Analysis of Turkey's Demographic and Social Structure Using PCA And K-Means Clustering

Özel SEBETCİ[1], & Doruk AYBERKİN[2]

Abstract

This study aims to investigate the demographic and socioeconomic inequalities in the provinces by taking advantage of Turkey's unique location and complex society and comprehensively analyzing the demographic and socioeconomic structure of the provinces in Turkey. A multidimensional data set containing the data of 81 provinces between 2010 and 2022, provided as open data by TUIK (Turkish Statistical Institute) for the purpose, uses advanced data such as Exploratory Data Analysis (EDA), Principal Component Analysis (PCA) and K-Means clustering. Were examined using data analysis techniques. The study discussed fundamental socioeconomic indicators such as internal migration, population growth, foreign population density, and divorce rates. The analysis results reveal significant differences between cities regarding migration patterns, population growth rate, and socioeconomic characteristics. PCA and cluster analysis provided a more detailed understanding of these differences by grouping towns according to their demographic and socioeconomic profiles. The study provides essential information regarding Turkey's population dynamics and social fabric, developing effective local and national policies and strengthening social cohesion. Looking at the results, it was revealed that some precautions should be taken. Particular attention should be given to urban planning, education, health, and social services, considering the multidimensional structure of demographic and socioeconomic profiles. In regions with intense migration flows, priority should be given to strengthening social services, increasing economic opportunities, and developing integration policies. The population growth rate is low or negative. Measures to encourage population growth and strategies to support economic development should be implemented in the regions where it is located.

Anahtar Kelimeler: Demographic Analysis, Socioeconomic Differences, Migration Patterns, Statistical Methods, Exploratory Data Analysis(EDA)

Türkiye'nin Demografik ve Sosyal Yapısının PCA ve K-Means Kümeleme Yöntemiyle Analizi

Öz

Bu çalışmanın amacı, Türkiye'nin benzersiz konumundan ve karmaşık toplumundan yararlanıp, Türkiye'deki illerin demografik ve sosyoekonomik yapısını kapsamlı bir şekilde analiz ederek illerdeki demografik ve sosyoekonomik eşitsizlikleri araştırmaktır. Amaca uygun olarak TÜİK (Türkiye İstatistik Kurumu) tarafından açık veri olarak sağlanan 2010-2022 yılları arasında 81 ilin verilerini içeren çok boyutlu bir veri seti, Keşif Amaçlı Veri Analizi (EDA), Temel Bileşen Analizi (PCA) ve K-Ortalamalar kümelemesi gibi ileri veri analizi teknikleri kullanılarak incelenmiştir. Çalışmada, iç göç, nüfus artışı, yabancı nüfus yoğunluğu ve boşanma oranları gibi temel sosyoekonomik göstergeler ele alınmıştır. Analiz sonuçları, şehirler arasında göç kalıpları, nüfus artış hızı ve sosyoekonomik özellikler açısından önemli farklılıklar olduğunu ortaya koymaktadır. PCA ve kümeleme analizi, şehirleri demografik ve sosyoekonomik profillerine göre gruplandırarak bu farklılıkları daha detaylı bir şekilde anlamayı sağlamıştır. Çalışma, Türkiye'nin nüfus dinamikleri ve sosyal dokusunun anlaşılması, etkili yerel ve ulusal politikaların geliştirilmesi ve sosyal uyumun güçlendirilmesi açısından önemli bilgiler sunmaktadır. Sonuçlara bakılarak bazı önlemlerin alınmasının gerekliliği ortaya konulmuştur. Demografik ve sosyoekonomik profillerin çok boyutlu yapısı dikkate alınarak kentsel planlama, eğitim, sağlık ve sosyal hizmetler gibi alanlara özel önem verilmelidir.Göç akışının yoğun olduğu bölgelerde, sosyal hizmetlerin güçlendirilmesine, ekonomik fırsatların artırılmasına ve entegrasyon politikalarının geliştirilmesine öncelik vermelidir.Nüfus artış hızının düşük veya negatif olduğu bölgelerde nüfus artışını teşvik edecek önlemler ve ekonomik kalkınmayı destekleyici stratejiler uygulanmalıdır.

Keywords: Demografik Analiz, Sosyoekonomik Farklılıklar, Göç Modelleri, İstatistik Metotlar, Keşif Amaçlı Veri Analizi(EDA)

[1] Doç. Dr. – Adnan Menderes University, Computer Technologies Department, Aydın, Türkiye, osebetci@adu.edu.tr,
ORCID: 0000-0002-2996-0270

[2] Asst. Prof. Dr. – Bayburt University, Management Information Systems Department, Bayburt, Türkiye, doruk@bayburt.edu.tr,
ORCID: 0000-0003-3409-8926

Introduction

The ongoing acceleration of worldwide globalization and the consequent changes in economic conditions make the understanding of the demographic and socio-economic profiles of cities highly important. Turkey, with its complex demographic structure affected by its geopolitical location and socio-economic dynamics, provides a rich study area for such an examination. Understanding the demographic and social structure of Turkey is of great importance in shaping effective policies related to demography and social areas in Turkey, as well as in strengthening social cohesion.

