Master Thesis

Feature Selection by Machine Learning to Predict Gentrification in Istanbul

by

Elvan Kahraman

International Urban Development Program

International School of Urban Sciences

University of Seoul

August 2023

Feature Selection by Machine Learning to Predict Gentrification in Istanbul

by

Elvan Kahraman

A Master Thesis submitted to the International School of Urban Sciences of University of Seoul in partial fulfillment of the requirements for the degree of Master of Urban Administration and Planning

June 2023

Approved by

Seungil LEE

advisor

This certifies that the master thesis of Elvan KAHRAMAN is approved.

Thesis Committee Chair: Myungje WOO

Chung Ho KIM (Thesis Committee Member #1)

Seungil LEE (Thesis Committee Member #2)

International School of Urban Sciences
University of Seoul
June 2023

ABSTRACT

Gentrification tends to increase the economic value in neighborhoods. Finally, it creates socioeconomic stress there and ends up with displacement due to a decrease in affordability for old inhabitants. It has started to affect almost all districts in Istanbul. This study discusses the negative effects of gentrification in Istanbul as a unique case and how to mitigate it.

Recent studies in the field of machine learning are allowing us to predict how and where complex urban problems will occur in the future. This study focuses on feature selection which is a key problem and the essential part of machine learning studies. Feature selection is a data preprocessing approach which is effective to prepare especially high-dimensional data. The objectives of feature selection are preparing clean and understandable data, allowing to study with high-dimensional data and, improving performance of models by building more comprehensive machine learning models.

The datasets are collected for eleven districts named Ataşehir, Beyoğlu, Bakırköy, Başakşehir, Beşiktaş, Kadıköy, Kartal, Küçükçekmece, Sarıyer, Ümraniye, Üsküdar. Gentrification is measured by census track data which consists of net migration rate and change of education level. Since the triggering relationship between gentrification and increasing economic value, a point of interest dataset which consists of 17000 rows approximately is used to select the features which identify gentrification most.

In conclusion, a machine learning model is provided which scores features of POI dataset to identify the most important ones which will give a strong idea for further studies to predict gentrification. It sets out the effects on policies which will combat gentrification.

Keywords: Gentrification, Feature Selection, Machine Learning, Istanbul, Prediction

Master Thesis

Feature Selection by Machine Learning to Predict Gentrification in Istanbul

by

Elvan Kahraman

International Urban Development Program
International School of Urban Sciences
University of Seoul

August 2023

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1 Background

1.1.1 Gentrification in Istanbul

Istanbul is the socio-economic, historical, and cultural center of Turkiye. Since prehistoric times, the city has hosted several nations, cultures and religions. Consequently, the neighborhoods reflect a historical pattern through their architectural and cultural structures. Therefore, protecting these structures in the city has been an essential concern of urban planning (Tekeli, 2013; Ergun, 2004).

The city has felt the effects of gentrifying neighborhoods for almost twenty-five years. Although gentrification is related to increasing the economic value of neighborhoods by renewing and rebuilding them, it affects the cultural sustainability by damaging both the cultural identity and socio-economical heterogeneity of the city.

According to Islam (2005) gentrification in Istanbul occurred by three waves. In 1980s, neo-liberal development policies drove the first wave of gentrification which first signs took place in neighborhoods named Kuzguncuk (Uskudar), Arnavutkoy (Besiktas) and Ortakoy (Besiktas). However, the process in these three neighborhoods began with housing stock rehabilitation, the process in Ortakoy continued as reorganization of Ortakoy Square by the local municipality. Consequently, being an attraction point by bars, cafes and restaurants led displacement for old habitants.

The second wave began in the late 80s in Cihangir, Galata and Asmalimescit neighborhood in Beyoglu. The process started in Istiklal Street first by the revitalization projects. While the gentrification proceeds as residential and commercial which drove a massive displacement in Asmalimescit, it was more prevalent and slow change in Galata and Cihangir. Cihangir faced the massive moves after the mid-1990s by dramatic increases in house prices.

The third wave first took place in Fener and Balat neighborhoods where are the parts of Golden Horn and have similar historical pattern with the first wave neighborhoods in the late 90s. The most habitants were poor live in buildings which initially built for single families (old Ottoman style wooden houses) then cut up into small units and became cheap rental flats. An investment project funded by the European Commission in partnership with Fatih Municipality to mitigate unpleasant smells caused by industrial pollution finished by rehabilitation of around one-seventh of the total housing stock in the two neighborhoods between 2004 and 2006. The gentrification process in early 2000s began with an active role of the central government through urban revitalization programs and state-private partnerships. In this sense, the Urban Regeneration Law, enacted in 2012, played a significant role in the housing structure and city life of Istanbul. It can be easily said institutional involvement was an invitatory of gentrification in the third wave. (Daş, Özşahin 2021)

1.1.2 Urban Regeneration in Istanbul

By the end of 90s central government, local governments and market actors began to discuss the necessity of the urban regeneration on TV channels (year). They defined the necessity of regeneration in Istanbul and pointed out earthquake, slum areas, low quality housing and environment, deteriorated pattern of historical architectures and they expected urban regeneration to solve these problems. Some arguments were putted forward to ensure that urban regeneration projects would be legitimate. First argument was to revive many industries by attracting international capital correspondingly the job opportunities could be increased. In the other words, developments in the construction industry and real estate market could prevent economic crisis. As a result, it could produce solutions for benefit of the society. The project of regeneration in the historical center could protect structure and sustainability of the historical center. The most legitimate argument was the threat of the predicted earthquake in Istanbul. Therefore, Istanbul has become a city where the most obvious examples of this regeneration were put into practice. Consequently, a bitter rant struggle took place in the city. (Türkün,2014)

1.1.3 Effects of Urban Regeneration on Gentrification

Urban regeneration is a complex process that involves the physical, social, economic, and environmental revitalization and renovation of urban areas. It can lead both positive and negative effects on gentrification, depending on how it is implemented and the context in which it occurs. (Turkun, 2014)

Turkun (2015) argues that since the big urban regeneration projects such as building trade centers, shopping malls, luxury housing estates has become an essential priority for municipalities, huge construction firms, and development agencies in Turkiye, urban areas with high rent – gaining has become attraction point for investments. Therefore, this has created a pressure on the housing areas of low-income group. It is a process ending up with displacement. In this context it is intended to demolish these squatter housing areas especially in in central locations and move the inhabitants out of the city center to live in the housing blocks built by Housing Development Administration of Turkiye (TOKI) and the condition of paying by means of bank loans for 15 years. On the other hand, in the historic areas, regeneration projects are imposed, leading to the gentrification of these areas and eviction of the inhabitants.

1.1.4 Academic Background and Practical Experience to Solve the Problem

In the context of urban regeneration projects in Istanbul, unless plan before applying regeneration projects and involve participants who is affected by these projects, the regeneration processes will lead to waste of sources, will effect negatively development of environment and architecture, will lead to risk of inequity (Tekeli,2013).

Tekeli (2013) underlines that after economic development of Turkiye has reached middle – income group level, it is natural planning to regenerate slum areas and strengthen building stock which is under risk of being destroyed by an earthquake. However, the implementation of the plans have had some problems which will lead negativities to be more irreversible than before. Therefore, the regeneration projects need to be discussed in the political and social public sphere transparently. In order to plan such a development to take place, civil society organizations and professional chambers have big role. Tekeli (2013) argues four policies to combat negative effects of urban regeneration projects: The first, the regeneration projects cannot be implemented without taking any idea and will of habitants who live in those neighborhoods where will be subject a regeneration project. Secondly, the process of participant must be examined well. Thirdly, The consistency between the financial resources of habitants and the model of the project. Lastly, the process must not be run by overly authorized governments which implement the projects by force.

After all, gaining lessons from past and tracking the displacement can be seen as a practical approach. For instance, Seattle's Office of Planning & Community Development (Seattle OPCD) developed the Displacement Risk Index for Seattle 2035 Comprehensive Equity Plan Analysis (Seattle OPCD, 2016). The office has given areas with higher displacement risk as lower growth targets and designed specific mitigation strategies for each village at risk.

1.2 Research Objectives

Even though neighborhood change starts off slowly, results keep follow an exponential growth in poverty. The local governments need to design of displacement strategies to identify vulnerable neighborhoods in the gentrifying areas.

For instance, reduction in quality and quantity of public facilities such as schools, transportation creates the need to identify neighborhood change in early phases. Local governments are working to maintain opportunity parity to prevent inequity, and to stem disinvestment or to ensure that existing residents benefit from new investments. Therefore, they need to get ahead of these neighborhood changes, take an action with a preventive approach, and to be able to assess situation in a well-timed approach. In this context, taking an action before gentrification occurs makes prediction one of the keys for the management of gentrification. Eventually it has been seen that predicting gentrification makes itself an essential policy.

The purpose of this research is to evaluate a machine learning model to predict features of data sets in order to score them that will increase accuracy of the prediction. Therefore, it will be a potent beginning before evaluating prediction of neighborhood changes for the future.

1.3 Scope

First, this study covers the definition and diagnostic elements of “gentrification” and its historical background in Istanbul and in global according to relevant literature. Actors and forces which trigger its existence and the urban problems occurred from gentrification in Istanbul is discussed distinguishing between universal and unique phenomena. This study describes academic background and practical experience to solve the problems caused by urban gentrification. Also, it is discussed how to deal with gentrification in Istanbul based on relevant literature.

Secondly, it covers definitions of methodology which this study conducts. It produces an evaluation of machine learning model predicting the most relevant feature among datasets. In order to evaluate a model for feature selection, point of interest data and census tracks of net migration rate and education level change data between 2013 and 2021 years in eleven districts are used.

Lastly, policy recommendations for the gentrification in Istanbul is proposed through this analyze.

1.4 Significance of the Research

Although there are many studies on identifying and understanding the gentrification, these studies remain limited for the decision makers. On the other hand, lack of sufficiently successful use cases among the predictive gentrification methods, and the lack of trust in the statistical process of neighborhood change warning systems are their key downfall. Predicting neighborhood almost real time has been a challenge in past due to unavailability of current and instantly updated data.

The main key point of this study is Feature Selection Problem (FSP) which means the selection of the most important independent variables in order to improve the prediction capability of a given system and help to develop accurate models for the future.

1.5 Organization of the Thesis

This study consists of five chapters. The first chapter is introduction part which describes gentrification in Istanbul and problems occurred from it by distinguishing universal and unique phenomena. Academic background and practical experiences to solve these problems are described. It discusses research objectives of the thesis and sets out the scope. Lastly it presents the purpose and objectives of the research and the expected outcomes.

The second chapter covers the definition and the diagnostic elements of the urban gentrification according to the relevant literature. It discusses how to deal with gentrification in a concentrated city as Istanbul. Lastly it describes the policy recommendations for the gentrification in Istanbul.

The third chapter sets out methodological framework, introducing the methods that were used and how the study was conducted. It covers research design and data validity sections.

The fourth chapter which named data analysis and interpretations dives deep into model that is produced by machine learning algorithm. In this chapter methodology which is used to train data set and data preparation steps is described, at the end a model and evaluation is produced.

The last chapter concludes the ideas through this research and recommend policies to combat with gentrification.

2. LITERATURE REVIEW

2.1 Defining Gentrification

The term of “gentrification” was coined by sociologist Ruth Glass in 1964. According to definition of Glass (1964); gentrification is a process that could be explain as arrival of a middle-high income class in a neighborhood and improving that neighborhood where a low-income class lives at that time. She observes that “Once this process of ‘gentrification’ starts in a district it goes on rapidly until all or most of the working-class occupiers are displaced and the whole social character of the district is changed.” (1964,26). Ozden (2016) defines that gentrification is a result of urban regeneration process and, gentrification is creating a city model by a middle-high class who occupies in center of cities. Ozden (2016) considers gentrification as socio-economical regeneration. In this context, we could consider that urban regeneration process and urban gentrification might affect each other. The arrival of new class helps to improve urban area by socially, economically, and physically. However, it leads old residents to move from their neighborhood. While Neo-liberal economies support this gentrification process since they benefit from luxury residents, Marxist policies criticize it.

In 1950s the gentrification occurred in New York and London for the first time, and in 1970s it occurred in Turkiye. The cities might trigger social inequity result of isolation, and ruined identities. This process might bring out social exclusion. The social exclusion deprives citizens to bind with society. Therefore, the excluded citizens prefer to be extraordinary and, to build a new environment to live integrated. Moreover, this triggers a poverty circle that the citizens can not break. In this context, the city policies produce inequities. After finishing industrialization in the city centers, the citizens began to return to the cities. Therefore, residents gained in value and, the citizens preferred to buy a house rather than rent it. Simultaneously, a social separation has occurred between low-income and middle-high income group. Special architects and designs of residents has set in motion a social transformation (Armagan,2014).

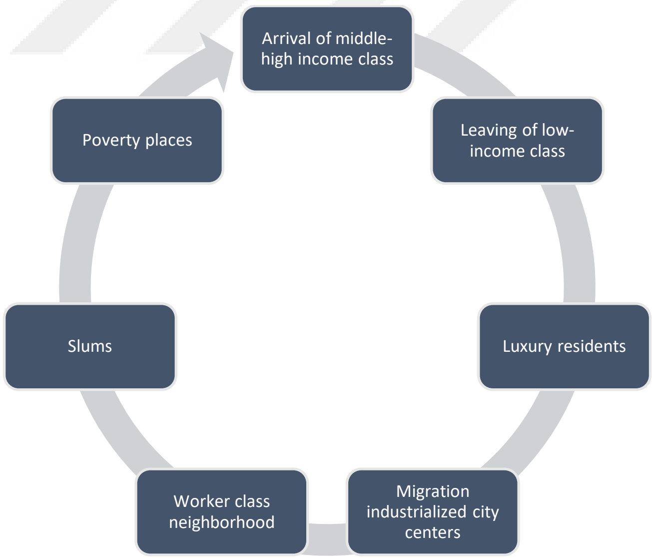


Figure 1: Life Cycle of A Gentrification

According to Smith (2006), the gentrification is one of the strategies in global cities and, luxury residents take place in the center of these cities. As high-middle classes take place as actors of gentrification in definition of Glass (1963), governments and companies take place as actors of this process in 20th century. Smith (2006) discusses that the gentrification has been planned by governments since urban gentrification and urban regeneration are similar concepts. As the centers of capital cities have begun to serve just the high-middle classes by the policies of neo-liberal economies, one of the questions and problems, as results of gentrification, occurs: where low-income classes are moving. In this context, the gentrification is not a permanent methodology for urban regeneration, but a shifting of slum areas.

