistik Kurumu (TÜİK) tarafından yayınlanan Adrese Dayalı Nüfus Sistemi ile mağazaların yer aldığı mahallelerdeki demografik yapı ile ilgili veriler elde edilmiştir. TÜİK'ten İstanbul ilinde bulunan bütün mahallelere ait istatistikler satın alınmış, mağazaların bulunduğu 175 mahallenin verileri ham veri kümesinden süzölmüştür.

Adrese Dayalı Nüfus Sisteminde, yaş dağılımı (0-4 yaşındaki insan sayısı, 5-9 yaşındaki insan sayısı, 10-14 yaşındaki insan sayısı...) medeni durum (bekâr insan sayısı, dul insan sayısı...) ve eğitim durumu (okuryazar olmayan insan sayısı, ilkokul mezunu olan insan sayısı, ilköğretim mezunu olan insan sayısı...) gibi mahalle sakinlerinin demografik özelliklerine ait istatistikler mevcuttur.

TÜİK'ten elde edilen verilerde bulunan değişkenler, firmanın pazarlama faaliyetleri ve segmentasyon analizi sonrası elde edilecek sonuçların kolay yorumlanabilir olması açısından tekrar düzenlenmiştir. TÜİK verilerine göre, 14 gruba ayrılmış olan yaş değişkeni (0-4, 5-9, 10-14...65 ve üzeri), pazarlama literatüründe hakkında birçok çalışma yapılmış ve halen yapılmakta olan X, Y ve Z kuşakları göz önüne alınarak, beş grupta (0-4, 5-14, 15-34, 35-54, 55 ve üzeri) birleştirilmiştir. Aynı şekilde eğitim seviyesi değişkeni de TÜİK ham verilerinde on grup iken, yine sonuçların kolay yorumlanabilmesi açısından dört grupta (eğitim yok, eğitim seviyesi düşük, eğitim seviyesi orta, eğitim seviyesi yüksek) toplanmıştır. Her bir gruba ait değerler TÜİK verilerinde insan sayısı olarak hesaplanmasının aksine, oran şeklinde hesaplanmıştır. Böylece Tablo 9'da gösterildiği gibi demografiyle ilgili değişkenlere ait değerler elde edilmiştir.

Mağazaların bulunduğu mahallelerde yaşayan insanların ekonomik durumları hakkında TÜİK vb. kurumlar aracılığıyla oluşturulmuş herhangi bir veri kümesi yoktur. Mahalle sakinlerinin ekonomik durumlarının bölgede bulunan konutların aylık kira bedelleri ile doğru orantılı olduğu düşünülmüş ve bu bedellerin tespiti ise bazı emlak şirketlerinin web siteleri aracılığıyla yapılmıştır. Tablo 9'da ev kirası değişkeni Türk Lirası cinsinden oluşturulmuştur.

Farklı farklı kaynaklardan bir araya getirilen bu sosyoekonomik değişkenlerle ilk yaklaşımda kullanılacak olan kümeleme analizinin girdisini oluşturacak toplamda 17 değişken elde edilmiştir.

Kullanılacak olan değişkenlerin tipleri incelendiğinde, kümeleme analizi için ihtiyaç duyulacak uzaklık matrisinin, farklı tipteki değişkenleri içeren veri kümelerine uygun bir uzaklık ölçüsü ile elde edilmesi gerektiği anlaşılmaktadır.

Tablo 10 Uygulamada Kullanılacak Olan Değişkenlerin Tipleri

Değişken	Değişken Tipi
<i>Ev Kirası bedeli (TL)</i>	<i>Nümerik</i>
<i>Fabrika Mevcut mu?</i>	<i>Kategorik</i>
<i>Üniversite Mevcut mu?</i>	<i>Kategorik</i>
<i>Ticaret Merkezi Mevcut mu?</i>	<i>Kategorik</i>
<i>Turistik Bölge Mevcut mu?</i>	<i>Kategorik</i>
<i>Rakip Sayısı</i>	<i>Nümerik</i>
<i>0-4 yaş aralığındaki insanların oranı</i>	<i>Nümerik</i>
<i>5-14 yaş aralığındaki insanların oranı</i>	<i>Nümerik</i>
<i>15-34 yaş aralığındaki insanların oranı</i>	<i>Nümerik</i>
<i>35-54 yaş aralığındaki insanların oranı</i>	<i>Nümerik</i>
<i>55 ve üzeri yaş aralığındaki insanların oranı</i>	<i>Nümerik</i>
<i>Bekâr insanların oranı</i>	<i>Nümerik</i>
<i>Evli insanların oranı</i>	<i>Nümerik</i>
<i>Hiç eğitim görmemiş insanların oranı</i>	<i>Nümerik</i>
<i>Düşük seviyede eğitim görmüş insanların oranı</i>	<i>Nümerik</i>
<i>Orta seviyede eğitim görmüş insanların oranı</i>	<i>Nümerik</i>
<i>Yüksek seviyede eğitim görmüş insanların oranı</i>	<i>Nümerik</i>

Segmentasyon için başvurulacak olan ikinci yaklaşımda kümeleme analizine girdi olacak olan değişkenler ise firmanın 175 mağazasının tamamına ait bir günlük satış verilerden oluşmuş olan veri kümesinden elde edilmiştir. Veri kümesine uygulanan Birliktelik Kuralları Analizi sonucunda elde edilecek bazı ürün ilişkileri kurallarının değerleri değişkenlerimizi oluşturmaktadır.

Bu amaçla ilk önce 175 mağazanın tamamının satış işlemlerinden oluşan genel veri kümesinde bazı ön işlemler yapılması gerekmektedir. Veri kümesinin ön işlemesi sadece segmentasyon işi için gerekli olmayıp aynı zamanda kümeleme analizi sonucunda elde edilecek olan segmentler için pazarlama stratejileri geliştirmek amacıyla yapılacak olan Sık Satılan Ürün Grupları ve Birliktelik Kuralları Madenciliği için de gerekmektedir.

Veri kümesinde 175 mağazanın bir günlük satışlarında yer alan toplam 23 bin 910 farklı ürün bulunmaktadır. Fakat aynı markaya ait örneğin yarım litrelik süt ile 330 ml’lik sütün barkodunun farklı olmasından ötürü bu ürünlerin ayrı ürünler olarak yer aldığı gözden kaçırılmamalıdır. Hâlbuki bu ürünler “süt” başlığı altında yer alabilir. Aksi takdirde bu kadar fazla sayıdaki ürüne birliktelik kuralları analizi uygulandığında destek ve güven değerleri yüksek güçlü ve anlamlı kurallar elde edilemeyecektir. Bu amaçla ürünler tek tek yeniden kodlanarak toplam 103 ürün grubunda sınıflandırılmıştır.

Firmaların veri tabanlarında genel olarak ürünlerin tam olarak tanımlayıcı isminin (örneğin X markasına ait 200 ml. muzlu süt) yanı sıra ürünün hangi kategoride (örneğin süt ürünleri) yer aldığı bilgisi de yer almaktadır. Fakat ürünlerin bu kadar genel gruplarda toplanması ve Birliktelik Kuralları Analizlerinin bu genel gruplarla yapılması ise çok genel ilişki kurallarının elde edilmesine sebep olmaktadır (örneğin; *süt ürünleri* => *kuru gıda* ilişkisi).

Bu nedenle daha açık ilişki kurallarının elde edilebilmesi açısından oluşturulan 103 ürün grubu genel kategorilerin aksine biraz daha ayrıntılı olarak oluşturulmuştur. Tablo 11’de bahsi geçen 103 ürün grubu sunulmuştur. Gruplar oluşturulurken izlenen yol genel olarak şu şekildedir. Farklı markalara ait bütün ebatlardaki sade sütler “süt” olarak, meyveli sütler “meyveli süt” olarak, farklı markalara ait bütün tipteki çikolatalar “çikolata”, farklı şekerli bisküvi çeşitleri “tatlı bisküvi”, tuzluları ise “tuzlu bisküvi” olarak kodlanmıştır. Aynı şekilde her çeşit sebze “sebze” olarak, meyveler “meyve” olarak kodlanmıştır.

Gruplandırma yapılırken analiz sonuçlarında farklı ilişkilerin elde edilebilmesi açısından bazı ürün grupları daha önce belirtildiği gibi detaylandırılmıştır. Örneğin meyveli sütler belirtildiği gibi süt başlığı altında birleştirilmeyip “meyveli süt” olarak, temizlik ile ilgili ürünler “çamaşır deterjanı”, “bulaşık deterjanı”, ayrıca çamaşır suyu ve yer temizleme ürünleri gibi sıvı veya toz diğer temizlik ürünleri “temizlik”, temizlik için kullanılan bez, paspas, fırça gibi alet ve malzemeler de “temizlik malzemesi” adı altında sınıflandırılmıştır. Ketçap, mayonez salça gibi bütün çeşit soslar “sos”, tenekedeki ürünler “konserve”, donmuş hazır gıdalar “donmuş gıda”, diş temizliği ile ilgili ürünler “diş”, tıraş ürünleri “tıras” şeklinde kodlanmıştır.

Tablo 11 Ürünlerin Sınıflandırılması

Alkol	Çay	Gevrek	Kömür	Pasta malzemesi	Spor	Tereyağı
Altın	Çerez	Giyim	Kuru meyve	Peçete	Su	Toz tatlı
Ayakkabı	Çikolata	Hayvan yemi vb.	Makarna	Ped vb.	Sucuk	Tıraş
Ayran	Çorba	Hazır yemek	Mama	Peynir	Süt	Turşu
Baharat	Diş ürünü	Hindi	Margarin	Poğaça	Süt meyveli	Tuvalet kâğıdı veya havlu
Bakliyat	Dondurma	Islak mendil	Meyve	Poşet	Şampuan	Tuz
Bal	Donmuş gıda	İç giyim	Meyve suyu	Reçel vb	Şeker	Tuzlu bisküvi
Balık	Duş jeli vb.	Kahve	Meyveli yoğurt	Sabun	Şekerleme	Yağ
Bebek ürünü	Eczane	Kaşar	Meze	Sakız	Takı	Yoğurt
Bira	Ekmek	Kek	Mobilya	Salam	Tatlı	Yumurta
Bira şişesi	Elektronik	Kırtasiye	Kondom	Sebze	Tatlı bisküvi	Zeytinyağı
Bitki çay	Et	Kitap	Oto gereç	Sıcak içecek	Tavuk	Zeytin
Bulaşık deterjan	Ev gereçleri	Konserve	Oyuncak	Sigara	Tekstil	Zücaciye
Cips	Gazete	Kontör	Özel ekmek	Sos	Temizlik	
Çamaşır deterjan	Gazlı içecek	Kozmetik	Parfüm	Sosis	Temizlik malzemesi	

4.3 MAĞAZA SEGMENTASYONU UYGULAMASI

Uygulamada kullanılacak her iki segmentasyon yaklaşımı da ardıl segmentasyon (post hoc) yaklaşımına uymaktadır.

4.3.1 Sosyoekonomik Bazlı Segmentasyon Yaklaşımı Uygulaması

Segmentasyon analizi için başvurulacak olan ilk yaklaşımda, Kümeleme Analizinde kullanılacak olan sosyoekonomik değişkenlerin belirlenip her bir mağaza için aldığı değerler de tespit edildikten sonra, uzaklık ölçüsünün belirlenmesi aşamasına geçilir. Değişkenlerin tipinin farklı olmasından ötürü Gower'in ortaya çıkarmış olduğu

uzaklık ölçüsü kullanılarak hesaplanması daha uygun olan uzaklık matrisi *R Programlama Dilindeki* bazı işlevler sayesinde gerçekleştirilmiştir.

R Programlama Dili herkesin erişimine açık bir sistemdir ve ücretsizdir. Gönüllüler tarafından oluşturulmuştur. Günümüzde Veri Analizi konusunda oldukça yaygın kullanılmakta ve günden güne içeriği gelişmektedir. Erişime açık olduğu için katkı yapmak isteyen herkes sistemde program paylaşılabilmektedir. Birçok veri madenciliği tekniğinin hali hazırda programda yazılmış olması *R Programlama Dili*'nin kullanımını programlama dili bilmeyenler için bile kolaylaştırmaktadır. Kullanımı konusunda sayısız kitaplar, el kitapçıkları, videolar ve internet ortamındaki forumlar vardır.

R'daki *cluster* paketinde yer alan *daisy* işlevi yardımıyla ihtiyaç duyulan uzaklık matrisi kolayca elde edilmiştir. Bu işlev sayesinde veri kümeleri için Öklidyen ve Manhattan uzaklık ölçüleri de hesaplanabilmektedir. İlk etapta 17 değişkenin 175 mağazanın her biri için tespit edilen değerlerinin oluşturduğu veri matrisi R programına aktarılmış, daha sonra da değişkenlerin tipleri programa tanıtılmıştır.

Değişkenler arasındaki korelasyonun tespiti için, literatürde birçok uygulamada kullanılmış olan R programlama dilinin *polycor* paketindeki *hetcor* işlevi sayesinde farklı tipte değişken tiplerini barındıran veri kümeleri için korelasyon değerleri hesaplanmaktadır. *Hetcor* işlevi nümerik tipteki değişkenler arasında Pearson çarpım-moment korelasyonu, nümerik ile kategorik değişkenler arasında Polyserial korelasyonu, kategorik değişkenler arasında ise, Polychoric korelasyonu hesaplamaktadır. Bu hesaplamalar sonucunda bekar oranı değişkeni ile evli oranı değişkeninin beklendiği gibi -1 korelasyona sahip olduğu, yaş 0-4 ile yaş 5-14 değişkenlerinin ayrıca düşük eğitimli oranı ile yüksek eğitimli oranı değişkenlerinin ise 0,90 korelasyon değerinden daha yüksek değerlere sahip olduğu tespit edilmiştir. Bu sebeple yaş 5-14 değişkeni, bekâr oranı değişkeni ve düşük eğitimli oranı değişkeni, uzaklık matrisi hesaplanmadan önce veri kümesinden çıkartılmıştır. Daha sonra segmentlerin özelliklerinin yorumlanması aşamasında bu değişkenler veri kümesinde tekrar göz önünde bulundurulmuştur.

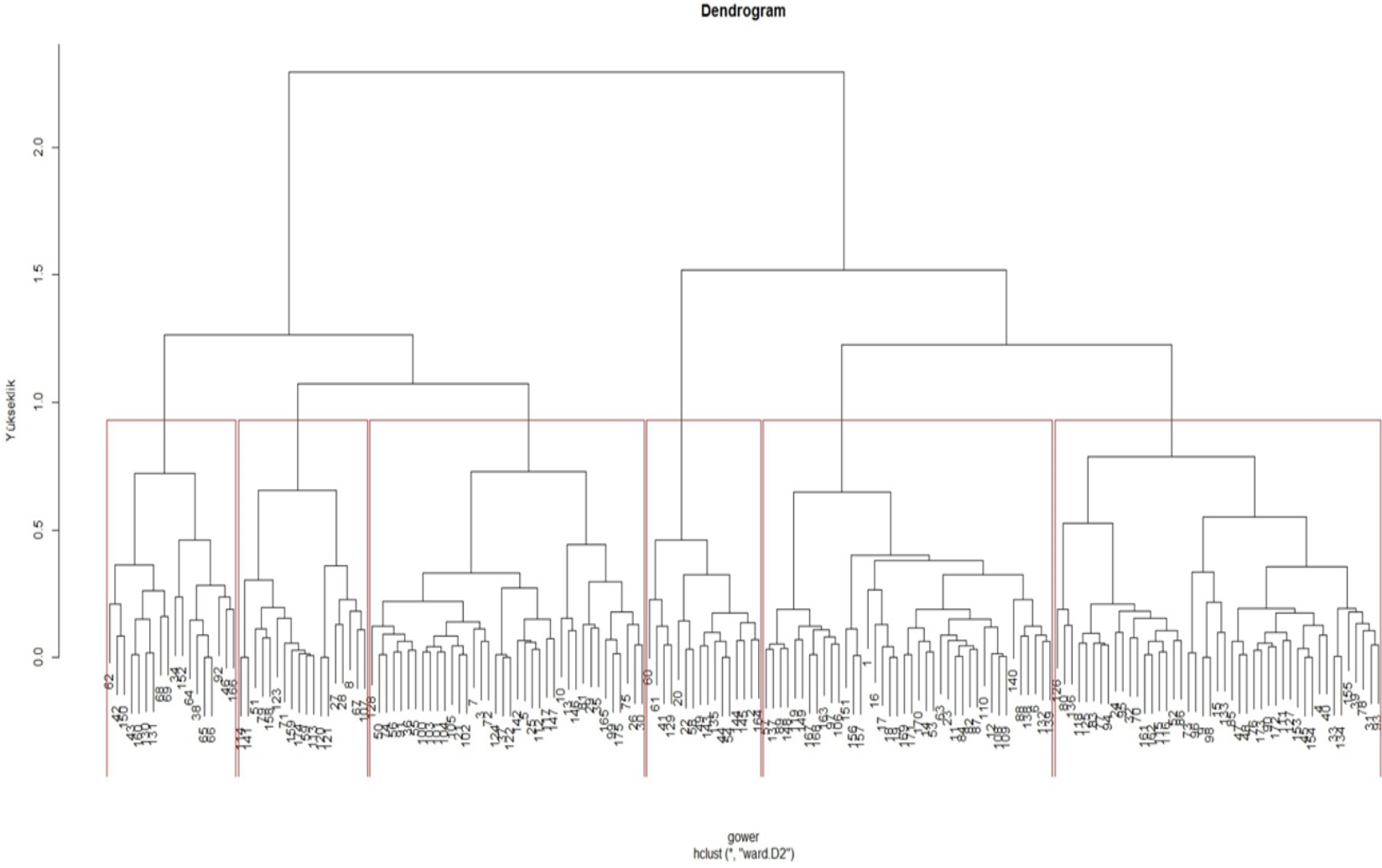
Hesaplanan 175×175 boyutundaki uzaklık matrisi Hiyerarşik Kümeleme Tekniklerinden biri olan Ward'ın Kümeleme Algoritması için girdi olarak kullanılmıştır.

Ward'ın algoritması için yine R'daki *stats* paketinde yer alan *hclust* işlevi kullanılmıştır. Bu işlev sayesinde uzaklık matrislerine Hiyerarşik Kümeleme Yöntemleri uygulanabilmekte ve sonuçlar dendrogramda görselleştirilebilmektedir. Ward'ın algoritmasının oluşturduğu dendrogram Şekil 12'de sunulmuştur.

Yine R Programlama Dilindeki *cluster* paketinde yer alan *agnes* işlevi sayesinde Birleştirici Katsayı %97 olarak hesaplanmıştır. Bu yüksek değer bize kullanılmış olan Ward'ın algoritmasının veri kümemizde açık küme yapıları bulunduğunu ispatlamaktadır.

Diğer taraftan programdaki *stats* paketinde yer alan *cophenetic* işlevi ile Kophenetic Korelasyon Katsayısı % 55 (Spearman Korelasyon ile) olarak tespit edilmiştir. Bu değer hiyerarşik algoritmanın ortaya çıkardığı sonucun orijinal veri kümesini orta derecede yansıtabildiğini göstermektedir. Fakat dendrogramın ortaya çıkardığı kümelerin özellikleri incelendiğinde ve çalışmamızın amacı kümeleme algoritmalarının performanslarını karşılaştırmanın aksine pazarlama alanında bir uygulama yapmak olduğu için bu sonuç iyi bir değer olarak değerlendirilebilir.

Gerçekten de dendrogramdan elde edilebilecek farklı kümeler incelendiğinde 175 mağazanın altı segmente ayrılması her bir grupta farklı değişkenlerin baskın olması ve dolayısıyla her bir grubun farklı özelliklerinin bulunması sebebiyle anlamlı görülmüştür. Dendrogramda ağacın kökünde firmanın sahip olduğu 175 mağaza görülmektedir.



Şekil 11: Sosyoekonomik Bazlı Segmentasyon Yaklaşımıyla Elde Edilen Dendrogram

Dendrogramdan alınan sonuçla mağazalar ait oldukları segmentlerde toplanmıştır. Segmentlerin özelliklerinin tespit edilmesi amacıyla her bir sosyoekonomik değişkenin segmentler için ortalama değerleri hesaplanmış ve Tablo 12’de sunulmuştur. Daha sonra bu değerler yardımıyla segmentlerdeki baskın değişkenler tespit edilmiş ve segmentlerin özellikleri Tablo 13’de incelenmiştir.

Tablo 12 Değişkenlerin Segmentler Bazındaki Ortalama Değerleri

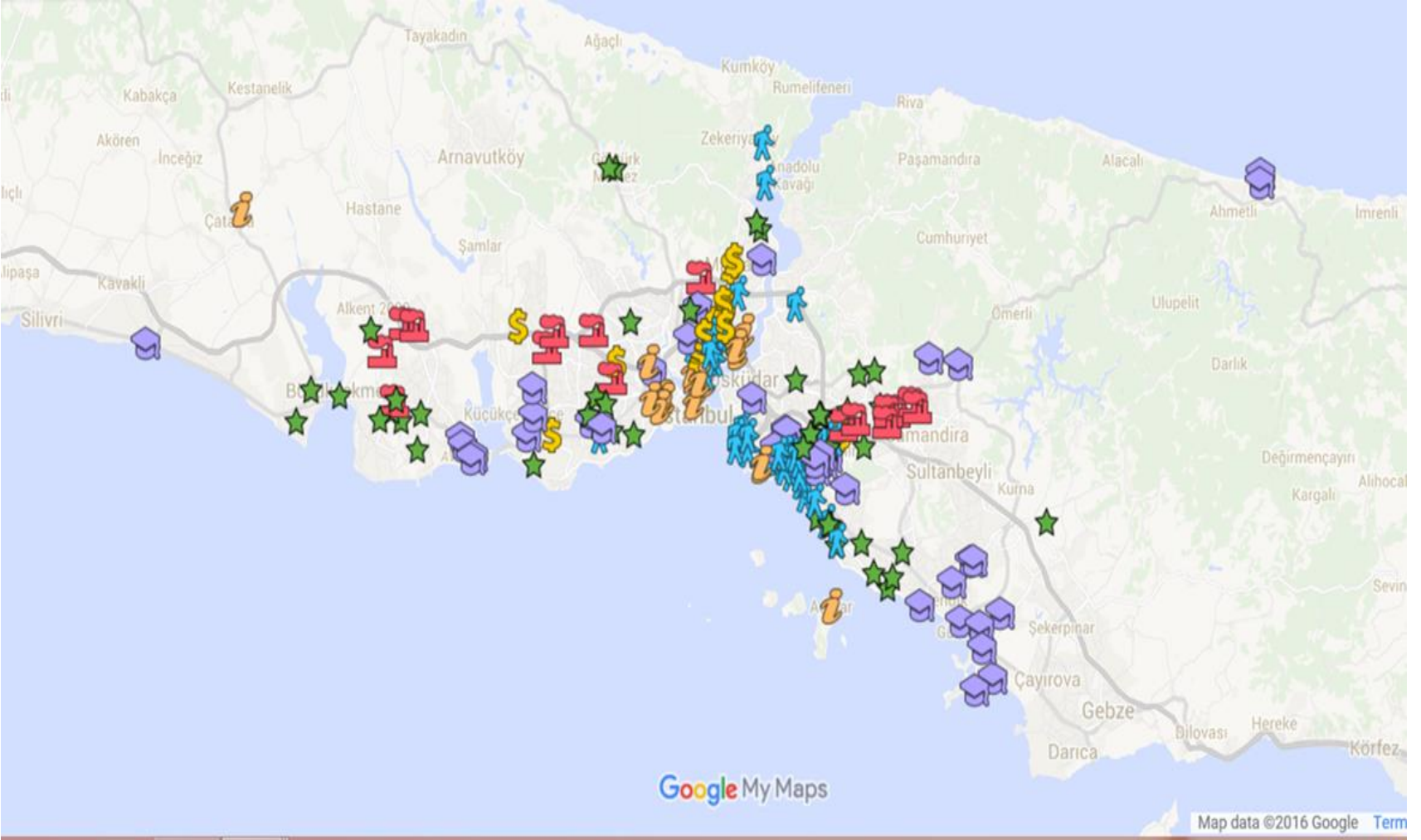
Değişkenler	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5	Segment 6
<i>Ev Kirası (TL)</i>	1096	1090	1833	1159	1775	1791
<i>Fabrika bölgesindeki mağaza sayısı</i>	0	15	0	2	0	1
<i>Üniversite bölgesindeki mağaza sayısı</i>	40	3	4	1	7	3
<i>Ticaret Merkezine yakın mağaza sayısı</i>	0	12	0	1	18	0
<i>Turistik Bölgedeki mağaza sayısı</i>	3	0	0	1	0	18
<i>Ortalama Rakip</i>	2,7	2,3	2,6	2,5	1,3	1,8
<i>Yaş 0-4</i>	0.07	0.10	0.04	0.07	0.05	0.05
<i>Yaş 5-14</i>	0.13	0.17	0.09	0.14	0.10	0.10
<i>Yaş 15-34</i>	0.34	0.37	0.28	0.33	0.33	0.33
<i>Yaş 35-54</i>	0.30	0.28	0.32	0.32	0.32	0.31
<i>Yaş 55+</i>	0.16	0.08	0.27	0.14	0.20	0.23
<i>Bekâr</i>	0.44	0.38	0.53	0.42	0.53	0.59
<i>Evli</i>	0.56	0.62	0.47	0.58	0.47	0.41
<i>Eğitim Yok</i>	0.11	0.15	0.07	0.11	0.08	0.10
<i>Eğitim Düşük</i>	0.44	0.56	0.26	0.42	0.33	0.37
<i>Eğitim Orta</i>	0.25	0.18	0.28	0.25	0.26	0.25
<i>Eğitim Yüksek</i>	0.20	0.12	0.41	0.23	0.34	0.29

Tablo 13 Segmentlerin Özellikleri

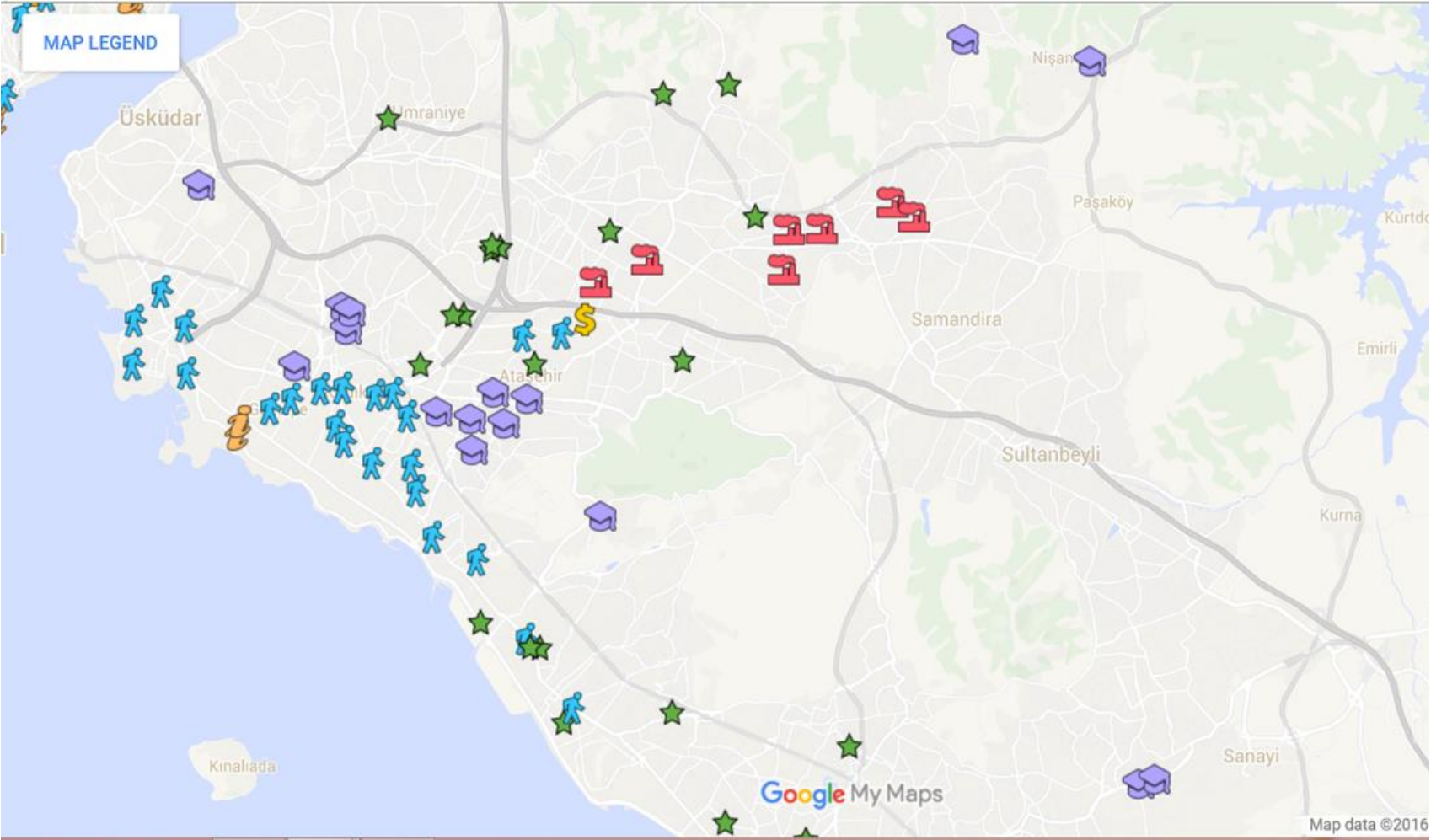
Segment	Sembol	Mağaza sayısı	Özellikler (BASKIN DEĞİŞKENLER)
1		40	Düşük düzeydeki ev kiralari, mağazaların hepsi üniversite bölgesinde, rekabet yüksek düzeyde, 5-14 yaş ve 15-34 yaş oranı yüksek, eğitim seviyesi düşük
2		16	Ev kiralari düşük, mağazalar fabrika ve ticaret bölgesinde, 0-4 yaş, 5-14 ve 15-34 yaş oranı çok yüksek, yüksek evli oranı, eğitim seviyesi çok düşük
3		38	Yüksek ev kiralari, rekabet orta düzeyde, çok yüksek 55+ yaş oranı, yüksek bekâr oranı, eğitim seviyesi çok yüksek
4		45	Orta ev kirası, yüksek evli oranı, rekabet yüksek, eğitim seviyesi düşük
5		18	Yüksek ev kirası, düşük rekabet düzeyi, mağazalar ticaret bölgesinde
6		18	Ev kiralari yüksek düzeyde, mağazalar turistik bölgelerde, rekabet düzeyi düşük, 55+ yaş oranı yüksek, yüksek bekâr oranı

Firmanın, 175 mağazasının hepsine birden mağaza ve müşteri ihtiyaçlarını gözetmeksizin kitle pazarlama stratejileri geliştirmesi günümüz şartları açısından uygun değildir. Günümüze uygun olan strateji ise küresel rekabet ortamında avantaj sağlamaya ve firmanın pazarda tutunabilmesine destek olacak olan hedef pazarlama stratejisidir. Tablo 13’den de görüldüğü üzere, mağazaların oluşturduğu altı grubun birbirinden farklı özellikleri mevcuttur.