Demographic structures represent some of the most dynamic dimensions of socio-economic transformation, reflecting fertility, mortality, migration, and educational trends. In this context, Turkey provides a compelling case: a middle-income country undergoing rapid urbanization, educational expansion, and migration-driven diversification over the past two decades. Although numerous studies have analyzed Turkish demographics using descriptive or econometric frameworks (Yıldırım & Öcal, 2014; Filiztekin, 2020), empirical research integrating multivariate statistical learning to expose spatial heterogeneity remains limited.

The geopolitical position of Turkey, combined with the coexistence of various cultures and rapid urbanization, requires a nuanced understanding of the country's population dynamics and social fabric. Analyzing demographic trends such as internal migration rates, age distribution, and ethnic diversity will provide crucial information that can be used in public activities aimed at overcoming social challenges.

This study contributes to the demographic analytics literature by combining Principal Component Analysis (PCA) and K-Means clustering to identify province-level socio-demographic typologies across Turkey from 2010 to 2022. PCA reduces the dimensional complexity of correlated indicators—such as fertility, mortality, education, urbanization, and migration—into a few interpretable components (Jolliffe&Cadima, 2016). Subsequently, K-Means classifies provinces into homogeneous clusters that exhibit distinct demographic signatures.

To comprehensively understand the complex interaction of demographic and social factors in Turkey, it is imperative to use techniques such as Principal Component Analysis (PCA) and clustering. Principal Component Analysis allows researchers to condense large datasets into a smaller set of variables while preserving the features they consider important in the dataset. In the context of Turkey's demographic and social structure, PCA can facilitate a more detailed interpretation of data by identifying correlations among various variables that needs to be analyzed.

Additionally, clustering techniques allow for the grouping of similar data points based on specific characteristics, uncovering hidden relationships, and revealing different population clusters in Turkey. This study can utilize PCA and clustering to uncover complex data groups, determine how demographic and social characteristics interact with each other and provide detailed information on Turkey's population structure. These techniques are effective in deriving meaningful insights from complex data sets, guiding evidence-based policy formulation, and providing information for targeted interventions to address various challenges faced by the Turkish society.

Methodologically, this dual approach builds upon European regional typology studies (Marques et al., 2020) and East Asian demographic clustering (Wang&Zhao, 2020), adapting their insights to an emerging-economy context. The integration of these unsupervised-learning techniques enhances the explanatory potential of regional demographic analysis and complements classical theories such as the Demographic Transition Theory (Notestein, 1945; Kirk, 1996).

The focus of this research is to conduct a comprehensive analysis of the demographic and socio-economic characteristics of Turkey's cities. Turkey has a complex demographic structure influenced by its geopolitical location, economic, and social dynamics. This study aims to present a detailed analysis of various parameters such as internal migration mobility, immigration, population growth rates, foreign population density, and divorce rates on a city basis, examining these dynamics.

In a world where globalization and local economic conditions are rapidly changing, understanding the demographic and socio-economic profiles of cities is crucial for policymakers and planners. This analysis aims to provide the essential information needed to better understand the challenges and opportunities faced by cities, thereby developing more effective local and national policies.

Another aim of this study is to demonstrate the applicability and effectiveness of various statistical and machine learning techniques commonly used in the analysis of large datasets. Among the methods used are correlation analysis, clustering, principal component analysis (PCA), and visualization techniques. These techniques aid in a better understanding of the dataset and help in uncovering significant trends and patterns.

In conclusion, this research aims to provide an in-depth view of Turkey's demographic and socio-economic structure and to show how advanced data analysis techniques can be effectively used in such an analysis.

Literature Review

Demographic and socio-economic analyses have long held significant importance in the fields of urban planning, policy-making, and social sciences. These studies are critical for understanding the development dynamics of cities and regions and predicting future trends. Research on Turkey's urban and demographic structure reflects the unique challenges and opportunities the country faces in both regional and international contexts.

The demographic and social changes observed in Türkiye can be explained by the Demographic Transition Theory (Notestein, 1945; Kirk, 1996), which posits a transition from high fertility and mortality rates to lower fertility and mortality rates, paralleling increasing rates of urbanization and industrialization in societies. Although similar demographic structures are observed in Southern Europe (Reher, 1998) and East Asia (Lee&Mason, 2014), value differences in migration movements and patterns are particularly prominent in Turkey.

In recent years, Turkey's rapid urban growth and demographic change have attracted the attention of many researchers. Specifically, studies like Göçmen and Ekmekçi (2015) have thoroughly examined the economic and social impacts of internal migration mobility. These studies have revealed that migration is dependent on factors such as economic opportunities, educational facilities, and quality of life (Öncel, H., & Levend, S., 2023).

Regarding immigration, Turkey's strategic location has made it a significant center for immigration, especially in the last decade. In this context, research such as Akçapar and Şimşek (2016) has explored the socio-economic impacts of immigration and its effects on the demographic structure (Karataş, A., 2023).

Studies on population growth rates and divorce statistics are also noteworthy. For instance, Karabatak & Sarı (2019) have investigated how Turkey's population growth rates can be related to economic growth, education level, and access to health services. Regarding divorce rates, the study by Kaya and Çolak (2020) has meticulously analyzed the effects of socio-economic factors and cultural changes on divorce rates (Thornes, B., & Collard, J., 2023).