2.2 Defining Machine Learning and Feature Selection

Utilizing quantitative methods to study an urban science such as neighborhood change has been controversial. Even though there are studies utilizing quantitative methods to measure rent gap or to analyze change through census tracks, it is still being avoided to utilize a quantitative method to study on gentrification in literature. However, mostly the literature utilizes qualitative methods based on media analysis, interviews and other observational methods to collect data to understand the gentrification, this method may overlook areas that experienced similar changes to those more widely recognized as gentrified (Reades et al.,2019).

Statistical methods can provide an answer to find the pattern which addresses gentrification among the datasets. However, studying on high volume and great diversity make it almost impossible to create linear methods. Therefore, it is inevitable to employ Machine Learning models since they provide to learn from and make predictions from large datasets. There are two optimization of data set: the first one is training some part of data to measure accuracy of prediction and the second one is testing for robustness using the trained data set.

Most recently, several gentrification studies have employed ML models evaluate the complex and dynamic nature of gentrification (Reades et al.,2019). ML models have significant strengths in their robustness and flexibility to work with complex spatial-temporal data inputs.

Feature Selection is an essential part of Machine Learning which is responsible for how much a dataset contributes to training the model. Features are accuracy coefficient of the ML algorithm. The more chosen the most relevant features among a feature set, the more accurate prediction.

3. RESEARCH METHODOLOGY

3.1 Selection of Eleven Districts

For this research eleven districts in Istanbul are identified. The criteria while choosing the districts was the diversity. Each eleven districts have their own cultural, economic and natural patterns.

a. Atasehir

Ataşehir is relatively a new district which is located on Anatolian side and composed of 17 neighborhoods. Most buildings in the district are high-rise towers, while some of them are among the tallest skyscrapers in Turkiye. The district is home for Turkish Central Bank Tower which is the tallest (1,154 ft. 10 in) skyscraper in Turkiye after Metropol Istanbul Tower (987ft. 6 in). Beside Istanbul Financial Center opened in Ataşehir in 2023. Unlike in Başakşehir, another satellite city, no single-family house types were designed for Ataşehir's residential projects.

The population of the district is 423,127 as of 2022.

Figure 2: Location of Ataşehir in Istanbul

Figure 3: Ataşehir District

b. Beyoğlu

Beyoğlu is a district on European Side of Istanbul separated from the old city (historic peninsula of Constantinople) by the Golden Horn.

In accordance with the character of the Tanzimat period, the first example of the application of western urbanism was started in Beyoğlu. The district could have been a starting point for westernization with the region’s history, geography, ethnic and cultural characteristics. Beyoğlu was a window to the world as a trade center with its harbor, quay and Golden Horn. In addition to this feature, its dominance over the capital and the Bosphorus attracted all foreign embassies. The local people were wealthy and open to the west.

Reforms started by regulations Istanbul separated for fourteen municipality in 29th December in 1857. Although Beyoğlu was the first application it named as “6. Office”. The reason to take this name was the most prosperous region in Paris named “Sixeme Arrondissement”. The first job of this office was to draw cadastral map of Beyoğlu. A Hospital was built and lighting of the streets with kerosene-fired lanterns was started for the first time in Cadde-I Kebir (Istiklal Street).

In 1924, Beyoğlu Municipality established. Municipality building was built by architect Barborini in 1879-1883 when Edouard Blacque was mayor of Beyoğlu Municipality.

In 1984, it was restructured within the framework of the “Local Administrations Law” enacted for Metropolitan and District Municipalities and, “Beyoğlu Municipality” consists of 45 neighborhoods received its current status.

The population of the district is 225,420 as of 2022.

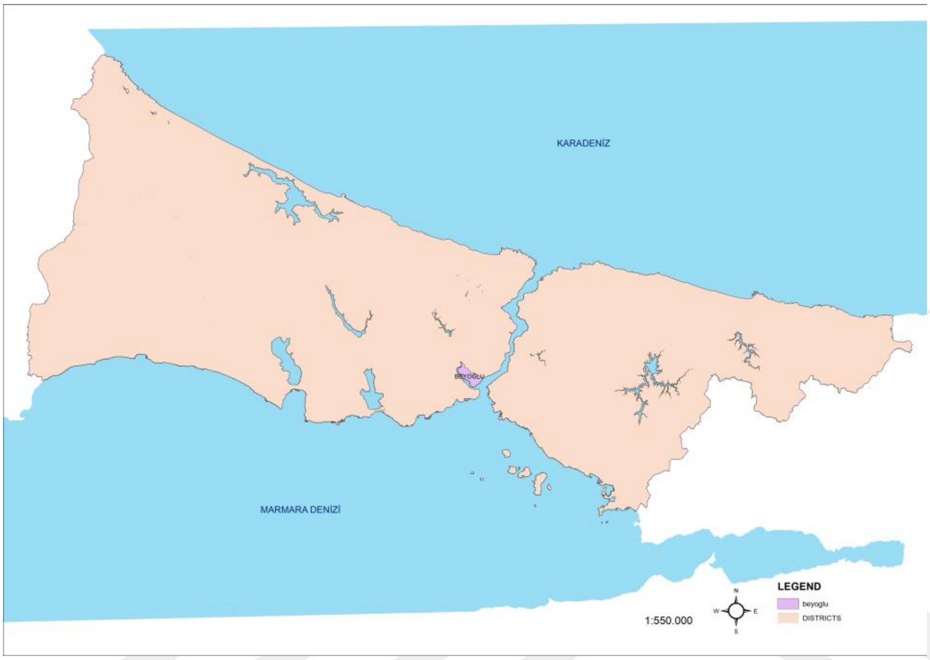


Figure 4: Location of Beyoğlu in Istanbul

Figure 5: Cihangir Neighborhood in Beyoğlu

c. Bakırköy

Bakırköy is a district on the European side of Istanbul. The quarter is densely populated. It has a residential character and is inhabited by an upper middle-class population.

In the Byzantine period Bakırköy was a separate community outside Constantinople, a well-watered pleasant seaside retreat from the city, and was called Hebdomon.

The center of Bakırköy is an important commercial area indicates shopping malls, bars and cafes, as well as conversion of streets to pedestrian malls.

The population of the district is 226,685 as of 2022.

Figure 6: Location of Bakırköy

Figure 7: Bakırköy District

d. Başakşehir

Başakşehir is one of the second-level districts located on European side. The district has been hosting many public or private corporate housing projects since 1994. The population of the district is 514,900 as of 2022.

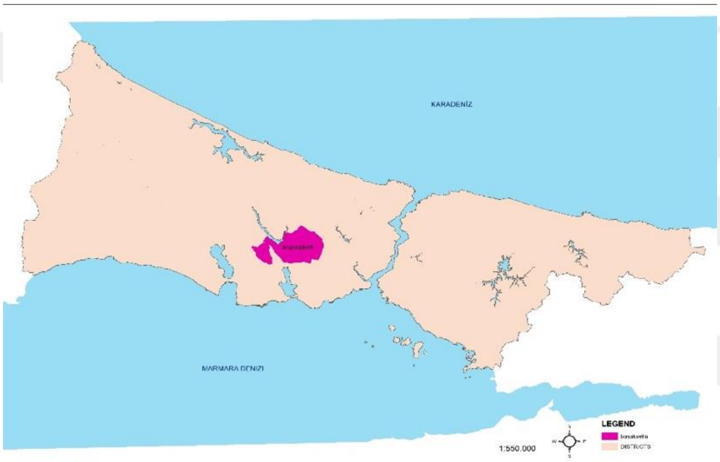


Figure 8: Location of Başakşehir

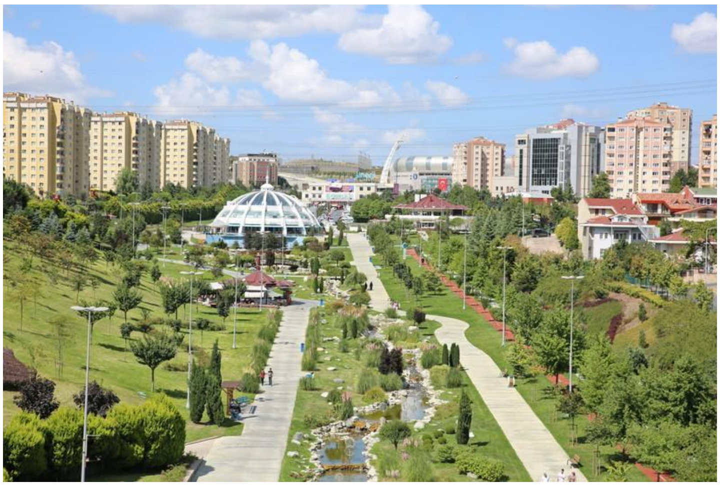


Figure 9: Başakşehir District

e. Beşiktaş

Beşiktaş is a district located on European shore of the Bosporus strait. It is known that during the reign of Suleyman the Magnificent, the region developed further and in the following years, this development continued and over time, it became a place where slave markets were established, soldiers and commercial caravans passing from Rumelia to Anatolia gathered and rested. With the construction of Çırağan Palace (1871), Dolmabahçe Palace (1856), Yıldız Palace (1898), and pavilions and mansions, it became the administrative center of the Ottoman Empire.

In 1930, it became a district by leaving from Beyoğlu. The fact that the Atatürk and his mother, Zübeyde Hanım, resided in Akaretler and his passing away in Dolmabahçe Palace gave Beşiktaş an exceptional place in the history of the Republic.

Beşiktaş is one of the central districts of Istanbul. During the Ottoman period, it was a place where various building types were seen together, as it was a district where members of the dynasty and high state officials as well as ordinary people lived. This diversity has added a different beauty to Beşiktaş district.

Figure 10: Location of Beşiktaş

Figure 11: Beşiktaş District

f. **Kadıköy**

The history of Kadıköy dates back to 5000 BC. Kadıköy known in classical antiquity and during the Roman and Byzantine eras as Chalcedon, is a large, and cosmopolitan district on the Anatolian side of Istanbul. It hosts major Haydarpaşa Terminal of the Turkish State Railways which opened in 1908. It was the terminus of the Istanbul-Baghdad and Istanbul-Damascus-Medina railways.

The population of the district is 483,064 as of 2022.

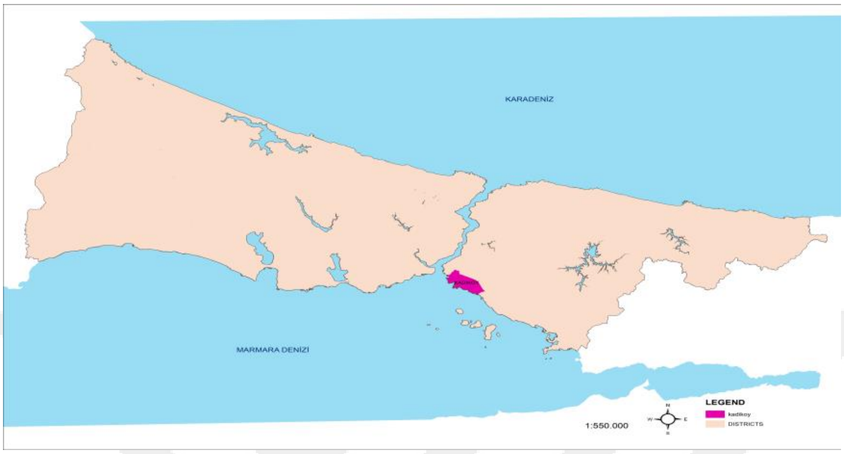


Figure 12: Location of Kadıköy

Figure 13: Kadıköy District

g. Kartal

During the Byzantine Empire Kartal (‘eagle’ in Turkish) was a fishing village on the shore of the Marmara Sea called Kartalimen in Greek. The district was founded at the beginning of the 6th century.

The population of the district is 483,418 as of 2022.

Figure 14: Location of Kartal

Figure 15: Kartal district

h. Küçükçekmece

Küçükçekmece lies on the European shore of the Sea of Marmara, on a lagoon named Lake Küçükçekmece. It is the second most populous district in Istanbul and the population of the district is growing rapidly day by day. New schools, shopping malls, modern living spaces and other facilities are being built rapidly parallel of this.

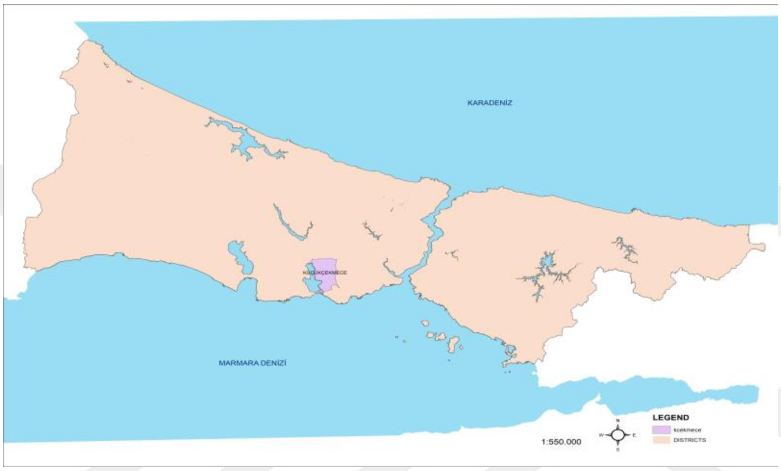


Figure 16: Location of Küçükçekmece

Figure 17: Küçükçekmece District

i. Sarıyer

Sarıyer is on the European side of Istanbul where consists of neighborhoods of Rumelifeneri, Tarabya, Yeniköy, İstinye, Emirgan and Rumelihisarı. Sarıyer's villages, backed by steep hills were once rural fishing communities. At the Ottoman period these villages were the places where sultans came for relaxing. During 18th and 19th centuries several foreign embassies built summer residences here.