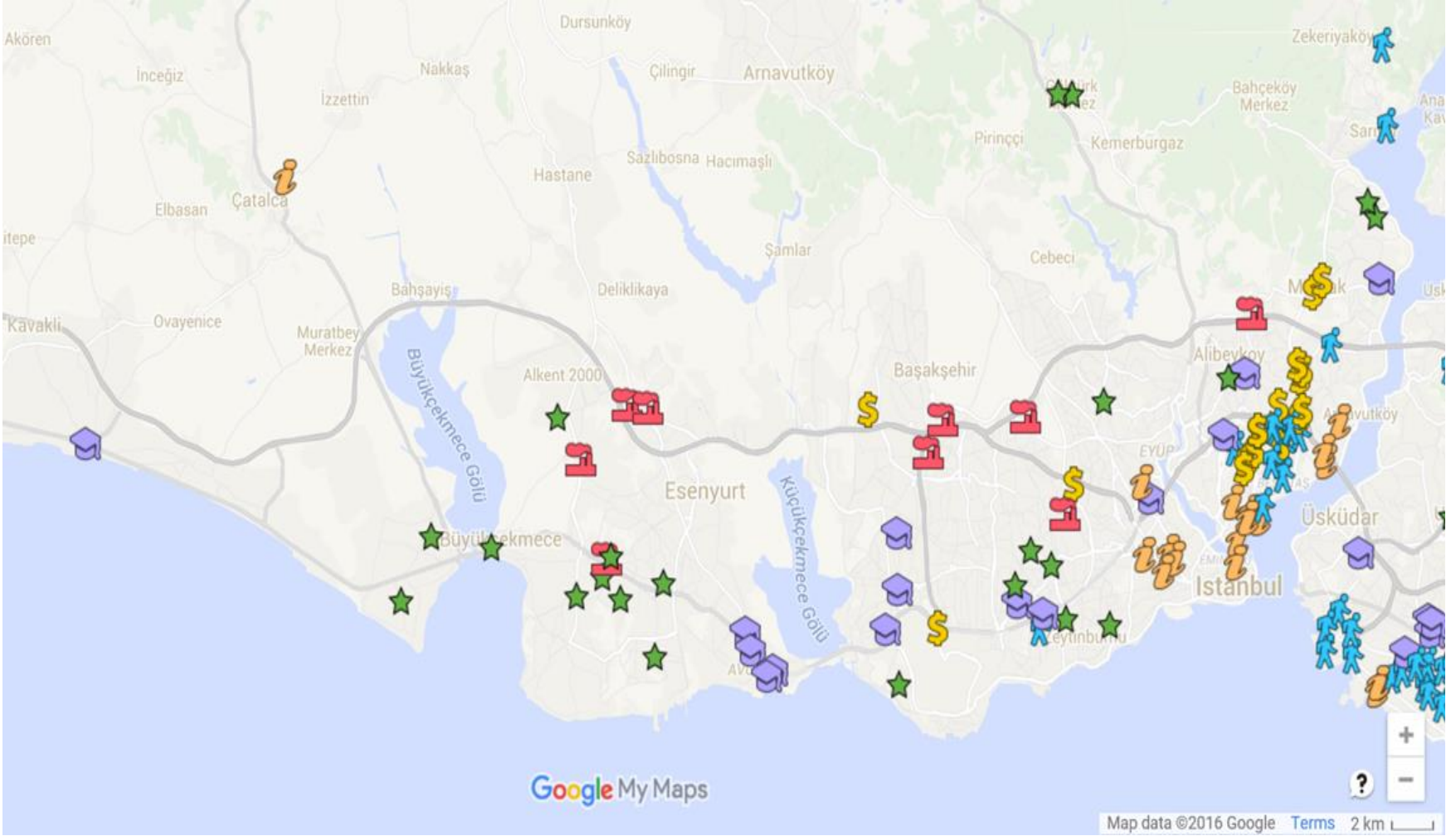
Aşağıdaki şekillerde ise sırasıyla mağazaların İstanbul haritasındaki konumları, Asya tarafındaki ve Avrupa tarafındaki konumları, ait oldukları segmentin sembolleri ile gösterilmiştir. Dikkat edileceği üzere konum olarak birbirine çok yakın olan mağazalar bile farklı segmentlerde yer alabilmektedir.



Şekil 12: İstanbul Genelindeki 175 Mağazanın Ait Oldukları Segment Sembolleriyile Gösterilmiş Konumları



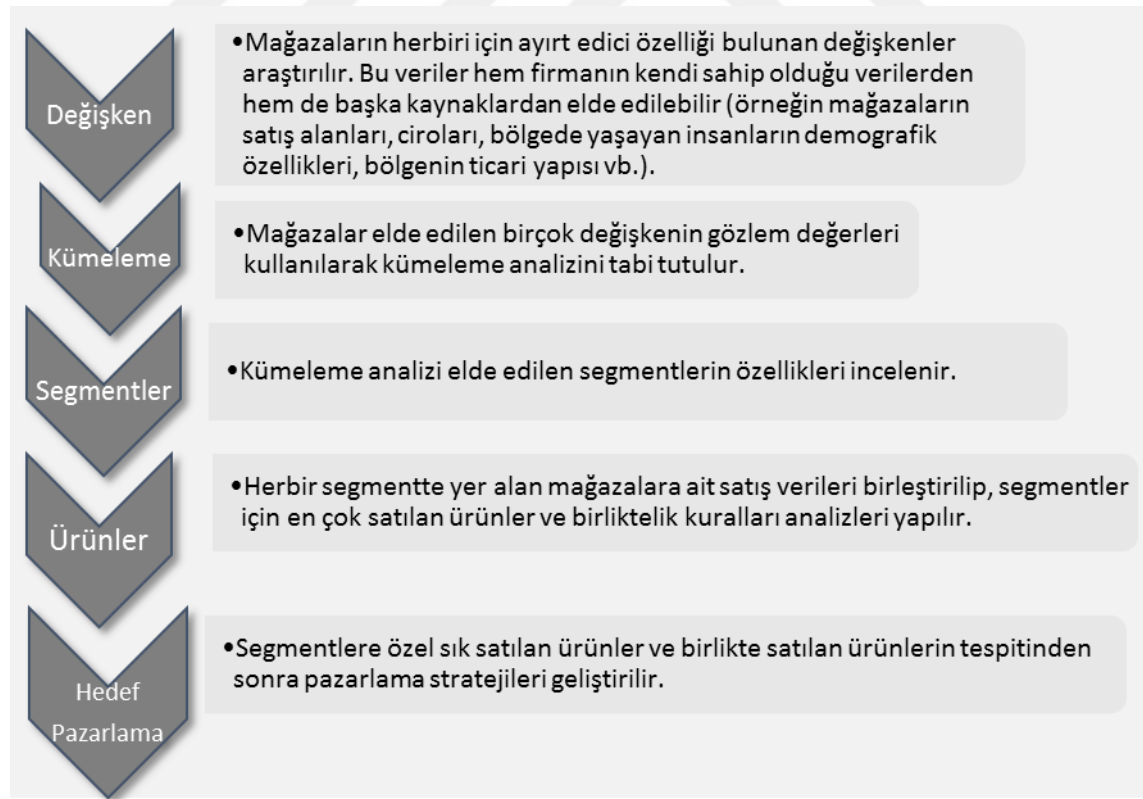
Şekil 13: Asya Yakasinda Mağazaların Sembollerile Dağılımı



Şekil 14: Avrupa Yakasında Yer Alan Mağazaların Sembolleriyile Dağılımı

İlk segmentteki mağazalar incelendiğinde mağazaların hepsinin üniversite bölgesinde olduğu (sembol olarak kep seçilmiştir) ve mağazaların bulunduğu mahallelerde yaşayan insanlar arasında gençlerin oranının yüksek olduğu, 2. Segmentteki mağazaların çoğunun ya ticari bölge ya da fabrika bölgesinde olduğu (sembol olarak fabrika seçilmiştir), bu mağazaların bulunduğu mahallelerdeki insanlar arasında evli oranının yüksek olduğu görülmektedir. Eğitim seviyesi ve emekli insan sayısının yüksek olduğu 3. Segmentte (sembol olarak yaşlı insan seçilmiştir) ayrıca insanların ekonomik durumlarının da iyi olduğu söylenebilir. 5. Segmentte insanların ekonomik durumu iyi ve mağazaların çevresinde ticaret merkezleri (sembol olarak dolar işareti seçilmiştir) mevcuttur. Rekabetin yoğun olmadığı 6. segmentteki mağazaların çoğu turistik bölgelerde (sembol olarak turizm ofisi seçilmiştir) yer almakta, bu bölgelerde yaşayan insanlardan bekâr oranı ve emekli olanların oranı yüksek görülmektedir.

Birinci yaklaşımla yapılmış olan segmentasyon çalışmasında izlenmiş olan yol haritası Şekil 16 'da gösterilmiştir.



Şekil 15: Farklı Kaynaklardan Elde Edilmiş Verilere Dayalı Segmentasyon Yaklaşımı için Yol Haritası

Uygulamanın dördüncü bölümünde (Başlık 4.4) bu segmentlerin bir günlük satış verileri Birliktelik Kuralları Analizi ile analiz edilecektir. Segmentlerin satış verileri, her bir segmentte yer alan mağazaların satış verilerinin birleştirilmesi ile elde edilmiştir. Örneğin birinci segmentte yer alan 40 mağazanın satış verileri tek bir veri kümesine birleştirilmiş ve oluşan veri kümesi vurgulandığı gibi Birliktelik Kuralları Analizi kullanılarak, pazarlama faaliyetleri için analiz edilmiştir. Bu inceleme her bir segment için ayrı ayrı yapılmıştır.

4.3.2 Kural Bazlı Segmentasyon Yaklaşımı Uygulaması

Segmentasyon analizi için başvurulacak olan ikinci yaklaşımda kümeleme analizinde girdi olarak kullanılacak olan değişkenler, 175 mağazanın tamamına ait satış verilerine Birliktelik Kuralları Analizi uygulanarak elde edilecek olan ürünler arasındaki ilişki kurallarının destek, güven ve kaldırma değerleridir.

Bu bağlamda, ilk etapta, mağazaların tamamına ait satışların oluşturduğu genel veri kümesinden, Birliktelik Kuralları Analizi yoluyla belirli bir destek ve güven değerinin üzerindeki güçlü ilişki kuralları tespit edilecektir. Daha sonra bu kuralların performansları; yani bu kuralların ilgili mağazalarda var olup olmadığı, eğer var ise bu kuralların destek, güven ve kaldırma değerlerinin ne olduğu, tek tek bütün mağazalar için araştırılacaktır. Bu kuralların aynı zamanda müşteri satın alma davranışları olduğu söylenebilir. Çünkü sebze alan bir müşterinin şekerleme alması (*sebze => şekerleme* kuralı) bir satın alma davranışıdır. Kümeleme analizi sonucunda performansları bu kurallara göre benzer olan, yani müşterileri benzer satın alma davranışı sergileyen mağazaların aynı grupta yer alması hedeflenmektedir.

Kuralları belirlemede *R* programlama dilindeki *arules* paketi kullanılacaktır. Daha önce ayrıntılı olarak açıklandığı üzere, ürünler tekrar kodlanarak gruplandırılmıştır. *R* programı yardımıyla veri kümesi birliktelik kuralları yapılabilecek hale; 0 ve 1'lerden oluşan (eğer herhangi bir satış fişinde herhangi bir ürün varsa 1, yoksa 0 değeri alacak şekilde) veri kümesine dönüştürülmüştür. Bu işlemi sağlayan işlevler *split* ve *as* işlevleridir. Kuralların tespit edilmesi ise *apriori* işlevi ile yapılmaktadır.

Kümeleme Analizi için kaç kuralın temel alınacağı yani kaç kuralın Kümeleme Analizine değişken olacağı ve kurallarda kaç öge (2 ürünlü, 3 ürünlü veya 4 ürünlü) yer alacağı soruları için yapılan denemeler sonucunda şu bulgulara rastlanmıştır: üç ürüne sahip kuralların çoğunun mağazaların büyük bir kısmında yer almadığı dolayısıyla kuralların çok az sayıdaki mağazada geçerli olduğu saptanmıştır. Bu amaçla ilk önce kararı verilen nokta, genel veri kümesinden sadece iki ürünlü kuralların elde edilmesi gerektiğidir. Yine yapılan denemeler sonucunda destek değeri 0,001 ve güven değeri 0,2'nin üzerinde ayrıca kaldırıcı değerlerinin ise bütün kurallar için bir değerinin üzerinde olduğu toplam 412 kuralın Kümeleme Analizi için temel olarak alınabileceği tespit edilmiştir.

Bu kadar fazla sayıdaki kuralın temel olarak alınmasının sebebi mümkün olduğunca farklı kuralların elde edilip, satışlar açısından birbirine benzer olan mağazaların daha doğru tespit edilebilmesini sağlamaktır. Çünkü bazı ürünlerin mağazaların tamamında raflarda olmaması veya bazı ürünlerin satış analizlerinin yapıldığı tarihte satılamamış olması, o ürünle ilgili kuralların bazı mağazalarda destek güven ve kaldırıcı değerlerinin 0 olmasına sebebiyet vermektedir. Tablo 14'de genel veri kümesinden elde edilmiş bahsi geçen 412 kuraldan 12'sine yer verilmiştir.

Kitap => çikolata kuralından başlayarak geri kalan 411 kural 175 mağazanın her birinin satış verilerinde araştırılmıştır. Bu 412 kural sütunlarda yer alacak ve her bir mağaza da satırlarda yer alacak şekilde, bu kuralların aldığı değerler (destek, güven ve kaldırıcı değerleri) her bir mağaza için hesaplanmış ve 175×412 boyutlu üç ayrı veri matrisi elde edilmiştir. Eğer herhangi bir kural veya kurallar mağazalarda yoksa destek, güven ve kaldırıcı değerleri 0 olarak kabul edilmiştir.

Veri matrislerinde yer alan değerlerin hepsinin nümerik olması sebebiyle uzaklık ölçüsü olarak Öklidyen uzaklık ölçüsünün kullanımı uygun görülmüştür.

Kuralların mağazalardaki performansını ölçen destek, güven ve kaldırıcı değerlerinden hangisinin Kümeleme Analizinde kullanılacağı hususuna, her üç değer kullanılarak uygulanan Kümeleme Analizleri sonucunda karar verilmiştir. Her bir mağazadaki bahsi geçen 412 kuralın güven değerleri ve kaldırıcı değerlerinin oluşturduğu

veri kümeleriyle yapılan Kümeleme Analizleri sonucunda bu değerlerin istenilen şekilde kümeler oluşturamadığı, ya sadece iki ya da onlarca sayıdaki kümenin oluştuğu gözlemlenmiştir.

Tablo 14 Segmentasyon için Kullanılacak Kural Örnekleri

Kural	Destek Değeri	Güven Değeri	Kaldıraç Değeri
$\{kitap\} \Rightarrow \{çikolata\}$	0.0010	0.2095	1.5708
$\{bal\} \Rightarrow \{meyve\}$	0.0016	0.3009	2.2835
$\{dondurma\} \Rightarrow \{çikolata\}$	0.0025	0.2140	1.6046
$\{traş malzemesi\} \Rightarrow \{sebze\}$	0.0016	0.2000	1.1479
$\{zeytinyağı\} \Rightarrow \{çikolata\}$	0.0018	0.2273	1.7046
$\{mama\} \Rightarrow \{meyve\}$	0.0026	0.2076	1.5752
$\{çorba\} \Rightarrow \{tatlı bisk.\}$	0.0023	0.2037	2.1371
$\{un\} \Rightarrow \{süt\}$	0.0038	0.3377	4.0250
$\{margarin\} \Rightarrow \{gazlı içe\}$	0.0025	0.2206	1.9894
$\{donmuş gıda\} \Rightarrow \{yoğurt\}$	0.0032	0.2005	3.6590
$\{gevrek\} \Rightarrow \{süt\}$	0.0074	0.4017	4.7871
$\{çamaşır det.\} \Rightarrow \{bulaşık det.\}$	0.0063	0.2487	11.8619

Kuralların mağazalardaki destek değerlerinin oluşturduğu veri matrisi ile yapılan Kümeleme Analizi sonucunda ise uygun sayılarda kümeler elde edilebileceği görülmüş, bu sebeple sadece kuralların mağazalardaki destek değerlerinden oluşan veri matrisi kullanılarak Kümeleme Analizi yapılmasına karar verilmiştir.

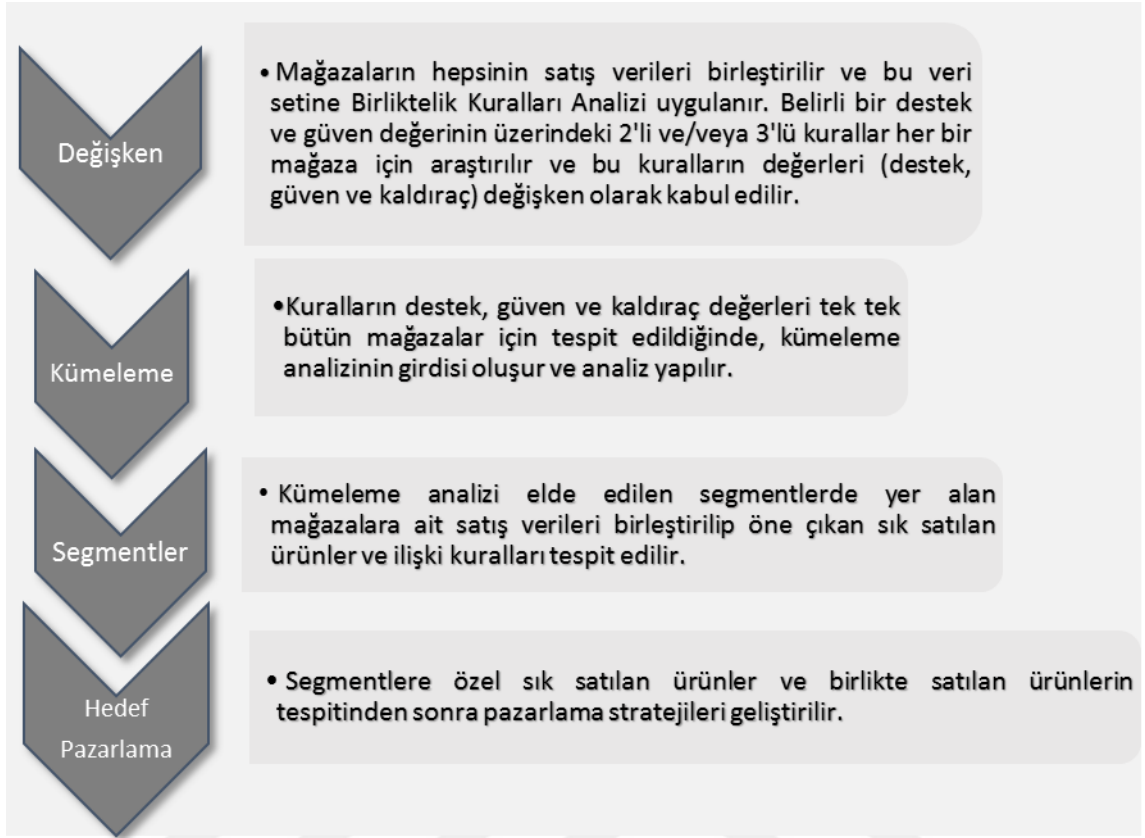
Kümeleme analizi yapmadan önce, mağazalar için 412 kuralın destek değerlerinin birbirleri ile olan korelasyonları tespit edilmiştir. Korelasyonu diğer kurallarla 0,90 değerinden yüksek 12 kural, R programlama dilinde yer alan *stats* paketinin içerisindeki *cor* işlevi ile tespit edilip veri kümesinden kaldırılmıştır. *Cor* işlevi Pearson'ın korelasyon katsayısını hesaplamaktadır.

Yine, R programlama dilinde yer alan *stats* paketindeki *dist* işlevi ise uzaklık matrisini Öklidyen ölçüsünü kullanarak hesaplayabilmektedir. Elde edilen 175×175 boyutlu uzaklık matrisi, kümeleme algoritmasının girdisini oluşturmuş, daha önce olduğu gibi Ward'ın tekniği kullanılarak Şekil 18'deki dendrogram elde edilmiştir. Dendrogramda ağacın kökünde numaralandırılan 175 mağaza görülmektedir.

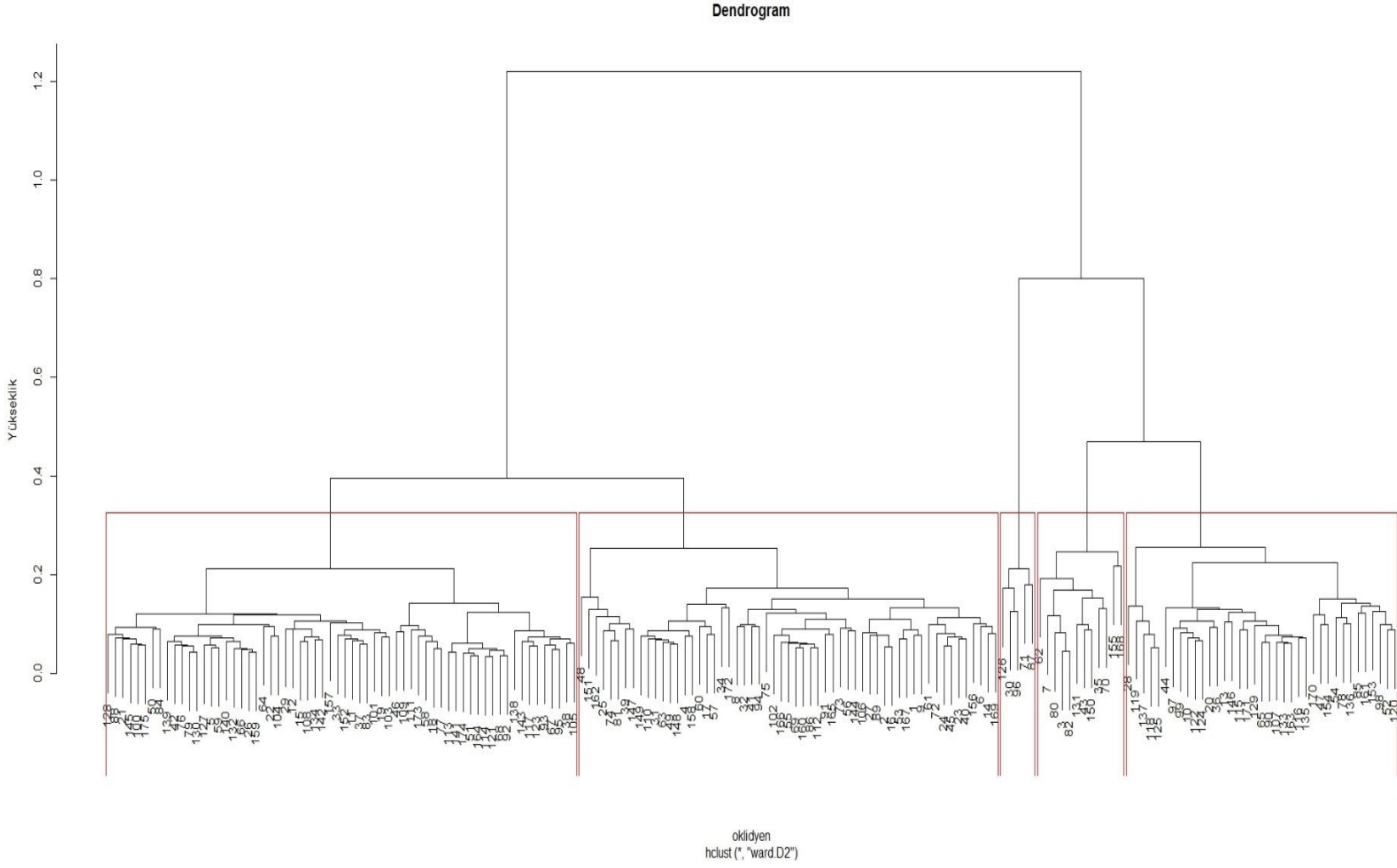
Birleştirici Katsayı bu yaklaşımda ilk yaklaşıma göre biraz daha düşük fakat yine de çok yüksek bir değer olan % 93 olarak hesaplanmıştır. Bu yüksek değer bize yine Ward'ın algoritmasının veri kümemizde açık küme yapıları bulunduğunu ispatlamaktadır.

Kophenetic Korelasyon Katsayısı ise ilk yaklaşıma nazaran daha yüksek olarak % 70 (Spearman Korelasyon ile) tespit edilmiştir. Bu değer hiyerarşik algoritmanın ortaya çıkardığı sonucun orijinal veri kümesini iyi derecede yansıtabildiğini göstermektedir.

Segmentasyon için temel alınan 400 kuralın mağazalardaki performanslarına göre diğer bir ifade ile sahip oldukları destek değerlerine göre mağazaların beş segmente bölünmesi uygun görülmektedir. Böylece bu beş ayrı grupta yer alan mağazalar, bahsi geçen 400 kuralın satış/satın alma sıklığına göre birbirine benzerdir sonucuna varılabilir. Daha açık ifade ile örneğin {kitap} => {çikolata}, {bal} => {meyve} gibi birbirinden farklı 400 kuralın, her bir mağazada ne kadar müşteri tarafından gerçekleştirilmiş olduğu (sıklık) hesaplanmış ve bu sıklıkların birbirine benzer olanları aynı segmentte yer almıştır.



Şekil 16: Kural Bazlı Segmentasyon Yaklaşımı için Yol Haritası



Şekil 17: İkinci Yaklaşımla Elde Edilen Kümeleme Analizine Ait Dendrogram

4.3.3 İki Yaklaşımın Karşılaştırılması

Kümeleme sonuçlarının karşılaştırılması için kullanılan ölçülerden olan çiftleri sayma yöntemindeki endekslerin hesaplanması Microsoft Excel programında yapılabilmektedir. İhtiyaç olunan, her iki kümeleme yaklaşımında mağazaların hangi kümelerde yer aldığının sayılması ile hazırlanan kontenjans tablosudur.

Tablo 15 İki Yaklaşımın Elde Edilen Kümelere Ait Eşleştirme Tablosu

	1	2	3	4	5	6	Toplam
1	18	6	12	14	3	4	57
2	2	0	3	3	0	4	12
3	13	5	15	11	11	9	64
4	6	5	7	15	3	1	37
5	1	0	1	2	1	0	5
Toplam	40	16	38	45	18	18	175

Rand Endeksi, Düzeltilmiş Rand Endeksi, Fowlkes-Mallows'un Endeksi Excel yardımıyla hesaplanmıştır. Fakat bütün bu hesaplamalar ve hesaplanması görece karmaşık olan bir çok ölçünün R programlama Dilinde de karşılıkları mevcuttur.

R'de yer alan "mclust" paketindeki "arandi" fonksiyonu sayesinde Rand Endeksi ve Düzeltilmiş Rand Endeksi değerleri hesaplanmıştır. Hesaplamalara göre Rand Endeksi değeri **0,63** iken Düzeltilmiş Rand Endeksi ise **0,0043** bulunmuştur. Rand Endeksinin, düzeltilmişinden çok daha yüksek çıkması rastlanan bir durumdur. Bunun sebebi şu şekilde açıklanabilir: Rand Endeksi sıfır ile bir ($0 \leq R \leq 1$) arasında değer aldığı için endeksin beklenen değeri ya sıfıra eşit ya da sıfırdan büyük olmaktadır. Diğer taraftan Düzeltilmiş Rand Endeksinin beklenen değeri sıfırdır ve endeksin maksimum değeri ise Rand Endeksinde olduğu gibi birdir. Dolayısıyla Düzeltilmiş Rand Endeksinin daha geniş bir aralıkta değer alması olağandır ve bu durum endeksin hassasiyetini arttırmaktadır (Yeung ve Ruzzo, 2001, s.2). Bu ölçünün anlamlılığı hakkında soru işaretleri de mevcuttur çünkü dağılımlara ilişkin güçlü varsayımlara dayanmaktadır (Wagner ve Wagner, 2007, s.5). Çiftleri sayma yöntemindeki ölçüler hakkında problemlerin var olduğu çeşitli çalışmalarda vurgulanmaktadır (Meilă, 2007, s.875). Literatürde en iyi uygulama yapabilmenin yolu henüz gösterilmemiş ve çalışmalar

yöntemleri uygulayıp yorumunu okuyucuya bırakmışlardır (Romano ve diğerleri, 2015, s.2).

“Clues” paketindeki “adjustedRand” işlevi sayesinde Rand Endeksinin aynı değeri verdiği, ayrıca Jaccard Endeksi ve Fowlkes-Mallows endekslerinin ise sırasıyla **0,13** ile **0,235** değerlerini hesaplandığı görülmüştür.

Fowlkes-Mallows değerini hesaplamak için hazırlanan başka bir paket olan “dendextend” paketindeki “FM_index” işlevi sayesinde hem endeks **0,2367** olarak tekrar hesaplanmış hem de endeksin Beklenen Değeri ve Varyansı da sırasıyla **0,2333** ile **0,0000493** olarak hesaplanmıştır.

Daha önce vurgulandığı üzere iki kümeleme tekniğinin karşılaştırılması ile ilgili akademik yazında, ele alınan veri kümeleri hem veri noktaları açısından hem de kullanılan değişkenler açısından birbirinin aynısıdır (Rabbany ve Zaiane, 2015; Romano ve diğerleri, 2014; Romano ve diğerleri, 2015; Yeung ve Ruzzo, 2001). Bu çalışmaların ortak amacı aynı veri kümesi üzerinde uygulanan birden çok tekniği karşılaştırarak bu tekniklerin performanslarını görelî olarak ölçmektir. Çalışmamız bu açıdan ilgili yazınla farklılık göstermektedir. Fakat daha önce de vurgulandığı gibi, Xiang ve diğerleri (2012) bu yöntemlerin farklı veri setlerinde de kullanılabilmesi gerektiğini beklemektedir.

Çalışmamızın hem sosyoekonomik değişkenlere dayalı ilk yaklaşımında hem de birliktelik kurallarına dayalı ikinci yaklaşımında veri noktaları (nesneler) süpermarket mağazalarıdır. Her iki yaklaşımda da kümelemesi yapılan mağaza sayısı aynı ($n = 175$) olmasına karşın kümelemelere girdi olan değişkenler birbirinden çok farklıdır.

Bu sebeple iki yöntemin karşılaştırılmasındaki amaç, uygulamamızda kullanılan kümeleme algoritmalarının performanslarını değerlendirmek değil sosyoekonomik yapı ile tüketim davranışı arasında bir ilişki olup olmadığını iki yaklaşımla elde edilen sonuçların uyumu üzerinden açıklamaya çalışmaktır. Beklentimiz, iki yaklaşım ile elde ettiğimiz kümelerde zayıf da olsa bir uyum gözlemlemektir. Çünkü bu tarz bir uyum benzer sosyoekonomik çevreye sahip mağazalarda benzer tüketim davranışlarının gerçekleştiği anlamına gelecektir.

Çalışmamızın sonuçlarına göre Rand endeksi iki yaklaşımımız arasında yüksek bir uyum olduğuna işaret etmektedir. Ancak doğrudan Rand endeksi ile yapılacak yorum yanıltıcı olacak bu yüzden şaştan arındırılmış uyumu veren Düzeltilmiş Rand endeksini ölçü almak daha uygun olacaktır. Bu ölçüye göre iki kümeleme yaklaşımımızın sonuçları arasında bir uyumdan söz etmek oldukça güçtür.

Fowlkes-Mallows endeksi dikkate alınacak olursa hesaplanan beklenen değer ve varyans doğrultusunda iki kümeleme yaklaşımımız sonuçları arasında zayıf da olsa bir uyumdan söz etmek mümkündür.