A general assessment of this literature indicates that demographic and socio-economic trends paint a complex picture and could have profound effects on the future of cities and regions. However, the current literature reveals gaps, particularly in the use of multidimensional datasets and the application of advanced data analysis techniques. This study aims to fill these gaps and provide a more comprehensive perspective on Türkiye's demographic and socio-economic structure by integrating existing theories and providing a comparative understanding of Türkiye's position in global demographic transformations and their effects on social policies.

Dataset and Method

Dataset

This study is conducted on a dataset that includes demographic and socio-economic data of Turkey's 81 provinces. The dataset is provided by [TUIK] as archival data and covers the years 2010-2022. The main variables in the dataset are: internal migration inflow, internal migration outflow, immigration inflow, immigration outflow, total foreign population, population per square kilometer, population growth rate, and the number of divorces. These variables have been used to analyze the demographic and socio-economic profiles of cities.

Data Collection Tools

The analysis process began with the dataset undergoing a standard scaling operation (Feature Scaling). This procedure was performed to normalize the effects of variables measured on different scales. Subsequently, analysis techniques such as K-Means and PCA were applied, and the results were presented using various visualization tools. These visualization tools include scatter plots, bar charts, and heat maps.

In this study, a mixed dataset encompassing fertility, mortality, mobility, and education patterns, combining demographic and socioeconomic dimensions, was used to represent the multidimensional structure of Türkiye's population dynamics. The dataset, provided by the Turkish Statistical Institute [TUIK], includes eight standardized variables measured annually between 2010 and 2022. Missing values were handled using linear interpolation methods consistent with demographic trend studies (Little and Rubin, 2019). To enable comparisons between metrics measured in different units, all continuous variables were standardized using z-scores. Although various techniques are available, PCA and K-Means were used due to their efficiency in processing and interpreting medium-sized datasets with mixed demographic indicators. K-Means provides clearer clustering structures in socio-demographic contexts (Jolliffe & Cadima, 2016; Xu & Tian, 2015).

Exploratory Data Analysis (EDA)

The EDA process is crucial for the initial examination and understanding of the dataset (Tukey, 1977). At this stage, the basic statistical properties of the dataset have been examined, missing data and outliers have been evaluated, and the fundamental distribution of data has been visualized (Da Poian et al., 2023). Methods used in the EDA process include histograms, box plots, correlation matrices, and heat maps (Behrens, 1997).

Exploratory data analysis is a process that aims to understand the structure, properties, and relationships of a data set. Thanks to the methods used, the data set is understood and summarized, data quality is evaluated, preliminary cleaning is performed, relationships between variables are determined, and as a result, hypotheses are produced, and the appropriate analysis methods is selected.

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a method that simplifies complex data by transforming it into a new set of variables called principal components, which are uncorrelated and capture the most important information from the original data. This technique facilitates the analysis and interpretation by reducing data dimensionality without losing significant information (Saccenti, 2024).

K-Means Clustering

K-Means clustering is a clustering algorithm that separates data into K clusters. The algorithm assigns each data point to the nearest cluster center and iteratively updates the cluster centers. This allows data to be divided into meaningful groups(MacQueen, 1967).

Analysis Process

The study first carried out the Exploratory Data Analysis (EDA) process on the data set. EDA is a process that aims to understand the structure, properties, and relationships of the data set. It is used to discover patterns, outliers, missing data, and correlations between variables in the dataset. EDA often includes visualization techniques (histograms, scatter plots, boxplots, etc.), summary statistics (mean, standard deviation, quartiles, etc.), and data transformations(Eken, 2020). Then, analyses involving various statistical and visualization techniques were performed on the data set. The K-Means algorithm (MacQueen, 1967) was employed for cluster analysis, and PCA (Principal Component Analysis) was utilized to reduce multidimensional data to two dimensions (Jolliffe, 2002). The EDA, PCA, and K-Means clustering methods employed in the study complement each other in the data analysis and modeling processes. The EDA is an essential preliminary step before the PCA and K-Means clustering.

Regarding the PCA, comprehending the correlations and variance between variables is paramount. The PCA is utilized for dimensionality reduction before the K-Means clustering. This approach helps to reduce the computational cost by improving the performance of the clustering method. The K-Means clustering method offers scaling of variables and the handling of outliers. The K-Means clustering method is used to validate patterns and relationships discovered with EDA or to better understand the reduced

dataset with PCA. Furthermore, correlation analysis was conducted to examine the relationships between variables in the dataset (Pearson, 1895).

Results

The findings of this study provide significant insights into the demographic and socio-economic characteristics of cities in Turkey. The analyses reveal significant differences among cities in terms of migration movements, population growth rates, and socio-economic characteristics. The main statistical result of the data set is given in the tables below.