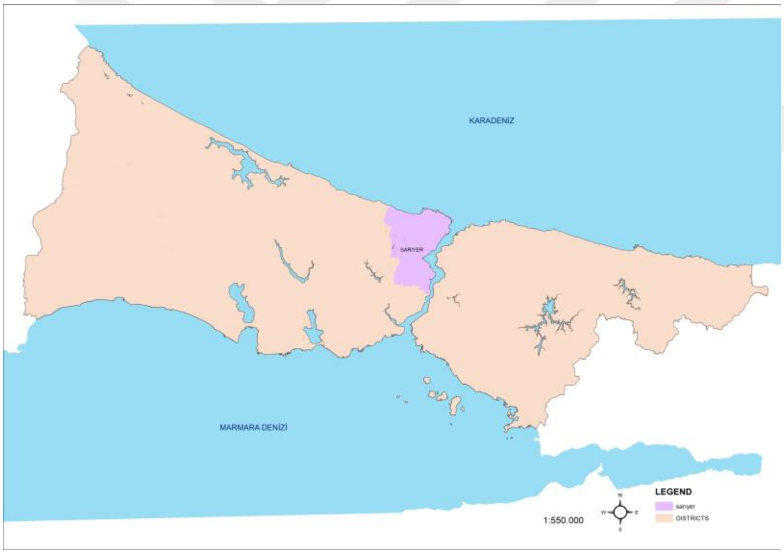


Figure 18: Location of Sarıyer

Figure 19: Sarıyer District

j. Ümraniye

Ümraniye is very large working-class district located on Anatolian part of the city. It consists of 38 neighborhoods. It was a rudimentary district short time ago. Recently, there has been a great deal of change all around the district. With the projects such as building complex, large shopping malls, entertainment centers and environmental urbanization, Ümraniye is fastest growing district in Istanbul.

Ümraniye began as a small village and population was not more than 1000. Due to migration to cities from rural areas of Anatolia, its population grew same as all the other areas in Istanbul. Consequently, it lost its green nature and numerous illegal houses built quickly without caring any of urbanization. Black Sea contractors built wherever was empty. In conclusion, aspect of Ümraniye is not different than far Anatolian city.

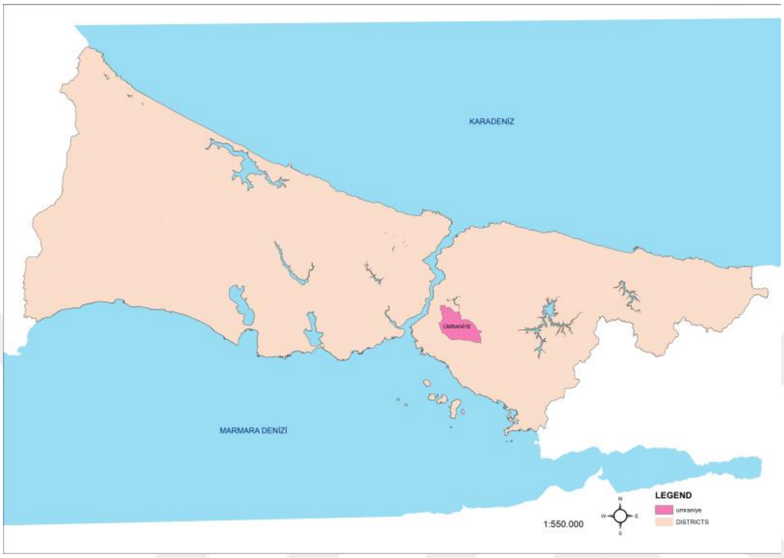


Figure 20: Location of Ümraniye

Figure 21: Ümraniye District

k. Üsküdar

Üsküdar is located on Anatolian side of the city and it is considered the gateway to Istanbul Asia. It includes highly populated 33 neighborhoods.

Among the most famous features of Uskudar is the Maiden Tower which is a small tower on a small island. The historic city center Fatih district is facing it on the opposite shore to the west. The district was called Skuotarion during Byzantine Empire. It has been a conservative cultural center of the Anatolian side since Ottoman times.

Figure 22: Location of Üsküdar

Figure 23: Üsküdar District

3.2 Research Design

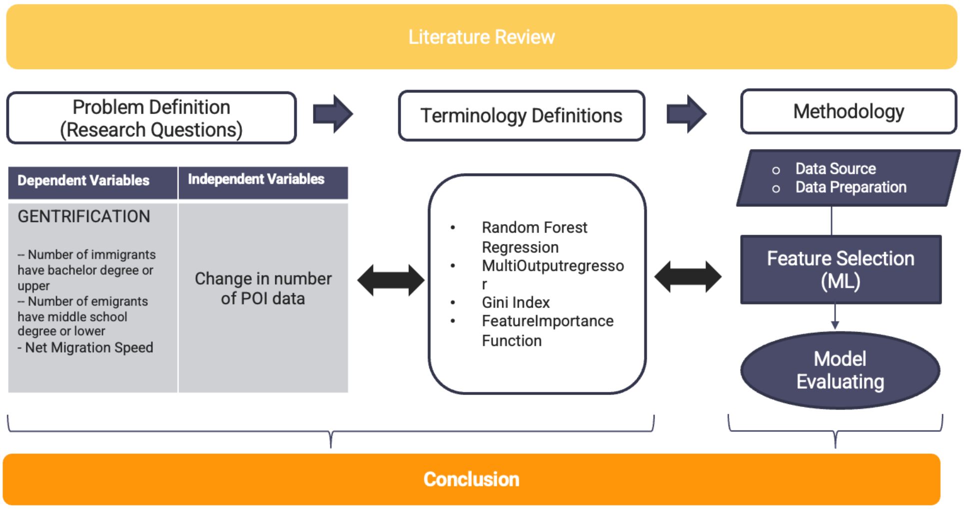


Figure 24: Research Design Scheme

3.3 Random Forest Regression Method

Random Forest is a machine learning method which creates a forest consists of trees. It starts with one node and creates a forest by branching out new nodes (ends). Trees are created by m predictors chosen randomly at every node. The count of m must be significantly smaller than the number of total predictors. While tree nodes include members of just one class for the classification method, they continue to branch out until a small amount of unit in the regression method.

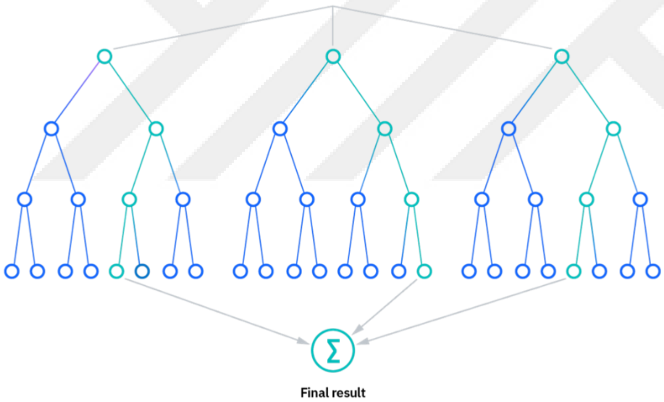


Figure 25: Random Forest Method

Random Forest provides valid predictions since it uses improved versions of random sampling and ensemble methods. Since low correlation between trees and low bias this method produces more accurate predictions. The method makes itself the best option since it makes easy to measure relatively importance of features in a dataset.

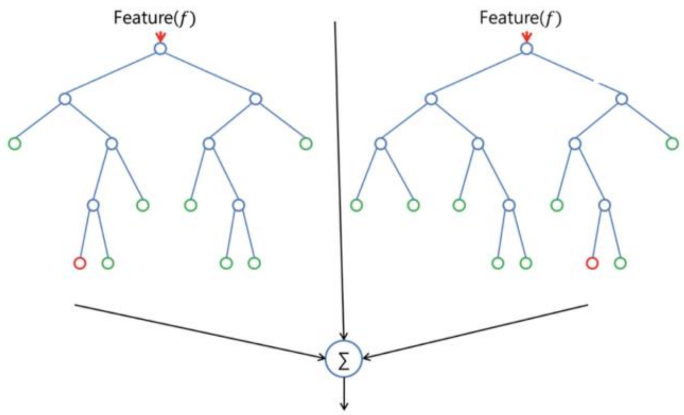
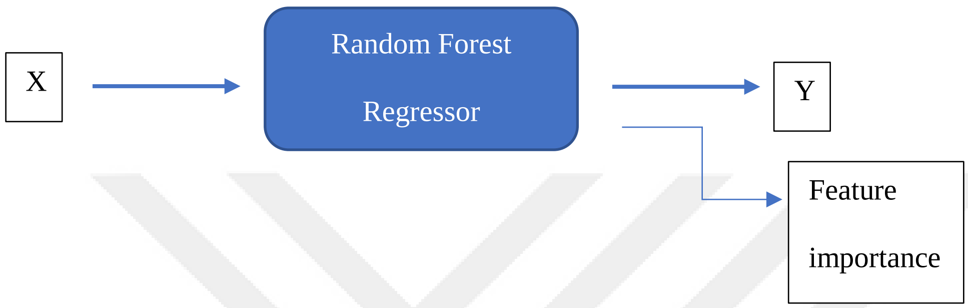


Figure 26: Random Forest Algorithm to Select Features

"Sklearn" library provides a useful tool which choose a feature by looking how much it decrease the pollution between trees in the forest. It trains this score for each feature automatically and scales evaluations, sum of all importance of feature are 1.

3.4 Feature Selection

Feature importance measures effect on the prediction. Feature selection and scoring importance of features are useful to interpret the structure of forest. Although principal component analysis is employed to reduce high volume data before statistical analysis, it remains to be failed.



Gini Index

One of the calculation methods to score the importance of features is Gini importance. Gini importance is derived from Gini index that is used when creating random forest trees. Gini index measures the inequality level of a sample that is assigned to a node. As p shows a classification problem, Gini index belongs to a k node is calculated by a formula as below:

$$Gk=2p(1-p)$$

The more purified a node is, the smaller Gini value becomes. When a node split over variable v the Gini value of the two nodes obtained will be smaller than the Gini value of the split node. The significance of Gini value of the variable v for each single tree is obtained by calculating the difference between these two values. After all the trees in the forest in the forest are created, the importance of the variable v is determined by summing the Gini significance of the trees in which the variable v takes place.

Multi-Output Regressor

If the data to be predicted is a vector, multi-output regression is required since the data will be in one row. Multiple output regression support can be added to any regressor by MultiOutputRegressor method. This strategy involves placing one regressor per target. Since each target is represented by one regressor, it is possible to obtain information about the target by examining the corresponding regressor. Since MultiOutputRegressor fits one regressor per target, it cannot take advantage of correlations between targets.

Feature- Importance Function

In order to select the features that are the subject of this study, the feature-importance function is used. By ordering the effect ratio of the features, the important features are listed.

4. DATA ANALYSIS / INTERPRETATIONS

This research looks for the answer which subcategories of POI dataset address gentrification which is defined by census tracks consists of migration rate, change in number of education levels.

4.1 Data Sources

4.1.1 TUIK (Turkish Statistical Institute) Data

The primary source of data is Turkish Statistical Institute (TUIK) which provides census track data. For this research two kinds of TUIK census track data sets are used:

a. <u>Net Migration Rate (per 1000)</u>

This data gives information on "net migration rate" data that defines the number of immigrants minus the number of emigrants divided by population. It expresses net number of migrants per 1000 population.

As it shows below, net migration rate maps of eleven districts – Ataşehir, Başakşehir, Bakırköy, Beyoğlu, Beşiktaş, Kadıköy, Kartal, Küçükçekmece, Sarıyer, Üsküdar, Ümraniye- between 2013-2021 is produced by GIS.