Tüm ölçütlerden elde edilen sonuçları genelleştirecek olursak kullanmış olduğumuz veri kümesi üzerinde uygulanan iki farklı yaklaşımın uyum içinde olduğunu iddia edebilmek için yeterli delil olmadığını söyleyebiliriz. Ancak veri kalitesini daha üst seviyelere taşıyarak ya da her iki yaklaşımda da girdi değişkenlerinin sayısını ve niteliğini arttırarak yeterli delil elde edilebileceğine dair ipuçları söz konusudur.

Unutulmamalıdır ki, gerek sosyoekonomik yaklaşımda kullandığımız değişkenlere ait verilerin farklı kaynaklardan ve el yordamı ile toplanmış olması nedeniyle, gerekse kural bazlı yaklaşımda kullanmış olduğumuz verilerin süpermarket zincirinin sadece bir günlük işlemlerine ait verileri olması nedeniyle ulaşacağımız bulguların çok yüksek kalitede olması beklenemezdi. Bununla birlikte, her iki kümeleme yönteminin de belirli düzeyde başarılı sonuçlar üreteceği yani bazı veri noktalarının hatalı kümelenmiş olabileceği de açıktır. Ve son olarak, sosyoekonomik çevre ile tüketim davranışı arasında bir ilişki olsa bile bu ilişkinin kusursuz bir açıklayıcılık göstermesini beklemek rasyonel bir yaklaşım olmazdı.

Bu gerçekler ışığında iki yaklaşımın ortaya koyduğu sonuçlar arasındaki uyumun farklı ölçütlerde farklı düzeylerde ortaya çıkmış olması şaşırtıcı değildir. Kaldı ki, iki kümeleme yöntemi arasındaki uyuma ilişkin ölçütlerin halen geliştirilmeye çalışıldığı ve akademik yazında bu ölçütlerden herhangi birinin önerilmesi yönünde görüş birliği bulunmadığı da bir başka gerçektir.

4.4 PAZARLAMA STRATEJİLERİ İÇİN SIK ÖĞE KÜME MADENCİLİĞİ VE BİRLİKTELİK KURALLARI MADENCİLİĞİ

Uygulama bölümünün bu alt bölümünde iki farklı yaklaşımla elde edilmiş olan biri altı segment, diğeri beş segment için, sık satılan ürünlerin ve ürünler arası ilişkilerin belirlenebilmesi amacıyla Birliktelik Kuralları Analizi uygulanacaktır. Bu sayede hem firmanın segmentlere uygulayacağı pazarlama faaliyetlerine yardımcı olunacak hem de satışlar/satın almalar açısından segmentler arasındaki farklılıklar da tespit edilmeye çalışılacaktır.

Bu analizler için izlenecek olan yol şu şekildedir. İlk etapta her bir segment için Sık Öğe Küme Madenciliği ile sık satılan ürün ve ürün grupları tespit edilecek, daha sonra yine her bir segment için ürünlerin ve ürün gruplarının ilişkileri Birliktelik Kuralları Analizi ile tespit edilecektir. Tablolar halinde sunulacak olan bu analizlerde, her defasında en çok satılan ilk 20 tek öğeli, çift öğeli ve üç öğeli ürün kümeleri ile ayrıca güven ve kaldırmaç değerleri en yüksek ilk 20 iki ürünlü ve üç ürünlü kurallar araştırılacaktır.

Bu amaçla elde edilen sonuçların hepsi ayrıntılı bir şekilde Ek 1 ve Ek 2’de yer almaktadır.

4.4.1 Sosyoekonomik Bazlı Segmentlere Ait Analizler

İlk segmentasyon yaklaşımı olan farklı kaynaklardan elde edilmiş sosyoekonomik verilerle segmentasyon sonucunda ortaya altı farklı segment çıkmış ve bu segmentlerin özellikleri Tablo 13’de gösterilmiştir.

Bu segmentleri oluşturan mağazalara ait satış verileri birleştirilerek, örneğin birinci segmente yer alan 40 mağazanın satış verileri tek bir veri kümesi altında toplanmış ve birinci segment için Sık Öğe Küme Madenciliği ve Birliktelik Kuralları Analizleri uygulanmıştır. Bu uygulamalar her bir segment için yapılmış ve segmentler arasında tespit edilen farklılıklar aşağıdaki tablolarda gösterilmiştir.

İlk tablomuz segmentlere özel sık satılan ürün kümeleri tablosudur (Tablo 16). Bu tablo oluşturulurken, her bir segmentin satış verileri, en çok satılan ilk 20 ürün veya ürün grubunu (sık ürün küme) elde edebilmek için analiz edilmiş (destek değerleri sık satılan 20 ürünü elde edebilmek için ayarlanmıştır) ve sadece bir segmentte olup

diğerlerinde hiç olmayan ürün/ürünler, segmente has ürün/ürünler olarak tabloya eklenmiştir.

Tablo 16 Sosyoekonomik Bazlı Segmentlere Özel Sık Öğe Kümeleri Tablosu

Segment	Mağaza Sayısı	Toplam Fiş Sayısı	Segmente has sık ürün	Segmente has sık iki öğeli küme	Segmente has sık üç öğeli küme
1	40	34998	Tavuk Et	-	Meyve, Sebze, Tavuk Gazlı içecek, Peynir, Sebze
2	16	9779	Cips	Gazlı İçecek, Sigara Çikolata, Sigara	Ekmek, Meyve, Sebze
3	38	29980	Yumurta	Meyve, Yoğurt Peynir, Süt	Meyve, Peynir, Süt Özel ekmek, Peynir, Süt
4	45	42193	-	-	-
5	18	17783	Hazır yemek	Çikolata, Su Çikolata, Kek Çerez, Çikolata	Özel ekmek, Sebze, Süt Çikolata, Süt, Tatlı bisküvi Gazlı içecek, Sebze, Süt Çerez, Çikolata, Gazlı içecek
6	18	14669	-	Meyve, Tatlı bisk.	Peynir, Sebze, Yoğurt

Tablo 16'dan da görüldüğü gibi, birinci segment için tavuk ürünü ve et ürünü sadece bu segmentte sık satılan ürünlerdir. Daha açık bir ifade ile diğer segmentlerde tavuk veya et ürünü sık satılmış ilk 20 ürün arasında yer almamıştır. Aynı şekilde 2. Segment için cips ürünü, 3. Segment için yumurta ürünü ve 5. Segment için hazır yemek ürünü sadece bu segmentlerde sık satılmıştır. 4. ve 6. segmentlerde hiçbir ürünün yer alması, bu segmentlerde sık satılan ürünlerin diğer segmentlerde de yer almasından kaynaklanmaktadır. Segmentlere özel sık satılmış iki öğeli ürün kümeleri ve üç öğeli ürün kümeleri tablodaki yerin almıştır. Örneğin 2. Segment incelendiğinde, gazlı içecek ve sigara ürünü sadece bu segmentte sık satılmış 2'li ürünler listesinde yer almıştır. Yine çikolata ve sigara ürünleri de bu segmente has sık satılmış 2'li ürünlerdir. Ekmek, meyve ve sebze ürünleri ise yine sadece 2. Segmente özel sık satılan üç öğeli ürün kümeleridir. 5. Segmentte çikolata ürünü ile su, kek ve çerez ürünlerinin satışına diğer segmentlerin

sık satışlarında rastlanmamıştır. Bu sonuçlar firmanın pazarlama faaliyetleri için destekleyici olabilirler.

4. segmentte hiçbir özel durumun söz konusu olmaması bu segmentin diğer segmentlerden ürün satışları konusunda farklılık göstermediğini ortaya koymaktadır. Yine de firma 4. Segmente ait Ek 1’de yer alan analizleri inceleyerek pazarlama faaliyetlerini destekleyebilir.

Bu bölüme ait aşağıda yer alan ikinci tablo ise, her bir segment için yapılan Birliktelik Kuralları Analizinden sonra segmentlerde ortaya çıkan kural farklılıklarını, diğer bir ifade ile segmentlere özel ürün ilişkileri kurallarını göstermektedir. Segmentler için Birliktelik Kuralları Analizi yapılırken, güven değerleri en yüksek ilk 20 iki ve üç ürünlü kurallar elde edilmiştir. Destek değeri ise kuralların daha güçlü olması için, her defasında 100 fişin altına düşmeyecek şekilde ayarlanmıştır. Segmentler için tek tek yapılmış olan bu analizler ayrıntılı bir şekilde Ek 1’de yer almaktadır.

Her bir segment için, güven değerleri yüksek kuralların tespit edilmesi ve segmentler arasındaki farklılıkların bu kurallara göre belirlenmesi aslında, segmentler arasındaki satın alma davranış farklılıklarını da ortaya koymaktadır. Örneğin Tablo 17’de görüldüğü gibi 1. Segment için tespit edilen $\{tuz\} \Rightarrow \{sebze\}$ kuralı, tuz almış müşterilerin % 42’sinin sebze de aldığını ve bu satın alma davranışının diğer hiçbir segmentte bu kadar yüksek güven değerinde bulunmadığını belirtmektedir.

Tablo 17 incelendiğinde her bir segment için ortaya çıkan kurallar her ne kadar güven değerleri yüksek kurallar olsa da kuralların ilgi çekiciliği kuşkuludur. Güven değeri yüksek kuralların sağ taraflarında yer alan ürünlerin çok bilindik bir ürün olması beklendiğinden, kaldırıcı değeri yüksek kuralların da incelenmesi firma tarafından ilgi çekiciliği olan kuralların tespiti açısından önemli görülmektedir.

Tablo 17 Sosyoekonomik Segmentlere Özel İlişkiler Tablosu (Güven Değeri Bazlı)

Segment	Segmente has 2 ürünlü kurallar	Segmente has 3 ürünlü kurallar
1	{tuz} => {sebze} {sosis} => {peynir}	{bakliyat, et} => {sebze} {meyve, temizlik} => {sebze} {meyve, temizlik malzemesi} => {sebze} {et, yoğurt} => {sebze}
2	{tuzlu bisk.} => {tatlı bisk.} {kek} => {tatlı bisk.} {cips} => {gazlı içe.} {ekmek} => {sebze} {yumurta} => {süt} {süt} => {sebze}	{çikolata, kek} => {tatlı bisk.} {çikolata, meyve} => {sebze} {sebze, süt} => {meyve} {sebze, tatlı bisk.} => {meyve} {çerez, meyve} => {sebze} {meyve, tatlı bisk.} => {sebze} {çikolata, sebze} => {meyve}
3	{yağ} => {sebze}	{konserve, yumurta} => {sebze} {meyve, poşet} => {sebze} {çamaşır det., meyve} => {sebze} {meyve, pasta malz.} => {sebze} {özel ekmek, zeytin} => {peynir} {meyve, zeytin} => {peynir}
4	{sosis} => {sebze} {margarin} => {sebze}	{yoğurt, zeytin} => {peynir} {meyve, yağ} => {sebze} {meyve, un} => {sebze} {donmuş gıda, meyve} => {sebze}
5	{çamaşır det.} => {temizlik} {şekerleme} => {çikolata}	{özel ekmek, sebze} => {meyve} {çikolata, tatlı bisk.} => {tuzlu bisk.} {meyve, sebze} => {süt} {meyve, sebze} => {peynir} {meyve, sebze} => {özel ekmek} {çikolata, tatlı bisk.} => {gazlı içe} {meyve, sebze} => {gazlı içe} {meyve, sebze} => {yoğurt}

Segment	Segmente has 2 ürünlü kurallar	Segmente has 3 ürünlü kurallar
6	-	{peynir, yumurta} => {sebze} {yoğurt, yumurta} => {sebze} {peynir, yoğurt} => {sebze} {özel ekmek, peynir} => {sebze} {peynir, süt} => {sebze} {süt, yumurta} => {sebze} {özel ekmek, yumurta} => {sebze} {özel ekmek, yoğurt} => {sebze} {özel ekmek, süt} => {sebze} {süt, yoğurt} => {sebze}

Bu bölüme ait üçüncü tablo olan Tablo 18’de bu defa segmentlerdeki en güçlü kurallar, yani her bir segment için kaldırıcı değerleri en yüksek ilk 20 kural tespit edilmiş ve güçlü kurallara dayalı farklılıklar aşağıdaki tabloda gösterilmiştir. Görüldüğü gibi her bir segmentin birçok farklı ve güçlü kuralı mevcuttur. Ayrıca güven değerleri yüksek kurallardan farklı olarak ilgi çekici kuralların da özellikle üç ürünlü kurallarda ortaya çıktığı söylenebilir.

Tablo 18 Sosyoekonomik Segmentlere Özel İlişkiler Tablosu (Kaldıraç Değeri Bazlı)

Segment	Segmente has 2 ürünlü kurallar	Segmente has 3 ürünlü kurallar
1	{pasta malz.} => {un} {yağ} => {şeker} {peçete} => {poşet} {şampuan} => {sabun} {zeytin} => {sucuk} {zeytin} => {salam} {bebek ürünü} => {mama} {kaşar} => {zeytin} {temizlik} => {çamaşır det.} {un} => {bakliyat} {çorba} => {makarna}	{kaşar, peynir} => {zeytin} {çamaşır det, makarna} => {bakliyat} {çay, peynir} => {şeker} {bakliyat, temizlik} => {çamaşır det} {temizlik, tuvhav} => {çamaşır det} {makarna, peynir} => {sos} {makarna, yumurta} => {bakliyat} {makarna, temizlik} => {bakliyat} {çay, makarna} => {bakliyat} {bakliyat, konserve} => {makarna} {çamaşır det, tuv.hav.} => {temizlik} {peynir, salam} => {zeytin} {süt, temizlik} => {çamaşır det} {konserve, makarna} => {bakliyat} {peynir, temizlik} => {bulaşık det} {çamaşır det, süt} => {temizlik} {bakliyat, makarna} => {konserve}
2	{tatlı bisk} => {tuzlu bisk} {makarna} => {sebze} {poğaç} => {peynir} {cips} => {gazlı içe} {yoğurt} => {sebze} {tatlı bisk} => {çikolata} {süt} => {peynir} {meyve} => {sebze}	{çikolata, tatlı bisk} => {kek} {çikolata, kek} => {tatlı bisk} {çerez, sebze} => {meyve} {sebze, tatlı bisk} => {meyve} {çikolata, sebze} => {meyve} {kek, tatlı bisk} => {çikolata} {çikolata, meyve} => {sebze}
3	{sabun} => {temizlik} {çamaşır det} => {tuhav} {şeker} => {bakliyat} {poşet} => {tuhav} {poşet} => {temizlik} {bakliyat} => {makarna} {bakliyat} => {konserve} {bulaşık det} => {tuhav} {peçete} => {temizlik}	{meyve, temizlik} => {çamaşır det} {çamaşır det, meyve} => {temizlik} {peynir, süt} => {zeytin} {peynir, yumurta} => {kaşar} {bakliyat, sebze} => {konserve} {süt, zeytin} => {peynir} {özel ekmek, peynir} => {zeytin} {özel ekmek, zeytin} => {peynir} {konserve, sebze} => {bakliyat} {kaşar, yumurta} => {peynir} {meyve, peynir} => {kaşar}

Segment	Segmente has 2 ürünli kurallar	Segmente has 3 ürünli kurallar
4	{çamaşır det} => {bulaşık det} {şeker} => {yağ} {mama} => {bebek ürünü} {margarin} => {şeker} {sucuk} => {salam} {sucuk} => {kaşar} {sucuk} => {zeytin} {yağ} => {konserve} {temizlik malz} => {bulaşık det} {parfüm} => {kozmetik} {zeytin} => {kaşar}	{bulaşık det, çamaşır det} => {temizlik} {bakliyat, peynir} => {makarna} {makarna, peynir} => {bakliyat} {makarna, süt} => {bakliyat} {bakliyat, süt} => {makarna} {makarna, meyve} => {bakliyat} {bakliyat, meyve} => {makarna} {yumurta, zeytin} => {peynir}
5	{yumurta} => {peynir} {balık} => {sebze} {baharat} => {sebze} {süt} => {yoğurt} {sos} => {sebze}	{meyve, özel ekmek} => {sebze} {meyve, sebze} => {peynir} {peynir, sebze} => {meyve} {meyve, sebze} => {özel ekmek} {özel ekmek, sebze} => {meyve} {tatlı bisk, tuzlu bisk} => {çikolata} {çikolata, tatlı bisk} => {gazlı içe}
6	{kaşar} => {peynir} {yumurta} => {yoğurt} {kaşar} => {özel ekmek} {tatlı bisk} => {kek} {kek} => {tuzlu bisk} {et} => {peynir} {zeytin} => {sebze} {şekerleme} => {çikolata} {peynir} => {özel ekmek}	{peynir, sebze} => {yumurta} {sebze, yoğurt} => {yumurta} {sebze, süt} => {yumurta} {sebze, yumurta} => {peynir} {sebze, yumurta} => {yoğurt} {özel ekmek, sebze} => {yumurta} {sebze, süt} => {yoğurt} {meyve, özel ekmek} => {peynir} {özel ekmek, sebze} => {peynir} {sebze, yoğurt} => {peynir} {sebze, yumurta} => {süt} {peynir, sebze} => {yoğurt} {sebze, süt} => {peynir} {meyve, peynir} => {özel ekmek} {meyve, tavuk} => {sebze} {meyve, süt} => {peynir} {meyve, yumurta} => {sebze} {peynir, sebze} => {özel ekmek}

4.4.2 Kural Bazlı Segmentlere Ait Analizler

İkinci segmentasyon yaklaşımı olan kural bazlı segmentasyon analizi sonucunda ortaya beş farklı segment çıkmıştı. Bu segmentleri oluşturan mağazaların satış verileri birleştirilmiş ve her bir segment için yine Sık Öğe Küme Madenciliği ve Birliktelik Kuralları Analizleri uygulanmıştır. Segmentler arasındaki farklılıklar aşağıdaki tablolarda, segmentlere ait detaylı analizler ise yine Ek 2’de yer almaktadır.

İlk tablo olan Tablo 19’da, daha önceki bölümde olduğu gibi segmentlere özel sık satılan ürün ve ürün kümeleri yer almaktadır. Her bir segmentin satış verileri analiz edilmiş ve en çok satılan ilk 20 ürün veya ürün grubu elde edilmiştir. Sadece bir segmentte olup diğerlerinde olmayan ürün/ürünler segmente has ürün/ürünler olarak tabloya eklenmiştir.

Tablo 19 Kural Bazlı Segmentlere Özel Sık Öğe Kümeleri Tablosu

Segment	Mağaza Sayısı	Toplam Fiş Sayısı	Segmente has sık ürün	Segmente has sık iki öğeli küme	Segmente has sık üç öğeli küme
1	57	39142		Sebze, Tatlı bisküvi Meyve, Tatlı bisküvi	Ekmek, Meyve, Sebze
2	12	24715	Hazır yemek	Meyve, Peynir Özel ekmek, Peynir Çerez, Meyve Gazlı içecek, Meyve	Peynir, Sebze, Yoğurt Makarna, Meyve, Sebze
3	64	39516		Çikolata, Sigara Çikolata, Kek Gazlı içe, Sigara Çikolata, Su Çerez, Çikolata Kek, Tatlı bisküvi	Kek, Tatlı bisküvi, Tuzlu bisk. Çikolata, Sigara, Tatlı bisk. Cips, Çikolata, Gazlı içe. Çikolata, Kek, Tuzlu bisk. Gazlı içe., Tatlı bisk., Tuzlu bisk. Meyve, Tatlı bisküvi, Tuzlu bisk. Çikolata, Gazlı içe, Sigara
4	37	40012			Çikolata, Süt, Tatlı bisk. Gazlı içecek, Peynir, Sebze
5	5	6017	Temizlik Balık	Gazlı içecek, Sebze Balık, Sebze Peynir, Süt Süt, Yumurta Meyve, Yumurta	Gazlı içe., Meyve, Sebze Özel ekmek, Sebze, Yoğurt Çerez, Meyve, Sebze Özel ekmek, Sebze, Yumurta Balık, Meyve, Sebze Gazlı içe., Özel ekmek, Sebze

Tablo 19’den da görüleceği üzere, hazır yemek ürünü, sadece 2. Segmentte özel sık satılan üründür. 5. Segmentte ise temizlik ürünleri ile balık ürünü sadece bu segmentte sık olarak gözlemlenmiştir. 3. Segmentte ise sık satılan 2’li ürünler analizinde diğer segmentlerden farklı olarak çikolata ürünü bazı ürünlerle birlikte sıkça satış fişlerindeki yerini almıştır.

Tablo 20 Kural Bazlı Segmentlere Özel İlişkiler Tablosu (Güven değerine göre)

Segment	Segmente has 2 ürünlü kurallar	Segmente has 3 ürünlü kurallar
1	-	{şekerleme, tatlı bisküvi} => {çikolata} {özel ekmek, tavuk} => {sebze} {yoğurt, yumurta} => {sebze}
2	{hindi} => {sebze} {tuz} => {sebze} {bal} => {peynir}	{meyve, özel ekmek} => {sebze} {çerez, sebze} => {meyve} {sebze, yumurta} => {meyve} {sebze, süt} => {meyve} {özel ekmek, süt} => {sebze} {sebze, yoğurt} => {meyve} {gazlı içecek, sebze} => {meyve}
3	{gevrek} => {süt} {kuru meyve} => {çerez} {kek} => {çikolata} {şekerleme} => {çikolata} {tatlı bisk} => {çikolata} {cips} => {gazlı içe} {tuzlu bisk} => {tatlı bisk} {cips} => {çikolata} {kaşar} => {peynir} {kek} => {tatlı bisk}	{çikolata, tuzlu bisk} => {tatlı bisk} {kek, tuzlu bisk} => {çikolata} {çerez, tatlı bisk} => {çikolata} {tatlı bisk, tuzlu bisk} => {çikolata} {çikolata, kek} => {tatlı bisk} {gazlı içe, tatlı bisk} => {çikolata} {sebze, süt} => {meyve} {sigara, tatlı bisk} => {çikolata} {cips, çikolata} => {gazlı içe} {cips, gazlı içe} => {çikolata} {meyve, tatlı bisk} => {çikolata}
4	{tereyağı} => {sebze} {kaşar} => {peynir} {tereyağı} => {peynir}	{kahve, meyve} => {sebze} {et, süt} => {sebze} {et, gazlı içe} => {sebze} {et, peynir} => {sebze}
5	-	{özel ekmek, yumurta} => {sebze} {süt, yoğurt} => {sebze} {peynir, yoğurt} => {sebze} {özel ekmek, yoğurt} => {sebze} {süt, yumurta} => {sebze} {gazlı içe, süt} => {sebze}

Bu bölüme ait ikinci tablo olan Tablo 20’de ise, her bir segment için yapılan Birliktelik Kuralları Analizi sonucunda elde edilen kuralların, bütün segmentler için aynı olup olmadığı ve segmentlere özel kuralların varlığı tespit edilmiştir. Segmentlere özel Birliktelik Kuralları Analizi yapılırken ilk önce güven değerleri en yüksek ilk 20, iki ve üç ürünlü kurallar elde edilmiştir. Destek değeri ise, daha önce olduğu gibi her defasında 100 fişin altına düşmeyecek şekilde ayarlanmıştır. Segmentler için yapılmış olan bu analizler ayrıntılı bir şekilde Ek 2’de yer almaktadır.

Üçüncü tablo Tablo 21’de, Birliktelik Kuralları Analizinde kaldırıcı değerleri en yüksek ilk 20, iki ve üç ürünlü kurallar elde edilip kümeler arası farklılıklar göz önüne alınmıştır. Destek değeri ise, kuralların anlamlı olabilmesi için, her defasında yine 100 fişin altına düşmeyecek şekilde ayarlanmıştır. Segmentler için yapılmış olan bu analizler de ayrıntılı bir şekilde Ek 2’de yer almaktadır.

Kaldırıcı değeri bir kuralın gücü olarak yorumlandığından, kaldırıcı değerlerine göre kuralların sıralaması yapılırken en güçlü kurallar ilk sıralarda yer alır. Segmentlerin sahip oldukları güçlü kurallara göre farklılıkları Tablo 21’de gösterilmiştir. Tablodan görüleceği üzere her bir segment için birçok farklı kural mevcuttur. Bu kuralların hepsi segmentlere özel güçlü kurallar olup, firmanın pazarlama stratejileri için destekleyici niteliktedir.

Tablo 21 Kural Bazlı Segmentlere Özel İlişkiler Tablosu (Kaldıraç değerlerine göre)

Seg.	Segmente has 2 ürünlü kurallar	Segmente has 3 ürünlü kurallar
1	{sabun} => {temizlik} {konserve} => {bakliyat} {sos} => {makarna} {zeytin} => {yumurta} {sucuk} => {peynir} {konserve} => {makarna} {kaşar} => {yumurta} {reçel vb} => {peynir}	{peynir, sebze} => {zeytin} {sebze, zeytin} => {peynir} {salam, sebze} => {peynir} {peynir, sebze} => {salam} {peynir, sebze} => {yumurta} {kek, tatlı bisk} => {tuzlu bisk} {kek, tuzlu bisk} => {tatlı bisk} {sebze, yumurta} => {peynir} {süt, tatlı bisk} => {tuzlu bisk} {süt, tuzlu bisk} => {tatlı bisk} {tatlı bisk, tuzlu bisk} => {kek} {çikolata, kek} => {tatlı bisk} {sebze, süt} => {yoğurt} {gazlı içe, tuzlu bisk} => {tatlı bisk} {şekerleme, tatlı bisk} => {çikolata}
2	{şeker} => {yağ} {tuz} => {konserve} {sabun} => {diş} {parfüm} => {diş} {yağ} => {konserve} {tuvhav} => {peçete} {tuz} => {baharat} {sos} => {tuz} {tuz} => {bakliyat} {sabun} => {şampuan}	{çay, temizlik} => {şeker} {poşet, temizlik} => {bulaşık det} {çamaşır det, makarna} => {bulaşık det} {bulaşık det, makarna} => {çamaşır det} {çamaşır det, tuvhav} => {bulaşık det} {bulaşık det, tuvhav} => {çamaşır det} {temizlik, tuvhav} => {peçete} {çay, peynir} => {şeker} {çay, süt} => {şeker} {baharat, makarna} => {konserve} {çamaşır det, temizlik malz} => {temizlik} {temizlik, tuvhav} => {poşet} {makarna, temizlik} => {bakliyat}
3	{kuru meyve} => {çerez} {yumurta} => {yoğurt} {yumurta} => {özel ekmek} {çay} => {kahve} {peynir} => {özel ekmek} {hazır yemek} => {meyve suyu} {et} => {sebze} {konserve} => {sebze} {cips} => {çerez} {cips} => {gazlı içe} {yumurta} => {sebze} {tavuk} => {sebze} {tatlı bisk} => {kek}	{meyve, özel ekmek} => {sebze} {çikolata, kek} => {tatlı bisk} {çikolata, tatlı bisk} => {tuzlu bisk} {çikolata, sebze} => {meyve} {kek, tatlı bisk} => {çikolata} {peynir, sebze} => {meyve} {meyve, sebze} => {peynir} {özel ekmek, sebze} => {meyve} {meyve, süt} => {sebze} {çerez, çikolata} => {tatlı bisk} {çikolata, meyve} => {tatlı bisk} {çerez, tatlı bisk} => {çikolata} {tatlı bisk, tuzlu bisk} => {çikolata} {gazlı içe, tatlı bisk} => {çikolata} {çikolata, gazlı içe} => {tatlı bisk}

Seg.	Segmente has 2 ürünlü kurallar	Segmente has 3 ürünlü kurallar
4	{toz tatlı} => {pasta malz} {yağ} => {şeker} {margarin} => {şeker} {konserve} => {yağ} {bulaşık det} => {temizlik} {parfüm} => {kozmetik} {kaşar} => {zeytin} {çorba} => {makarna} {ıslak mendil} => {bebek ürünü}	{peynir, şeker} => {yağ} {bakliyat, çay} => {şeker} {peynir, yağ} => {şeker} {bakliyat, konserve} => {şeker} {konserve, şeker} => {bakliyat} {bakliyat, peynir} => {yağ} {bakliyat, temizlik} => {çamaşır det} {makarna, şeker} => {konserve} {çamaşır det, temizlik} => {bulaşık det} {makarna, yağ} => {bakliyat} {bakliyat, makarna} => {yağ} {konserve, makarna} => {şeker} {çay, temizlik} => {çamaşır det}
5	{peynir} => {zeytin} {tereyağ} => {peynir} {kek} => {çikolata} {yumurta} => {kaşar} {tatlı bisk} => {çikolata} {tavuk} => {et} {bakliyat} => {temizlik} {yumurta} => {makarna} {pasta malz} => {süt}	{peynir, süt} => {yumurta} {süt, yoğurt} => {yumurta} {peynir, sebze} => {kaşar} {yoğurt, yumurta} => {süt} {peynir, süt} => {yoğurt} {kaşar, sebze} => {peynir} {özel ekmek, süt} => {yumurta} {meyve, yumurta} => {süt} {süt, yumurta} => {yoğurt} {çikolata, sebze} => {tatlı bisk} {meyve, yoğurt} => {yumurta} {sebze, tatlı bisk} => {çikolata} {meyve, peynir} => {yumurta} {meyve, yumurta} => {yoğurt} {özel ekmek, süt} => {yoğurt} {özel ekmek, yumurta} => {süt} {özel ekmek, sebze} => {kaşar}

4.5 Tartışma ve Öneriler

Firma özellikle pazarlama ile ilgili karar verme süreçlerinde, segmentasyon analizi sonucunda ortaya çıkan segmentlerin farklı özelliklerini göz önünde bulundurmalıdır. Örneğin eğitim seviyesinin yüksek olduğu yerlerde veya üniversite bölgelerinde internet ve akıllı telefon kullanımının daha yaygın olduğu düşünülürse diğer segmentlerden farklı bir strateji olarak internet aracılığıyla pazarlama faaliyetleri yürütülebilir. Mağazalara özel web siteleri oluşturularak kampanyalar buradan duyurulabilir, mobil indirim kuponları hazırlanabilir, üye müşterilere kısa mesaj yoluyla özel indirimler uygulanabilir. Ekonomik gelirin yüksek olduğu segmentlerde yine diğer segmentlerden farklı olarak pahalı ürünlerin pazarlanması, üniversite bölgesindeki mağazalarda da ucuz ve pratik ürünlerin pazarlanması yoluna gidilebilir. Segmentlerde ortaya çıkan yüksek evli oranı ve yüksek çocuk yaş oranı değerleri dikkate alınarak diğer segmentlerden farklı olarak ailelere yönelik kampanyalar geliştirilebilir.

Hangi ürünlerin pazarlanacağı konusunda ise firmanın segmentler için yapacağı Birliktelik Kuralları Analizleri çok önem taşımaktadır. Firma hedefleri doğrultusunda ürünlerini sınıflandırıp, örneğin hedef hangi markaların birlikte satıldığını tespit etmekse markalara göre sınıflandırma yapılır ve bu sınıflandırma doğrultusunda sonuçlar elde edilebilir. Ürünlerin ayrıntılı olarak sınıflandırma işi firma tarafından otomatik hale getirebilir. Ürün satıldığı anda, ürünün veri tabanı sistemine geçen tam açıklayıcı ismiyle (örneğin X markasına ait yağsız 330 ml. süt) birlikte satılan ürünün firmanın hedefleri doğrultusunda daha önceden belirlediği ürün grubuna da (örneğin süt ürünleri grubu veya süt grubu veya yağsız süt grubu gibi) dâhil edilmesi sağlanabilir. Daha açık bir ifade ile ürün sınıflarını belirleme ve tekrar kodlama işi; genel ilişki kuralları elde edilmek isteniyorsa genel ürün grupları altında (süt ürünleri gibi), daha özel ilişki kuralları elde edilmek isteniyorsa da özel ürün grupları (yağlı/yağsız sütler) altında sınıflandırılabilir.