Table 1. *Total Migration*

|	Incoming Internal Migration	Outbound Internal Migration	Immigration	Emigrate	Total Foreign Population
Count	1053.000000	1053.000000	1053.000000	1053.000000	1053.000000
Mean	32316.901235	32316.901235	6512.797721	3850.670465	11171.303894
STD	50726.815255	48105.976156	25520.484141	14742.334790	46707.404844
Min	3664.000000	4085.000000	80.000000	69.000000	12.000000
25%	12337.000000	13142.000000	785.000000	544.000000	844.000000
50%	18543.000000	21370.000000	1978.000000	1136.000000	2992.000000
75%	35103.000000	35442.000000	4525.000000	2513.000000	7779.000000
Max	498676.000000	595803.000000	353141.000000	200211.000000	740954.000000

Table 2. *Divorce and population growth*

	Number of People Per Km^2	Population Growth Rate	Divorce
Count	1053.000000	1053.000000	1053.000000
Mean	32316.901235	32316.901235	6512.797721
STD	50726.815255	48105.976156	25520.484141
Min	3664.000000	4085.000000	80.000000
25%	12337.000000	13142.000000	785.000000
50%	18543.000000	21370.000000	1978.000000
75%	35103.000000	35442.000000	4525.000000
Max	498676.000000	595803.000000	353141.000000

Correlation Analysis

The heatmap of variables with high correlations has demonstrated that there are strong correlations greater than 0.5 between some variables. Specifically, a significant positive relationship has been found between internal migration inflow and population growth rate (See Visual 1: Heatmap of Variables with High Correlation), (Westfall, Arias, & Fulton, 2017).
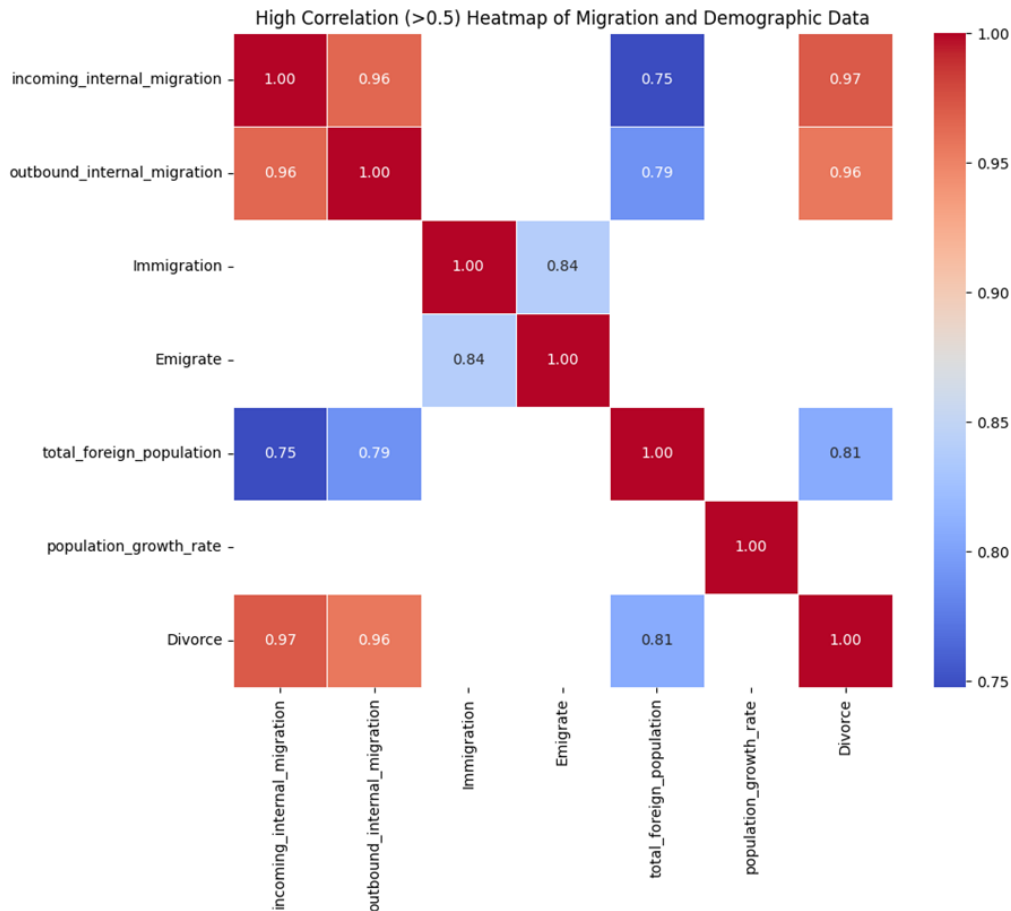
Figure 1. *Heat Map of Highly Correlated Variables*

The correlation matrix and heatmap display the strength and direction of the relationships between variables. Correlation values range between -1 and 1; values close to 1 indicate a strong positive relationship, values near -1 indicate a strong negative relationship, and values around 0 suggest a weak or no relationship. Below are significant findings and interpretations:

Internal Migration (Incoming and Outgoing): The correlation between incoming and outgoing internal migration is quite high (0.96), indicating that migration movements are generally balanced; there is a strong relationship between the number of people moving into a region and the number of people leaving that region.

Total Foreign Population: There are strong positive correlations between the total foreign population and incoming internal migration (0.75), outgoing internal migration (0.79), and the number of people per square kilometer (0.78). This shows that migration movements and population density are associated with the size of the foreign population.

People per Square Kilometer: This variable shows high correlations with incoming internal migration (0.89) and outgoing internal migration (0.91), indicating that population density in certain areas is closely related to internal migration movements.

Divorce: The number of divorces shows high correlations with incoming internal migration (0.97) and outgoing internal migration (0.96). This situation suggests that socioeconomic factors and regional demographic changes can have significant effects on family structures.

The high correlation between migration and unemployment suggests an endogenous loop where labor mobility both results from and perpetuates regional disparities. These findings align with Todaro's (1969) migration theory and recent empirical validations by Docquier et al. (2020), which show that internal migration is influenced not only by income differentials but also by institutional quality and service access gradients.