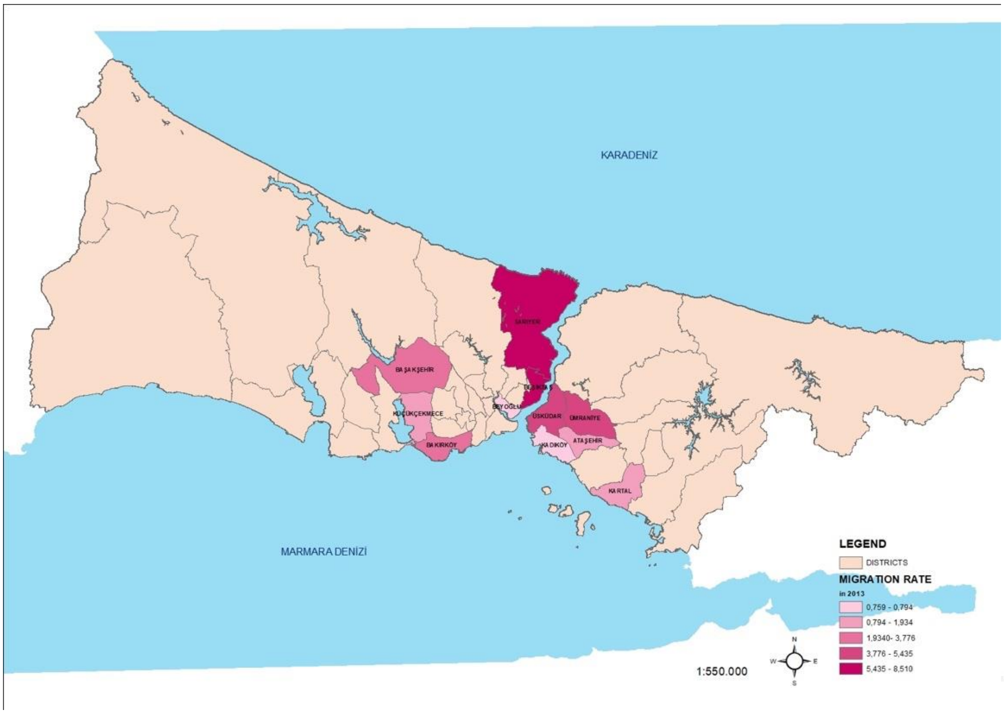


Figure 27: Map of Net Migration Rate in 2013

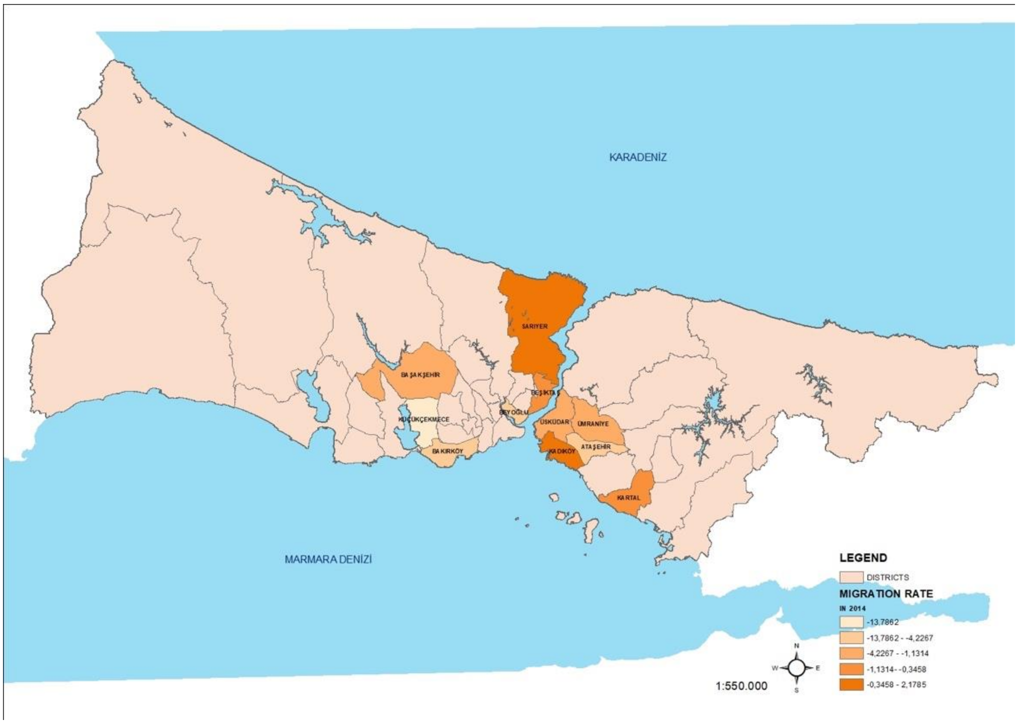


Figure 28: Map of Net Migration Rate in 2014

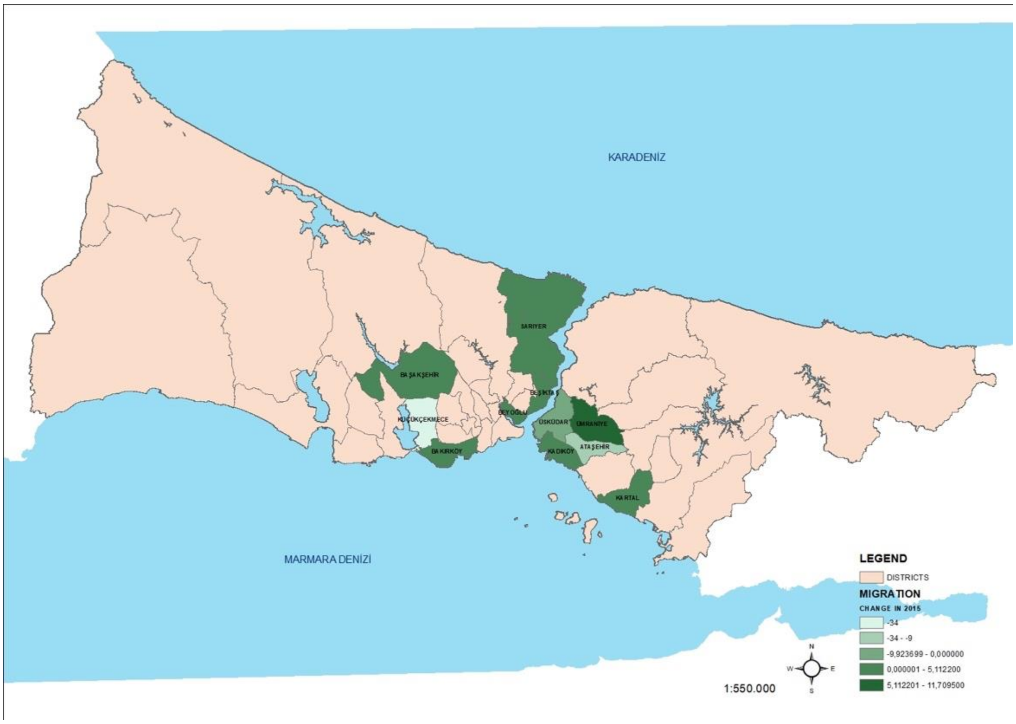


Figure 29: Map of Net Migration Rate in 2015

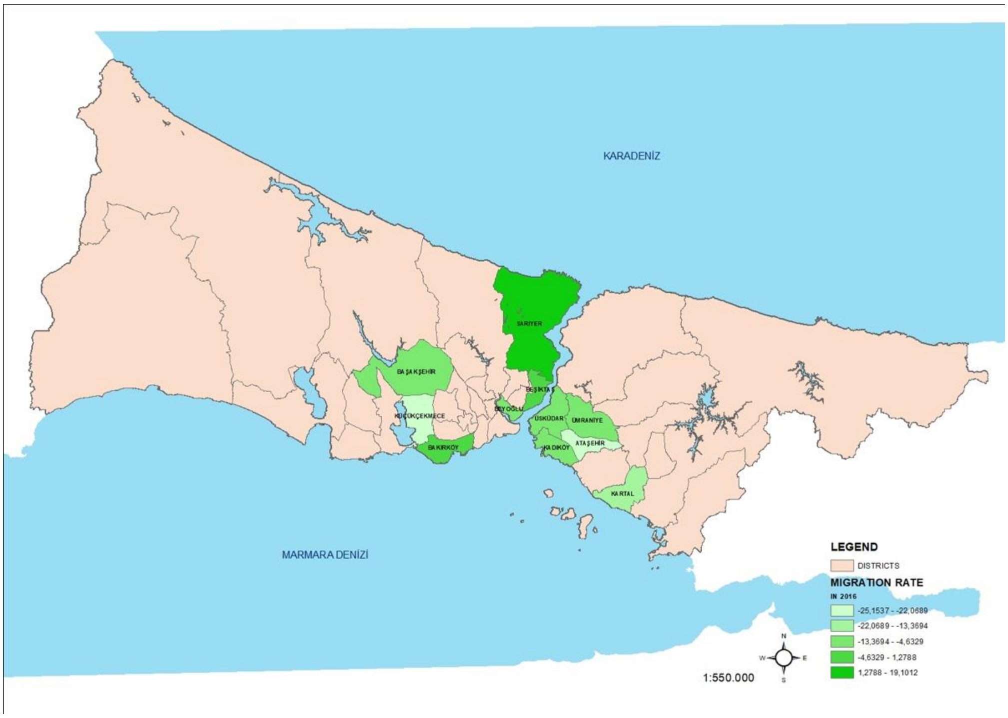


Figure 30: Map of Net Migration Rate in 2016

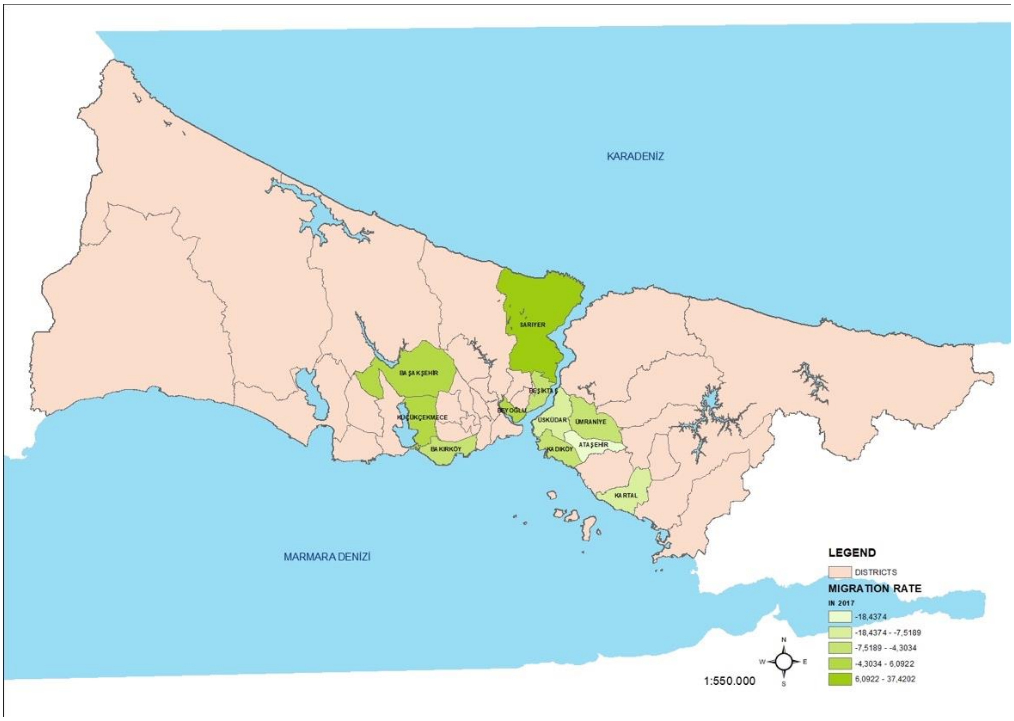


Figure 31: Map of Net Migration Rate in 2017

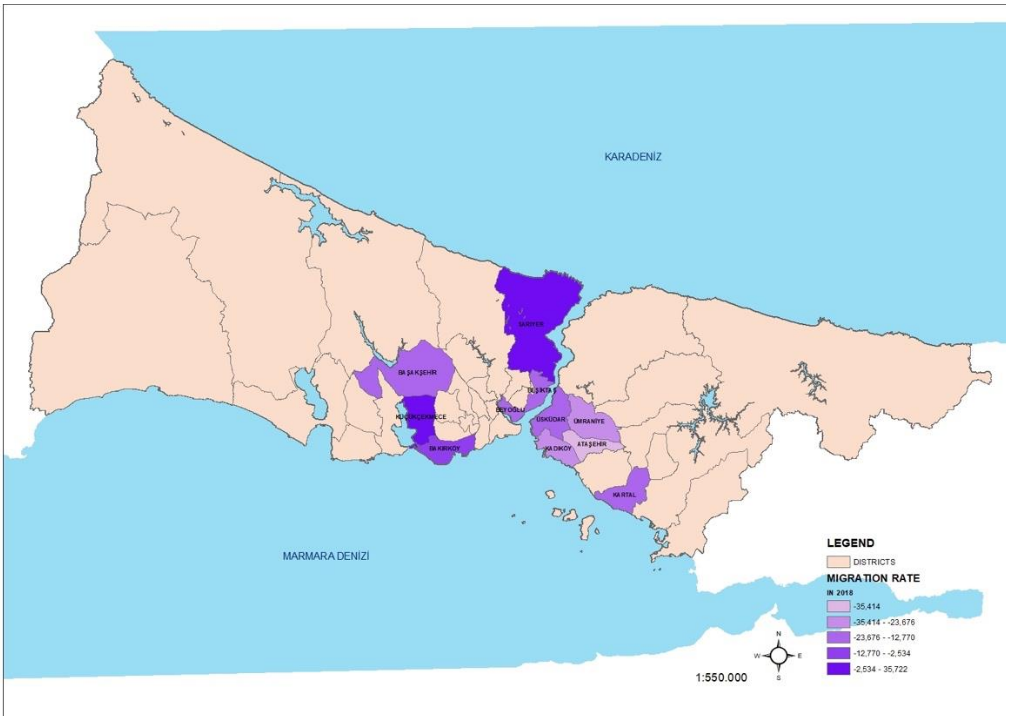


Figure 32: Map of Net Migration Rate in 2018

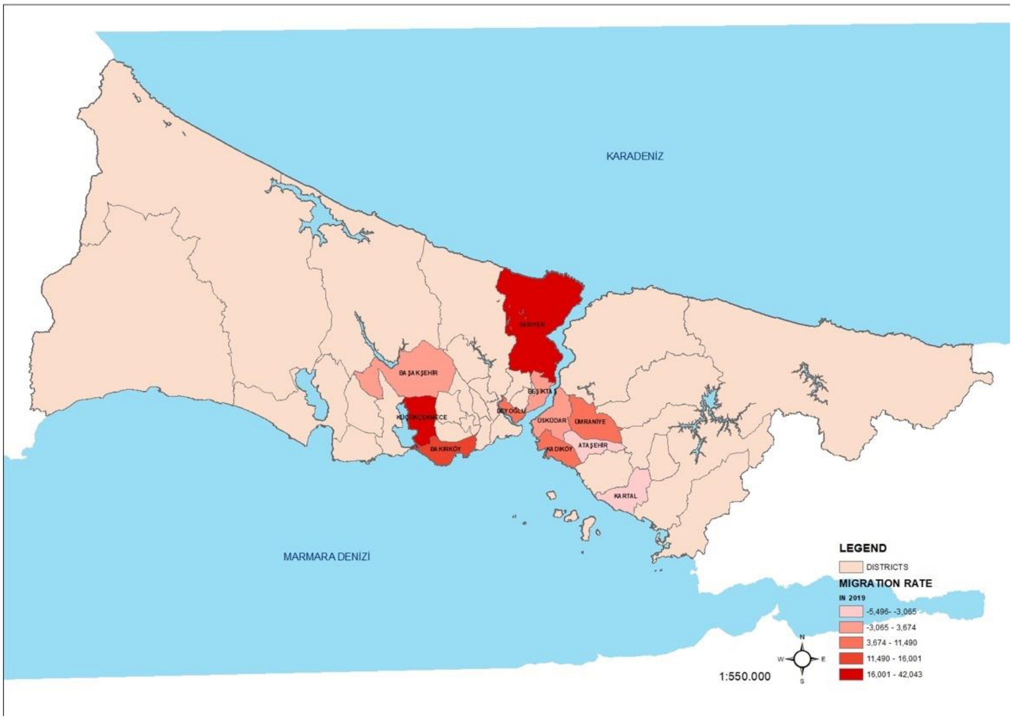


Figure 33: Map of Net Migration Rate in 2019

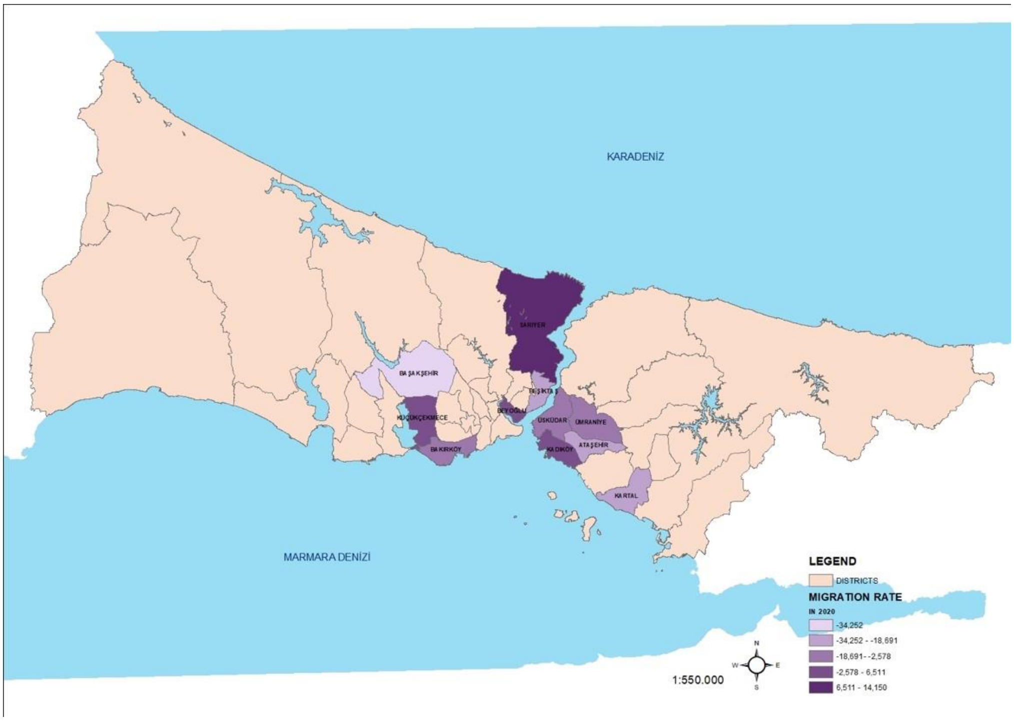


Figure 34: Map of Net Migration Rate in 2020

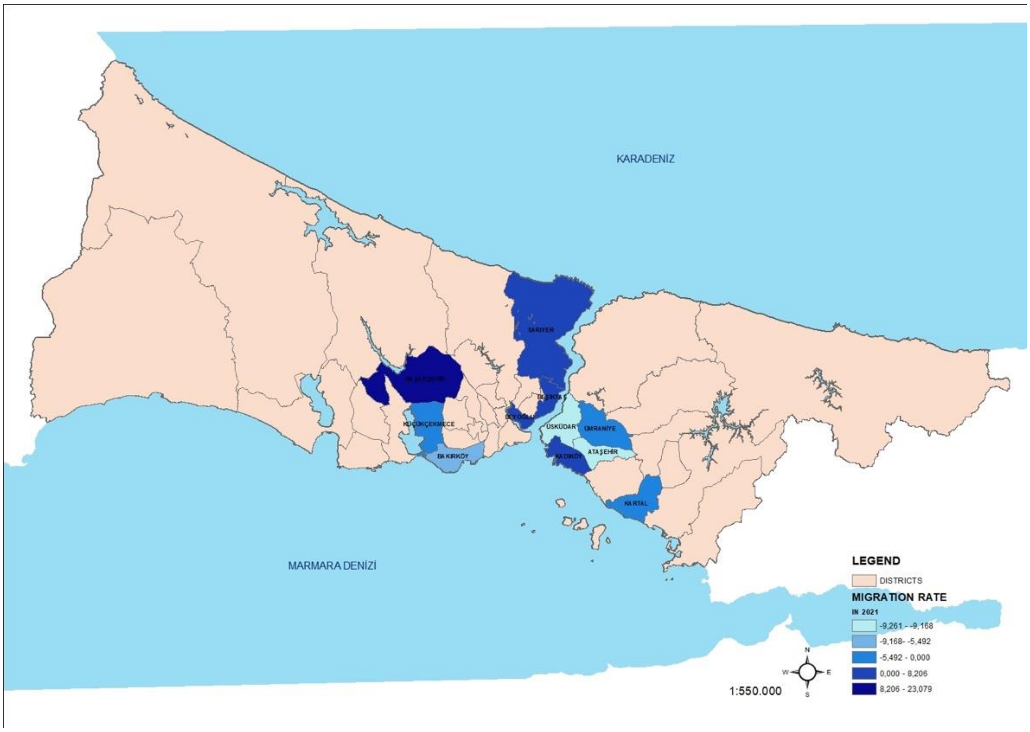


Figure 35: Map of Net Migration Rate in 2021

b. <u>Change in number of education level</u>

Change in number of education level dataset consist of six variables:

1. ALG 7: Number of immigrants have bachelor's degree.
2. ALG 8: Number of immigrants have master's degree.
3. ALG 9: Number of immigrants have doctor's degree.
4. VERG 1: Number of emigrants have no education degree.
5. VERG 2: Number of emigrants have elementary school degree.
6. VERG 3: Number of emigrants have middle school degree.

The graphics below that beginning from Figure 18 until Figure 28, show change in numbers of education level for each eleven districts which explains education level of the displacement:

Figure 36 shows that number of immigrants have bachelor’s degree, master’s degree, doctor’s degree have begun to increase significantly in Bakırköy since 2017.

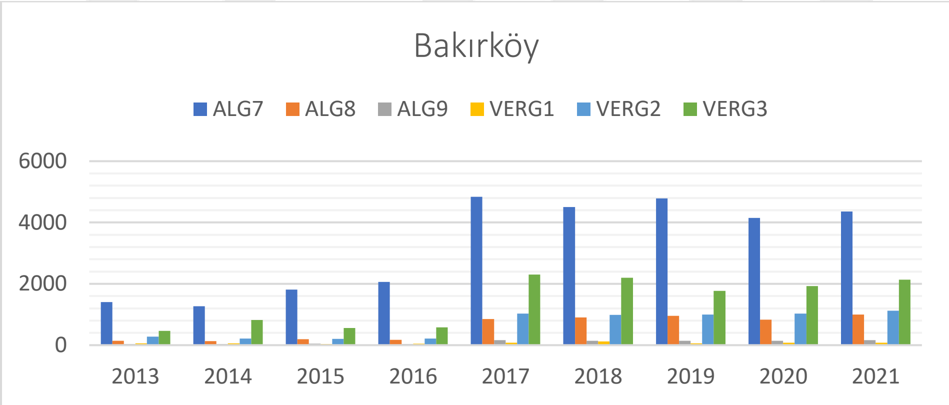