Ürünleri daha detaylı bir şekilde sınıflandırmak daha zengin ilişki kurallarının elde edilmesine olanak sağlayacaktır. Fakat bu ilişkilerin elde edilmesi için mağazalara ait sadece bir günlük satış verileri ile değil, birkaç haftalık hatta birkaç aylık verilere ihtiyaç duyulabilir.

Çok daha büyük veri kümeleriyle analiz yapma imkânı bulunan firmaların, detaylı bir şekilde oluşturduğu ürün grupları ile segmentlerin farklı özelliklerinin tutarlılık göstermesi mümkündür. Bir günlük veri kümesiyle yapılmış olan bu çalışmada bile segmentlerin özelliklerinin özetlendiği Tablo 13 ile segmentlere özel sık satılan ürünlerin yer aldığı Tablo 16 birlikte incelendiğinde anlamlı çıkarsamalarda bulunmak mümkün olabilir. Örneğin 2. Segmentte yer alan cips ürününün bu segmentte çocuk ve genç yaştaki nüfusun çok yüksek olmasına, 5. Segmentte yer alan hazır yemek ürününün (bu ürün hazır ev yemekleri reyonundaki ürünlerdir) ise bu segmentin ekonomik gelir durumunun yüksek olmasına bağlanabilir. Bu gibi ilişkilerin daha açık bir şekilde belirlenebilmesi açısından büyük veri kümeleri ile analiz yapmak gereklidir.

Firmanın uygulayacak olduğu Birliktelik Kuralları Analizi sonucu ortaya çıkan segmentlere özel kurallar pazarlama faaliyetlerinde kullanılabilecek, satışların ve kârlılığın arttırılmasını sağlayabilen ürün ilişkileridir. Yüksek güven değerli kurallar direkt olarak pazarlama faaliyetlerinde; örneğin çapraz satışlarda (cross sell), yukarı satışlarda (up sell), indirim kuponlarında veya başka türlü tekliflerde kullanılabilir. Kaldıraç değeri yüksek kurallarla ise segmentlere özel pazarlama faaliyetleri gerçekleştirilebilir. Bazı örnekler verilecek olunursa, birbirini tamamlayan ürünlerin beraber paketlenip sunulması veya fiyatta yapılacak olan bir indirimle satışların arttırılması stratejisi yürütülebilir. Kurallardan tespit edilen ilişkili ürünlerin mümkünse raflarının birbirine yakın olması, mümkün değilse bir ürün küçük bir sepette ilişkisi olduğu diğer ürünün rafının yakınına konumlandırılabilir. Yine ilişkisi olan ürünlerin indirim kataloglarında birlikte yer alması, hatta birinin fiyatının düşürülüp diğerinin fiyatının yükseltilmesi düşünülebilir.

SONUÇ

İşletmelerin devamlılığını sağlayabilmesi açısından kâr etmek önemli bir amaçtır. Bu amacı sağlamak için müşterilerin ihtiyaçlarını bilmek ve müşteri tatmin düzeyini arttırmak şarttır. Bu sebeple işletmeler müşterileri hakkındaki her türlü malumatı, verileri edinmeye çalışırlar. Bu veriler gerek müşterilerin demografileri ile ilgili, gerek satın alma davranışları gerekse de anket yoluyla elde edilebilecek memnuniyet, talep, şikâyet vb. ile ilgili olabilir. Önemli olan bu verileri etkin bir şekilde analiz ederek faydalı çıkarsamalarda bulunup, müşteriye tatmin etmek ve dolayısıyla sadakatini arttırabilmektir. Vurgulandığı gibi ilk amaç olan kâr bu yolla gerçekleştirilebilir.

İşletmelerin sahip oldukları verileri analiz edip faydalı çıkarsamalarda bulunması işini Veri Madenciliği süreci belirli bir sistematığa yerleştirip kolay hale getirmektedir. Ayrıca Veri Madenciliği tarafından önerilen istatistiksel tekniklerle de sahip olunan veriler etkin bir şekilde çözümlenip kullanışlı hale gelmektedir. Bu çalışmada işletmelerin Veri Madenciliği tekniklerinden olan Kümeleme Analizi ve Birliktelik Kuralları Analizini özellikle pazarlama faaliyetlerinde nasıl kullanabilecekleri ele alınmıştır.

Çalışmaya konu olan bir süpermarket zinciri firmasının, ülkenin farklı şehirlerinde yüzlerce, hatta bir şehirde bile onlarca mağazası bulunmaktadır. Çalışmamızın sınırlandırıldığı İstanbul ilinde ise firmanın 175 mağazası mevcuttur. İstanbul ili kozmopolit yapısıyla her bir semti, hatta her bir mahallesi için farklı demografik ve sosyoekonomik yapıya sahiptir. Böyle bir ortamda, farklı semtlerde mağazaları bulunan perakende firmalarının hedef pazarlamadan ziyade eski anlayıştaki kitle pazarlama stratejisini gütmesi anlamsızdır. Bu yüzden günümüz şartlarında perakende firmalarının mağazalarını segmente etmesi dünya çapındaki birçok danışmanlık şirketleri tarafından da uygun ve tavsiye edilen bir strateji olarak görülmektedir.

Bu amaçla çalışmamızda mağazalar, birinci yaklaşımımızda mağazaların bulunduğu bölgelerin ekonomik yapıları ve bu bölgelerde ikamet eden insanların demografik özellikleri de göz önüne alınarak Kümeleme Analizi yoluyla Ward'ın

algoritması kullanılarak altı segmente ayrılmıştır. Daha sonra her bir segment için pazarlama faaliyetlerinde kullanılmak üzere sık satılan ürünler ve birlikte satılan ürünlerin tespiti amacıyla yine Veri Madenciliği tekniklerinden olan Birliktelik Kuralları Analizi uygulanmıştır.

Mağaza segmentasyonu için uygulanacak ikinci yaklaşımımızda ise bu defa her bir mağaza müşterilerinin satın alma davranışları göz önüne alınmış ve mağazalar yine Ward'ın kümeleme algoritması kullanılarak beş segmente ayrılmıştır. Bu segmentler için de Birliktelik Kuralları Analizi ile pazarlama faaliyetlerinde kullanılabilecek ürün/ürün grupları tespit edilmiştir.

İlk yaklaşımda görülmüştür ki gerçekten her bir segment farklı özelliklere sahiptir. Ayrıca İstanbul haritasında, konumları ait oldukları segmentlerin sembolleriyle gösterilen mağazalar dikkatlice incelendiğinde, birbirine yakın olanların bile farklı segmentlerde yer alabilecekleri görülmüştür. Bu durum bize birbirine yakın olan mağazalara aynı pazarlama stratejilerinin uygulanabileceği tezinin doğru olmadığını göstermektedir.

İkinci yaklaşımda mağazalar, müşterilerinin sergiledikleri farklı satın alma davranışları yardımıyla beş segmente ayrılmıştır. Segmentler için yapılan ürün analizlerinde, her ne kadar her bir segmente has sık satılan veya birlikte satılan bazı ürün/ürünler olsa da segmentleri açıkça birbirinden ayırt edebilecek şekilde farklılıklar elde edilemediği düşünülebilir. Bu durumun ortaya çıkma sebebi, analizi yapılan veri kümesinin bir günlük satış işlemlerinden oluşmasıdır. Uzun dönemde segmentler arasındaki açık farklılıkların ortaya çıkması muhtemeldir.

Birçok mağazası bulunan perakende firmalarının mağaza segmentasyonu işini bir sistematığa bağlayıp pazarlama ve satış süreçleri için önemli kararlar almadan önce belirli aylarda veya mevsimlerde mağazalarını segmente etmesi firmaların küresel rekabet ortamında daha başarılı olabilmelerine destek olacaktır.

KAYNAKÇA

Kitaplar

- Aggarwal, C. C. (2015). *Data mining: The textbook*. Switzerland: Springer.
- Ahlemeyer-Stubbe, A. ve S. Coleman. (2014). *A practical guide to data mining for business and industry*: John Wiley & Sons.
- Alex, B., S. Stephen ve T. Kurt. (2000). *Building data mining applications for CRM. New York (etc.): McGraw-Hill*.
- Alpaydın, E. (2011). *Yapay Öğrenme*. İstanbul: Boğaziçi. Ün v Yayınları.
- Bennett, P. D. (1995). *Dictionary of Marketing Terms*: NTC Business Books.
- Berry, M. J. ve G. Linoff. (1997). *Data mining techniques: for marketing, sales, and customer support*: John Wiley & Sons, Inc.
- Camm, J. ve J. Cochran. (2016). *Essentials of Business Analytics*: Nelson Education.
- Chen, M., S. Mao, Y. Zhang ve V. C. Leung. (2014). *Big data: related technologies, challenges and future prospects*: Springer.
- Cios, K. J., W. Pedrycz ve R. W. Swiniarski. (2012). *Data mining methods for knowledge discovery* (Vol. 458): Springer Science & Business Media.
- Dean, J. (2014). *Big data, data mining, and machine learning: value creation for business leaders and practitioners*: John Wiley & Sons.
- Doyle, C. (2011). *A dictionary of marketing*: Oxford University Press.
- Everitt, B. (1974). *Cluster analysis* 122: Heinemann, London.
- Gan, G., C. Ma ve J. Wu. (2007). *Data clustering: theory, algorithms, and applications* (Vol. 20): Siam.
- Giudici, P. (2005). *Applied data mining: statistical methods for business and industry*: John Wiley & Sons.
- Han, J., M. Kamber ve J. Pei. (2011). *Data mining: concepts and techniques*: Elsevier.
- Hastie, T., R. Tibshirani ve J. Friedman. (2009). *The Elements of Statistical Learning*. New York: Springer-Verlag.
- Jain, A. K. ve R. C. Dubes. (1988). *Algorithms for clustering data*: Prentice-Hall, Inc.

- Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms*: John Wiley & Sons.
- Kaufman, L. ve P. Roussew. (1990). *Finding Groups in Data-An Introduction to Cluster Analysis*. A Wiley-Science Publication John Wiley & Sons: Inc.
- Kotler, P. ve G. Armstrong. (2004). *Principles of marketing*. Upper Saddle River, NJ: Prentice Hall.
- Ledolter, J. (2013). *Data mining and business analytics with R*: John Wiley & Sons.
- Legendre, P. ve L. F. Legendre. (2012). *Numerical ecology* (Vol. 24): Elsevier.
- Leskovec, J., A. Rajaraman ve J. D. Ullman. (2014). *Mining of massive datasets*: Cambridge University Press.
- Lilien, G. L. ve P. Kotler. (1983). *Marketing decision making: A model-building approach*: Harper & Row New York, NY.
- Myers, J. H. (1996). *Segmentation and positioning for strategic marketing decisions*.
- Nakip, M. (2006). *Pazarlama arařtırmaları teknikler ve (SPSS destekli) uygulamalar*. Ankara: Seçkin Yayıncılık.
- Olson, D. L. ve D. Delen. (2008). *Advanced data mining techniques*: Springer Science & Business Media.
- Wedel, M. ve W. A. Kamakura. (2012). *Market segmentation: Conceptual and methodological foundations* (Vol. 8): Springer Science & Business Media.
- Zaki, M. J. ve W. Meira Jr. (2014). *Data mining and analysis: fundamental concepts and algorithms*: Cambridge University Press.

Sürekli Yayınlar

- Aalst, W. v. d. ve E. Damiani. (2015). Processes Meet Big Data: Connecting Data Science with Process Science. *Services Computing, IEEE Transactions on*. 8.6, 810-819.
- Agrawal, R., H. Mannila, R. Srikant, H. Toivonen ve A. I. Verkamo. (1996). Fast Discovery of Association Rules. *Advances in knowledge discovery and data mining*. 12.1, 307-328.
- Ahmet, Ş. ve E. Cengiz. (2010). Tüketicilerin Perakende Market Zincirlerine Yönelik Yapmış Oldukları Satın Alma Sıklıklarına Göre Bölümlendirilmesi. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*. 24.2.
- Aiguo, Z., J. Lanling ve S. Ping. (2005). *Application of Data Mining in supermarket*. 2011 International Conference on Consumer Electronics, Communications and Networks (CECNet).
- Annie, L. C. M. ve A. D. Kumar. (2012). Market Basket Analysis for a Supermarket based on Frequent Itemset Mining. *International Journal of Computer Science Issues(IJCSI)*. 9.5.
- Atkins, K. G., A. Kumar ve Y.-K. Kim. (2016). Smart grocery shopper segments. *Journal of International Consumer Marketing*. 28.1, 42-53.
- Azizi, S. ve V. Hosseinabadi. (2015). Segmenting Medical Market Based on Banking Expectations: Evidence from Iran. *Journal of Competitiveness*. 7.2.
- Bacila, M. F., A. Radulescu ve I. L. Marar. (2013). *Customer segmentation based on the value of consumption patterns in telecommunications*. The Proceedings of the International Conference" Marketing-from Information to Decision", 36.
- Badea, L. M. (2014). Predicting consumer behavior with artificial neural networks. *Procedia Economics and Finance*. 15, 238-246.
- Bai, X., Z. Zhang, Q. Tian, W. Yin ve J. Dong. (2009). *A novel customer segmentation framework for china's supermarkets*. Service Operations, Logistics and Informatics, 2009. SOLI'09. IEEE/INFORMS International Conference on, 471-476.
- Bayus, B. L. ve R. Mehta. (1995). A segmentation model for the targeted marketing of consumer durables. *Journal of marketing research*, 463-469.
- Behe, B. K., T. A. Prince ve H. K. Tayama. (1992). Market segmentation of supermarket floral customers. *Hortscience*. 27.5, 459-462.

- Ben Ali, B. ve Y. Massmoudi. (2013). *K-means clustering based on gower similarity coefficient: A comparative study*. Modeling, Simulation and Applied Optimization (ICMSAO), 1-5.
- Berg, C. J., P. M. Ling, H. Guo, C. M. Klatt, J. L. Thomas, J. S. Ahluwalia ve L. C. An. (2011). Using market research to understand health behaviors among college students. *College Student Journal*. 45.4, 726.
- Bermingham, P., T. Hernandez ve I. Clarke. (2013). Network Planning and Retail Store Segmentation: A Spatial Clustering Approach. *International Journal of Applied Geospatial Research (IJAGR)*. 4.1, 67-79.
- Bigné, E., J. Aldas-Manzano, I. Küster ve N. Vila. (2010). Mature market segmentation: a comparison of artificial neural networks and traditional methods. *Neural Computing and Applications*. 19.1, 1-11.
- Bijak, K. ve L. C. Thomas. (2012). Does segmentation always improve model performance in credit scoring? *Expert Systems with Applications*. 39.3, 2433-2442.
- Bilgiç, E., M. Kantardzic ve Ö. Çakir. (2015). Retail Store Segmentation for Target Marketing. *Advances in Data Mining: Applications and Theoretical Aspects*, 32-44.
- Bittmann, R. M. ve R. Gelbard. (2009). Visualization of multi-algorithm clustering for better economic decisions—the case of car pricing. *Decision Support Systems*. 47.1, 42-50.
- Boote, A. S. (1982). Psychographic segmentation in Europe. *Journal of Advertising Research*.
- Brentari, E., L. Dancelli ve M. Manisera. (2016). Clustering ranking data in market segmentation: a case study on the Italian McDonald's customers' preferences. *Journal of Applied Statistics*, 1-18.
- Brijs, T., G. Swinnen, K. Vanhoof ve G. Wets. (1999). *Using association rules for product assortment decisions: A case study*. Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, 254-260.
- Cao, F., J. Liang, D. Li, L. Bai ve C. Dang. (2012). A dissimilarity measure for the k-Modes clustering algorithm. *Knowledge-Based Systems*. 26, 120-127.
- Carvalho, J. C., P. Cardoso, L. C. Crespo, S. Henriques, R. Carvalho ve P. Gomes. (2011). Biogeographic patterns of spiders in coastal dunes along a gradient of mediterraneity. *Biodiversity and conservation*. 20.4, 873-894.

- Ceylan, H. H. (2013). Perakende Sektöründe Konjoint ve Kümeleme Analizi ile Fayda Temelli Pazar Bölümlendirme. *Yönetim ve Ekonomi: Celal Bayar Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*. 20.1, 141-154.
- Chen, I. J. ve K. Popovich. (2003). Understanding customer relationship management (CRM) People, process and technology. *Business process management journal*. 9.5, 672-688.
- Chen, M.-C., A.-L. Chiu ve H.-H. Chang. (2005). Mining changes in customer behavior in retail marketing. *Expert Systems with Applications*. 28.4, 773-781.
- Chen, M. C., C. M. Chao ve K. T. Wu. (2012). Pattern filtering and classification for market basket analysis with profit-based measures. *Expert Systems*. 29.2, 170-182.
- Chen, M. C. ve C. P. Lin. (2007). A data mining approach to product assortment and shelf space allocation. *Expert Systems with Applications*. 32.4, 976-986.
- Chen, Y. L., K. Tang, R. J. Shen ve Y. H. Hu. (2005). Market basket analysis in a multiple store environment. *Decision Support Systems*. 40.2, 339-354.
- Chengduan, W. (2015). Applying Bi-clustering Algorithm in Customer Segmentation for High-Value Customers.
- Cheung, Y. M. ve H. Jia. (2013). Categorical-and-numerical-attribute data clustering based on a unified similarity metric without knowing cluster number. *Pattern Recognition*. 46.8, 2228-2238.
- Clarke, I., W. Mackaness ve B. Ball. (2003). Modelling Intuition in Retail Site Assessment (MIRSA): making sense of retail location using retailers' intuitive judgements as a support for decision-making. *The International Review of Retail, Distribution and Consumer Research*. 13.2, 175-193.
- Cooil, B., L. Aksoy ve T. L. Keiningham. (2008). Approaches to customer segmentation. *Journal of Relationship Marketing*. 6.3-4, 9-39.
- Çakır, Ö. (2008). Veri Madenciliğinde Sınıflandırma Yöntemlerinin Karşılaştırılması: Bankacılık Müşteri Veri Tabanı Üzerinde Bir Uygulama. *Basılmamış Doktora Tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü*.
- Çil, İ. (2012). Consumption universes based supermarket layout through association rule mining and multidimensional scaling. *Expert Systems with Applications*. 39.10, 8611-8625.
- De Godoy, R. C., R. Deliza, L. B. Gheno, S. Licodiedoff, C. N. Frizon, R. H. Ribani ve G. G. dos Santos. (2013). Consumer perceptions, attitudes and acceptance of new and traditional mate tea products. *Food research international*. 53.2, 801-807.

- De Mauro, A., M. Greco ve M. Grimaldi. (2015). *What is big data? A consensual definition and a review of key research topics*. AIP Conference Proceedings, 97-104.
- De Oña, R. ve J. De Oña. (2015). Analysis of transit quality of service through segmentation and classification tree techniques. *Transportmetrica A: Transport Science*. 11.5, 365-387.
- Devi, A. ve J. Sharma. (2013). Investigating role stress in frontline bank employees: A cluster based approach. *IIMB Management Review*. 25.3, 171-178.
- Díaz-Pérez, F. M. ve M. Bethencourt-Cejas. (2016). CHAID algorithm as an appropriate analytical method for tourism market segmentation. *Journal of Destination Marketing & Management*.
- Dibb, S. ve P. Stern. (1995). Questioning the reliability of market segmentation techniques. *Omega*. 23.6, 625-636.
- Djokic, N., S. Salai, R. Kovac-Znidarsic, I. Djokic ve G. Tomic. (2013). The use of conjoint and cluster analysis for preference-based market segmentation. *Engineering Economics*. 24.4, 343-355.
- Doyle, P. ve J. Saunders. (1985). Market segmentation and positioning in specialized industrial markets. *The Journal of Marketing*, 24-32.
- Durkin, M. (2004). In search of the Internet-banking customer: exploring the use of decision styles. *International Journal of Bank Marketing*. 22.7, 484-503.
- Ekinci, Y., F. Ulengin ve N. Uray. (2014). Using customer lifetime value to plan optimal promotions. *The Service Industries Journal*. 34.2, 103-122.
- Eusébio, C. ve M. João Carneiro. (2015). How diverse is the youth tourism market? An activity-based segmentation study. *Turizam: znanstveno-stručni časopis*. 63.3, 295-316.
- Everitt, B. ve T. Hothorn. (2011). Cluster analysis *An Introduction to Applied Multivariate Analysis with R* (pp. 163-200): Springer.
- Everitt, B. S. ve C. Merette. (1990). The clustering of mixed-mode data: a comparison of possible approaches. *Journal of Applied Statistics*. 17.3, 283-297.
- Fayyad, U., G. Piatetsky-Shapiro ve P. Smyth. (1996). From data mining to knowledge discovery in databases. *AI magazine*. 17.3, 37.
- Fonseca, J. R. (2013). Clustering in the field of social sciences: that is your choice. *International Journal of Social Research Methodology*. 16.5, 403-428.
- Fowlkes, E. B. ve C. L. Mallows. (1983). A method for comparing two hierarchical clusterings. *Journal of the American Statistical association*. 78.383, 553-569.

- Frees, E. W., R. A. Derrig ve G. Meyers. (2014). *Predictive modeling applications in actuarial science* (Vol. 1): Cambridge University Press.
- Füller, J. ve K. Matzler. (2008). Customer delight and market segmentation: An application of the three-factor theory of customer satisfaction on life style groups. *Tourism management*. 29.1, 116-126.
- Garver, M. S., R. L. Divine ve S. A. Spralls. (2014). Choice-Based Conjoint Analysis of the Local Coupon Preferences of Millennials. *Journal of Promotion Management*. 20.2, 240-249.
- Gonzalez-Benito, B. ve J. González-Benito. (2005). The role of geodemographic segmentation in retail location strategy. *International Journal of Market Research*. 47.3, 295-316.
- Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, 857-871.
- Green, P. E. ve V. R. Rao. (1971). Conjoint measurement for quantifying judgmental data. *Journal of marketing research*, 355-363.
- Gulluoglu, S. S. (2015). *Segmenting customers with data mining techniques*. Digital Information, Networking, and Wireless Communications (DINWC), 2015 Third International Conference on, 154-159.
- Guo, Y., B. Denizci Guillet, D. Kucukusta ve R. Law. (2015). Segmenting Spa Customers Based on Rate Fences Using Conjoint and Cluster Analyses. *Asia Pacific Journal of Tourism Research*, 1-19.
- Güçdemir, H. ve H. Selim. (2015). Integrating multi-criteria decision making and clustering for business customer segmentation. *Industrial Management & Data Systems*. 115.6, 1022-1040.
- H. Davenport, T. (2014). How strategists use “big data” to support internal business decisions, discovery and production. *Strategy & Leadership*. 42.4, 45-50.
- Haddad, Y., J. Haddad, A. Olabi, N. Shuayto, T. Haddad ve I. Toufeili. (2007). Mapping determinants of purchase intent of concentrated yogurt (Labneh) by conjoint analysis. *Food Quality and Preference*. 18.5, 795-802.
- Hahsler, M. ve K. Hornik. (2007). Building on the arules infrastructure for analyzing transaction data with R *Advances in Data Analysis* (pp. 449-456): Springer.
- Halkidi, M., Y. Batistakis ve M. Vazirgiannis. (2001). On clustering validation techniques. *Journal of intelligent information systems*. 17.2-3, 107-145.
- Hall, M., I. Witten ve E. Frank. (2011). Data mining: Practical machine learning tools and techniques. *Kaufmann, Burlington*.

- Han, Z., Z. Xiao-hang, W. Qi, Z. Ze-cong ve W. Cen-yue. (2014). *Customer segmentation on mobile online behavior*. Management Science & Engineering (ICMSE), 2014 International Conference on, 103-109.
- Haverila, M., M. Rod ve N. Ashill. (2013). Cell phone product-market segments using product features as a cluster variate: a multi-country study. *Journal of Strategic Marketing*. 21.2, 101-124.
- Hawkes, G. F. ve E. W. McLaughlin. (1994). *STARS: Segment Targeting at Retail Stores*: Department of Agricultural, Resource, and Managerial Economics, Cornell University.
- He, Z., X. Xu ve S. Deng. (2005). Clustering mixed numeric and categorical data: A cluster ensemble approach. *arXiv preprint cs/0509011*.
- Hendrickson, J. L. (2014). Methods for Clustering Mixed Data.
- Hennig, C. (2015). Clustering strategy and method selection. *arXiv preprint arXiv:1503.02059*.
- Hennig, C. ve T. F. Liao. (2013). How to find an appropriate clustering for mixed-type variables with application to socio-economic stratification. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 62.3, 309-369.
- Hjort, K., B. Lantz, D. Ericsson ve J. Gattorna. (2013). Customer segmentation based on buying and returning behaviour. *International Journal of Physical Distribution & Logistics Management*. 43.10, 852-865.
- Hubert, L. ve P. Arabie. (1985). Comparing partitions. *Journal of Classification*. 2.1, 193-218.
- Hung, C. ve C.-F. Tsai. (2008). Market segmentation based on hierarchical self-organizing map for markets of multimedia on demand. *Expert Systems with Applications*. 34.1, 780-787.
- Husić-Mehmedović, M., M. Čičić ve E. Agić. (2015). Regional Lifestyle Segmentation in the Western Balkans. *South East European Journal of Economics and Business*. 9.2, 46-53.
- Ismail El-Adly, M. (2007). Shopping malls attractiveness: a segmentation approach. *International Journal of Retail & Distribution Management*. 35.11, 936-950.
- Jaccard, P. (1908). *Nouvelles recherches sur la distribution florale*.
- James, G., D. Witten, T. Hastie ve R. Tibshirani. (2013). *An introduction to statistical learning* (Vol. 112). New York: Springer-Verlag.
- Johnson, R. A. ve D. W. Wichern. (1992). *Applied multivariate statistical analysis* (Vol. 4): Prentice hall Englewood Cliffs, NJ.

- Kargari, M. ve M. M. Sepehri. (2012). Stores clustering using a data mining approach for distributing automotive spare-parts to reduce transportation costs. *Expert Systems with Applications*. 39.5, 4740-4748.
- Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied statistics*, 119-127.
- Kennedy, J., M. Worosz, E. C. Todd ve M. K. Lapinski. (2008). Segmentation of US consumers based on food safety attitudes. *British Food Journal*. 110.7, 691-705.
- Khatami, A., S. Mirghasemi, A. Khosravi ve S. Nahavandi. (2015). *A New Color Space Based on K-Medoids Clustering for Fire Detection*. Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on, 2755-2760.
- Kimiloglu, H., V. Aslihan Nasir ve S. Nasir. (2010). Discovering behavioral segments in the mobile phone market. *Journal of Consumer Marketing*. 27.5, 401-413.
- Kitchin, R. ve T. P. Lauriault. (2015). Small data in the era of big data. *GeoJournal*. 80.4, 463-475.
- Koehn, N. F. (2001). Howard Schultz and Starbucks Coffee Company. *Harvard Business School Cases*.
- Kolyshkina, I., E. Nankani, S. Simoff ve S. Denize. (2010). *Retail analytics in the context of "Segmentation, Targeting, Optimisation" of the operations of convenience store franchises*.
- Kosters, W. A., E. Marchiori ve A. A. Oerlemans. (1999). Mining clusters with association rules *Advances in Intelligent Data Analysis* (pp. 39-50): Springer.
- Kronberger, G. ve M. Affenzeller. (2011). Market basket analysis of retail data: supervised learning approach *Computer Aided Systems Theory–EUROCAST 2011* (pp. 464-471): Springer.
- Lance, G. N. ve W. T. Williams. (1967). A general theory of classificatory sorting strategies II. Clustering systems. *The Computer Journal*. 10.3, 271-277.
- Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*. 6, 70.
- Lefait, G. ve T. Kechadi. (2010). *Customer segmentation architecture based on clustering techniques*. Digital Society, 2010. ICDS'10. , 243-248.
- Lent, B., A. Swami ve J. Widom. (1997). *Clustering association rules*. Data Engineering, 1997. Proceedings. 13th International Conference on, 220-231.
- Lesakova, D. (2014). Senior's lifestyle and their store choice. *Megatrend revija*. 11.4, 71-82.

- Li, H. (2005). *Applications of data warehousing and data mining in the retail industry*. Services Systems and Services Management, 2005. Proceedings of ICSSSM'05. 2005 International Conference on, 1047-1050.
- Li, W., X. Wu, Y. Sun ve Q. Zhang. (2010). *Credit card customer segmentation and target marketing based on data mining*. Computational Intelligence and Security (CIS), 2010 International Conference on, 73-76.
- Liao, S.-h., Y.-j. Chen ve Y.-t. Lin. (2011). Mining customer knowledge to implement online shopping and home delivery for hypermarkets. *Expert Systems with Applications*. 38.4, 3982-3991.
- Lippman, B. W. (2003). Retail revenue management—Competitive strategy for grocery retailers. *Journal of revenue and pricing management*. 2.3, 229-233.
- Machauer, A. ve S. Morgner. (2001). Segmentation of bank customers by expected benefits and attitudes. *International Journal of Bank Marketing*. 19.1, 6-18.
- Mangiameli, P., S. K. Chen ve D. West. (1996). A comparison of SOM neural network and hierarchical clustering methods. *European Journal of Operational Research*. 93.2, 402-417.
- Mao, S., L. Nie ve G. R. Thoma. (2005). *Unsupervised style classification of document page images*. Image Processing, 2005. ICIP 2005. IEEE International Conference on, II-510-513.
- Mehrotra, A. ve R. Agarwal. (2009). Classifying customers on the basis of their attitudes towards telemarketing. *Journal of Targeting, Measurement and analysis for Marketing*. 17.3, 171-193.
- Meilă, M. (2007). Comparing clusterings—an information based distance. *Journal of multivariate analysis*. 98.5, 873-895.
- Memery, J., P. Megicks, R. Angell ve J. Williams. (2012). Understanding ethical grocery shoppers. *Journal of Business Research*. 65.9, 1283-1289.
- Mendes, A. B. ve M. G. M. S. Cardoso. (2006). Clustering supermarkets: the role of experts. *Journal of Retailing and Consumer Services*. 13.4, 231-247.
- Migueis, V. L., A. S. Camanho ve J. F. E. Cunha. (2012). Customer data mining for lifestyle segmentation. *Expert Systems with Applications*. 39.10, 9359-9366.
- Milenova, B. L. ve M. M. Campos. (1997). Clustering large databases with numeric and nominal values using orthogonal projections. *MA, USA*.
- Mingoti, S. A. ve J. O. Lima. (2006). Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms. *European Journal of Operational Research*. 174.3, 1742-1759.