This analysis reveals that migration movements, population density, and social dynamics (for example, divorce rates) are in strong relationships with each other. The relationships between migrant inflows and outflows and the total foreign population indicate that migration flows have significant impacts on local demographic structures. Moreover, the relationships between the population growth rate and other variables highlight the complex nature of demographic changes and their influence by numerous factors. These findings provide important insights for local governments, policymakers, and social scientists to better understand the effects of migration movements and demographic changes on social structures.

Time Series Analyses

The time series analysis of annual average values has revealed trends in the variables over the years. Specifically, variations in migration amounts and population growth rates over time have been observed (See Visual 2: Time Series Analysis of Annual Average Values).
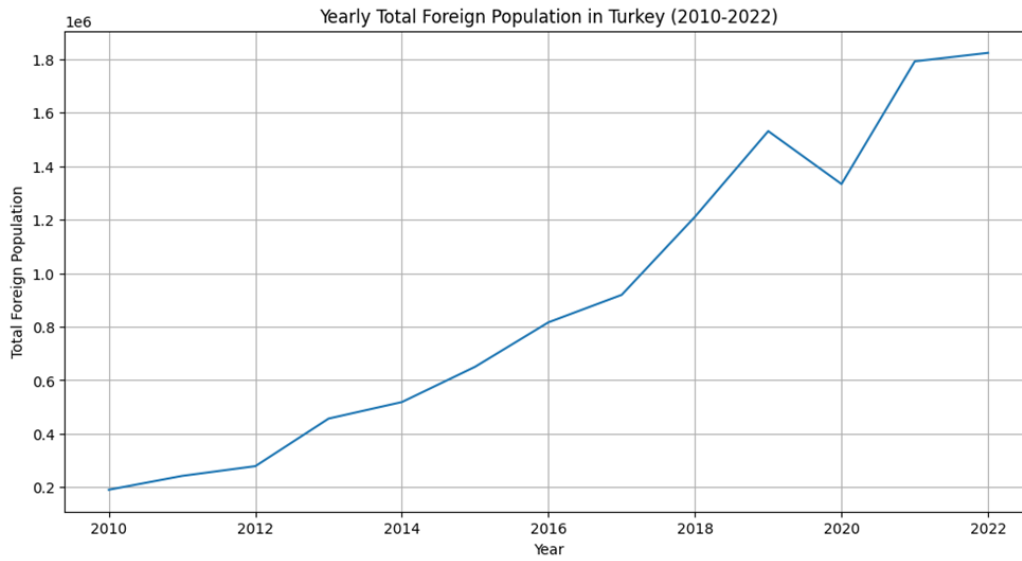
Figure 2. *Time Series Analysis of Annual Average Values*

The time series graph shows how the total foreign population in Turkey has changed between the years 2010 and 2022. According to the data, there has been a significant increase in the total foreign population in Turkey during this period. The main findings are as follows:

2010-2013: During this period, a gradual increase in the foreign population is observed, with the foreign population rising from 190,531 in 2010 to 456,506 in 2013.

2013-2018: After 2013, a period of rapid increase in the total foreign population is entered. From 2013 to 2018, the population has almost more than doubled, reaching 1,211,034.

2018-2022: Significant increases in the foreign population are again observed from 2018, reaching 1,823,836 by 2022. Especially in 2021, a significant jump (1,792,036) compared to the previous year is observed.

This analysis demonstrates that Turkey has experienced a significant increase in its foreign population over the last twelve years. Possible reasons for this increase include Turkey's geographical location, economic opportunities, regional conflicts, and general trends in immigration movements.

The increasing foreign population in Turkey creates various effects both socially and economically. This situation presents new challenges and opportunities in areas such as service provision to migrants, social integration, the labor market, and housing demand.

Especially the continuous increase in the foreign population after 2018 indicates that Turkey plays a significant role as a host country for migrants, and this will significantly affect future socio-economic planning. This increasing trend is an important indicator that should be considered in planning policies and services for migrants at both national and local levels.