Figure 36: Change in Number of Education Level in Bakirkoy

Figure 37 shows that number of immigrants have bachelor’s degree, master’s degree, doctor’s degree have begun to increase significantly in Beşiktaş since 2017.

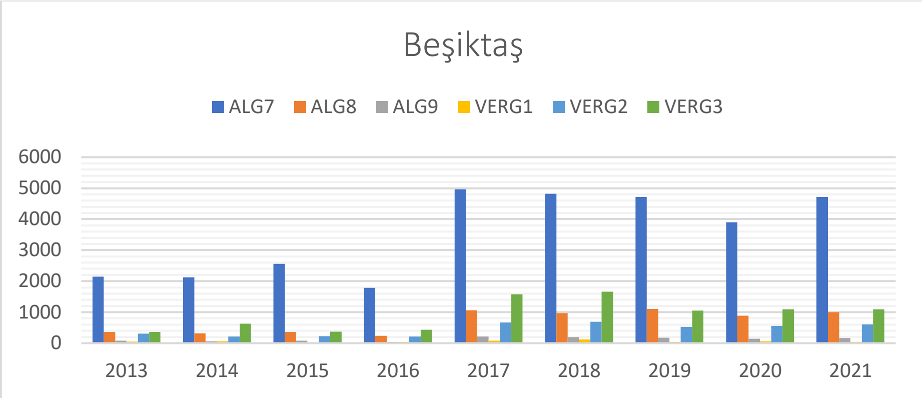


Figure 37: Change in Number of Education Level in Beşiktaş

Figure 38 shows that number of emigrants have middle school degree reached a pick point in 2018 and number of immigrants have bachelor's degree has begun to increase in Beyoğlu.

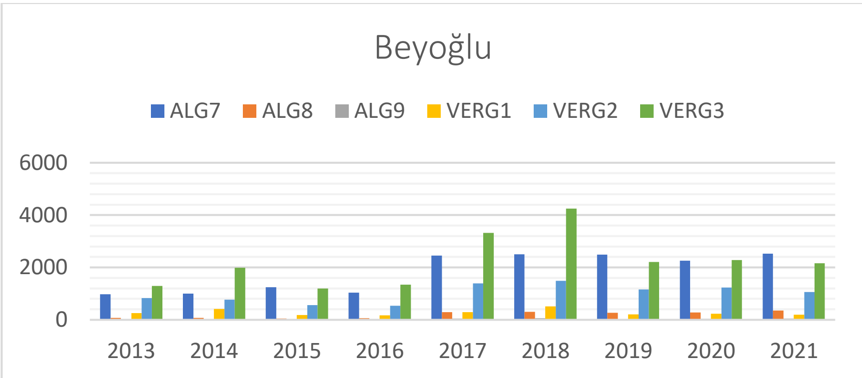


Figure 38: Change in Number of Education Level in Beyoğlu

Figure 39 shows that number of immigrants have bachelor's degree and master's degree reached a pick point in 2019 in Kadıköy while number of emigrants have middle school degree has begun to decrease.

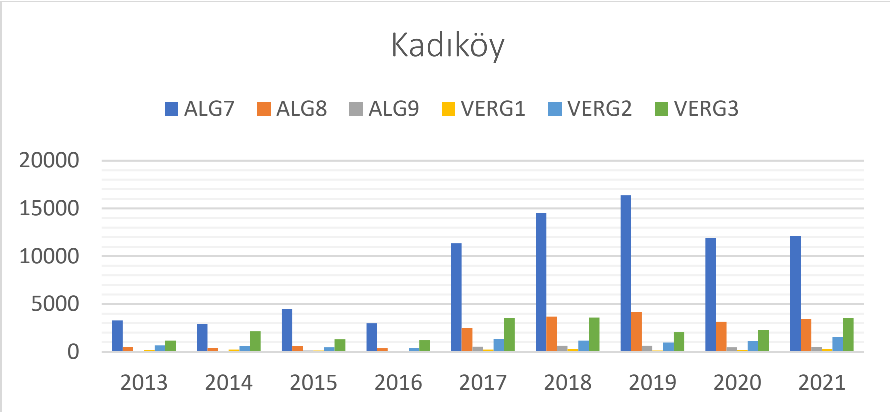


Figure 39: Change in Number of Education Level in Kadıköy

Figure 40 shows that number of immigrants have bachelor's degree, master's degree, doctor's degree have begun to increase significantly in Kartal since 2017.

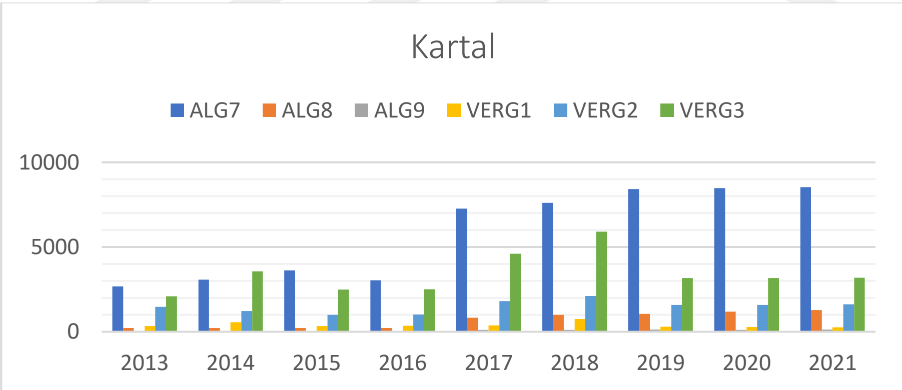


Figure 40: Change in Number of Education Level in Kartal

Figure 41 shows that number of immigrants have bachelor's degree, master's degree, doctor's degree have begun to increase significantly in Sarıyer since 2017.

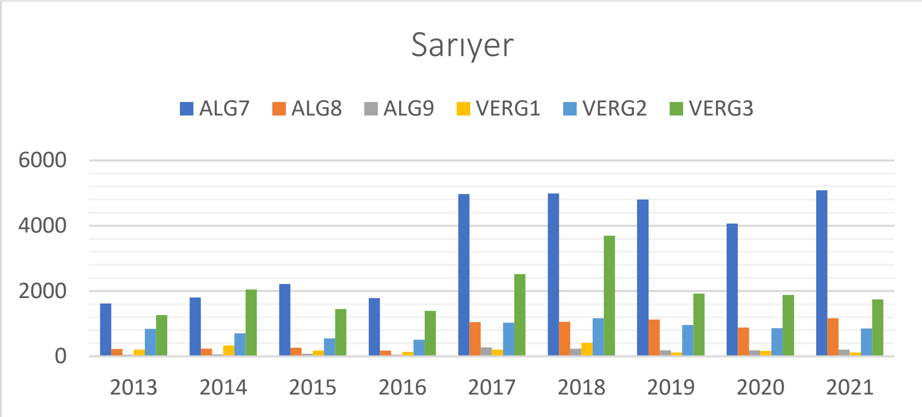


Figure 41: Change in Numbers of Education Level in Sarıyer

Figure 42 shows that the number of immigrants have bachelor's degree, master's degree, doctor's degree and the number of emigrants have middle school degree or less have begun to increase significantly in Üsküdar by 2017.