- Morlini, I. ve S. Zani. (2010). Comparing approaches for clustering mixed mode data: an application in marketing research *Data Analysis and Classification* (pp. 49-57): Springer.
- Murtagh, F. ve P. Legendre. (2011). Ward's hierarchical clustering method: clustering criterion and agglomerative algorithm. *arXiv preprint arXiv:1111.6285*.
- Nakahara, T. ve K. Yada. (2012). Analyzing consumers' shopping behavior using RFID data and pattern mining. *Advances in Data Analysis and Classification*. 6.4, 355-365.
- Namvar, M., M. R. Gholamian ve S. KhakAbi. (2010). *A two phase clustering method for intelligent customer segmentation*. Intelligent Systems, Modelling and Simulation (ISMS), 2010 International Conference on, 215-219.
- Ngai, E. W., L. Xiu ve D. C. Chau. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*. 36.2, 2592-2602.
- Norusis, M. J. (2010). *PASW statistics 18 guide to data analysis*: Prentice Hall Press.
- Ozansoy, Ç. T. (2012). Tüketicilerin Sosyopsikolojik Ve Demografik Özellikleri, İlgiilenim, Subjektif Bilgi Ve Güven Düzeyine Bağlı Olarak Moda Giysi Pazarının Bölümlendirilmesi. *Öneri Dergisi*. 9.33, 143-152.
- Özcan, T. ve Ş. Esnaf. (2010). Perakende Endüstrisinde Raf Alanı Tahsis ve Mağaza Yerleşim Optimizasyonuna Bütünleşik Bir Model Önerisi. *İÜ Mühendislik Bilimleri Dergisi*. 1.1, 55-63.
- Panwar, J. (2008). Benefit Segmentation of Gutkha Market: A Study in the State of Gujarat. *Vision: The Journal of Business Perspective*. 12.3, 31-38.
- Pardeshi, B. ve D. Toshniwal. (2010). *Improved k-medoids clustering based on cluster validity index and object density*. Advance Computing Conference (IACC), 2010 IEEE 2nd International, 379-384.
- Park, S.-B., Y.-K. Lee ve N. Chung. (2013). Why don't consumers go internet shopping in Korea? Segmentation of consumer lifestyle approach. *Behaviour & Information Technology*. 32.5, 468-479.
- Partners, B. R. (2012). 1st Annual Merchandise Planning & Allocation Benchmark Survey. *Boston ABD*.
- Pavoine, S., J. Vallet, A. B. Dufour, S. Gachet ve H. Daniel. (2009). On the challenge of treating various types of variables: application for improving the measurement of functional diversity. *Oikos*. 118.3, 391-402.
- Plasse, M., N. Niang, G. Saporta, A. Villemint ve L. Leblond. (2007). Combined use of association rules mining and clustering methods to find relevant links between

- binary rare attributes in a large data set. *Computational Statistics & Data Analysis*. 52.1, 596-613.
- Punj, G. ve D. W. Stewart. (1983). Cluster analysis in marketing research: Review and suggestions for application. *Journal of marketing research*, 134-148.
- Rabbany, R. ve O. R. Zaiane. (2015). Generalization of clustering agreements and distances for overlapping clusters and network communities. *Data mining and knowledge discovery*. 29.5, 1458-1485.
- Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*. 66.336, 846-850.
- Rangan, V. K., R. T. Moriarty ve G. S. Swartz. (1992). Segmenting customers in mature industrial markets. *The Journal of Marketing*, 72-82.
- Reynolds, A. P., G. Richards ve V. J. Rayward-Smith. (2004). The application of k-medoids and pam to the clustering of rules *Intelligent Data Engineering and Automated Learning-IDEAL 2004* (pp. 173-178): Springer.
- Riaz, M., A. Arooj, M. T. Hassan ve J.-B. Kim. (2014). *Clustering based association rule mining on online stores for optimized cross product recommendation*. Control, Automation and Information Sciences (ICCAIS), 2014 International Conference on, 176-181.
- Robinson, H. W. (1938). The Equilibrium Price in a Perfect Intertemporal Market. *Econometrica: Journal of the Econometric Society*, 48-62.
- Romano, S., J. Bailey, X. V. Nguyen ve K. Verspoor. (2014). *Standardized Mutual Information for Clustering Comparisons: One Step Further in Adjustment for Chance*. ICML, 1143-1151.
- Romano, S., N. X. Vinh, J. Bailey ve K. Verspoor. (2015). Adjusting for Chance Clustering Comparison Measures. *arXiv preprint arXiv:1512.01286*.
- Rygielski, C., J.-C. Wang ve D. C. Yen. (2002). Data mining techniques for customer relationship management. *Technology in society*. 24.4, 483-502.
- Seegebarth, B., S. H. Behrens, C. Klarmann, N. Hennigs, L. L. Scribner ve C. Griffith. (2016). Customer value perception of organic food: cultural differences and cross-national segments. *British Food Journal*. 118.2.
- Segal, M. N. ve R. W. Giacobbe. (1994). Market segmentation and competitive analysis for supermarket retailing. *International Journal of Retail & Distribution Management*. 22.1, 38-48.
- Shahrabi, J., S. M. Najafabadi ve M. Sahebnasagh. (2009). *Applying cluster analysis to segment construction steel market*. Industrial Engineering and Engineering Management, 2009. IEEM 2009. IEEE International Conference on, 792-797.

- Smith, W. R. (1956). Product Differentiation And Market Segmentation As Alternative Marketing Strategies. *Journal of Marketing*. 21.1, 3-8.
- Sokal, R. R. ve F. J. Rohlf. (1962). The comparison of dendrograms by objective methods. *Taxon*, 33-40.
- Sreedharan, S., G. Radhakrishnan, D. Gupta ve T. Sudarshan. (2014). *Analysis of robotic environment using low resolution image sequence*. Contemporary Computing and Informatics (IC3I), 2014 International Conference on, 495-499.
- Strehl, A., G. K. Gupta ve J. Ghosh. (1999). *Distance based clustering of association rules*. Proceedings ANNIE, 759-764.
- Struyf, A., M. Hubert ve P. Rousseeuw. (1997). Clustering in an object-oriented environment. *Journal of Statistical Software*. 1.4, 1-30.
- Swenson, E. R., N. D. Bastian ve H. B. Nembhard. (2016). Data Analytics in Health Promotion: Health Market Segmentation and Classification of Total Joint Replacement Surgery Patients. *Expert Systems with Applications*.
- Tekin, M., Y. Köse, Ö. Koyuncuoğlu ve E. Tekin. (2015). *The Analysis of Product Categories and Sales Relationships among Valuable Customers through Data Mining and Its Application to a National Retailer through Association Rules and Cluster Analysis*. International Interdisciplinary Business-Economics Advancement Conference, 180.
- Tien, J. M. (2013). Big data: Unleashing information. *Journal of Systems Science and Systems Engineering*. 22.2, 127-151.
- Timor, M. ve U. Şimşek. (2008). Veri Madenciliğinde Sepet Analizi ile Tüketici Davranışı Modellemesi. *Yönetim*, 19 (59), 3-10.
- Vakharia, A. J. ve J. Mahajan. (2000). Clustering of objects and attributes for manufacturing and marketing applications. *European Journal of Operational Research*. 123.3, 640-651.
- Vermeulen-Jourdan, L., C. Dhaenens ve E.-G. Talbi. (2004). Clustering nominal and numerical data: A new distance concept for a hybrid genetic algorithm *Evolutionary Computation in Combinatorial Optimization* (pp. 220-229): Springer.
- Vinh, N. X., J. Epps ve J. Bailey. (2009). *Information theoretic measures for clusterings comparison: is a correction for chance necessary?* Proceedings of the 26th Annual International Conference on Machine Learning, 1073-1080.
- Wagner, S. ve D. Wagner. (2007). *Comparing clusterings: an overview*: Universität Karlsruhe, Fakultät für Informatik Karlsruhe.

- Wei, W., K. N. Ngan ve N. Habili. (2004). *Multiple feature clustering algorithm for automatic video object segmentation*. Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on, iii-625-628 vol. 623.
- Weinstein, A. (2004). *Handbook of market segmentation: Strategic targeting for business and technology firms*: Psychology Press.
- Wind, Y. (1978). Issues and advances in segmentation research. *Journal of marketing research*, 317-337.
- Wishart, D. (2003). K-means clustering with outlier detection, mixed variables and missing values *Exploratory Data Analysis in Empirical Research* (pp. 216-226): Springer.
- Xiang, Q., Q. Mao, K. M. Chai, H. L. Chieu, I. Tsang ve Z. Zhao. (2012). A Split-Merge Framework for Comparing Clusterings. *arXiv preprint arXiv:1206.6475*.
- Yeung, K. Y. ve W. L. Ruzzo. (2001). Details of the adjusted Rand index and clustering algorithms, supplement to the paper "An empirical study on principal component analysis for clustering gene expression data". *Bioinformatics*. 17.9, 763-774.
- Ying, L. ve W. Yuanyuan. (2010). *Application of clustering on credit card customer segmentation based on AHP*. Logistics Systems and Intelligent Management, 2010 International Conference on, 1869-1873.
- Zabih, R. ve V. Kolmogorov. (2004). *Spatially coherent clustering using graph cuts*. Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, II-437-II-444 Vol. 432.
- Zakrzewska, D. ve J. Murlewski. (2005). *Clustering algorithms for bank customer segmentation*. Intelligent Systems Design and Applications, 2005. ISDA'05. Proceedings. 5th International Conference on, 197-202.
- Zhiyu, Z. ve L. Congdong. (2009). *Research on Application to Customer Classification in Management Decision-Making Based on Multivariate Statistics*. Information Technology and Applications, 2009. IFITA'09. International Forum on, 323-326.
- Ziff, R. (1971). Psychographics for market segmentation. *Journal of Advertising Research*. 11.2, 3-9.

İnternet Kaynakları

- Bornac, G. (2015). The Power of Store Clustering. <http://www.manh.com/resources/articles/2015/08/27/power-store-clustering>. 30.05.2016.
- Company, W. P. (2013). A Simple Approach to Retail Clustering. http://www.wilsonperumal.com/media/publications/PDFs/Vantage_Point_2013_Issue3.pdf. 30.05.2016.
- Donofrio, T. J. (2009). Advanced Planning and Optimization Part 3: Store Clustering. *Retail Systems and Services* <http://risnews.edgl.com/retail-news/Advanced-Planning-and-Optimization-Part-3--Store-Clustering38904> 26.05.2016.
- Gartner. (2016). Big Data. <http://www.gartner.com/it-glossary/big-data/>. 30.05.2016.
- Hürriyet Emlak Endeksi, www.hurriyetemlak.com/Emlak-Endeksi/Detayli-Analiz.20.01.2016
- Marr, B. (2016). A Brief History of Big Data Everyone Should Read. <https://www.linkedin.com/pulse/brief-history-big-data-everyone-should-read-bernard-marr>. 30.05.2016.
- Milliyet Emlak, www.milliyetemlak.com. 20.01.2016
- Neal, W. D. (2016). Principles of market segmentation: www.mktgsensei.com. 30.05.2016.
- Press, G. (2016). A Very Short History Of Big Data. <http://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/#1e89a9ce55da>.
- Singh, W. (2008). Robustness of three hierarchical agglomerative clustering techniques for ecological data. Access mode: http://www.unuftp.is/static/files/rannsoknarritegrdir/WarshaSingh_MastersThesis.pdf. 1.09.2016
- Vohra, G. (2011). Store Clustering. <http://analyticstraining.com/2011/store-clustering/>. 26.05.2016.
- Xu, R. ve D. C. Wunsch. (2009). Clustering. Retrieved from <http://site.ebrary.com/id/10257659>. 30.05.2016

EKLER



EK 1: İLK SEGMENTASYON YAKLAŞIMI İÇİN ANALİZLER

Farklı kaynaklardan elde edilmiş verilerle yapılmış olan segmentasyon sonucu 6 segment elde edilmişti. Bu segmentler için yapılan sık ürün küme ve birliktelik kuralları analizleri birinci ekte verilmiştir.

1. Segmente Ait Analizler

Birinci segmentte yer alan 40 mağazaya ait veriler, toplam 34 bin 998 fişten ve bu fişlere ait, her bir fişte bir ürün sadece bir kez yer alacak şekilde, toplam 112 bin 742 işlemden oluşmaktadır. Analiz öncesi bir fişte birden fazla aynı ürün yer alıyorsa; örneğin müşteri 3 adet makarna almışsa, ilgili aynı ürünün bir fişte bir kere yer alması sağlanmıştır. Her bir satış fişte ortalama 3 farklı ürün olduğu düşünülebilir. Tablodan görüldüğü gibi, sebze ürünü en çok satılan üründür ve 5696 fişte yer almıştır.

Tablo 1.1

1. Segmentte Sık Satılan Tek Öğeli Ürün Küme*

Ürün	Destek	Fiş sayısı
Sebze	0.1628	5696
Çikolata	0.1361	4764
Meyve	0.1262	4416
Sigara	0.1189	4163
Gazlı içecek	0.1083	3791
Tatlı bisküvi	0.0963	3371
Özel ekmek	0.0955	3341
Su	0.0863	3020
Süt	0.0748	2618
Peynir	0.0717	2510
Ekmek	0.0711	2489
Çerez	0.0650	2274
Tatlı	0.0622	2178
Ev gereçleri	0.0575	2014
Tuzlu bisküvi	0.0573	2004
Poğaç	0.0540	1890
Tavuk	0.0518	1813
Meyve suyu	0.0515	1803
Yoğurt	0.0486	1702
Et	0.0469	1642

* En çok satılan ürünler analizi yapılırken, sadece tek ürün elde edilecek şekilde kısıtlamaya gidilmiştir.

Tablo 2’de ise en çok birlikte satılan “ikili” ürünlerin tespiti amacıyla, analiz 2 ürün elde etmek kısıtlanmış ve yine sık 20 ürün kümesi elde edilmiştir.

Tablo 1.2

1. Segmentte Sık Satılan İki Öğeli Ürün Küme*

Ürün	Destek	Fiş sayısı
Meyve, sebze	0.05272	1845
Çikolata, tatlı bisküvi	0.03580	1253
Özel ekmek, sebze	0.03037	1063
Peynir, sebze	0.02986	1045
Çikolata, gazlı içe	0.02466	863
Sebze, süt	0.02380	833
Gazlı içecek, sebze	0.02372	830
Meyve, özel ekmek	0.02312	809
Tatlı bisküvi, tuzlu bisküvi	0.02266	793
Çikolata, sebze	0.02252	788
Çikolata, meyve	0.02146	751
Et, sebze	0.02117	741
Meyve, peynir	0.02083	729
Sebze, tavuk	0.02043	715
Meyve, süt	0.02034	712
Sebze, yumurta	0.02009	703
Ekmek, sebze	0.01977	692
Özel ekmek, peynir	0.01946	681
Çikolata, tuzlu bisküvi	0.01940	679
Sebze, yoğurt	0.01920	672

* Analiz yapılırken iki ürün elde edilecek şekilde kısıtlama yapılmıştır.

Tablo 3’te de en çok birlikte satılan “üçlü” ürünlerin tespiti amacıyla, analiz 3 ürün elde etmek için ve en çok satılan ilk 20 ürünün tespiti için yapılmıştır.

Tablo 1.3

1. Segmentte Sık Satılan Üç Öğeli Ürün Küme*

Ürün	Destek	Fiş sayısı
Meyve, peynir, sebze	0.0139	486
Meyve, özel ekmek, sebze	0.0136	476
Meyve, sebze, süt	0.0123	429
Çikolata, tatlı bisküvi, tuzlu bisküvi	0.0106	371
Çikolata, meyve, sebze	0.0106	371
Gazlı içecek, meyve, sebze	0.0104	365
Meyve, sebze, yumurta	0.0097	340
Et, meyve, sebze	0.0097	339
Özel ekmek, peynir, sebze	0.0097	338
Çerez, meyve, sebze	0.0093	327
Meyve, sebze, tatlı bisküvi	0.0089	313
Meyve, sebze, yoğurt	0.0088	309
Peynir, sebze, yumurta	0.0085	296
Peynir, sebze, süt	0.0083	290
Çikolata, gazlı içecek, tatlı bisküvi	0.0082	287
Meyve, sebze, tavuk	0.0081	282
Çikolata, sebze, tatlı bisküvi	0.0079	275
Çikolata, meyve, tatlı bisküvi	0.0078	273
Gazlı içecek, peynir, sebze	0.0075	264
Sebze, süt, yumurta	0.0075	262

* Analiz yapılırken üç ürün elde edilecek şekilde kısıtlama yapılmıştır.

Aşağıdaki tablo ve bundan sonraki tablolarda ise birliktelik kuralları analizleri mevcuttur. İlk etapta sadece iki ürünlü kurallar elde edilmiştir. İki ürünlü kural analizinde güven değeri yüksek tutulmuştur ancak en yüksek güven değeri % 56 olarak bulunmuştur. Destek değerinin de 0.003'ün altında olmaması istenmiştir, bu şekilde oluşan iki ürünlü ilk 20 kural aşağıdadır.

Tablo 1.4

1. Segmente Ait Güveni Yüksek İki Ürünlü İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{zeytin} => {peynir}	0.0115	0.5678	7.9170
{turşu} => {sebze}	0.0042	0.5345	3.2844
{tereyağı} => {sebze}	0.0066	0.5055	3.1059
{konserve} => {sebze}	0.0127	0.4837	2.9718
{zeytin} => {sebze}	0.0096	0.4760	2.9246
{yumurta} => {sebze}	0.0201	0.4628	2.8436
{tereyağı} => {peynir}	0.0059	0.4571	6.3741
{et} => {sebze}	0.0212	0.4513	2.7728
{balık} => {sebze}	0.0105	0.4482	2.7541
{bakliyat} => {sebze}	0.0159	0.4444	2.7303
{salam} => {peynir}	0.0126	0.4432	6.1800
{z yağı} => {sebze}	0.0037	0.4411	2.7101
{baharat} => {sebze}	0.0119	0.4357	2.6773
{kaşar} => {peynir}	0.0109	0.4318	6.0207
{gevrek} => {süt}	0.0069	0.4226	5.6495
{tuz} => {sebze}	0.0041	0.4194	2.5766
{meyve} => {sebze}	0.0527	0.4178	2.5671
{kaşar} => {sebze}	0.0106	0.4171	2.5630
{peynir} => {sebze}	0.0299	0.4163	2.5581
{sosis} => {peynir}	0.0035	0.4150	5.7860

* Analiz yapılırken sadece iki ürünlü kurallar ürün elde edilecek şekilde kısıtlama yapılmıştır.

Tablo 1.5

1. Segmente Ait Kaldırıcı Yüksek İki Ürünlü İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{un} => {şeker}	0.0038	0.3109	12.4071
{pasta malz.} => {un}	0.0042	0.1407	11.4238
{yağ} => {şeker}	0.0048	0.2522	10.0660
{peçete} => {poşet}	0.0032	0.1854	9.9231
{bulaşık det.} => {çamaşır det.}	0.0080	0.3125	9.7651
{bebek ürünü} => {mama}	0.0035	0.1485	9.7486
{kaşar} => {zeytin}	0.0047	0.1871	9.2511
{peçete} => {tuv.hav.}	0.0043	0.2483	8.8060
{şeker} => {çay}	0.0089	0.3558	8.7806
{temizlik} => {çamaşır det.}	0.0121	0.2751	8.5963
{temizlik malz.} => {temizlik}	0.0065	0.3709	8.4624
{un} => {bakliyat}	0.0037	0.2970	8.3217
{sabun} => {diş}	0.0040	0.1983	8.1166
{bulaşık det.} => {temizlik}	0.0090	0.3516	8.0209
{şampuan} => {sabun}	0.0043	0.1596	7.9669
{zeytin} => {peynir}	0.0115	0.5678	7.9170
{zeytin} => {sucuk}	0.0037	0.1836	7.7705
{zeytin} => {salam}	0.0044	0.2189	7.7005
{sos} => {konserve}	0.0050	0.2019	7.6956
{çorba} => {makarna}	0.0038	0.3037	7.6898

Aşağıdaki tabloda ve bir sonraki tabloda bu defa, üç ürün içeren güven değeri ve kaldırıcı değeri yüksek kurallar sunulmuştur.

Tablo 1.6

1. Segmente Ait Güveni Yüksek Üç Ürünlü İlişki Kuralları*

Kural	Destek	Güven	Kaldıraç
{meyve, peçete} => {sebze}	0.0033	0.7733	4.7516
{meyve, tereyağı} => {sebze}	0.0033	0.7550	4.6388
{salam, zeytin} => {peynir}	0.0033	0.7484	10.4351
{yumurta, zeytin} => {peynir}	0.0039	0.7459	10.3998
{konserve, meyve} => {sebze}	0.0052	0.7418	4.5579
{kaşar, zeytin} => {peynir}	0.0035	0.7410	10.3316
{bakliyat, meyve} => {sebze}	0.0074	0.7407	4.5513
{bakliyat, et} => {sebze}	0.0048	0.7293	4.4808
{kaşar, meyve} => {sebze}	0.0051	0.7247	4.4528
{meyve, temizlik} => {sebze}	0.0069	0.7147	4.3914
{et, meyve} => {sebze}	0.0097	0.7137	4.3851
{balık, meyve} => {sebze}	0.0050	0.7114	4.3710
{meyve, sos} => {sebze}	0.0041	0.7108	4.3673
{meyve, temizlik malzemesi} => {sebze}	0.0030	0.7020	4.3132
{meyve, yumurta} => {sebze}	0.0097	0.6967	4.2809
{et, yoğurt} => {sebze}	0.0047	0.6966	4.2800
{süt, zeytin} => {peynir}	0.0036	0.6902	9.6240
{meyve, salam} => {sebze}	0.0045	0.6886	4.2310
{et, yumurta} => {sebze}	0.0046	0.6864	4.2177
{makarna, meyve} => {sebze}	0.0067	0.6841	4.2031

* Sınırlama sadece üç ürünlü ilişkiler elde etmek için yapılmıştır. Destek değeri en az 100 fişin üstündedir.

Tablo 1.7

1. Segmente Ait Kaldırıcı Yüksek Üç Ürünlü İlişki Kuralları*

Kural	Destek	Güven	Kaldıraç
{kaşar, peynir} => {zeytin}	0.0035	0.3211	15.8751
{çamaşır det., makarna} => {bakliyat}	0.0032	0.5657	15.8502
{çay, peynir} => {şeker}	0.0034	0.3871	15.4477
{bulaşık det., temizlik} => {çamaşır det.}	0.0044	0.4857	15.1777
{bakliyat, temizlik} => {çamaşır det.}	0.0036	0.4846	15.1434
{temizlik, tuv.hav.} => {çamaşır det.}	0.0033	0.4691	14.6597
{makarna, peynir} => {sos}	0.0035	0.3494	14.1873
{çamaşır det., temizlik} => {bulaşık det.}	0.0044	0.3626	14.1617
{makarna, yumurta} => {bakliyat}	0.0034	0.5000	14.0104
{makarna, temizlik} => {bakliyat}	0.0033	0.4958	13.8917
{çay, makarna} => {bakliyat}	0.0031	0.4846	13.5784
{bakliyat, konserve} => {makarna}	0.0035	0.5354	13.5585
{çamaşır det., tuv.hav.} => {temizlik}	0.0033	0.5729	13.0698
{peynir, salam} => {zeytin}	0.0033	0.2630	13.0026
{süt, temizlik} => {çamaşır det.}	0.0037	0.4146	12.9542
{konserve, makarna} => {bakliyat}	0.0035	0.4601	12.8917
{peynir, yumurta} => {zeytin}	0.0039	0.2596	12.8334
{peynir, temizlik} => {bulaşık det.}	0.0031	0.3244	12.6713
{çamaşır det., süt} => {temizlik}	0.0037	0.5551	12.6642
{bakliyat, makarna} => {konserve}	0.0035	0.3315	12.6384

* Sınırlama sadece üç ürünlü ilişkiler elde etmek için yapılmıştır.

2. Segmente Ait Analizler

16 mağazaya ait 1 günlük satış verisi toplam 9 bin 779 fişten ve bu fişlere ait toplam 27 bin 554 işlemden oluşmaktadır.

Tablo 2.1

2. Segmentte Sık Satılan Tek Öğeli Ürün Küme

Ürün	Destek	Fiş sayısı
Sigara	0.1689	1686
Sebze	0.1436	1433
Çikolata	0.1285	1283
Gazlı içe	0.1137	1135
Meyve	0.1057	1055
Tatlı bisküvi	0.0974	971
Ekmek	0.0798	796
Özel ekmek	0.0775	774
Su	0.0759	757
Süt	0.0703	701
Poğaç	0.0652	651
Peynir	0.0649	648
Çerez	0.0629	628
Tuzlu bisküvi	0.0569	567
Gazete	0.0544	543
Ev gereçleri	0.0512	511
Meyve suyu	0.0491	490
Kek	0.0442	441
Kahve	0.0431	430
Cips	0.0425	425

Tablo 2.2**2. Segmentte Sık Satılan İki Öğeli Ürün Küme**

Ürün	Destek	Fiş sayısı
Meyve, Sebze	0.0384	384
Çikolata, Tatlı bisküvi	0.0323	322
Ekmek, Sebze	0.0234	234
Çikolata, Gazlı içecek	0.0234	234
Özel ekmek, Sebze	0.0227	227
Peynir, Sebze	0.0212	211
Gazlı içecek, Sebze	0.0212	211
Sebze, Süt	0.0198	198
Tatlı bisküvi, Tuzlu bisküvi	0.0194	194
Gazlı içecek, Sigara	0.0178	178
Çikolata, Sigara	0.0173	172
Et, Sebze	0.0172	171
Çikolata, Sebze	0.0169	168
Meyve, Süt	0.0164	163
Sebze, Yumurta	0.0160	159
Çikolata, Meyve	0.0160	159
Sebze, Yoğurt	0.0157	157
Çikolata, Tuzlu bisküvi	0.0157	157
Gazlı içe, Tatlı bisküvi	0.0154	154
Sebze, Tavuk	0.0154	154

Tablo 2.3**2. Segmentte Sık Satılan Üç Ögeli Ürün Küme**

Ürün	Destek	Fiş sayısı
Meyve, Sebze, Süt	0.0089	89
Çikolata, Kek, Tatlı bisküvi	0.0075	74
Çikolata, Gazlı içecek, Tatlı bisküvi	0.0074	73
Gazlı içecek, Meyve, Sebze	0.0073	72
Çikolata, Meyve, Sebze	0.0073	72
Çikolata, Tatlı bisküvi, Tuzlu bisküvi	0.0072	71
Meyve, Peynir, Sebze	0.0070	69
Meyve, Özel ekmek, Sebze	0.0065	65
Meyve, Sebze, Yoğurt	0.0064	64
Meyve, Sebze, Tatlı bisküvi	0.0063	63
Çerez, Meyve, Sebze	0.0061	61
Çikolata, Meyve, Tatlı bisküvi	0.0058	58
Sebze, Süt, Yumurta	0.0056	56
Sebze, Süt, Yoğurt	0.0056	56
Ekmek, Meyve, Sebze	0.0056	56
Özel ekmek, Peynir, Sebze	0.0056	56
Çikolata, Sebze, Tatlı bisküvi	0.0055	55
Et, Meyve, Sebze	0.0053	53
Meyve, Sebze, Yumurta	0.0053	53
Peynir, Sebze, Süt	0.0053	53

Tablo 2.4

2. Segmente Ait Güveni Yüksek İki Ürün İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{salam} => {peynir}	0.0102	0.5000	7.7000
{et} => {sebze}	0.0172	0.4941	3.4416
{konserve} => {sebze}	0.0101	0.4459	3.1061
{yumurta} => {sebze}	0.0160	0.4394	3.0607
{bakliyat} => {sebze}	0.0108	0.4206	2.9298
{tavuk} => {sebze}	0.0152	0.4116	2.8669
{makarna} => {sebze}	0.0116	0.4036	2.8109
{yoğurt} => {sebze}	0.0157	0.3747	2.6098
{meyve} => {sebze}	0.0384	0.3636	2.5328
{tuzlu bisk.} => {tatlı bisk.}	0.0194	0.3417	3.5102
{kek} => {tatlı bisk.}	0.0149	0.3380	3.4716
{kek} => {çikolata}	0.0148	0.3356	2.6112
{tatlı bisk.} => {çikolata}	0.0323	0.3319	2.5823
{peynir} => {sebze}	0.0212	0.3260	2.2705
{cips} => {gazlı içe}	0.0133	0.3125	2.7481
{ekmek} => {sebze}	0.0234	0.2936	2.0449
{özel ekmek} => {sebze}	0.0227	0.2929	2.0399
{yumurta} => {süt}	0.0104	0.2873	4.0899
{cips} => {çikolata}	0.0122	0.2861	2.2254
{süt} => {sebze}	0.0198	0.2824	1.9669

Tablo 2.5

2. Segmente Ait Kaldırıcı Yüksek İki Ürünlü İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{salam} => {peynir}	0.0102	0.5000	7.7000
{peynir} => {yumurta}	0.0101	0.1559	4.2946
{yumurta} => {süt}	0.0104	0.2873	4.0899
{yoğurt} => {süt}	0.0119	0.2822	4.0175
{tatlı bisk.} => {tuzlu bisk.}	0.0194	0.1996	3.5102
{kek} => {tatlı bisk.}	0.0149	0.3380	3.4716
{et} => {sebze}	0.0172	0.4941	3.4416
{konserve} => {sebze}	0.0101	0.4459	3.1061
{yumurta} => {sebze}	0.0160	0.4394	3.0607
{bakliyat} => {sebze}	0.0108	0.4206	2.9298
{tavuk} => {sebze}	0.0152	0.4116	2.8669
{makarna} => {sebze}	0.0116	0.4036	2.8109
{poğaç} => {peynir}	0.0119	0.1818	2.8000
{cips} => {gazlı içe.}	0.0133	0.3125	2.7481
{kek} => {çikolata}	0.0148	0.3356	2.6112
{yoğurt} => {sebze}	0.0157	0.3747	2.6098
{özel ekmek} => {peynir}	0.0131	0.1689	2.6005
{tatlı bisk.} => {çikolata}	0.0323	0.3319	2.5823
{süt} => {peynir}	0.0118	0.1674	2.5779
{meyve} => {sebze}	0.0384	0.3636	2.5328

Tablo 2.6

2. Segmente Ait Güveni Yüksek Üç Ürün İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{meyve, yoğurt} => {sebze}	0.0064	0.6000	4.1791
{meyve, peynir} => {sebze}	0.0070	0.5620	3.9143
{meyve, süt} => {sebze}	0.0089	0.5438	3.7873
{çikolata, kek} => {tatlı bisk.}	0.0075	0.5034	5.1715
{gazlı içe, meyve} => {sebze}	0.0073	0.4830	3.3641
{gazlı içe, tatlı bisk.} => {çikolata}	0.0074	0.4768	3.7095
{çerez, sebze} => {meyve}	0.0061	0.4615	4.3650
{çikolata, meyve} => {sebze}	0.0073	0.4551	3.1700
{çikolata, tuzlu bisk.} => {tatlı bisk.}	0.0072	0.4545	4.6691
{sebze, süt} => {meyve}	0.0089	0.4485	4.2412
{meyve, özel ekmek} => {sebze}	0.0065	0.4384	3.0532
{sebze, tatlı bisk.} => {meyve}	0.0063	0.4366	4.1293
{çerez, meyve} => {sebze}	0.0061	0.4348	3.0283
{meyve, tatlı bisk.} => {sebze}	0.0063	0.4306	2.9989
{çikolata, sebze} => {meyve}	0.0073	0.4303	4.0696
{sebze, yoğurt} => {meyve}	0.0064	0.4091	3.8690
{tatlı bisk., tuzlu bisk.} => {çikolata}	0.0072	0.3684	2.8662
{gazlı içe, sebze} => {meyve}	0.0073	0.3430	3.2439
{peynir, sebze} => {meyve}	0.0070	0.3285	3.1068
{çikolata, gazlı içe.} => {tatlı bisk.}	0.0074	0.3144	3.2296