Time series graphs show changes in the annual averages of various demographic and social indicators in Turkey between 2010-2022. Below are significant observations and interpretations for each indicator: (See Visual 3: Time Series Analysis of Annual Average Values).
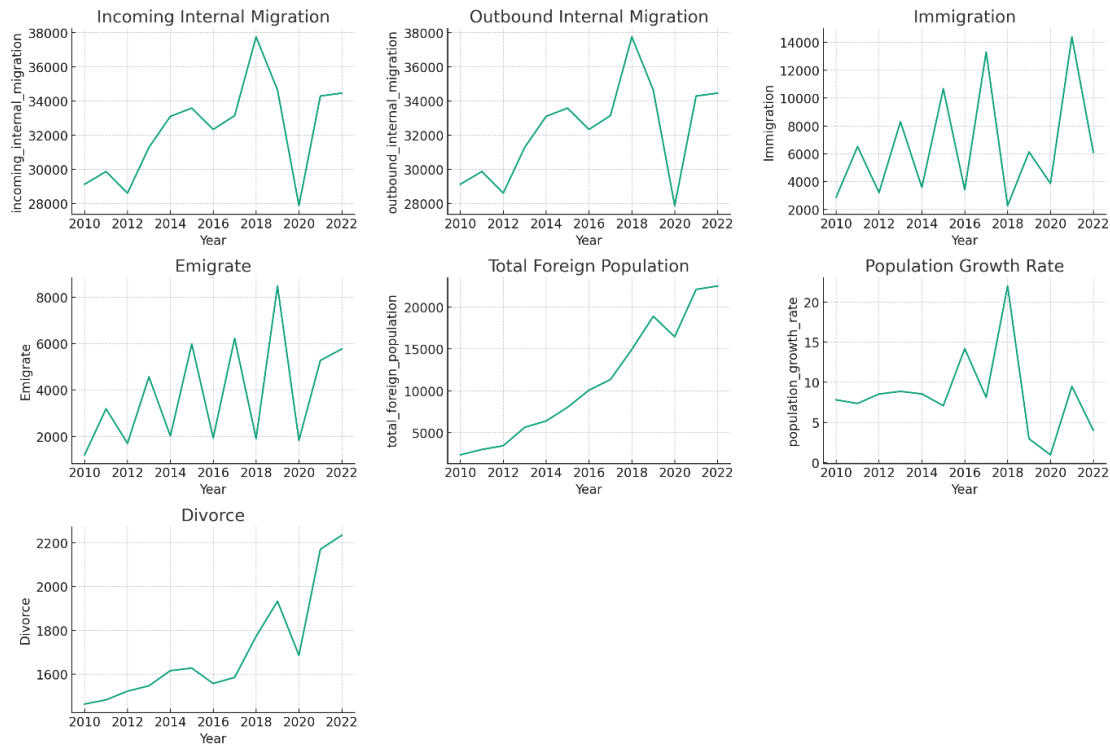
Figure 3. *Time Series Analys Internal Migration (Incoming and Outgoing)*

Internal Migration (Incoming and Outgoing): A general increasing trend is observed over time for both indicators. Particularly, a noticeable increase in 2018 highlights periods of intensified internal migration movements.

Immigrant Arrival (Immigration): A sharp increase in immigrant arrivals was observed in 2017, followed by 2015 and 2021. This indicates significant increases in immigration flows to Turkey during certain periods.

Immigrant Departure (Emigrate): There is a fluctuating trend in immigrant departures, with a notable increase in 2019.

Total Foreign Population: A continuous and marked increasing trend is observed from 2010 to 2022. This demonstrates that Turkey is hosting an increasingly larger foreign population.

Population Growth Rate: While this rate varies over the years, there was a significant peak in 2018. This indicates that there have been substantial changes in the population growth rate in certain years.

Divorce: A generally slow increasing trend in the number of divorces is observed. There is a continuous increasing trend up to 2022. This may suggest changes in social dynamics within the society and an increase in divorce incidents over time.

This analysis shows that Turkey has experienced significant changes in various demographic and social indicators over the last twelve years. The changes in internal migration, immigration, population growth rates, and divorce numbers highlight significant trends in the country's socio-economic structure and social dynamics. Such analyses can help policymakers and researchers understand societal changes and predict future trends.

When looking at time series graphs, it can be observed that the value of each variable studied in the analyses shows a numerical decrease in the year 2020. This decrease can be explained by COVID-19. COVID-19 has the potential to affect migration movements in societies (Fielding & Ishikawa, 2021). This effect is observed as a decrease in these graphs.
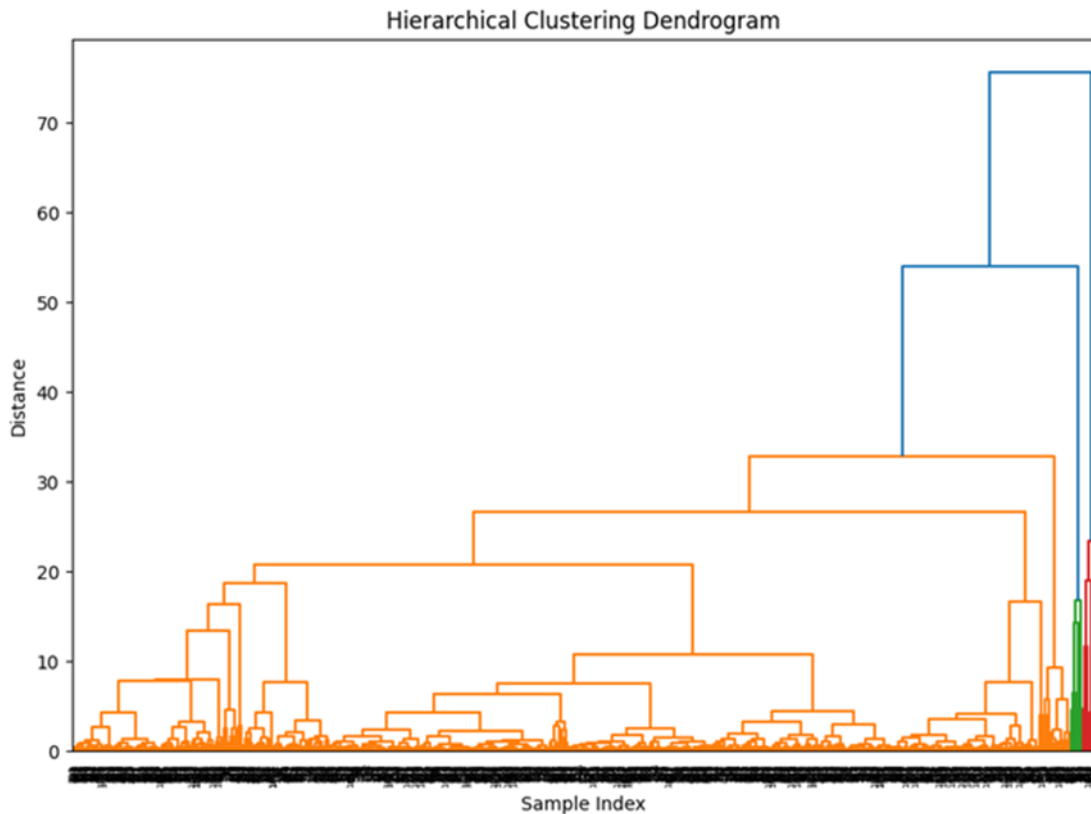
Hierarchical Clustering