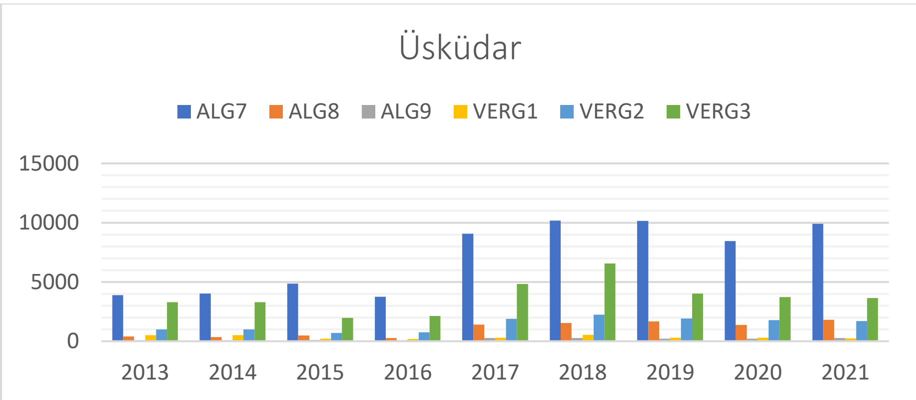


Figure 42: Change in Numbers of Education Level in Üsküdar

Figure 43 shows that the number of immigrants have bachelor's degree, master's degree, doctor's degree and the number of emigrants have middle school degree or less have begun to increase significantly in Küçükçekmece by 2017.

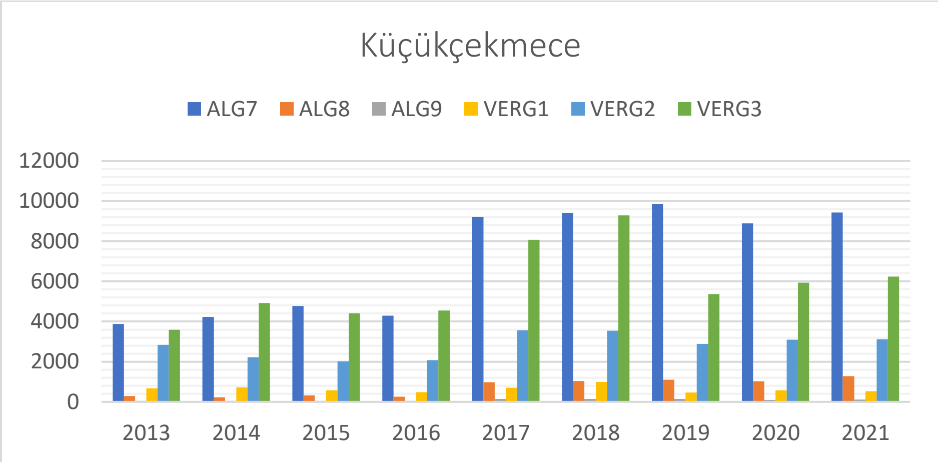


Figure 43: Change in Number of Education Level in Küçükçekmece

Figure 44 shows that the number of immigrants have bachelor's degree, master's degree, doctor's degree and the number of emigrants have middle school degree or less have begun to increase significantly in Ümraniye by 2017.

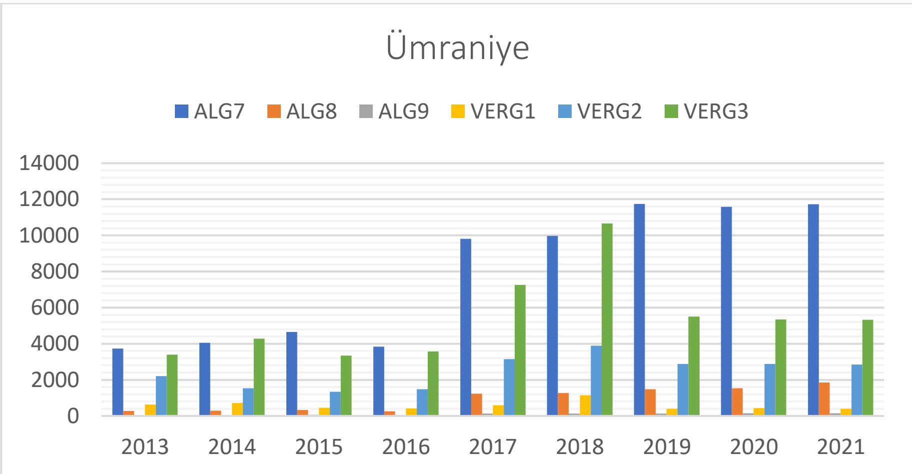


Figure 44: Change in number of education level in Ümraniye

Figure 45 shows that the number of immigrants have bachelor's degree, master's degree, doctor's degree and the number of emigrants have middle school degree or less have begun to increase significantly in Ataşehir by 2017.

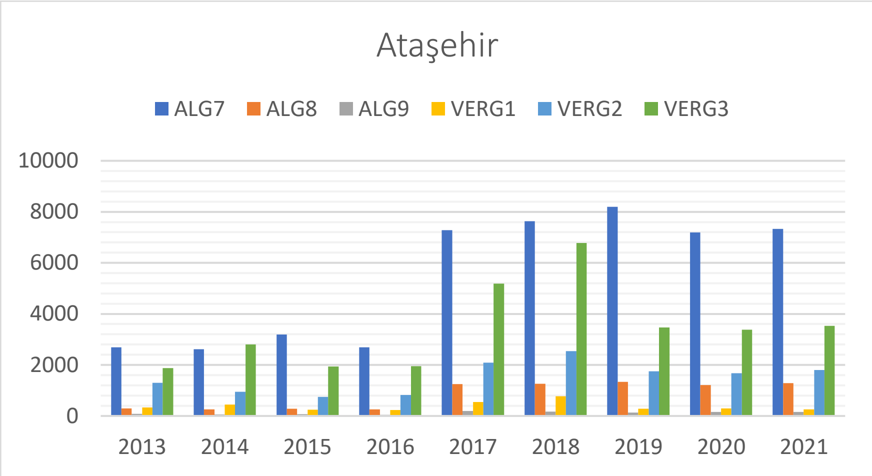


Figure 45:Change in number of education level in Ataşehir

Figure 46 shows that the number of immigrants have bachelor's degree, master's degree, doctor's degree and the number of emigrants have middle school degree or less have begun to increase significantly in Başakşehir by 2017.

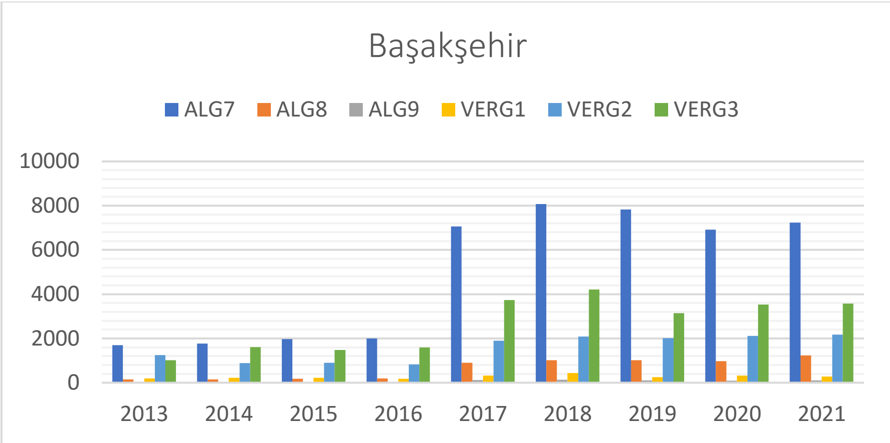


Figure 46: Change in Number of Education Level in Başakşehir

4.1.2 Point of Interest Data (Basarsoft)

The POI data is a broad data that includes all commercial or non-commercial place names or businesses. The dataset such contains 4 million POI data in nearly 500 categories. It has been selected data that can be related to gentrification such as counts of restaurants, retail stores, health centers-organizations, pharmacies, historical and touristic places, religious facilities, cultural places, entertainment places, settlement areas, official institutions, financial organizations, trade areas, industry and manufacturing areas, civil society organizations, parking areas, traffic lights, traffic radars, CCTVs, highway cameras, metro stations, train stations, taxi stops, terminals, marinas, sport facilities, base stations. It is nearly consisted of 17560 rows. This data set is used as independent variables to choose the most relevant features that might be a sign of gentrification.

Table 1: POI Data Table

Category	Sub-Category	Count
Shopping	Antique Shop	176524
	Big Stores	
	Outfit Stores	
	Gross Convenience Store	
	Book Store /Stationary	
	Organic Products Stores	
	Pet shop	
	High-end markets	
	Chain electronic Convenience Stores	
	Chain Convenience Stores	
	Chain Building Supply Stores	
	Deli Stores	
Railway	Terminals	1184
	Tramway stations	
	Train stations / Urban	
	Train stations / Interurban	
	Metro Stations	
Seaway	Sea Taxi	1323
	Ticket Offices	
	Port	
	Yacht Port	
	Quays	
	Quay (Sea Bus)	
	Quay (Car Ferry)	
	Quay (Urban Routes)	
Religious Facilities	Church	3261
	Synagogue	
	Others	
Education Facilities	Colleges	14022
	High Schools	
Entertainment	Bar	12498
	Meyhane	
	Fun Places / Other	
Financial Facilities	ATM	82879
	Branch bank	
Road	Gas Station	58200
	Taxi Stop	
	Red Traffic Sign Camera	
	City Surveillance Camreas	
	Parking Lots (Outdoor-Public)	
	Parking Lots (Outdoor-Private)	
	Parking Lost (Indoor)	
	Parking Areas	
	Radar	
	Traffic Lights	

	Transportation Structures	
Housing Areas	Residence	566023
	Villa	
	Apartments Blocks	
	Apart Residential Blocks	
Cultural Facilities	Museum	2182
	Movie Theater	
	Theaters	
Government Agencies	Courthouse	16446
	Ministry Offices	
	Ministry Units	
	Municipality buildings	
	Municipality Units	
	Police Department	
	Government Hosue	
	Notary	
	Post Office	
	Telecommunication Office	
	Fire Department	
Industrial Areas	Factory	17100
Health Facilities	Dental healthcare center	34073
	Dental Clinic	
	Private Clinic	
	Optic Shop	
	Psychologist office	
	Spa center	
	Vet	
Non-Governmental Organizations	Association/Foundation	33609
Sport facilities	Tennis Court	572
Historical and Touristic Facilities	Ancient Cities	22823
	Monuments	
	Apart hotel	
	Balloon tour	
	Turkish bath	
	Castle	
	Camping areas	
	Hot spa	
	Guesthouses/ rest facilities	
	Hotel	
	Lodgings	
	Beach (free)	
	Beach	
	Palaces and pavilions	
	Protected areas	
	Historical artifacts	
	Touristic information offices	
Agricultural Production Areas	Farm	677
Telecommunication	Transmitting stations	7256

	Law Firm	
	Bike Renting	
	Bike/Motorbike Shops	
	Airline Companies	
	Hairdressers	
	Dry Cleaning Shops	
	Laser Epilation Centers	
	Architect Offices	
Trade Areas	Engineering Offices	218797
	rent a car firms	
	Auto galleries	
	Auto authorized stores	
	Translation office	
	Tourism firms	
	Catering firms	
	Bazaar and arcades	
	Business center blocks	
	Ice Cream shops	
	Doner-kebab shop	
	Burrito kebab shop	
	Fast food	
	Bakery	
	Coffeeshops	
	Kıraathane	
	Restaurant/ Sea food	
	Restaurant / Home food	
	Restaurant / Indian	
	Restaurant/Kebab	
	Restaurant/Kofte- Barbeque	
Refreshments Areas	Restaurant /Mexican	296127
	Restaurant/Russian	
	Restaurant/Sakatat	
	Restaurant/Vegan	
	Restaurant/Local Foods	
	Restaurant / Chinese Japanese	
	Restaurant/Italian	
	Pizza Shop	
	Simit Bakery	
	Dessert Shop	
	Tea Garden Cafe	
	Çiğ Köfte Shop	
	Soup Shop	
	Lake	
Prominent Points	Places with view	3556
	Locality	

4.2 Data Preparation

First, two different datasets merged by referencing to "DISTRICT_CODE" column as it shows below:

Table 2: Merged Table

DISTRICT_CODE	SUB_CATEGORY	VERSION	COUNT	NET_MIG_RATE	ALG7	ALG8	ALG9	VERG1	VERG2	VERG3
1166	Theatre	2013	2	2,7326	1410	140	41	58	283	465
1852	…	…	…	…	…	…	…	…	…	…
1449	…	…	…	…	…	…	…	…	…	…

In order to run ML algorithms, the features of datasets must be numeric values. Categorical data with qualitative characteristics need to be digitized. Nominal data includes classified data that do not have numerical characteristics and are not superior to each other. There are several methods to digitize nominal data. Ordinal Encoder, Label Encoder and One Hot Encoder are the most used methods. Our dataset has two nominal data: DISTRICT_CODE, SUB_CATEGORY. Although DISTRICT_CODE has a numerical value, it has no numerical superiority other than the district identification. Since it has no superiority or dependency, it is necessary to transform the data in the rows into column headers with One Hot Encoder method.