Tablo 2.7

2. Kümedeki Mağazalara Ait Kaldırıcı Yüksek Üç Ürün İlişki Kuralları*

Kural	Destek	Güven	Kaldıraç
{çikolata, tatlı bisk} => {kek}	0.0075	0.2310	5.2293
{çikolata, kek} => {tatlı bisk}	0.0075	0.5034	5.1715
{çikolata, tuzlu bisk} => {tatlı bisk}	0.0072	0.4545	4.6691
{çerez, sebze} => {meyve}	0.0061	0.4615	4.3650
{sebze, süt} => {meyve}	0.0089	0.4485	4.2412
{meyve, yoğurt} => {sebze}	0.0064	0.6000	4.1791
{sebze, tatlı bisk} => {meyve}	0.0063	0.4366	4.1293
{çikolata, sebze} => {meyve}	0.0073	0.4303	4.0696
{meyve, sebze} => {yoğurt}	0.0064	0.1676	3.9866
{meyve, peynir} => {sebze}	0.0070	0.5620	3.9143
{çikolata, tatlı bisk} => {tuzlu bisk}	0.0072	0.2215	3.8961
{kek, tatlı bisk} => {çikolata}	0.0075	0.5000	3.8898
{sebze, yoğurt} => {meyve}	0.0064	0.4091	3.8690
{meyve, süt} => {sebze}	0.0089	0.5438	3.7873
{gazlı içe, tatlı bisk} => {çikolata}	0.0074	0.4768	3.7095
{gazlı içe, meyve} => {sebze}	0.0073	0.4830	3.3641
{meyve, sebze} => {süt}	0.0089	0.2314	3.2936
{gazlı içe, sebze} => {meyve}	0.0073	0.3430	3.2439
{çikolata, gazlı içe} => {tatlı bisk}	0.0074	0.3144	3.2296
{çikolata, meyve} => {sebze}	0.0073	0.4551	3.1700

3. SEGMENTE AİT ANALİZLER

Üçüncü segmentte yer alan 38 mağazaya ait veriler, toplam 29 bin 980 fişten ve bu fişlere ait her bir fişte bir ürün sadece bir kez yer alacak şekilde, toplam 92 bin 680 işlemden oluşmaktadır

Tablo 3.1

3. Segmentte Sık Satılan Tek Öğeli Ürün Küme

Ürün	Destek	Fiş sayısı
Sebze	0.1990	5966
Meyve	0.1514	4540
Sigara	0.1274	3818
Çikolata	0.1188	3561
Özel ekmek	0.1123	3368
Gazlı içecek	0.1044	3130
Süt	0.0937	2808
Tatlı bisküvi	0.0855	2564
Peynir	0.0773	2316
Su	0.0758	2272
Çerez	0.0657	1970
Yoğurt	0.0655	1964
Tuzlu bisküvi	0.0621	1861
Gazete	0.0599	1797
Ekmek	0.0570	1708
Ev gereçleri	0.0556	1666
Yumurta	0.0512	1534
Kahve	0.0481	1441
Meyve suyu	0.0435	1304
Tatlı	0.0432	1295

Tablo 3.2**3. Segmentte Sık Satılan İki Ögeli Ürün Küme**

Ürün	Destek	Fiş sayısı
Meyve, Sebze	0.0654	1961
Özel ekmek, Sebze	0.0383	1147
Peynir, Sebze	0.0342	1026
Sebze, Süt	0.0319	956
Çikolata, Tatlı bisküvi	0.0300	900
Meyve, Özel ekmek	0.0299	895
Meyve, Süt	0.0263	789
Sebze, Yoğurt	0.0258	774
Özel ekmek, Peynir	0.0253	759
Meyve, Peynir	0.0246	739
Sebze, Yumurta	0.0243	728
Gazlı içecek, Sebze	0.0236	709
Çikolata, Sebze	0.0232	695
Özel ekmek, Süt	0.0230	689
Çikolata, Meyve	0.0225	674
Çikolata, Gazlı içe	0.0224	673
Tatlı bisküvi, Tuzlu bisküvi	0.0194	582
Meyve, Yoğurt	0.0192	575
Peynir, Süt	0.0190	570
Çikolata, Süt	0.0189	568

Tablo 3.3**3. Segmentte Sık Satılan Üç Öğeli Ürün Küme**

Ürün	Destek	Fiş sayısı
Meyve, Özel ekmek, Sebze	0.0176	529
Meyve, Peynir, Sebze	0.0170	510
Meyve, Sebze, Süt	0.0154	461
Özel ekmek, Peynir, Sebze	0.0129	387
Meyve, Sebze, Yoğurt	0.0122	365
Meyve, Sebze, Yumurta	0.0116	348
Çikolata, Meyve, Sebze	0.0111	334
Özel ekmek, Sebze, Süt	0.0107	320
Meyve, Özel ekmek, Peynir	0.0105	316
Peynir, Sebze, Süt	0.0102	307
Gazlı içe, Meyve, Sebze	0.0102	305
Peynir, Sebze, Yumurta	0.0098	295
Çerez, Meyve, Sebze	0.0096	287
Meyve, Özel ekmek, Süt	0.0094	283
Sebze, Süt, Yumurta	0.0093	278
Özel ekmek, Sebze, Yoğurt	0.0090	269
Sebze, Süt, Yoğurt	0.0089	268
Meyve, Peynir, Süt	0.0089	268
Özel ekmek, Peynir, Süt	0.0085	254
Özel ekmek, Sebze, Yumurta	0.0083	250

Tablo 3.4

3. Segmente Ait Güveni Yüksek İki Ürün İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{zeytin} => {peynir}	0.0102	0.5881	7.6131
{konserve} => {sebze}	0.0137	0.5269	2.6479
{zeytin} => {sebze}	0.0088	0.5057	2.5415
{et} => {sebze}	0.0182	0.5023	2.5241
{un} => {sebze}	0.0055	0.4911	2.4680
{baharat} => {sebze}	0.0123	0.4856	2.4400
{balık} => {sebze}	0.0145	0.4807	2.4154
{bakliyat} => {sebze}	0.0148	0.4805	2.4145
{yumurta} => {sebze}	0.0243	0.4746	2.3848
{tere yağ} => {sebze}	0.0079	0.4436	2.2292
{peynir} => {sebze}	0.0342	0.4430	2.2262
{tavuk} => {sebze}	0.0181	0.4407	2.2143
{kaşar} => {sebze}	0.0111	0.4405	2.2135
{salam} => {peynir}	0.0093	0.4321	5.5934
{meyve} => {sebze}	0.0654	0.4319	2.1706
{kuru meyve} => {çerez}	0.0057	0.4286	6.5221
{yağ} => {sebze}	0.0057	0.4286	2.1536
{sos} => {sebze}	0.0110	0.4273	2.1471
{pasta malzemesi} => {sebze}	0.0115	0.4266	2.1439
{makarna} => {sebze}	0.0156	0.4264	2.1425

Tablo 3.5

3. Segmente Ait Kaldırıcı Yüksek İki Ürünlü İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{un} => {pasta malz.}	0.0038	0.3373	12.4681
{bulaşık det.} => {çamaşır det.}	0.0054	0.2588	11.3927
{peçete} => {tuv.hav.}	0.0046	0.2680	10.4603
{çamaşır det.} => {temizlik}	0.0090	0.3950	10.3426
{temizlik malz.} => {temizlik}	0.0046	0.3603	9.4342
{bulaşık det.} => {temizlik}	0.0074	0.3562	9.3273
{sabun} => {temizlik}	0.0051	0.2969	7.7732
{çamaşır det.} => {tuv.hav.}	0.0045	0.1968	7.6812
{şeker} => {çay}	0.0054	0.3051	7.6284
{peynir} => {zeytin}	0.0102	0.1326	7.6131
{şeker} => {bakliyat}	0.0041	0.2316	7.5320
{salam} => {kaşar}	0.0041	0.1898	7.5273
{poşet} => {tuv.hav.}	0.0035	0.1927	7.5208
{poşet} => {temizlik}	0.0050	0.2752	7.2064
{bakliyat} => {makarna}	0.0080	0.2603	7.0945
{bakliyat} => {konserve}	0.0056	0.1811	6.9618
{bulaşık det.} => {tuv.hav.}	0.0037	0.1773	6.9218
{kuru meyve} => {çerez}	0.0057	0.4286	6.5221
{peçete} => {temizlik}	0.0043	0.2485	6.5077
{sos} => {konserve}	0.0043	0.1688	6.4892

Tablo 3.6

3. Segmente Ait Güveni Yüksek Üç Ürün İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{meyve, peçete} => {sebze}	0.0037	0.7303	3.6697
{baharat, meyve} => {sebze}	0.0057	0.7257	3.6469
{süt, zeytin} => {peynir}	0.0036	0.7248	9.3828
{meyve, şeker} => {sebze}	0.0036	0.7248	3.6424
{bakliyat, meyve} => {sebze}	0.0070	0.7201	3.6188
{balık, meyve} => {sebze}	0.0074	0.7175	3.6057
{konserve, yumurta} => {sebze}	0.0038	0.7170	3.6029
{meyve, zeytin} => {sebze}	0.0046	0.7113	3.5746
{meyve, poşet} => {sebze}	0.0037	0.7044	3.5397
{konserve, meyve} => {sebze}	0.0057	0.7037	3.5362
{çamaşır det, meyve} => {sebze}	0.0055	0.7009	3.5219
{meyve, pasta malz} => {sebze}	0.0057	0.6980	3.5073
{meyve, sos} => {sebze}	0.0052	0.6964	3.4997
{et, yumurta} => {sebze}	0.0048	0.6938	3.4863
{et, meyve} => {sebze}	0.0080	0.6925	3.4801
{meyve, peynir} => {sebze}	0.0170	0.6901	3.4680
{kaşar, meyve} => {sebze}	0.0058	0.6890	3.4622
{sebze, zeytin} => {peynir}	0.0060	0.6856	8.8750
{özel ekmek, zeytin} => {peynir}	0.0041	0.6833	8.8456
{meyve, zeytin} => {peynir}	0.0044	0.6804	8.8078

Tablo 3.7

3. Segmente Ait Kaldırıcı Yüksek Üç Ürünlü İlişki Kuralları*

Kural	Destek	Güven	Kaldıraç
{meyve, temizlik} => {çamaşır det.}	0.0040	0.3552	15.6382
{sebze, temizlik} => {çamaşır det.}	0.0040	0.3193	14.0550
{çamaşır det., meyve} => {temizlik}	0.0040	0.5085	13.3155
{çamaşır det., sebze} => {temizlik}	0.0040	0.4672	12.2324
{peynir, süt} => {zeytin}	0.0036	0.1895	10.8820
{meyve, peynir} => {zeytin}	0.0044	0.1786	10.2587
{peynir, sebze} => {zeytin}	0.0060	0.1764	10.1319
{peynir, yumurta} => {kaşar}	0.0040	0.2526	10.0184
{makarna, sebze} => {bakliyat}	0.0047	0.3006	9.7757
{bakliyat, sebze} => {konserve}	0.0037	0.2506	9.6307
{süt, zeytin} => {peynir}	0.0036	0.7248	9.3828
{özel ekmek, peynir} => {zeytin}	0.0041	0.1621	9.3073
{sebze, zeytin} => {peynir}	0.0060	0.6856	8.8750
{özel ekmek, zeytin} => {peynir}	0.0041	0.6833	8.8456
{meyve, zeytin} => {peynir}	0.0044	0.6804	8.8078
{konserve, sebze} => {bakliyat}	0.0037	0.2701	8.7818
{bakliyat, sebze} => {makarna}	0.0047	0.3183	8.6747
{peynir, süt} => {kaşar}	0.0041	0.2175	8.6269
{kaşar, yumurta} => {peynir}	0.0040	0.6557	8.4883
{meyve, peynir} => {kaşar}	0.0050	0.2030	8.0493

4. SEGMENTE AİT ANALİZLER

Dördüncü segmentte yer alan 45 mağazaya ait veriler, toplam 42 bin 193 fişten ve bu fişlere ait her bir fişte bir ürün sadece bir kez yer alacak şekilde, toplam 135 bin 966 işlemden oluşmaktadır.

Tablo 4.1

4. Segmentte Sık Satılan Tek Öğeli Ürün Küme

Ürün	Destek	Fiş sayısı
Sebze	0.1806	7621
Çikolata	0.1363	5751
Sigara	0.1338	5645
Meyve	0.1292	5453
Gazlı içecek	0.1158	4886
Özel ekmek	0.1033	4359
Tatlı bisküvi	0.0958	4043
Süt	0.0891	3759
Su	0.0814	3434
Peynir	0.0740	3124
Ekmek	0.0669	2821
Çerez	0.0660	2786
Tuzlu bisküvi	0.0620	2617
Yoğurt	0.0566	2389
Gazete	0.0556	2348
Ev gereçleri	0.0550	2321
Meyve suyu	0.0512	2159
Tatlı	0.0504	2128
Yumurta	0.0492	2077
Poğaç	0.0477	2012

Tablo 4.2**4. Segmentte Sık Satılan Çift Öğeli Ürün Küme**

Ürün	Destek	Fiş sayısı
Meyve, Sebze	0.0616	2601
Çikolata, Tatlı bisküvi	0.0363	1531
Özel ekmek, Sebze	0.0358	1512
Sebze, Süt	0.0321	1354
Peynir, Sebze	0.0308	1298
Çikolata, Gazlı içe	0.0272	1148
Meyve, Süt	0.0264	1112
Çikolata, Sebze	0.0255	1077
Gazlı içe, Sebze	0.0249	1051
Meyve, Özel ekmek	0.0244	1029
Sebze, Yoğurt	0.0236	996
Tatlı bisküvi, Tuzlu bisküvi	0.0231	974
Sebze, Yumurta	0.0229	966
Meyve, Peynir	0.0216	910
Çikolata, Meyve	0.0212	893
Özel ekmek, Süt	0.0211	890
Özel ekmek, Peynir	0.0208	879
Ekmek, Sebze	0.0205	864
Çikolata, Tuzlu bisküvi	0.0203	856
Çikolata, Süt	0.0200	844

Tablo 4.3**4. Segmentte Sık Satılan Üç Ögeli Ürün Küme**

Ürün	Destek	Fiş sayısı
Meyve, Sebze, Süt	0.01716	724
Meyve, Özel ekmek, Sebze	0.01588	670
Meyve, Peynir, Sebze	0.01503	634
Çikolata, Meyve, Sebze	0.01164	491
Meyve, Sebze, Yumurta	0.01131	477
Meyve, Sebze, Yoğurt	0.01114	470
Özel ekmek, Peynir, Sebze	0.01076	454
Çikolata, Tatlı bisküvi, Tuzlu bisküvi	0.01071	452
Gazlı içe, Meyve, Sebze	0.01069	451
Özel ekmek, Sebze, Süt	0.01048	442
Peynir, Sebze, Süt	0.01000	422
Çerez, Meyve, Sebze	0.00988	417
Meyve, Sebze, Tatlı bisküvi	0.00917	387
Çikolata, Gazlı içe, Tatlı bisküvi	0.00901	380
Sebze, Süt, Yumurta	0.00894	377
Sebze, Süt, Yoğurt	0.00879	371
Çikolata, Sebze, Tatlı bisküvi	0.00865	365
Et, Meyve, Sebze	0.00851	359
Peynir, Sebze, Yumurta	0.00846	357
Meyve, Özel ekmek, Süt	0.00822	347

Tablo 4.4

4. Segmente Ait Güveni Yüksek İki Ürün İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{zeytin} => {peynir}	0.0118	0.5864	7.9195
{turşu} => {sebze}	0.0037	0.4936	2.7329
{zeytin} => {sebze}	0.0098	0.4877	2.6999
{baharat} => {sebze}	0.0119	0.4869	2.6956
{et} => {sebze}	0.0199	0.4855	2.6882
{meyve} => {sebze}	0.0616	0.4770	2.6408
{sosis} => {sebze}	0.0033	0.4747	2.6284
{balık} => {sebze}	0.0109	0.4722	2.6144
{tere yağ} => {sebze}	0.0068	0.4669	2.5849
{yumurta} => {sebze}	0.0229	0.4651	2.5750
{konserve} => {sebze}	0.0120	0.4604	2.5491
{z yağı} => {sebze}	0.0036	0.4481	2.4807
{bakliyat} => {sebze}	0.0154	0.4433	2.4543
{tere yağ} => {peynir}	0.0065	0.4410	5.9566
{kaşar} => {peynir}	0.0105	0.4332	5.8509
{un} => {sebze}	0.0053	0.4310	2.3864
{sos} => {sebze}	0.0111	0.4292	2.3764
{margarin} => {sebze}	0.0055	0.4291	2.3757
{tavuk} => {sebze}	0.0200	0.4253	2.3548
{pasta malzemesi} => {sebze}	0.0136	0.4210	2.3309

Tablo 4.5

4. Segmente Ait Kaldırıcı Yüksek İki Ürünlü İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{un} => {şeker}	0.0029	0.2356	11.5337
{çamaşır det.} => {bulaşık det.}	0.0075	0.2584	11.4636
{şeker} => {yağ}	0.0037	0.1798	11.4261
{mama} => {bebek ürünü}	0.0030	0.1994	10.9530
{un} => {pasta malz.}	0.0040	0.3257	10.0963
{margarin} => {şeker}	0.0026	0.2026	9.9158
{sucuk} => {salam}	0.0037	0.2155	9.5397
{temizlik malz.} => {temizlik}	0.0053	0.3633	8.8410
{çamaşır det.} => {temizlik}	0.0104	0.3590	8.7343
{sos} => {konserve}	0.0057	0.2197	8.4336
{salam} => {kaşar}	0.0046	0.2025	8.3938
{sucuk} => {kaşar}	0.0035	0.2017	8.3581
{sucuk} => {zeytin}	0.0029	0.1685	8.3547
{yağ} => {konserve}	0.0034	0.2169	8.3260
{sabun} => {diş}	0.0035	0.1806	8.2920
{bulaşık det.} => {temizlik}	0.0077	0.3396	8.2644
{temizlik malz.} => {bulaşık det.}	0.0027	0.1849	8.2054
{parfüm} => {kozmetik}	0.0025	0.1967	8.1763
{zeytin} => {kaşar}	0.0039	0.1951	8.0848
{peynir} => {zeytin}	0.0118	0.1597	7.9195

Tablo 4.6

4. Segmente Ait Güveni Yüksek Üç Ürünlü İlişki Kuralları*

Kural	Destek	Güven	Kaldıraç
{et, meyve} => {sebze}	0.0085	0.7754	4.2928
{baharat, meyve} => {sebze}	0.0054	0.7729	4.2790
{balık, meyve} => {sebze}	0.0050	0.7599	4.2069
{meyve, zeytin} => {sebze}	0.0051	0.7597	4.2061
{konserve, meyve} => {sebze}	0.0055	0.7557	4.1839
{meyve, tavuk} => {sebze}	0.0077	0.7414	4.1048
{salam, zeytin} => {peynir}	0.0026	0.7400	9.9945
{yoğurt, zeytin} => {peynir}	0.0029	0.7378	9.9649
{meyve, yumurta} => {sebze}	0.0113	0.7361	4.0754
{kaşar, zeytin} => {peynir}	0.0029	0.7349	9.9262
{bakliyat, meyve} => {sebze}	0.0072	0.7348	4.0681
{meyve, sos} => {sebze}	0.0050	0.7301	4.0422
{meyve, yağ} => {sebze}	0.0028	0.7273	4.0265
{makarna, meyve} => {sebze}	0.0081	0.7149	3.9579
{yumurta, zeytin} => {peynir}	0.0037	0.7136	9.6384
{meyve, şeker} => {sebze}	0.0037	0.7136	3.9510
{meyve, salam} => {sebze}	0.0043	0.7121	3.9423
{meyve, un} => {sebze}	0.0027	0.7063	3.9101
{donmuş gıda, meyve} => {sebze}	0.0033	0.7041	3.8981
{meyve, tere yağ} => {sebze}	0.0034	0.7039	3.8970

Tablo 4.7

4. Segmente Ait Kaldırıcı Yüksek Üç Ürün İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{bulaşık det., temizlik} => {çamaşır det.}	0.0038	0.4985	17.1964
{çamaşır det., temizlik} => {bulaşık det.}	0.0038	0.3667	16.2713
{bulaşık det., çamaşır det.} => {temizlik}	0.0038	0.5095	12.3974
{bakliyat, peynir} => {makarna}	0.0041	0.4958	12.2898
{peynir, yumurta} => {zeytin}	0.0037	0.2449	12.1437
{makarna, peynir} => {bakliyat}	0.0041	0.4137	11.9233
{makarna, süt} => {bakliyat}	0.0037	0.3921	11.2993
{bakliyat, süt} => {makarna}	0.0037	0.4093	10.1473
{makarna, meyve} => {bakliyat}	0.0040	0.3501	10.0901
{peynir, sebze} => {zeytin}	0.0063	0.2034	10.0842
{bakliyat, meyve} => {makarna}	0.0040	0.4063	10.0729
{sebze, temizlik} => {çamaşır det.}	0.0039	0.2916	10.0598
{meyve, peynir} => {zeytin}	0.0043	0.2011	9.9706
{yumurta, zeytin} => {peynir}	0.0037	0.7136	9.6384
{çamaşır det., sebze} => {temizlik}	0.0039	0.3909	9.5114
{makarna, sebze} => {bakliyat}	0.0054	0.3280	9.4541
{peynir, süt} => {kaşar}	0.0040	0.2163	8.9659
{meyve, zeytin} => {peynir}	0.0043	0.6466	8.7336
{bakliyat, sebze} => {makarna}	0.0054	0.3498	8.6709
{sebze, zeytin} => {peynir}	0.0063	0.6361	8.5918

5. SEGMENTE AİT ANALİZLER

Beşinci kümede yer alan 18 mağazaya ait veriler, toplam 17 bin 783 fişten ve bu fişlere ait her bir fişte bir ürün sadece bir kez yer alacak şekilde, toplam 47 bin 156 işlemden oluşmaktadır.

Tablo 5.1

5. Segmentte Sık Satılan Tek Öğeli Ürün Küme

Ürün	Destek	Fiş sayısı
Çikolata	0.1471	2616
Sigara	0.1425	2534
Sebze	0.1314	2336
Gazlı içe	0.1118	1988
Meyve	0.1103	1961
Tatlı bisküvi	0.1013	1802
Su	0.0901	1603
Tuzlu bisküvi	0.0754	1341
Poğaç	0.0744	1323
Süt	0.0733	1304
Meyve suyu	0.0643	1144
Özel ekmek	0.0626	1114
Çerez	0.0615	1093
Peynir	0.0557	991
Ekmek	0.0487	866
Yoğurt	0.0438	779
Hazır yemek	0.0437	777
Kek	0.0423	752
Kahve	0.0422	751
Tatlı	0.0388	690

Tablo 5.2**5. Segmentte Sık Satılan Çift Ögeli Ürün Küme**

Ürün	Destek	Fiş sayısı
Meyve, Sebze	0.0349	621
Çikolata, Tatlı bisküvi	0.0339	602
Çikolata, Gazlı içecek	0.0255	454
Tatlı bisküvi, Tuzlu bisküvi	0.0225	400
Çikolata, Tuzlu bisküvi	0.0222	394
Özel ekmek, Sebze	0.0201	358
Peynir, Sebze	0.0193	343
Sebze, Süt	0.0181	321
Meyve, Süt	0.0177	314
Gazlı içe, Sebze	0.0168	298
Gazlı içe, Tatlı bisküvi	0.0163	289
Çikolata, Meyve	0.0160	284
Çikolata, Su	0.0159	283
Çikolata, Sebze	0.0154	274
Çikolata, Kek	0.0152	270
Çikolata, Süt	0.0152	270
Çerez, Çikolata	0.0145	257
Ekmek, Sebze	0.0144	256
Gazlı içe, Meyve	0.0140	249
Sebze, Yoğurt	0.0136	241

* Analiz yapılırken iki ürün elde edilecek şekilde kısıtlama yapılmıştır.

Tablo 5.3**5. Segmentte Sık Satılan Üç Ögeli Ürün Küme**

Ürün	Destek	Fiş sayısı
Çikolata, Tatlı bisküvi, Tuzlu bisküvi	0.0094	167
Meyve, Sebze, Süt	0.0084	149
Meyve, Peynir, Sebze	0.0073	130
Meyve, Özel ekmek, Sebze	0.0073	129
Çikolata, Gazlı içe, Tatlı bisküvi	0.0069	122
Gazlı içe, Meyve, Sebze	0.0061	108
Meyve, Sebze, Yoğurt	0.0060	107
Çikolata, Meyve, Sebze	0.0059	105
Özel ekmek, Peynir, Sebze	0.0056	99
Çikolata, Kek, Tatlı bisküvi	0.0055	98
Özel ekmek, Sebze, Süt	0.0052	93
Çikolata, Süt, Tatlı bisküvi	0.0052	93
Meyve, Sebze, Yumurta	0.0052	92
Peynir, Sebze, Süt	0.0051	91
Peynir, Sebze, Yumurta	0.0047	84
Gazlı içecek, Sebze, Süt	0.0047	84
Sebze, Süt, Yumurta	0.0047	83
Çerez, Çikolata, Gazlı içecek	0.0047	83
Meyve, Sebze, Tatlı bisküvi	0.0047	83
Sebze, Süt, Yoğurt	0.0046	82

Tablo 5.4

5. Segmente Ait Güveni Yüksek İki Ürün İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{zeytin} => {peynir}	0.0062	0.5314	9.5357
{et} => {sebze}	0.0106	0.4701	3.5791
{gevrek} => {süt}	0.0065	0.4514	6.1553
{yumurta} => {sebze}	0.0127	0.4464	3.3985
{balık} => {sebze}	0.0073	0.4403	3.3516
{baharat} => {sebze}	0.0085	0.4393	3.3443
{konserve} => {sebze}	0.0084	0.4331	3.2973
{tavuk} => {sebze}	0.0111	0.4292	3.2673
{kuru meyve} => {çerez}	0.0066	0.4120	6.7027
{bakliyat} => {sebze}	0.0078	0.3850	2.9312
{sos} => {sebze}	0.0064	0.3787	2.8832
{çamaşır det} => {temizlik}	0.0064	0.3742	14.1272
{cips} => {çikolata}	0.0089	0.3597	2.4454
{kek} => {çikolata}	0.0152	0.3590	2.4407
{peynir} => {sebze}	0.0193	0.3461	2.6348
{makarna} => {sebze}	0.0091	0.3448	2.6245
{şekerleme} => {çikolata}	0.0085	0.3447	2.3435
{tatlı bisk} => {çikolata}	0.0339	0.3341	2.2710
{pasta malz} => {sebze}	0.0067	0.3324	2.5304
{özel ekmek} => {sebze}	0.0201	0.3214	2.4464

Tablo 5.5

5. Segmente Ait Kaldırıcı Yüksek İki Ürünlü İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{çamaşır det.} => {temizlik}	0.0064	0.3742	14.1272
{zeytin} => {peynir}	0.0062	0.5314	9.5357
{kuru meyve} => {çerez}	0.0066	0.4120	6.7027
{gevrek} => {süt}	0.0065	0.4514	6.1553
{yumurta} => {peynir}	0.0080	0.2837	5.0914
{makarna} => {peynir}	0.0064	0.2420	4.3420
{yumurta} => {süt}	0.0084	0.2976	4.0587
{yumurta} => {özel ekmek}	0.0070	0.2460	3.9275
{et} => {sebze}	0.0106	0.4701	3.5791
{özel ekmek} => {peynir}	0.0121	0.1930	3.4633
{yumurta} => {sebze}	0.0127	0.4464	3.3985
{balık} => {sebze}	0.0073	0.4403	3.3516
{baharat} => {sebze}	0.0085	0.4393	3.3443
{konserve} => {sebze}	0.0084	0.4331	3.2973
{tavuk} => {sebze}	0.0111	0.4292	3.2673
{süt} => {yoğurt}	0.0101	0.1373	3.1336
{tuzlu bisk.} => {tatlı bisk.}	0.0225	0.2983	2.9436
{bakliyat} => {sebze}	0.0078	0.3850	2.9312
{kek} => {tatlı bisk.}	0.0125	0.2965	2.9264
{sos} => {sebze}	0.0064	0.3787	2.8832

Tablo 5.6

5. Segmente Ait Güveni Yüksek Üç Ürün İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{meyve, peynir} => {sebze}	0.0073	0.6250	4.7579
{meyve, özel ekmek} => {sebze}	0.0073	0.5811	4.4235
{meyve, yoğurt} => {sebze}	0.0060	0.5573	4.2424
{meyve, süt} => {sebze}	0.0084	0.4745	3.6123
{sebze, yoğurt} => {meyve}	0.0060	0.4440	4.0262
{gazlı içe, meyve} => {sebze}	0.0061	0.4337	3.3018
{çikolata, tuzlu bisk.} => {tatlı bisk.}	0.0094	0.4239	4.1828
{gazlı içe, tatlı bisk.} => {çikolata}	0.0069	0.4221	2.8697
{tatlı bisk., tuzlu bisk.} => {çikolata}	0.0094	0.4175	2.8381
{peynir, sebze} => {meyve}	0.0073	0.3790	3.4370
{gazlı içe, sebze} => {meyve}	0.0061	0.3624	3.2865
{özel ekmek, sebze} => {meyve}	0.0073	0.3603	3.2676
{çikolata, tatlı bisk.} => {tuzlu bisk.}	0.0094	0.2774	3.6787
{çikolata, gazlı içe} => {tatlı bisk.}	0.0069	0.2687	2.6519
{meyve, sebze} => {süt}	0.0084	0.2399	3.2721
{meyve, sebze} => {peynir}	0.0073	0.2093	3.7565
{meyve, sebze} => {özel ekmek}	0.0073	0.2077	3.3160
{çikolata, tatlı bisk.} => {gazlı içe}	0.0069	0.2027	1.8128
{meyve, sebze} => {gazlı içe}	0.0061	0.1739	1.5557
{meyve, sebze} => {yoğurt}	0.0060	0.1723	3.9333

Tablo 5.7

5. Segmente Ait Kaldırıcı Yüksek Üç Ürünlü İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{meyve, peynir} => {sebze}	0.0073	0.6250	4.7579
{meyve, özel ekmek} => {sebze}	0.0073	0.5811	4.4235
{meyve, yoğurt} => {sebze}	0.0060	0.5573	4.2424
{sebze, süt} => {meyve}	0.0084	0.4642	4.2093
{çikolata, tuzlu bisk.} => {tatlı bisk.}	0.0094	0.4239	4.1828
{sebze, yoğurt} => {meyve}	0.0060	0.4440	4.0262
{meyve, sebze} => {yoğurt}	0.0060	0.1723	3.9333
{meyve, sebze} => {peynir}	0.0073	0.2093	3.7565
{çikolata, tatlı bisk.} => {tuzlu bisk.}	0.0094	0.2774	3.6787
{meyve, süt} => {sebze}	0.0084	0.4745	3.6123
{peynir, sebze} => {meyve}	0.0073	0.3790	3.4370
{meyve, sebze} => {özel ekmek}	0.0073	0.2077	3.3160
{gazlı içe, meyve} => {sebze}	0.0061	0.4337	3.3018
{gazlı içe, sebze} => {meyve}	0.0061	0.3624	3.2865
{meyve, sebze} => {süt}	0.0084	0.2399	3.2721
{özel ekmek, sebze} => {meyve}	0.0073	0.3603	3.2676
{gazlı içe, tatlı bisk.} => {çikolata}	0.0069	0.4221	2.8697
{tatlı bisk., tuzlu bisk.} => {çikolata}	0.0094	0.4175	2.8381
{çikolata, gazlı içe} => {tatlı bisk.}	0.0069	0.2687	2.6519
{çikolata, tatlı bisk.} => {gazlı içe}	0.0069	0.2027	1.8128

6. SEGMENTE AİT ANALİZLER

Altıncı kümede yer alan 18 mağazaya ait veriler, toplam 14 bin 669 fişten ve bu fişlere ait her bir fişte bir ürün sadece bir kez yer alacak şekilde, toplam 44 bin 417 işlemten oluşmaktadır.