Initially, after applying the Mean Shift clustering algorithm, a 'Cluster' column was added to the dataset, and each observation was assigned to a cluster (Wu & Yang, 2007). As seen in this example, all observations have been assigned to the same cluster (Cluster 0). This indicates that a large portion of the dataset shares similar characteristics, and the Mean Shift algorithm has identified a single dominant cluster. Since Mean Shift is a density-based clustering method, it clusters based on the density center of the data distribution, where a large part of the data points shows similar characteristics (Emekdaş, E. F.; 2010).

Single Cluster: The algorithm's assignment of all observations to a single cluster may indicate that the observations in the dataset exhibit similar characteristics regarding the variables studied and that there are no distinct sub-groups. This has indicated that the characteristics in the dataset have a homogeneous distribution.

Analysis: Assignment to a single cluster indicates that the analyzed characteristics (internal migration, immigration, total foreign population, population growth rate, and divorce numbers) generally show similar trends and distributions across all observations.

The hierarchical clustering analysis and the generated dendrogram have shown that cities can be grouped based on their demographic and socio-economic characteristics. This analysis has distinctly revealed the similarities and differences between cities (See Visual 4: Hierarchical Clustering Dendrogram).

Figure 4. *Hierarchical Clustering Dendrogram*

The hierarchical clustering dendrogram shows how observations in the dataset are grouped based on their similarities to each other. The dendrogram represents the distance or degree of similarity at each merging step with horizontal lines. The height of these lines represents the 'distance' or difference between the clusters being merged.

Analysis of the clustering structure: The dendrogram indicates that the dataset potentially contains different subgroups. Branch points that merge at greater distances represent groups of features that are more distinct from each other.

Number of Clusters: To determine the optimal number of clusters, it is useful to look for long vertical lines in the dendrogram. These lines are points where the merging distance (i.e., the difference between clusters) is significant, meaning the subgroups are significantly different from each other.

Clustering Decision: To decide on the optimal number of clusters, 'long' vertical lines indicating a significant increase in distance in the dendrogram are examined. These long lines can be used as a cutting point; thus, cutting the dendrogram at this point represents as many clusters as there are horizontal lines below it.

A detailed examination of the dendrogram provides a deeper understanding of the natural groupings contained in the dataset. Specifically, this analysis can be used to understand the similarities and differences in the dataset, identify subgroups, and conduct more targeted analyses.

This result can serve as a starting point for a more detailed analysis of the examined characteristics or the trial of different clustering techniques. The use of different parameters or algorithms (for example, K-Means, DBSCAN) would be beneficial in better understanding the potential subgroups in the dataset, leading to the implementation of the K-Means analysis.

K-Means Clustering

The K-Means clustering analysis (Likas, Vlassis, & Verbeek, 2003) has revealed that the cities are divided into five different groups. Each cluster contains cities with specific demographic and socio-economic characteristics (See Visual 5: K-Means Clustering Results).
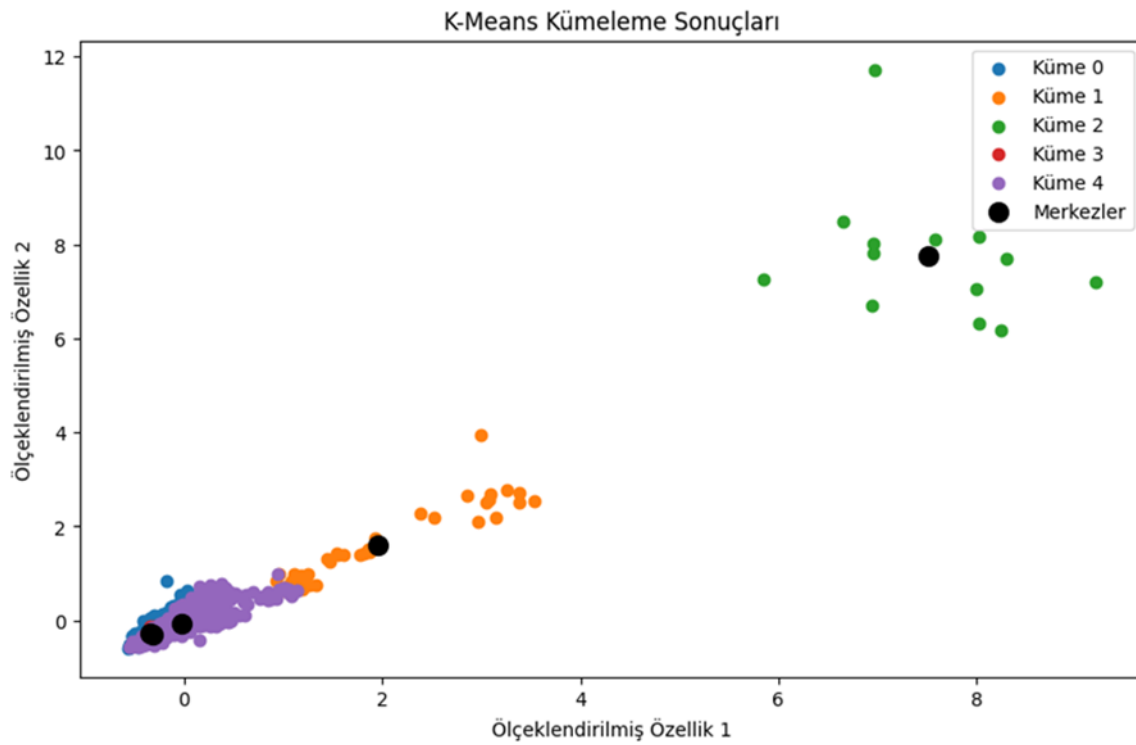
Şekil 5. *K-Means Clustering Results*