Therefore, the dataset table as shown in Table 2 transforms as below:

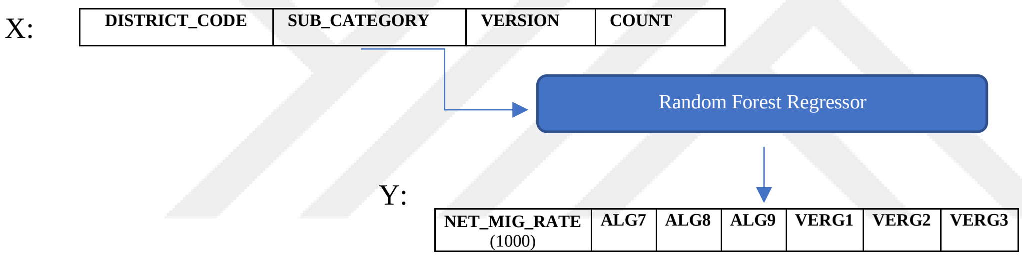
Table 3: Table after OneHotEncoder

1166	1852	1449	SUB_CATEGORY	VERSION	COUNT	NET_MIG_RATE	ALG7	ALG8	ALG9	VERG1	VERG2	VERG3
1	0	0	Theatre	2013	2	2,7326	1410	140	41	58	283	465
0	1	0	…	…	…	…	…	…	…	…	…	…

After the data is digitized, input data (X) table is created from a total of 139 columns including year, count, eleven districts and 126 sub_category columns.

According to column headings, we have 9711 input data as row. Totally, a training input matrix (X) of 9711*139 is obtained.

Output data consists of 7 columns is added to the code for estimation for the 9711 data rows we have. The Y is a matrix 9711*7 in size.



4.3 Train/Test Split

In order to run ML algorithms, the dataset must be randomly split for training and testing. For training 70 % of the total 9711 data is used for training and 30 % is used for the prediction test of the training.

Eventually, "X_train" matrix is created from 6797 data rows and "X_test" matrix from 2914 data rows.

4.4 Model Evaluation

The estimation success of the model developed by Random Forest Method according to our training set (X_Train,y_train) is achived by running the test data (X_test, y_test). According to developed model the success of estimation for each district separately and all is observed as 99%. The first 10-12 features according to results for each district are as follows:

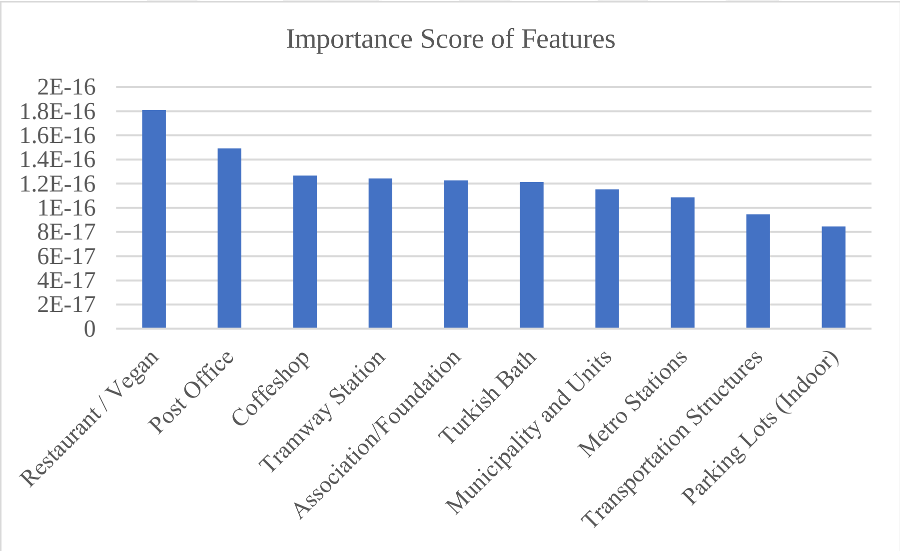


Figure 47: Features for Bakırköy

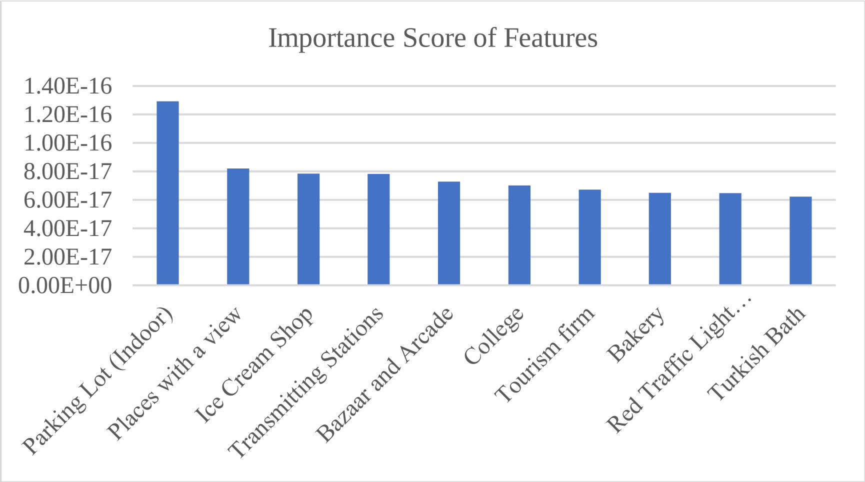


Figure 48: Features for Beşiktaş

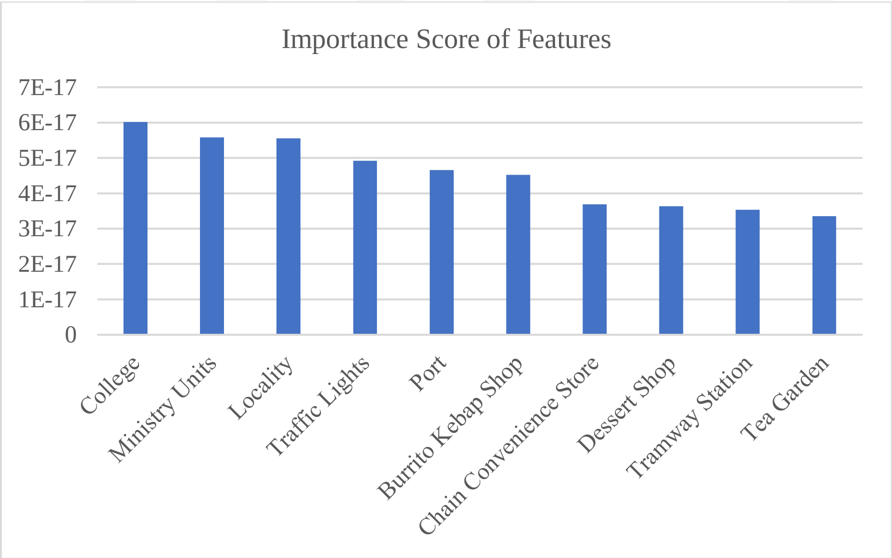


Figure 49: Features for Beyoğlu

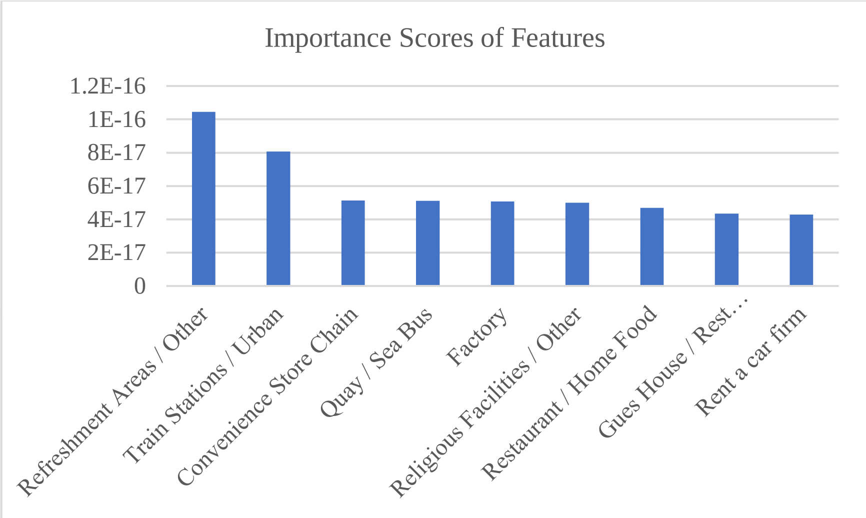


Figure 50: Features for Kadıköy

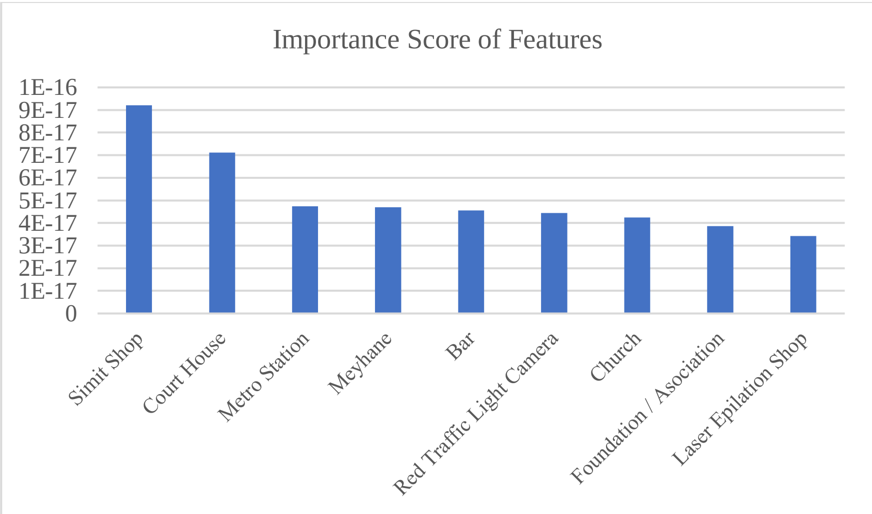


Figure 51: Features for Kartal

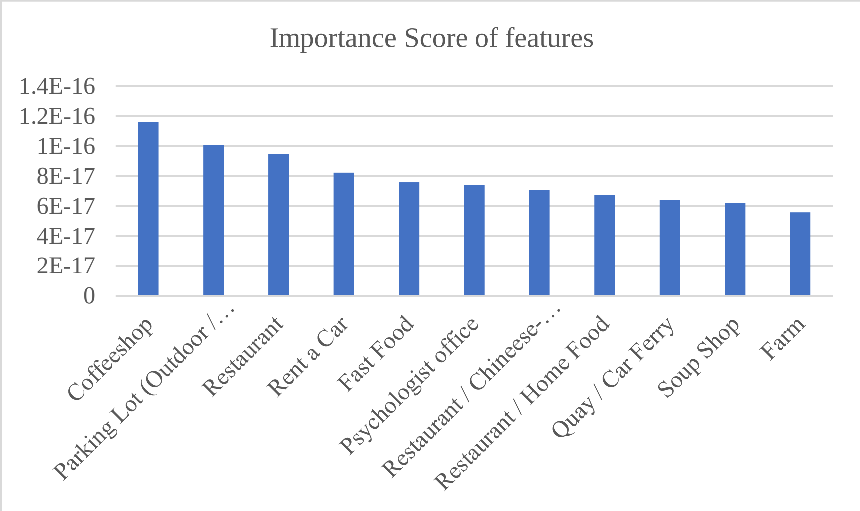


Figure 52: Features for Sarıyer

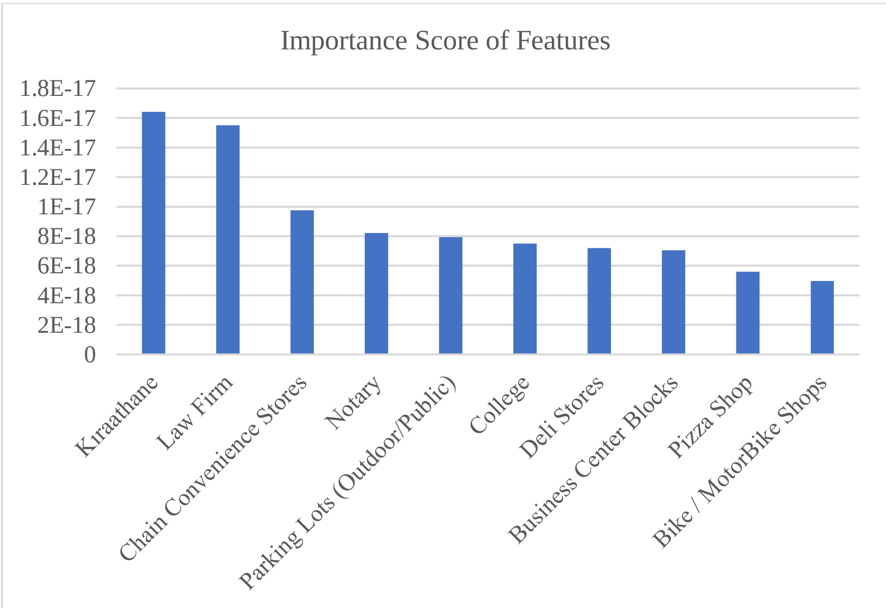


Figure 53: Features for Üsküdar

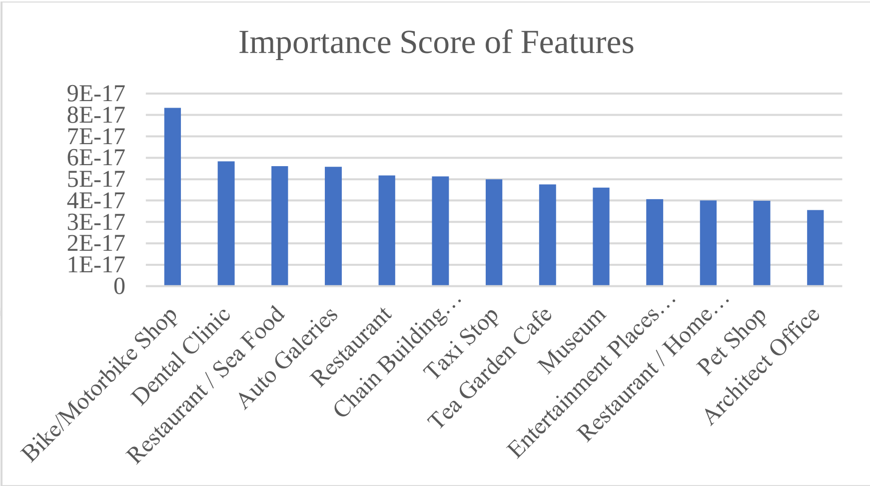


Figure 54: Features for Küçükçekmece

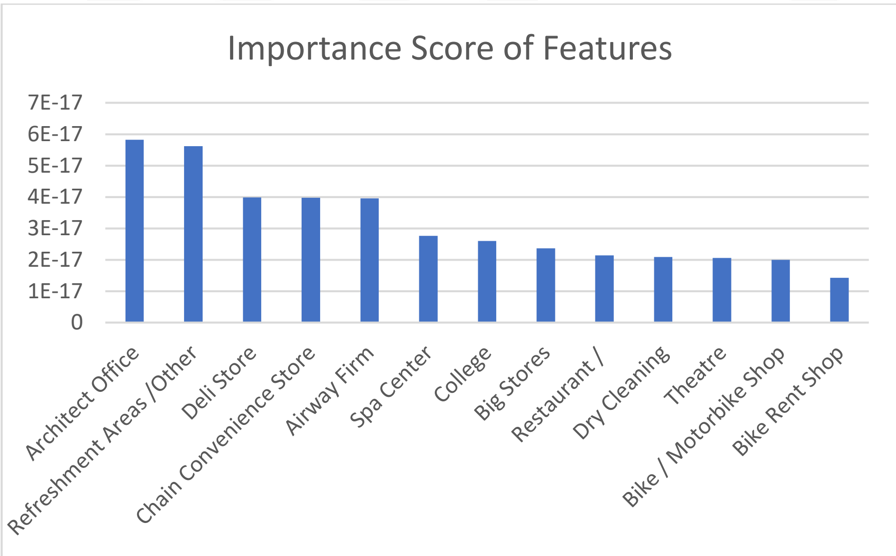


Figure 55: Features for Umraniye

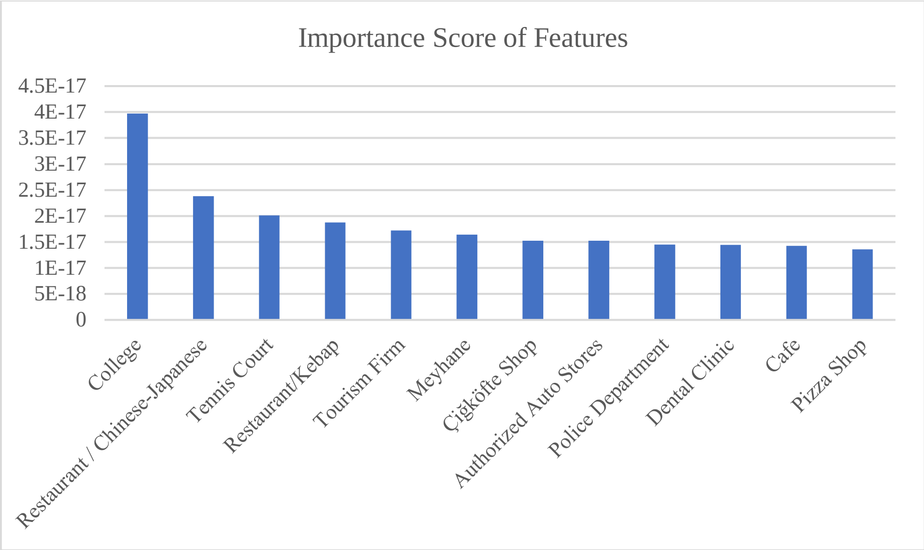


Figure 56: Features for Ataşehir

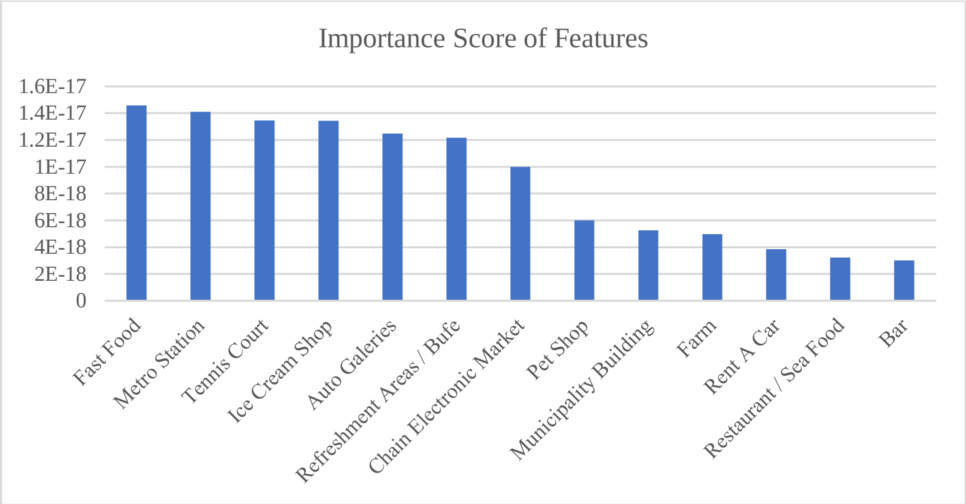


Figure 57: Features for Başakşehir

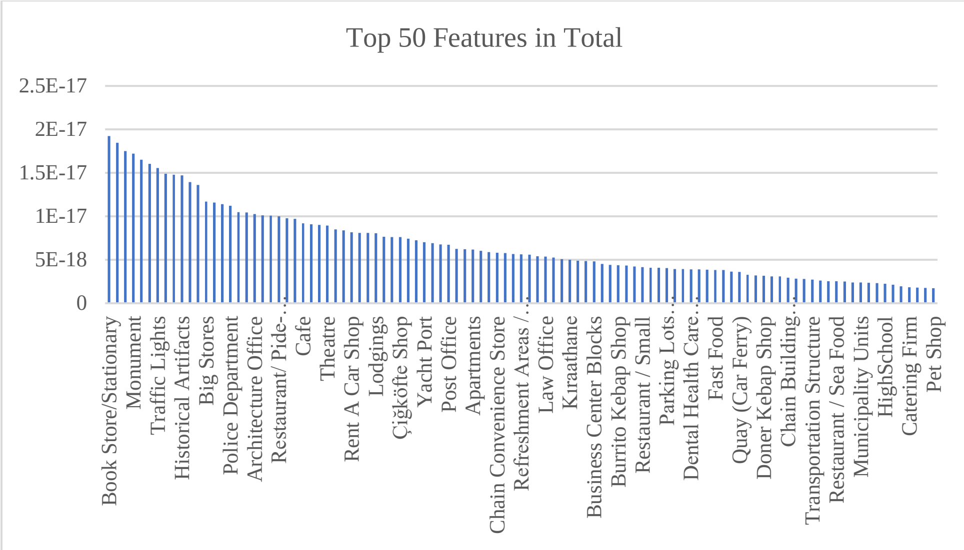


Figure 58: Top 50 Features in Total

4.5 Interpretations

Based on the analysis and the objectives of this research, the following recommendations are proposed:

1. Figures 29, 30, 31, 32, 33, 34, 35, 36, 37 show that different features have come into prominence for each district. For instance, selecting vegan restaurant, associations- foundations, and indoor parking lots as features in Bakırköy will maximize the accuracy of the prediction model.
2. As a result of the analysis in which all districts were included (Figure 40), bookstore / stationery, monuments, police departments, architect offices, cafes, theatres, rent a car shops, bufes, law offices, burrito

kebab shops, outdoor (public) parking lots, dental healthcare centers, sea food restaurants, municipality units, catering firms, pet shops have come into prominence. It can be said that selecting these features will maximize the accuracy of the prediction model for all districts.

5. CONCLUSIONS

5.1 Summary

In this research a regressor model with 99 % accuracy is developed to predict and order importance of the variables of the datasets. Gentrification is defined by census track dataset which involves net migration rate (Figures 9, 10, 11, 12, 13, 14, 15, 16, 17) and change in numbers of education levels (Figures 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28) by years since the character of displacement shows footprints of gentrification in neighborhoods. Since urban areas with high rent – gaining has become attraction point for investments and public services as a result, we used point of interest data as shown in Table 1 to select variables which will maximize the prediction model for future.

5.2 Future Studies

Urban planning is a public service that is based on benefit of societies. It must ensure a safe, healthy and sustainable urban environment for all income groups of societies. Urban planning systematically explains, predicts and scripts economic, demographical, socio-cultural, environmental dynamics and causes of spatial formation through a scientific criterion. The main function of planning is to set goals for the city, effect spatial processes, drive changes, and define tools. It is a teamwork involving different professions and actors. Although the role of planners is evolved by improving the planning system and politics, the distinctive and constant difference of planners is that they focus on settlement, the benefit of societies and the future.

Elimination of urban poverty and inequities is one of the principles of sustainable urbanization and settlement. Gentrifying neighborhoods show that planning processes need a well-timed and preventative approach. Predicting gentrification is the most appropriate step towards developing accurate and timely policies. This research can be improved as a further study by developing a model to predict gentrification by using the features which have been evaluated as a result of the feature selection model.

BIBLIOGRAPHY

Ergun, Nilgun. (2004). *Gentrification in Istanbul*. Cities. 21. 391-405. 10.1016/j.cities.2004.07.004.

(Journal Article, 1 Author)

Özbay Daş, Z, Özşahin, G. (2021) *An empirical analysis of gentrification in Istanbul*. *Reg Sci Policy*

Pract. 2021; 13: 501– 526. https://doi.org/10.1111/rsp3.12283

Coskun, N., & Yalcin, S. (2007). *Gentrification in a globalizing world, Case Study: Istanbul.* ENHR International Conference, Rotterdam. (Journal Article, 2 Authors)

İslam, T. (2005) *"Outside the Core: Gentrification in Istanbul." In: Gentrification in a Global Context,*

edited by R. Atkinson and G. Bridge, pp. 121–136. London: Routledge.

Turkun, A., Aslan, Ş., & Erdoğan, E. K. (2014). *Mülk, Mahal: İnsan İstanbul'da Kentsel Dönüşüm.*

Turkun, A. (2015). *Asuman Turkun- housing for urban poor. Housing the Urban Poor Policy and*

Practice in Developing Countries. (Journal Article, 1 Author)

Seattle OPCD, https://www.seattle.gov/opcd/ongoing-initiatives/seattle-2035-comprehensive-plan

Glass, R. (1964). *London: Aspects of Change.* London: MacGibbon & Kee.

Smith, N. (2006). *Gentrification Generalized: From Local Anomaly to "Urban Regeneration" as Global*

Urban Strategy, Frontiers of Capital: Ethnographic Reflections on The New Economy. pp192

Reades J., De Souza J., Hubbard P. (2019), *Understanding urban gentrification through machine learning*.

Urban Studies, do: 10.1177/0042098018789054.

https://data.tuik.gov.tr/

https://www.basarsoft.com.tr/

초록

젠트리피케이션은 대체로 인근 지역의 경제적 가치를 증가시킨다. 이는 해당 지역에 사회적 스트레스를 발생시키고 구매력 감소로 인한 원주민들의 이주를 야기한다. 이스탄불의 거의 모든 지역이 젠트리피케이션의 영향을 받기 시작했다. 이 연구는 이스탄불에서 일어나는 젠트리피케이션 사례들의 부정적인 영향을 살피고, 이를 완화하는 방법을 논의한다.

최근 머신러닝 분야의 연구를 통해 복잡한 도시 문제가 앞으로 어디서 어떻게 발생할지를 예측할 수 있게 되었다. 이 연구는 머신러닝 연구의 핵심 문제이자 필수적인 부분인 특성선택(feature selection)에 초점을 둔다. 특성선택은 데이터 전처리 방법으로, 특히 고차원 데이터를 준비하는 데 효과적이다. 특성선택은 깨끗하고 이해하기 쉬운 데이터를 준비함으로써 고차원 데이터를 통한 학습을 가능하게 하고, 더욱 포괄적인 머신러닝 모델을 구축해 모델의 성능을 향상시키는 것을 목적으로 한다.

데이터 세트는 아타셰히르(Ataşehir), 베욜루(Beyoğlu), 바크르쾨이(Bakırköy), 바샥셰히르(Başakşehir), 베식타슈(Beşiktaş), 카드쾨이(Kadıköy), 카르탈(Kartal), 퀴췩체크메제(Küçükçekmece), 사르예르(Sarıyer), 윔라니예(Ümraniye),

위스퀴다르(Üsküdar) 11 개 지역에 대해 수집한다. 젠트리피케이션의 측정은 순이동률과 교육 수준 변화로 구성된 인구조사 트랙 데이터를 통해 수행한다. 젠트리피케이션을 가장 잘 식별하는 기능을 선택할 때는 젠트리피케이션과 경제적 가치 증가 사이의 유발 관계를 고려하여 약 17,000 개의 행으로 구성된 관심 지점 데이터 집합을 사용한다.

결론에서는 POI 데이터 세트의 특징에 점수를 매겨 가장 중요한 특징을 식별하는 머신 러닝 모델의 제공을 통해, 젠트리피케이션을 예측하기 위한 추가 연구에 강력한 아이디어를 제공한다. 그리고 젠트리피케이션에 대처할 수 있는 정책의 영향을 제시한다.

주제어: 젠트리피케이션, 특성선택, 머신러닝, 이스탄불, 예측

ACKNOWLEDGEMENT

This thesis writing process would not have been accomplished without the support and cooperation of several individuals and organizations. It is my pleasure to present my sincere gratitude towards them. I would like to express my special thanks to my mentor Mr. Lee for his time he provided.

I would also like to thank Selma Yakut, Maya Arıkanlı, for their invaluable advice, feedbacks and supports.

I am grateful to my manager Fatih Dogan and colleagues in Urban Planning Department for their precious supports.

Finally, I would like to thank my mother for her unflagging support in my whole academic life. This study would not have been successful without their cooperation and inputs.

“This thesis was financially supported by Seoul Metropolitan Government under the International Cooperation Fund-ODA.”