Tablo 6.1

6. Segmentte Sık Satılan Tek Öğeli Ürün Küme

Ürün	Destek	Fiş sayısı
Sebze	0.2050	3007
Meyve	0.1556	2283
Çikolata	0.1348	1977
Sigara	0.1223	1794
Gazlı içecek	0.1130	1658
Tatlı bisküvi	0.1031	1512
Su	0.0963	1412
Özel ekmek	0.0948	1390
Süt	0.0927	1360
Peynir	0.0780	1144
Tuzlu bisküvi	0.0681	999
Yoğurt	0.0642	942
Çerez	0.0640	939
Ekmek	0.0588	862
Meyve suyu	0.0577	847
Gazete	0.0562	825
Yumurta	0.0501	735
Kahve	0.0460	675
Poğaç	0.0455	668
Tatlı	0.0416	610

Tablo 6.2**6. Segmentte Sık Satılan İki Öğeli Ürün Küme**

Ürün	Destek	Fiş sayısı
Meyve, Sebze	0.0677	993
Özel ekmek, Sebze	0.0366	537
Peynir, Sebze	0.0333	489
Sebze, Süt	0.0322	472
Çikolata, Tatlı bisküvi	0.0303	445
Sebze, Yoğurt	0.0266	390
Meyve, Özel ekmek	0.0260	381
Meyve, Süt	0.0258	378
Gazlı içe, Sebze	0.0256	376
Sebze, Yumurta	0.0244	358
Meyve, Peynir	0.0228	335
Çikolata, Sebze	0.0224	328
Çikolata, Meyve	0.0223	327
Çikolata, Gazlı içecek	0.0217	319
Tatlı bisküvi, Tuzlu bisküvi	0.0211	310
Gazlı içecek, Meyve	0.0202	296
Özel ekmek, Peynir	0.0191	280
Meyve, Tatlı bisküvi	0.0190	279
Ekmek, Sebze	0.0187	274
Et, Sebze	0.0185	272

Tablo 6.3**6. Segmentte Sık Satılan Üç Öğeli Ürün Küme**

Ürün	Destek	Fiş sayısı
Meyve, Özel ekmek, Sebze	0.0172	252
Meyve, Peynir, Sebze	0.0164	240
Meyve, Sebze, Süt	0.0151	221
Meyve, Sebze, Yumurta	0.0122	179
Meyve, Sebze, Yoğurt	0.0118	173
Özel ekmek, Peynir, Sebze	0.0115	169
Gazlı içe, Meyve, Sebze	0.0108	158
Çikolata, Meyve, Sebze	0.0104	152
Özel ekmek, Sebze, Süt	0.0097	143
Peynir, Sebze, Yumurta	0.0096	141
Peynir, Sebze, Süt	0.0096	141
Çerez, Meyve, Sebze	0.0093	136
Sebze, Süt, Yumurta	0.0089	130
Meyve, Sebze, Tatlı bisküvi	0.0087	128
Sebze, Süt, Yoğurt	0.0085	125
Özel ekmek, Sebze, Yoğurt	0.0083	122
Peynir, Sebze, Yoğurt	0.0082	121
Meyve, Özel ekmek, Peynir	0.0082	121
Özel ekmek, Sebze, Yumurta	0.0078	115
Meyve, Özel ekmek, Süt	0.0078	114

Tablo 6.4

6. Kümedeki Mağazalara Ait Güveni Yüksek İki Ürünlü İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{zeytin} => {peynir}	0.0110	0.5625	7.2127
{zeytin} => {sebze}	0.0106	0.5382	2.6255
{et} => {sebze}	0.0185	0.5354	2.6120
{baharat} => {sebze}	0.0151	0.5237	2.5547
{balık} => {sebze}	0.0121	0.5130	2.5024
{yumurta} => {sebze}	0.0244	0.4871	2.3761
{tere yağ} => {sebze}	0.0078	0.4831	2.3565
{konserve} => {sebze}	0.0128	0.4724	2.3043
{bakliyat} => {sebze}	0.0134	0.4635	2.2612
{tavuk} => {sebze}	0.0163	0.4434	2.1631
{meyve} => {sebze}	0.0677	0.4350	2.1218
{sos} => {sebze}	0.0106	0.4342	2.1180
{kaşar} => {sebze}	0.0112	0.4293	2.0943
{peynir} => {sebze}	0.0333	0.4274	2.0852
{yoğurt} => {sebze}	0.0266	0.4140	2.0197
{makarna} => {sebze}	0.0161	0.4111	2.0057
{gevrek} => {süt}	0.0074	0.4082	4.4033
{salam} => {peynir}	0.0077	0.4036	5.1748
{pasta malzemesi} => {sebze}	0.0089	0.3994	1.9483
{özel ekmek} => {sebze}	0.0366	0.3863	1.8846

Tablo 6.5

6. Segmente Ait Kaldırıcı Yüksek İki Ürünlü İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{zeytin} => {peynir}	0.0110	0.5625	7.2127
{salam} => {peynir}	0.0077	0.4036	5.1748
{gevrek} => {süt}	0.0074	0.4082	4.4033
{kaşar} => {peynir}	0.0081	0.3115	3.9945
{peynir} => {yumurta}	0.0145	0.1862	3.7159
{yumurta} => {yoğurt}	0.0116	0.2313	3.6017
{yumurta} => {süt}	0.0150	0.2993	3.2285
{kaşar} => {özel ekmek}	0.0077	0.2958	3.1218
{makarna} => {peynir}	0.0093	0.2369	3.0381
{tuzlu bisk.} => {tatlı bisk.}	0.0211	0.3103	3.0105
{tatlı bisk.} => {kek}	0.0119	0.1157	2.8439
{kek} => {tuzlu bisk.}	0.0078	0.1910	2.8039
{yumurta} => {özel ekmek}	0.0133	0.2653	2.7998
{et} => {peynir}	0.0073	0.2106	2.7008
{zeytin} => {sebze}	0.0106	0.5382	2.6255
{şekerleme} => {çikolata}	0.0082	0.3538	2.6251
{et} => {sebze}	0.0185	0.5354	2.6120
{kek} => {çikolata}	0.0143	0.3518	2.6100
{yoğurt} => {süt}	0.0154	0.2399	2.5877
{peynir} => {özel ekmek}	0.0191	0.2448	2.5830

Tablo 6.6

6. Segmente Ait Güveni Yüksek Üç Ürünlü İlişki Kuralları*

Kural	Destek	Güven	Kaldıraç
{meyve, tavuk} => {sebze}	0.0072	0.7681	3.7471
{meyve, yumurta} => {sebze}	0.0122	0.7553	3.6844
{meyve, peynir} => {sebze}	0.0164	0.7164	3.4949
{makarna, meyve} => {sebze}	0.0068	0.7042	3.4354
{meyve, yoğurt} => {sebze}	0.0118	0.6705	3.2711
{peynir, yumurta} => {sebze}	0.0096	0.6620	3.2293
{meyve, özel ekmek} => {sebze}	0.0172	0.6614	3.2266
{yoğurt, yumurta} => {sebze}	0.0074	0.6412	3.1278
{peynir, yoğurt} => {sebze}	0.0082	0.6402	3.1231
{sebze, zeytin} => {peynir}	0.0067	0.6387	8.1899
{özel ekmek, peynir} => {sebze}	0.0115	0.6036	2.9444
{peynir, süt} => {sebze}	0.0096	0.5924	2.8901
{süt, yumurta} => {sebze}	0.0089	0.5909	2.8826
{özel ekmek, yumurta} => {sebze}	0.0078	0.5897	2.8769
{meyve, süt} => {sebze}	0.0151	0.5847	2.8521
{özel ekmek, yoğurt} => {sebze}	0.0083	0.5837	2.8476
{özel ekmek, süt} => {sebze}	0.0097	0.5564	2.7144
{süt, yoğurt} => {sebze}	0.0085	0.5531	2.6982
{çerez, sebze} => {meyve}	0.0093	0.5440	3.4954
{gazlı içe, meyve} => {sebze}	0.0108	0.5338	2.6039

Tablo 6.7

6. Segmente Ait Kaldırıcı Yüksek Üç Ürün İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{peynir, sebze} => {yumurta}	0.0096	0.2883	5.7547
{sebze, yoğurt} => {yumurta}	0.0074	0.2795	5.5780
{sebze, süt} => {yumurta}	0.0089	0.2754	5.4969
{sebze, yumurta} => {peynir}	0.0096	0.3939	5.0502
{sebze, yumurta} => {yoğurt}	0.0074	0.3045	4.7413
{özel ekmek, sebze} => {yumurta}	0.0078	0.2142	4.2740
{sebze, süt} => {yoğurt}	0.0085	0.2648	4.1240
{çikolata, tuzlu bisk.} => {tatlı bisk.}	0.0075	0.4231	4.1046
{meyve, özel ekmek} => {peynir}	0.0082	0.3176	4.0723
{özel ekmek, sebze} => {peynir}	0.0115	0.3147	4.0354
{sebze, yoğurt} => {peynir}	0.0082	0.3103	3.9783
{sebze, yumurta} => {süt}	0.0089	0.3631	3.9167
{peynir, sebze} => {yoğurt}	0.0082	0.2474	3.8532
{sebze, süt} => {peynir}	0.0096	0.2987	3.8305
{meyve, peynir} => {özel ekmek}	0.0082	0.3612	3.8118
{meyve, tavuk} => {sebze}	0.0072	0.7681	3.7471
{meyve, süt} => {peynir}	0.0075	0.2910	3.7314
{meyve, yumurta} => {sebze}	0.0122	0.7553	3.6844
{peynir, sebze} => {özel ekmek}	0.0115	0.3456	3.6472
{çikolata, tatlı bisk} => {tuzlu bisk}	0.0075	0.2472	3.6297

EK 2: İKİNCİ SEGMENTASYON YAKLAŞIMI İÇİN ANALİZLER

Mağazalara ait satış verilerinden elde edilmiş olan ilişki kuralları ile yapılmış olan segmentasyon sonucu 5 segment elde edilmişti. Bu segmentler için yapılan sık ürün küme ve birliktelik kuralları analizleri ikinci ekte verilmiştir.

1. Segmente Ait Analizler

Birinci segmentte yer alan 57 mağazaya ait veriler, toplam 39 bin 142 fişten ve bu fişlere ait her bir fişte bir ürün sadece bir kez yer alacak şekilde, toplam 104 bin 867 işlemden oluşmaktadır. Analiz öncesi bir fişte birden fazla aynı ürün yer alıyorsa; örneğin müşteri 3 adet makarna almışsa, ilgili aynı ürünün bir fişte bir kere yer alması sağlanmıştır. Her bir satış fişte ortalama 3 farklı ürün olduğu düşünülebilir. Tablodan görüldüğü gibi, sebze ürünü en çok satılan üründür ve 39 bin 142 fişin 7 bin 389 adedinde yer almıştır.

Tablo 1.1

1. Segmentte Sık Satılan Tek Öğeli Ürün Küme

Ürün	Destek	Fiş sayısı
Sebze	0.1888	7389
Meyve	0.1299	5084
Sigara	0.1241	4856
Çikolata	0.1227	4803
Gazlı içecek	0.0998	3905
Tatlı bisküvi	0.0922	3607
Özel ekmek	0.0788	3085
Su	0.0771	3016
Ekmek	0.0759	2972
Süt	0.0727	2845
Peynir	0.0550	2152
Tuzlu bisküvi	0.0545	2134
Çerez	0.0541	2118
Gazete	0.0523	2049
Meyve suyu	0.0479	1874
Yoğurt	0.0477	1866
Tavuk	0.0420	1643
Ev gereçleri	0.0419	1640
Kek	0.0412	1611
Kahve	0.0396	1550

* En çok satılan ürünler analizi yapılırken, sadece tek ürün elde edilecek şekilde kısıtlamaya gidilmiştir.

Tablo 2’de ise en çok birlikte satılan “ikili” ürünlerin tespiti amacıyla, analiz 2 ürün elde etmek kısıtlanmış ve yine sık 20 ürün kümesi elde edilmiştir.

Tablo 1.2
1. Segmentte Sık Satılan İki Öğeli Ürün Küme*

Ürün	Destek	Fiş sayısı
Meyve, Sebze	0.05133	2009
Çikolata, Tatlı bisküvi	0.03020	1182
Özel ekmek, Sebze	0.02192	858
Çikolata, Gazlı içecek	0.02041	799
Sebze, Süt	0.02011	787
Peynir, Sebze	0.01983	776
Tatlı bisküvi, Tuzlu bisküvi	0.01957	766
Ekmek, Sebze	0.01926	754
Çikolata, Sebze	0.01873	733
Gazlı içecek, Sebze	0.01747	684
Çikolata, Meyve	0.01704	667
Sebze, Yoğurt	0.01592	623
Sebze, Yumurta	0.01587	621
Meyve, Süt	0.01579	618
Sebze, Tatlı bisküvi	0.01548	606
Sebze, Tavuk	0.01543	604
Çikolata, Tuzlu bisküvi	0.01512	592
Meyve, Tatlı bisküvi	0.01428	559
Meyve, Özel ekmek	0.01397	547
Gazlı içe, Tatlı bisküvi	0.01392	545

* Analiz yapılırken iki ürün elde edilecek şekilde kısıtlama yapılmıştır.

Tablo 3’te de en çok birlikte satılan “üçlü” ürünlerin tespiti amacıyla, analiz 3 ürün elde etmek için ve en çok satılan ilk 20 ürünün tespiti için yapılmıştır.

Tablo 1.3

1. Segmentte Sık Satılan Üç Öğeli Ürün Küme*

Ürün	Destek	Fiş sayısı
Meyve, Sebze, Süt	0.0087	339
Çikolata, Meyve, Sebze	0.0077	301
Meyve, Peynir, Sebze	0.0075	292
Meyve, Özel ekmek, Sebze	0.0075	292
Çikolata, Tatlı bisküvi, Tuzlu bisküvi	0.0074	291
Meyve, Sebze, Tatlı bisküvi	0.0069	270
Çikolata, Gazlı içecek, Tatlı bisküvi	0.0062	241
Çikolata, Kek, Tatlı bisküvi	0.0059	230
Meyve, Sebze, Yumurta	0.0059	230
Meyve, Sebze, Yoğurt	0.0058	227
Gazlı içecek, Meyve, Sebze	0.0057	224
Çikolata, Sebze, Tatlı bisküvi	0.0054	211
Çikolata, Meyve, Tatlı bisküvi	0.0052	203
Ekmek, Meyve, Sebze	0.0049	191
Peynir, Sebze, Yumurta	0.0048	189
Çerez, Meyve, Sebze	0.0048	186
Özel ekmek, Peynir, Sebze	0.0047	185
Özel ekmek, Sebze, Süt	0.0046	182
Çerez, Çikolata, Tatlı bisküvi	0.0045	177
Meyve, Sebze, Tavuk	0.0045	176

* Analiz yapılırken üç ürün elde edilecek şekilde kısıtlama yapılmıştır.

Aşağıdaki tablo ve bundan sonraki tablolarda ise birliktelik kuralları analizleri mevcuttur. İlk etapta sadece iki ürünlü kurallar elde edilmiştir. İki ürünlü kural analizinde güven değeri yüksek tutulmuştur ancak en yüksek güven değeri % 51 olarak bulunmuştur.

Genel olarak destek değeri 100 fişe karşılık gelen destek değerinin (0,0025) üzerinde kurallar üretilmeye çalışılmıştır.

Tablo 1.4

1. Segmente Ait Güveni Yüksek İki Ürünlü İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{zeytin} => {peynir}	0.0075	0.5139	9.3478
{zeytin} => {sebze}	0.0066	0.4477	2.3718
{baharat} => {sebze}	0.0080	0.4276	2.2651
{margarin} => {sebze}	0.0045	0.4224	2.2378
{et} => {sebze}	0.0132	0.4179	2.2140
{konserve} => {sebze}	0.0086	0.4176	2.2121
{salam} => {peynir}	0.0084	0.4098	7.4532
{yumurta} => {sebze}	0.0159	0.4032	2.1361
{tere yağ} => {sebze}	0.0037	0.3989	2.1131
{meyve} => {sebze}	0.0513	0.3952	2.0933
{bakliyat} => {sebze}	0.0103	0.3783	2.0039
{gevrek} => {süt}	0.0045	0.3779	5.1995
{kaşar} => {sebze}	0.0066	0.3709	1.9648
{tavuk} => {sebze}	0.0154	0.3676	1.9474
{balık} => {sebze}	0.0046	0.3666	1.9420
{peynir} => {sebze}	0.0198	0.3606	1.9102
{tuzlu bisküvi} => {tatlı bisküvi}	0.0196	0.3590	3.8952
{un} => {sebze}	0.0030	0.3587	1.9000
{sucuk} => {sebze}	0.0040	0.3405	1.8038
{kaşar} => {peynir}	0.0061	0.3395	6.1753

* Analiz yapılırken sadece iki ürünlü kurallar ürün elde edilecek şekilde kısıtlama yapılmıştır.

Tablo 1.5

1. Segmente Ait Kaldırıcı Yüksek İki Ürünlü İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{bulaşık det} => {çamaşır det}	0.0039	0.2429	11.7939
{kaşar} => {salam}	0.0036	0.1997	9.7960
{zeytin} => {peynir}	0.0075	0.5139	9.3478
{çamaşır det} => {temizlik}	0.0061	0.2953	9.2170
{bulaşık det} => {temizlik}	0.0047	0.2937	9.1659
{çay} => {şeker}	0.0049	0.1439	7.7382
{salam} => {peynir}	0.0084	0.4098	7.4532
{sabun} => {temizlik}	0.0034	0.2332	7.2795
{kaşar} => {peynir}	0.0061	0.3395	6.1753
{konserve} => {bakliyat}	0.0034	0.1648	6.0402
{bakliyat} => {makarna}	0.0058	0.2125	5.9595
{sos} => {makarna}	0.0039	0.2089	5.8559
{zeytin} => {yumurta}	0.0032	0.2213	5.6236
{sucuk} => {peynir}	0.0036	0.3017	5.4880
{konserve} => {makarna}	0.0040	0.1945	5.4549
{gevrek} => {süt}	0.0045	0.3779	5.1995
{tuvhav} => {temizlik}	0.0031	0.1603	5.0040
{kaşar} => {yumurta}	0.0034	0.1897	4.8223
{reçel vb} => {peynir}	0.0053	0.2512	4.5691
{peynir} => {yumurta}	0.0095	0.1729	4.3936

Aşağıdaki tabloda ve bir sonraki tabloda bu defa, üç ürün içeren güven değeri ve kaldırıcı değeri yüksek kurallar sunulmuştur.

Tablo 1.6

1. Segmente Ait Güveni Yüksek Üç Ürünlü İlişki Kuralları*

Kural	Destek	Güven	Kaldıraç
{et, meyve} => {sebze}	0.0040	0.6858	3.6331
{baharat, meyve} => {sebze}	0.0028	0.6708	3.5535
{bakliyat, meyve} => {sebze}	0.0036	0.6651	3.5231
{konserve, meyve} => {sebze}	0.0027	0.6646	3.5206
{meyve, peynir} => {sebze}	0.0075	0.6404	3.3922
{meyve, yumurta} => {sebze}	0.0059	0.6301	3.3380
{meyve, tavuk} => {sebze}	0.0045	0.5906	3.1286
{çay, meyve} => {sebze}	0.0032	0.5822	3.0839
{meyve, temizlik} => {sebze}	0.0032	0.5780	3.0618
{meyve, yoğurt} => {sebze}	0.0058	0.5704	3.0213
{sebze, zeytin} => {peynir}	0.0037	0.5642	10.2621
{şekerleme, tatlı bisküvi} => {çikolata}	0.0030	0.5640	4.5962
{meyve, pasta malzemesi} => {sebze}	0.0028	0.5635	2.9848
{makarna, meyve} => {sebze}	0.0037	0.5489	2.9076
{meyve, süt} => {sebze}	0.0087	0.5485	2.9058
{ekmek, meyve} => {sebze}	0.0049	0.5426	2.8744
{meyve, özel ekmek} => {sebze}	0.0075	0.5338	2.8278
{özel ekmek, tavuk} => {sebze}	0.0027	0.5300	2.8076
{yoğurt, yumurta} => {sebze}	0.0034	0.5194	2.7513
{kek, tuzlu bisküvi} => {tatlı bisküvi}	0.0039	0.5151	5.5892

* Sınırlama sadece üç ürünlü ilişkiler elde etmek için yapılmıştır. Destek değeri en az 100 fişin üstündedir.

Tablo 1.7

2. Segmente Ait Kaldırıcı Yüksek Üç Ürünlü İlişki Kuralları*

Kural	Destek	Güven	Kaldıraç
{peynir sebze} => {zeytin}	0.0037	0.1869	12.7420
{sebze zeytin} => {peynir}	0.0037	0.5642	10.2621
{salam sebze} => {peynir}	0.0033	0.4885	8.8861
{peynir sebze} => {salam}	0.0033	0.1649	8.0907
{peynir sebze} => {yumurta}	0.0048	0.2436	6.1905
{kek tatlı bisk} => {tuzlu bisk}	0.0039	0.3105	5.6949
{sebze süt} => {yumurta}	0.0045	0.2236	5.6841
{kek tuzlu bisk} => {tatlı bisk}	0.0039	0.5151	5.5892
{sebze yumurta} => {peynir}	0.0048	0.3043	5.5357
{sebze yoğurt} => {yumurta}	0.0034	0.2151	5.4669
{çikolata tuzlu bisk} => {tatlı bisk}	0.0074	0.4916	5.3342
{süt tatlı bisk} => {tuzlu bisk}	0.0031	0.2752	5.0472
{süt tuzlu bisk} => {tatlı bisk}	0.0031	0.4572	4.9619
{tatlı bisk tuzlu bisk} => {kek}	0.0039	0.2010	4.8847
{meyve süt} => {yumurta}	0.0030	0.1909	4.8531
{çikolata kek} => {tatlı bisk}	0.0059	0.4415	4.7906
{çikolata tatlı bisk} => {kek}	0.0059	0.1946	4.7278
{sebze süt} => {yoğurt}	0.0045	0.2236	4.6910
{gazlı içe tuzlu bisk} => {tatlı bisk}	0.0033	0.4309	4.6762
{şekerleme tatlı bisk} => {çikolata}	0.0030	0.5640	4.5962

* Sınırlama sadece üç ürünlü ilişkiler elde etmek için yapılmıştır.

2. Segmente Ait Analizler

13 mağazaya ait 1 günlük satış verisi toplam 24 bin 715 fişten ve bu fişlere ait toplam 98 bin 943 işlemden oluşmaktadır.

Tablodan görüldüğü gibi, sebze ürünü en çok satılan üründür ve 24 bin 715 fişin 4 bin 854 adedinde yer almıştır.

Tablo 2.1

2. Segmentte Sık Satılan Tek Ögeli Ürün Küme

Ürün	Destek	Fiş sayısı
Sebze	0.1964	4854
Meyve	0.1706	4216
Özel ekmek	0.1440	3559
Çikolata	0.1300	3213
Gazlı içecek	0.1118	2763
Peynir	0.1076	2659
Sigara	0.1037	2563
Süt	0.0957	2365
Tatlı bisküvi	0.0882	2180
Poğaç	0.0868	2145
Çerez	0.0806	1992
Su	0.0804	1987
Tatlı	0.0800	1977
Ev gereçleri	0.0781	1930
Hazır yemek	0.0689	1703
Yoğurt	0.0680	1681
Et	0.0661	1634
Tavuk	0.0637	1574
Yumurta	0.0636	1572
Tuzlu bisküvi	0.0628	1570

Tablo 2.2**2. Segmentte Sık Satılan İki Öğeli Ürün Küme**

Ürün	Destek	Fiş sayısı
Meyve, Sebze	0.0872	2155
Özel ekmek, Sebze	0.0557	1377
Peynir, Sebze	0.0525	1298
Meyve, Özel ekmek	0.0466	1152
Sebze, Süt	0.0419	1036
Meyve, Peynir	0.0407	1006
Özel ekmek, Peynir	0.0376	929
Çikolata, Tatlı bisküvi	0.0366	905
Meyve, Süt	0.0356	880
Et, Sebze	0.0344	850
Sebze, Yumurta	0.0344	850
Sebze, Yoğurt	0.0336	830
Gazlı içecek, Sebze	0.0329	813
Çikolata, Meyve	0.0324	801
Çikolata, Sebze	0.0323	798
Çerez, Meyve	0.0306	756
Özel ekmek, Süt	0.0305	754
Çerez, Sebze	0.0298	737
Gazlı içecek, Meyve	0.0294	727
Sebze, Tavuk	0.0291	719

Tablo 2.3**2. Segmentte Sık Satılan Üç Öğeli Ürün Küme**

Ürün	Destek	Fiş sayısı
Meyve, Peynir, Sebze	0.0298	737
Meyve, Özel ekmek, Sebze	0.0293	724
Meyve, Sebze, Süt	0.0236	583
Özel ekmek, Peynir, Sebze	0.0208	514
Meyve, Sebze, Yumurta	0.0195	482
Meyve, Sebze, Yoğurt	0.0185	457
Çerez, Meyve, Sebze	0.0185	457
Et, Meyve, Sebze	0.0181	447
Çikolata, Meyve, Sebze	0.0178	440
Peynir, Sebze, Süt	0.0175	433
Gazlı içecek, Meyve, Sebze	0.0174	430
Peynir, Sebze, Yumurta	0.0172	425
Meyve, Özel ekmek, Peynir	0.0169	418
Özel ekmek, Sebze, Süt	0.0168	415
Meyve, Sebze, Tavuk	0.0148	366
Sebze, Süt, Yumurta	0.0147	363
Meyve, Peynir, Süt	0.0146	361
Meyve, Özel ekmek, Süt	0.0144	356
Peynir, Sebze, Yoğurt	0.0141	348
Makarna, Meyve, Sebze	0.0141	348

Tablo 2.4

2. Segmente Ait Güveni Yüksek İki Ürünlü İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{zeytin} => {peynir}	0.0187	0.6106	5.6733
{turşu} => {sebze}	0.0085	0.5633	2.8689
{konserve} => {sebze}	0.0204	0.5430	2.7654
{yumurta} => {sebze}	0.0344	0.5410	2.7552
{baharat} => {sebze}	0.0212	0.5380	2.7398
{zeytin} => {sebze}	0.0163	0.5351	2.7251
{bakliyat} => {sebze}	0.0239	0.5320	2.7094
{et} => {sebze}	0.0344	0.5208	2.6523
{un} => {sebze}	0.0087	0.5181	2.6384
{meyve} => {sebze}	0.0872	0.5110	2.6025
{hindi} => {sebze}	0.0045	0.5091	2.5927
{makarna} => {sebze}	0.0255	0.5020	2.5565
{tere yağ} => {sebze}	0.0126	0.5016	2.5546
{balık} => {sebze}	0.0225	0.4991	2.5418
{zeytinyağı} => {sebze}	0.0078	0.4974	2.5332
{kaşar} => {sebze}	0.0192	0.4948	2.5198
{tuz} => {sebze}	0.0069	0.4942	2.5168
{yoğurt} => {sebze}	0.0336	0.4940	2.5160
{bal} => {peynir}	0.0041	0.4928	4.5783
{peynir} => {sebze}	0.0525	0.4880	2.4851

Tablo 2.5

2. Segmente Ait Kaldırıcı Yüksek İki Ürünlü İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{mama} => {bebek ürünü}	0.0043	0.3023	13.0601
{un} => {şeker}	0.0051	0.3036	11.0839
{şeker} => {yağ}	0.0054	0.1965	10.3748
{un} => {pasta malz}	0.0063	0.3735	9.8201
{çamaşır det} => {bulaşık det}	0.0124	0.2923	8.4582
{tuz} => {konserve}	0.0044	0.3169	8.4207
{sabun} => {diş}	0.0068	0.2784	8.2409
{parfüm} => {diş}	0.0049	0.2713	8.0302
{yağ} => {konserve}	0.0057	0.2991	7.9499
{yağ} => {bakliyat}	0.0067	0.3526	7.8572
{un} => {bakliyat}	0.0059	0.3494	7.7866
{tuvhav} => {peçete}	0.0080	0.2052	7.6028
{şeker} => {çay}	0.0106	0.3885	7.5010
{tuz} => {baharat}	0.0040	0.2907	7.3764
{sos} => {tuz}	0.0040	0.1025	7.3613
{tuz} => {bakliyat}	0.0045	0.3256	7.2559
{temizlik malz} => {poşet}	0.0052	0.2122	7.2328
{sabun} => {şampuan}	0.0050	0.2026	7.2060
{bakliyat} => {makarna}	0.0162	0.3607	7.0918
{temizlik} => {çamaşır det}	0.0182	0.2991	7.0612

Tablo 2.6

2. Segmente Ait Güveni Yüksek Üç Ürünli İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{meyve, yumurta} => {sebze}	0.0195	0.7454	3.7960
{meyve, peynir} => {sebze}	0.0298	0.7316	3.7259
{et, meyve} => {sebze}	0.0181	0.7256	3.6955
{meyve, yoğurt} => {sebze}	0.0185	0.6898	3.5128
{meyve, süt} => {sebze}	0.0236	0.6636	3.3797
{peynir, yumurta} => {sebze}	0.0172	0.6553	3.3374
{meyve, özel ekmek} => {sebze}	0.0293	0.6299	3.2078
{çerez, sebze} => {meyve}	0.0185	0.6209	3.6391
{peynir, süt} => {sebze}	0.0175	0.6081	3.0971
{çerez, meyve} => {sebze}	0.0185	0.6037	3.0745
{gazlı içecek, meyve} => {sebze}	0.0174	0.5928	3.0192
{sebze, yumurta} => {meyve}	0.0195	0.5676	3.3264
{peynir, sebze} => {meyve}	0.0298	0.5670	3.3232
{sebze, süt} => {meyve}	0.0236	0.5643	3.3070
{özel ekmek, peynir} => {sebze}	0.0208	0.5544	2.8232
{çikolata, sebze} => {meyve}	0.0178	0.5526	3.2389
{özel ekmek, süt} => {sebze}	0.0168	0.5525	2.8135
{sebze, yoğurt} => {meyve}	0.0185	0.5518	3.2340
{çikolata, meyve} => {sebze}	0.0178	0.5513	2.8074
{gazlı içecek, sebze} => {meyve}	0.0174	0.5308	3.1108

Tablo 2.7

2. Kümedeki Mağazalara Ait Kaldırıcı Yüksek Üç Ürün İlişki Kuralları*

Kural	Destek	Güven	Kaldıraç
{çay, temizlik} => {şeker}	0.0041	0.3769	13.7581
{poşet, temizlik} => {bulaşık det}	0.0044	0.4467	12.9282
{çamaşır det, makarna} => {bulaşık det}	0.0040	0.4444	12.8624
{bulaşık det, makarna} => {çamaşır det}	0.0040	0.5405	12.7598
{makarna, şeker} => {bakliyat}	0.0044	0.5625	12.5358
{çamaşır det, tuvhav} => {bulaşık det}	0.0041	0.4298	12.4382
{bulaşık det, tuvhav} => {çamaşır det}	0.0041	0.5233	12.3532
{temizlik, tuvhav} => {peçete}	0.0041	0.3333	12.3513
{çay, peynir} => {şeker}	0.0048	0.3362	12.2720
{çay, süt} => {şeker}	0.0044	0.3323	12.1318
{baharat, makarna} => {konserve}	0.0046	0.4520	12.0120
{çamaşır det, temizlik malz} => {temizlik}	0.0045	0.7273	11.9751
{temizlik, tuvhav} => {poşet}	0.0043	0.3497	11.9202
{temizlik, tuvhav} => {bulaşık det}	0.0051	0.4118	11.9166
{bulaşık det, temizlik} => {çamaşır det}	0.0072	0.5028	11.8695
{bulaşık det, peynir} => {çamaşır det}	0.0051	0.5020	11.8494
{çay, makarna} => {bakliyat}	0.0052	0.5311	11.8365
{çamaşır det, makarna} => {bakliyat}	0.0048	0.5289	11.7867
{makarna, temizlik} => {çamaşır det}	0.0050	0.4940	11.6606
{makarna, temizlik} => {bakliyat}	0.0052	0.5181	11.5457

3. SEGMENTE AİT ANALİZLER

64 mağazaya ait 1 günlük satış verisi toplam 39 bin 516 fişten ve bu fişlere ait toplam 88 bin 328 işlemden oluşmaktadır.

Tablo 3.1

3. Segmentte Sık Satılan Tek Öğeli Ürün Küme

Ürün	Destek	Fiş sayısı
Sigara	0.1473	5822
Çikolata	0.1369	5409
Sebze	0.1150	4543
Gazlı içe	0.1014	4007
Tatlı bisküvi	0.0971	3836
Meyve	0.0964	3811
Su	0.0895	3535
Tuzlu bisküvi	0.0678	2681
Süt	0.0615	2430
Çerez	0.0533	2108
Özel ekmek	0.0525	2074
Gazete	0.0507	2003
Meyve suyu	0.0480	1897
Ekmek	0.0447	1768
Poğaç	0.0423	1672
Peynir	0.0402	1590
Kahve	0.0398	1574
Kek	0.0396	1566
Yoğurt	0.0388	1532
Tatlı	0.0322	1274

Tablo 3.2**3. Segmentte Sık Satılan İki Öğeli Ürün Küme**

Ürün	Destek	Fiş sayısı
Çikolata, Tatlı bisküvi	0.0289	1142
Meyve, Sebze	0.0250	988
Çikolata, Gazlı içecek	0.0194	767
Tatlı bisküvi, Tuzlu bisküvi	0.0188	743
Çikolata, Tuzlu bisküvi	0.0175	692
Çikolata, Sigara	0.0135	533
Gazlı içe, Tatlı bisküvi	0.0135	533
Çikolata, Kek	0.0129	510
Gazlı içe, Sigara	0.0125	494
Çikolata, Su	0.0121	478
Özel ekmek, Sebze	0.0116	458
Çikolata, Meyve	0.0114	450
Çerez, Çikolata	0.0113	447
Meyve, Süt	0.0109	431
Sebze, Süt	0.0106	419
Kek, Tatlı bisküvi	0.0106	419
Çikolata, Süt	0.0103	407
Peynir, Sebze	0.0102	403
Gazlı içe, Sebze	0.0100	395
Ekmek, Sebze	0.0100	395

Tablo 3.3**3. Segmentte Sık Satılan Üç Ögeli Ürün Küme**

Ürün	Destek	Fiş sayısı
Çikolata, Tatlı bisküvi, Tuzlu bisküvi	0.0070	277
Çikolata, Gazlı içecek, Tatlı bisküvi	0.0048	190
Çikolata, Kek, Tatlı bisküvi	0.0046	182
Meyve, Sebze, Süt	0.0036	142
Meyve, Özel ekmek, Sebze	0.0033	130
Çikolata, Meyve, Sebze	0.0033	130
Meyve, Peynir, Sebze	0.0031	122
Çerez, Çikolata, Tatlı bisküvi	0.0031	122
Çikolata, Meyve, Tatlı bisküvi	0.0031	122
Kek, Tatlı bisküvi, Tuzlu bisküvi	0.0029	115
Çikolata, Sigara, Tatlı bisküvi	0.0029	115
Meyve, Sebze, Tatlı bisküvi	0.0029	115
Gazlı içecek, Meyve, Sebze	0.0028	111
Cips, Çikolata, Gazlı içecek	0.0027	107
Çikolata, Kek, Tuzlu bisküvi	0.0027	107
Gazlı içe, Tatlı bisküvi, Tuzlu bisküvi	0.0027	107
Meyve, Tatlı bisküvi, Tuzlu bisküvi	0.0027	107
Meyve, Sebze, Yumurta	0.0026	103
Çikolata, Gazlı içecek, Sigara	0.0026	103
Meyve, Sebze, Yoğurt	0.0026	103

Tablo 3.4

3. Segmente Ait Güveni Yüksek İki Ürünlü İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{zeytin} => {peynir}	0.0040	0.5113	12.7079
{gevrek} => {süt}	0.0047	0.3725	6.0577
{kuru meyve} => {çerez}	0.0041	0.3575	6.7008
{et} => {sebze}	0.0038	0.3470	3.0186
{konserve} => {sebze}	0.0046	0.3454	3.0045
{kek} => {çikolata}	0.0129	0.3250	2.3746
{yumurta} => {sebze}	0.0075	0.3239	2.8175
{tavuk} => {sebze}	0.0045	0.3207	2.7897
{şekerleme} => {çikolata}	0.0078	0.3160	2.3082
{salam} => {peynir}	0.0040	0.3107	7.7213
{tatlı bisküvi} => {çikolata}	0.0289	0.2977	2.1749
{cips} => {gazlı içe}	0.0086	0.2870	2.8307
{bakliyat} => {sebze}	0.0044	0.2848	2.4771
{baharat} => {sebze}	0.0038	0.2838	2.4686
{tuzlu bisküvi} => {tatlı bisküvi}	0.0188	0.2775	2.8587
{cips} => {çikolata}	0.0082	0.2736	1.9986
{kaşar} => {peynir}	0.0031	0.2711	6.7379
{kek} => {tatlı bisküvi}	0.0106	0.2663	2.7431
{makarna} => {sebze}	0.0054	0.2612	2.2723
{meyve} => {sebze}	0.0250	0.2592	2.2550

Tablo 3.5

3. Segmente Ait Kaldırıcı Yüksek İki Ürünlü İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{zeytin} => {peynir}	0.0040	0.5113	12.7079
{peynir} => {salam}	0.0040	0.1006	7.7213
{kaşar} => {peynir}	0.0031	0.2711	6.7379
{kuru meyve} => {çerez}	0.0041	0.3575	6.7008
{gevrek} => {süt}	0.0047	0.3725	6.0577
{yumurta} => {peynir}	0.0047	0.2000	4.9706
{yumurta} => {yoğurt}	0.0036	0.1554	4.0092
{yumurta} => {özel ekmek}	0.0045	0.1935	3.6863
{çay} => {kahve}	0.0035	0.1417	3.5568
{peynir} => {özel ekmek}	0.0073	0.1811	3.4511
{hazır yemek} => {meyve suyu}	0.0043	0.1591	3.3135
{yumurta} => {süt}	0.0047	0.2011	3.2700
{et} => {sebze}	0.0038	0.3470	3.0186
{konserve} => {sebze}	0.0046	0.3454	3.0045
{cips} => {çerez}	0.0048	0.1599	2.9981
{tatlı bisk} => {tuzlu bisk}	0.0188	0.1940	2.8587
{cips} => {gazlı içe}	0.0086	0.2870	2.8307
{yumurta} => {sebze}	0.0075	0.3239	2.8175
{tavuk} => {sebze}	0.0045	0.3207	2.7897
{tatlı bisk} => {kek}	0.0106	0.1087	2.7431

Tablo 3.6

3. Segmente Ait Güveni Yüksek Üç Ürünlü İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{meyve, peynir} => {sebze}	0.0031	0.4788	4.1644
{meyve, özel ekmek} => {sebze}	0.0033	0.4444	3.8659
{kek, tatlı bisk} => {çikolata}	0.0046	0.4365	3.1885
{sebze, tatlı bisk} => {meyve}	0.0029	0.4145	4.2984
{kek, tuzlu bisk} => {tatlı bisk}	0.0029	0.4070	4.1928
{çikolata, tuzlu bisk} => {tatlı bisk}	0.0070	0.3994	4.1146
{kek, tuzlu bisk} => {çikolata}	0.0027	0.3789	2.7684
{çerez, tatlı bisk} => {çikolata}	0.0031	0.3713	2.7123
{tatlı bisk, tuzlu bisk} => {çikolata}	0.0070	0.3710	2.7101
{çikolata, kek} => {tatlı bisk}	0.0046	0.3576	3.6834
{gazlı içe, tatlı bisk} => {çikolata}	0.0048	0.3570	2.6082
{sebze, süt} => {meyve}	0.0036	0.3445	3.5721
{sigara, tatlı bisk} => {çikolata}	0.0029	0.3353	2.4495
{gazlı içe, meyve} => {sebze}	0.0028	0.3343	2.9082
{meyve, süt} => {sebze}	0.0036	0.3341	2.9061
{cips, çikolata} => {gazlı içe}	0.0027	0.3323	3.2771
{çikolata, sebze} => {meyve}	0.0033	0.3316	3.4387
{cips, gazlı içe} => {çikolata}	0.0027	0.3167	2.3138
{meyve, tatlı bisk} => {çikolata}	0.0031	0.3152	2.3031
{peynir, sebze} => {meyve}	0.0031	0.3062	3.1747

Tablo 3.7

3. Segmente Ait Kaldırıcı Yüksek Üç Ürünlü İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{meyve, peynir} => {sebze}	0.0031	0.4788	4.1644
{çikolata, tuzlu bisk} => {tatlı bisk}	0.0070	0.3994	4.1146
{çikolata, tatlı bisk} => {kek}	0.0046	0.1594	4.0215
{meyve, özel ekmek} => {sebze}	0.0033	0.4444	3.8659
{çikolata, kek} => {tatlı bisk}	0.0046	0.3576	3.6834
{sebze, süt} => {meyve}	0.0036	0.3445	3.5721
{çikolata, tatlı bisk} => {tuzlu bisk}	0.0070	0.2417	3.5622
{çikolata, sebze} => {meyve}	0.0033	0.3316	3.4387
{kek, tatlı bisk} => {çikolata}	0.0046	0.4365	3.1885
{peynir, sebze} => {meyve}	0.0031	0.3062	3.1747
{meyve, sebze} => {peynir}	0.0031	0.1255	3.1192
{özel ekmek, sebze} => {meyve}	0.0033	0.2876	2.9819
{meyve, süt} => {sebze}	0.0036	0.3341	2.9061
{çerez, çikolata} => {tatlı bisk}	0.0031	0.2768	2.8513
{çikolata, meyve} => {tatlı bisk}	0.0031	0.2705	2.7866
{çerez, tatlı bisk} => {çikolata}	0.0031	0.3713	2.7123
{tatlı bisk, tuzlu bisk} => {çikolata}	0.0070	0.3710	2.7101
{gazlı içe, tatlı bisk} => {çikolata}	0.0048	0.3570	2.6082
{çikolata, gazlı içe} => {tatlı bisk}	0.0048	0.2493	2.5686
{meyve, sebze} => {özel ekmek}	0.0033	0.1336	2.5455

4. SEGMENTE AİT ANALİZLER

37 mağazaya ait 1 günlük satış verisi toplam 40 bin 12 fişten ve bu fişlere ait toplam 137 bin 866 işlemden oluşmaktadır.

Tablo 4.1

4. Segmentte Sık Satılan Tek Öğeli Ürün Küme

Ürün	Destek	Fiş sayısı
Sebze	0.1783	7134
Sigara	0.1375	5502
Çikolata	0.1375	5502
Meyve	0.1271	5086
Gazlı içe	0.1218	4873
Özel ekmek	0.1088	4353
Tatlı bisk	0.0980	3921
Süt	0.0953	3813
Su	0.0829	3317
Peynir	0.0829	3317
Ekmek	0.0692	2769
Çerez	0.0684	2737
Tuzlu bisk	0.0621	2485
Yoğurt	0.0595	2381
Ev gereçleri	0.0579	2317
Gazete	0.0574	2297
Poğaç	0.0565	2261
Tatlı	0.0541	2165
Meyve suyu	0.0521	2085
Tavuk	0.0512	2049

Tablo 4.2**4. Segmentte Sık Satılan Çift Ögeli Ürün Küme**

Ürün	Destek	Fiş sayısı
Meyve, Sebze	0.0572	2289
Çikolata, Tatlı bisküvi	0.0374	1496
Özel ekmek, Sebze	0.0367	1468
Peynir, Sebze	0.0333	1332
Sebze, Süt	0.0322	1288
Gazlı içe, Sebze	0.0288	1152
Çikolata, Gazlı içecek	0.0284	1136
Meyve, Süt	0.0264	1056
Çikolata, Sebze	0.0258	1032
Sebze, Yoğurt	0.0250	1000
Meyve, Özel ekmek	0.0250	1000
Sebze, Yumurta	0.0232	928
Tatlı bisküvi, Tuzlu bisküvi	0.0231	924
Özel ekmek, Peynir	0.0225	900
Meyve, Peynir	0.0225	900
Et, Sebze	0.0223	892
Özel ekmek, Süt	0.0220	880
Çikolata, Meyve	0.0217	868
Sebze, Tavuk	0.0215	860
Çikolata, Süt	0.0215	860

Tablo 4.3**4. Segmentte Sık Satılan Üç Ögeli Ürün Küme**

Ürün	Destek	Fiş sayısı
Meyve, Sebze, Süt	0.0158	632
Meyve, Özel ekmek, Sebze	0.0155	620
Meyve, Peynir, Sebze	0.0148	592
Gazlı içe, Meyve, Sebze	0.0117	468
Çikolata, Meyve, Sebze	0.0115	460
Meyve, Sebze, Yoğurt	0.0112	448
Özel ekmek, Peynir, Sebze	0.0110	440
Çikolata, Tatlı bisküvi, Tuzlu bisküvi	0.0104	416
Meyve, Sebze, Yumurta	0.0103	412
Özel ekmek, Sebze, Süt	0.0102	408
Peynir, Sebze, Süt	0.0100	400
Çerez, Meyve, Sebze	0.0094	376
Sebze, Süt, Yumurta	0.0091	364
Çikolata, Gazlı içe, Tatlı bisküvi	0.0088	352
Meyve, Sebze, Tatlı bisküvi	0.0087	348
Sebze, Süt, Yoğurt	0.0087	348
Et, Meyve, Sebze	0.0084	336
Çikolata, Sebze, Tatlı bisküvi	0.0084	336
Çikolata, Süt, Tatlı bisküvi	0.0084	336
Gazlı içecek, Peynir, Sebze	0.0084	336

Tablo 4.4

4. Segmente Ait Güveni Yüksek İki Ürünlü İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{zeytin} => {peynir}	0.0129	0.5942	7.1655
{baharat} => {sebze}	0.0131	0.4878	2.7355
{et} => {sebze}	0.0223	0.4851	2.7208
{turşu} => {sebze}	0.0038	0.4784	2.6830
{konserve} => {sebze}	0.0140	0.4779	2.6802
{zeytin} => {sebze}	0.0102	0.4675	2.6218
{yumurta} => {sebze}	0.0232	0.4648	2.6068
{tereyağ} => {sebze}	0.0071	0.4636	2.6000
{balık} => {sebze}	0.0116	0.4527	2.5386
{meyve} => {sebze}	0.0572	0.4502	2.5248
{bakliyat} => {sebze}	0.0158	0.4456	2.4990
{sos} => {sebze}	0.0110	0.4365	2.4478
{un} => {sebze}	0.0056	0.4356	2.4430
{kaşar} => {peynir}	0.0112	0.4356	5.2532
{z yağı} => {sebze}	0.0042	0.4336	2.4317
{salam} => {peynir}	0.0106	0.4333	5.2250
{tereyağ} => {peynir}	0.0065	0.4209	5.0757
{yoğurt} => {sebze}	0.0250	0.4208	2.3599
{tavuk} => {sebze}	0.0215	0.4191	2.3503
{makarna} => {sebze}	0.0166	0.4148	2.3265

Tablo 4.5

4. Segmente Ait Kaldırıcı Yüksek İki Ürünü İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{toz tatlı} => {pasta malz.}	0.0029	0.4370	13.1415
{bulaşık det.} => {çamaşır det.}	0.0078	0.3302	11.4665
{un} => {şeker}	0.0033	0.2538	10.9845
{yağ} => {şeker}	0.0044	0.2500	10.8206
{mama} => {bebek ürünü}	0.0033	0.2114	10.4470
{un} => {pasta malz.}	0.0043	0.3333	10.0232
{margarin} => {şeker}	0.0027	0.2181	9.4388
{temizlik malz.} => {temizlik}	0.0058	0.3756	9.1667
{çamaşır det.} => {temizlik}	0.0104	0.3609	8.8092
{temizlik malz.} => {poşet}	0.0027	0.1743	8.4739
{konserve} => {yağ}	0.0043	0.1476	8.4217
{bulaşık det.} => {temizlik}	0.0082	0.3447	8.4115
{yağ} => {bakliyat}	0.0050	0.2875	8.1330
{un} => {bakliyat}	0.0037	0.2860	8.0901
{parfüm} => {kozmetik}	0.0027	0.1964	8.0109
{kaşar} => {zeytin}	0.0045	0.1733	7.9799
{çorba} => {makarna}	0.0039	0.3187	7.9633
{ıslak mendil} => {bebek ürünü}	0.0037	0.1609	7.9540
{kaşar} => {salam}	0.0049	0.1922	7.8646
{çay} => {şeker}	0.0078	0.1814	7.8495

Tablo 4.6

4. Segmente Ait Güveni Yüksek Üç Ürün İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{et, meyve} => {sebze}	0.0084	0.7495	4.2032
{baharat, meyve} => {sebze}	0.0052	0.7388	4.1436
{konserve, meyve} => {sebze}	0.0056	0.7380	4.1390
{balık, meyve} => {sebze}	0.0051	0.7257	4.0699
{makarna, meyve} => {sebze}	0.0074	0.7242	4.0616
{bakliyat, meyve} => {sebze}	0.0067	0.7211	4.0439
{meyve, tavuk} => {sebze}	0.0078	0.7165	4.0184
{meyve, yumurta} => {sebze}	0.0103	0.6990	3.9203
{meyve, pasta malz} => {sebze}	0.0065	0.6963	3.9052
{meyve, temizlik} => {sebze}	0.0060	0.6758	3.7902
{kahve, meyve} => {sebze}	0.0058	0.6639	3.7233
{meyve, yoğurt} => {sebze}	0.0112	0.6638	3.7227
{sebze, zeytin} => {peynir}	0.0067	0.6595	7.9530
{meyve, peynir} => {sebze}	0.0148	0.6562	3.6802
{çay, meyve} => {sebze}	0.0054	0.6462	3.6241
{et, süt} => {sebze}	0.0056	0.6425	3.6031
{meyve, özel ekmek} => {sebze}	0.0155	0.6177	3.4643
{et, gazlı içe} => {sebze}	0.0053	0.6147	3.4476
{et, peynir} => {sebze}	0.0063	0.6119	3.4317
{ekmek, meyve} => {sebze}	0.0063	0.6009	3.3702

Tablo 4.7

4. Segmente Ait Kaldırıcı Yüksek Üç Ürünlü İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{peynir, şeker} $\Rightarrow$ {yağ}	0.0027	0.3978	22.6945
{bakliyat, çay} $\Rightarrow$ {şeker}	0.0028	0.4531	19.6123
{peynir, yağ} $\Rightarrow$ {şeker}	0.0027	0.4467	19.3352
{bakliyat, konserve} $\Rightarrow$ {şeker}	0.0026	0.4141	17.9216
{konserve, şeker} $\Rightarrow$ {bakliyat}	0.0026	0.6092	17.2333
{bakliyat, peynir} $\Rightarrow$ {yağ}	0.0028	0.3013	17.1907
{bulaşık det, temizlik} $\Rightarrow$ {çamaşır det}	0.0040	0.4925	17.1014
{makarna, şeker} $\Rightarrow$ {bakliyat}	0.0031	0.5981	16.9203
{bakliyat, temizlik} $\Rightarrow$ {çamaşır det}	0.0029	0.4818	16.7280
{makarna, şeker} $\Rightarrow$ {konserve}	0.0025	0.4813	16.4885
{çamaşır det, temizlik} $\Rightarrow$ {bulaşık det}	0.0040	0.3864	16.3293
{makarna, yağ} $\Rightarrow$ {bakliyat}	0.0028	0.5750	16.2659
{çay, makarna} $\Rightarrow$ {bakliyat}	0.0027	0.5522	15.6220
{çamaşır det, makarna} $\Rightarrow$ {bakliyat}	0.0029	0.5479	15.5006
{temizlik, tuvhav} $\Rightarrow$ {bulaşık det}	0.0026	0.3661	15.4708
{bulaşık det, peynir} $\Rightarrow$ {çamaşır det}	0.0026	0.4454	15.4640
{bakliyat, makarna} $\Rightarrow$ {yağ}	0.0028	0.2700	15.4005
{konserve, makarna} $\Rightarrow$ {şeker}	0.0025	0.3515	15.2153
{çay, temizlik} $\Rightarrow$ {çamaşır det}	0.0026	0.4372	15.1817
{makarna, temizlik} $\Rightarrow$ {çamaşır det}	0.0027	0.4360	15.1384

5. SEGMENTE AİT ANALİZLER

5 mağazaya ait 1 günlük satış verisi toplam 6 bin 17 fişten ve bu fişlere ait toplam 30 bin 661 işlemden oluşmaktadır.

Tablo 5.1

5. Kümedeki Mağazalara Ait Veriler için En Çok Satılan ilk 20 Ürün*

Ürün	Destek	Fiş sayısı
Sebze	0.3204	1928
Meyve	0.2272	1367
Özel ekmek	0.1898	1142
Süt	0.1640	987
Peynir	0.1527	919
Gazlı içe	0.1496	900
Çikolata	0.1451	873
Sigara	0.1218	733
Yoğurt	0.1110	668
Çerez	0.1082	651
Yumurta	0.1072	645
Tatlı bisküvi	0.1002	603
Su	0.0906	545
Et	0.0858	516
Temizlik	0.0854	514
Tavuk	0.0841	506
Tuzlu bisk	0.0771	500
Ekmek	0.0751	500
Balık	0.0751	500
Meyve suyu	0.0734	499

* En çok satılan ürünler analizi yapılırken, ilk olarak ürün sayısı kısıtlamasına gidilmemiştir. Bu yüzden bu tabloda en çok birlikte satılan 2'li 3'lü ürünler de yer alabilirdi. Fakat ikili ürünlerden sadece meyve-sebze yer almıştır.

Tablo 2’de ise en çok birlikte satılan ikili ürünlerin tespiti amacıyla analiz en az 2 ürün elde etmek için ve yine en çok satılan ilk 20 ürünü görmek için ayarlanmıştır.

Tablo 5.2

5. Kümedeki Mağazalara Ait Veriler için En Çok Birlikte Satılan İkili ilk 20 Ürün*

Ürün	Destek	Fiş sayısı
{meyve, sebze}	0.1464	881
{özel ekmek, sebze}	0.1050	632
{sebze, süt}	0.0929	559
{peynir, sebze}	0.0922	555
{meyve, özel ekmek}	0.0751	452
{meyve, süt}	0.0738	444
{sebze, yumurta}	0.0683	411
{sebze, yoğurt}	0.0673	405
{meyve, peynir}	0.0642	386
{gazlı içe, sebze}	0.0627	377
{özel ekmek, peynir}	0.0587	353
{özel ekmek, süt}	0.0573	345
{et, sebze}	0.0532	320
{çikolata, sebze}	0.0527	317
{çerez, sebze}	0.0509	306
{balık, sebze}	0.0499	300
{peynir, süt}	0.0497	299
{sebze, tavuk}	0.0494	297
{süt, yumurta}	0.0494	297
{meyve, yumurta}	0.0484	291

* Analiz yapılırken iki ürün elde edilecek şekilde kısıtlama yapılmıştır.

Tablo 3’te de en çok birlikte satılan üçlü ürünlerin tespiti amacıyla, analiz en az 3 ürün elde etmek için ve en çok satılan ilk 20 ürünü görmek için ayarlanmıştır.

Tablo 5.3

5. Kümedeki Mağazalara Ait Veriler için En Çok Birlikte Satılan Üçlü ilk 20 Ürün*

Ürün	Destek	Fiş sayısı
{meyve, sebze, süt}	0.0585	352
{meyve, özel ekmek, sebze}	0.0553	333
{meyve, peynir, sebze}	0.0509	306
{özel ekmek, peynir, sebze}	0.0409	246
{meyve, sebze, yumurta}	0.0406	244
{meyve, sebze, yoğurt}	0.0392	236
{özel ekmek, sebze, süt}	0.0367	221
{peynir, sebze, süt}	0.0361	217
{sebze, süt, yumurta}	0.0339	204
{gazlı içe, meyve, sebze}	0.0332	200
{sebze, süt, yoğurt}	0.0324	195
{meyve, özel ekmek, peynir}	0.0312	188
{peynir, sebze, yumurta}	0.0309	186
{özel ekmek, sebze, yoğurt}	0.0299	180
{çerez, meyve, sebze}	0.0297	179
{meyve, özel ekmek, süt}	0.0297	179
{özel ekmek, sebze, yumurta}	0.0291	175
{meyve, peynir, süt}	0.0288	173
{balık, meyve, sebze}	0.0279	168
{gazlı içe, özel ekmek, sebze}	0.0279	168

* Analiz yapılırken üç ürün elde edilecek şekilde kısıtlama yapılmıştır.

Tablo 5.4

5. Kümedeki Mağazalara Ait Güveni Yüksek İki Ürünli İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{zeytin} => {sebze}	0.0239	0.6698	2.0902
{konserve} => {sebze}	0.0332	0.6645	2.0737
{tereyağ} => {sebze}	0.0249	0.6637	2.0714
{balık} => {sebze}	0.0499	0.6637	2.0714
{meyve} => {sebze}	0.1464	0.6445	2.0113
{yumurta} => {sebze}	0.0683	0.6372	1.9886
{kaşar} => {sebze}	0.0346	0.6341	1.9791
{baharat} => {sebze}	0.0381	0.6326	1.9742
{bakliyat} => {sebze}	0.0454	0.6305	1.9676
{et} => {sebze}	0.0532	0.6202	1.9354
{poşet} => {sebze}	0.0259	0.6142	1.9167
{zeytin} => {peynir}	0.0218	0.6093	3.9893
{meyveli yoğ} => {sebze}	0.0294	0.6082	1.8982
{yoğurt} => {sebze}	0.0673	0.6063	1.8921
{peynir} => {sebze}	0.0922	0.6039	1.8847
{ayran} => {sebze}	0.0206	0.6019	1.8786
{pasta malz} => {sebze}	0.0381	0.5995	1.8709
{sos} => {sebze}	0.0331	0.5870	1.8320
{tavuk} => {sebze}	0.0494	0.5870	1.8318
{temizlik malz} => {sebze}	0.0203	0.5728	1.7875

Tablo 5.5

5. Kümedeki Mağazalara Ait Kaldırıcı Yüksek İki Ürünli İlişki Kuralları

Kural	Destek	Güven	Kaldıraç
{temizlik malz} => {temizlik}	0.0175	0.4930	5.7707
{çamaşır det} => {bulaşık det}	0.0175	0.2567	5.7211
{çamaşır det} => {temizlik}	0.0299	0.4401	5.1519
{makarna} => {bakliyat}	0.0233	0.3627	5.0400
{temizlik} => {bulaşık det}	0.0189	0.2218	4.9426
{tuvhav} => {temizlik}	0.0196	0.3491	4.0868
{peynir} => {zeytin}	0.0218	0.1425	3.9893
{tereyağ} => {peynir}	0.0206	0.5487	3.5923
{tuzlu bisk} => {tatlı bisk}	0.0269	0.3491	3.4839
{kek} => {çikolata}	0.0184	0.4723	3.2555
{yumurta} => {kaşar}	0.0188	0.1752	3.2138
{kaşar} => {peynir}	0.0264	0.4848	3.1739
{tatlı bisk} => {çikolata}	0.0440	0.4395	3.0290
{tavuk} => {et}	0.0216	0.2569	2.9959
{gevrek} => {süt}	0.0173	0.4749	2.8950
{bakliyat} => {temizlik}	0.0176	0.2448	2.8657
{bakliyat} => {yumurta}	0.0221	0.3072	2.8654
{yumurta} => {süt}	0.0494	0.4605	2.8071
{yumurta} => {makarna}	0.0189	0.1767	2.7551
{pasta malz} => {süt}	0.0278	0.4372	2.6651

Tablo 5.6

5. Kümedeki Mağazalara Ait Güveni Yüksek Üç ÜrünlÜ İlişki Kuralları*

Kural	Destek	Güven	Kaldıraç
{balık, meyve} => {sebze}	0.0279	0.8485	2.6480
{meyve, yumurta} => {sebze}	0.0406	0.8385	2.6168
{meyve, temizlik} => {sebze}	0.0261	0.8177	2.5519
{meyve, yoğurt} => {sebze}	0.0392	0.8138	2.5397
{meyve, süt} => {sebze}	0.0585	0.7928	2.4742
{meyve, peynir} => {sebze}	0.0509	0.7927	2.4740
{et, meyve} => {sebze}	0.0273	0.7885	2.4607
{peynir, yumurta} => {sebze}	0.0309	0.7530	2.3501
{meyve, özel ekmek} => {sebze}	0.0553	0.7367	2.2992
{gazlı içe, meyve} => {sebze}	0.0332	0.7273	2.2697
{peynir, süt} => {sebze}	0.0361	0.7258	2.2650
{özel ekmek, yumurta} => {sebze}	0.0291	0.7143	2.2292
{süt, yoğurt} => {sebze}	0.0324	0.7065	2.2049
{peynir, yoğurt} => {sebze}	0.0276	0.7034	2.1952
{özel ekmek, peynir} => {sebze}	0.0409	0.6969	2.1749
{özel ekmek, yoğurt} => {sebze}	0.0299	0.6897	2.1523
{süt, yumurta} => {sebze}	0.0339	0.6869	2.1436
{gazlı içe, süt} => {sebze}	0.0253	0.6756	2.1083
{çerez, meyve} => {sebze}	0.0297	0.6729	2.1001
{çikolata, meyve} => {sebze}	0.0276	0.6694	2.0890

Tablo 5.7

5. Kümedeki Mağazalara Ait Kaldırıcı Yüksek Üç Ürünlü İlişki Kuralları*

Kural	Destek	Güven	Kaldıraç
{peynir, süt} => {yumurta}	0.0211	0.4247	3.9624
{süt, yoğurt} => {yumurta}	0.0191	0.4167	3.8870
{peynir, sebze} => {kaşar}	0.0191	0.2072	3.8011
{yoğurt, yumurta} => {süt}	0.0191	0.6183	3.7692
{peynir, süt} => {yoğurt}	0.0206	0.4147	3.7355
{kaşar, sebze} => {peynir}	0.0191	0.5529	3.6199
{özel ekmek, süt} => {yumurta}	0.0219	0.3826	3.5692
{meyve, süt} => {yumurta}	0.0278	0.3761	3.5088
{meyve, yumurta} => {süt}	0.0278	0.5739	3.4985
{süt, yumurta} => {yoğurt}	0.0191	0.3872	3.4877
{çikolata, sebze} => {tatlı bisk}	0.0183	0.3470	3.4626
{meyve, yoğurt} => {yumurta}	0.0178	0.3690	3.4420
{sebze, süt} => {yumurta}	0.0339	0.3649	3.4044
{sebze, tatlı bisk} => {çikolata}	0.0183	0.4911	3.3846
{sebze, yoğurt} => {yumurta}	0.0243	0.3605	3.3629
{meyve, peynir} => {yumurta}	0.0231	0.3601	3.3593
{meyve, yumurta} => {yoğurt}	0.0178	0.3677	3.3120
{özel ekmek, süt} => {yoğurt}	0.0209	0.3652	3.2897
{özel ekmek, yumurta} => {süt}	0.0219	0.5388	3.2845
{özel ekmek, sebze} => {kaşar}	0.0188	0.1788	3.2800

* Sınırlama sadece üç ürünlü ilişkiler elde etmek için yapılmıştır.