The K-Means clustering results show that the dataset is divided into five different clusters using the scaled features. In the graph, each cluster is represented by a different color, and black dots indicate the cluster centers. According to this analysis:

Cluster Distribution: The visualization clearly shows how the clusters are distributed in space. Each cluster represents data points grouped according to a specific combination of features.

Cluster Centers: The cluster centers, indicated by black dots, represent the central tendency of the data points belonging to each cluster. These centers determine the positions of the clusters in the feature space.

Clustering Quality: The distance between clusters and the density within them provide information about the quality of clustering. A visual inspection can reveal that some clusters are close to each other, while others are more distinctly separated.

Application: K-Means clustering is an effective method for separating observations in the dataset into groups with similar features. This analysis can help in understanding around which features the observations cluster and how these features differentiate between clusters.

The clustering results demonstrate that observations in the dataset are divided into meaningful groups based on their similarities and differences. These groups can be used to understand how certain characteristics, such as internal migration, immigration, foreign population, population growth rate, and divorce rates, vary together. This type of analysis can provide valuable insights for policy-making, strategic planning, or targeted interventions.

The outcomes of this analysis are an important step in understanding the relationships between observations and identifying significant patterns, especially in large and complex datasets. Each cluster can represent a specific behavior, trend, or characteristic, and this information can be utilized in further analyses and decision-making processes.

According to the K-Means clustering results, the classification has categorized different provinces of Turkey according to similar features, taking into account the demographic and social structure characteristics of the country. Based on the characteristics each group represents, the following interpretations can be made:

Cluster 0: Wide Geographic Distribution

Cluster 0 includes provinces from many different regions of Turkey and has a wide geographic distribution. This cluster contains provinces outside of metropolitan areas, which have relatively lower rates of migration in and out and have a smaller foreign population. These provinces reflect Turkey's traditional socio-economic structure and are regions where economic activities in sectors such as agriculture and small-scale industry are predominant.

Cluster 1: Metropolises and Tourism Centers

Cluster 1 includes Turkey's economically, politically, and tourism-wise prominent metropolises such as Ankara, Antalya, Bursa, and İzmir. These provinces are characterized by high rates of internal and external migration, a dense foreign population, and a dynamic social structure. Additionally, these provinces are centers of the country's tourism, industry, and service sectors and have high economic activity.

Cluster 2: Istanbul

Cluster 2 solely includes Istanbul, indicating that Istanbul has a different demographic and social structure from all other provinces in Turkey. As the country's most populous city and the economic and cultural capital, Istanbul possesses unique features such as high migration intake, a dense foreign population, and social diversity.

Cluster 3: Yalova and Yozgat

Cluster 3 includes the provinces of Yalova and Yozgat, indicating that these provinces resemble each other in terms of certain demographic or social characteristics. This means that both provinces share common features in terms of specific migration rates, population growth rates, and social dynamics.

Cluster 4: Provinces Showing Diversity

Cluster 4 contains many provinces from various regions of Turkey, demonstrating that these provinces exhibit similar demographic and social structure characteristics. This cluster includes both metropolitan areas and other provinces, indicating that certain features (for example, migration rates, economic activity) bring these provinces together in the analysis.

The K-Means clustering results have laid bare the diversity and complexity of the demographic and social structure of Turkey's provinces. Each cluster represents a group of provinces with different economic, social, and cultural characteristics. This classification serves as an important resource for policymakers and local governments in developing policies and strategies tailored to the specific needs of each region. In particular, the formation of Istanbul as a cluster on its own emphasizes the city's unique

position and significance across the country. On the other hand, the grouping of metropolises and tourism centers highlights the similarities in the economic and social dynamics of these provinces.

Box Plots and Histograms

Box plots show the distribution and outliers of each variable, while histograms have revealed the frequency distributions of the variables (Dormann, 2020). These visualizations have clearly demonstrated the overall distribution and main characteristics of the variables in the dataset (See figure 6: Box Plots).
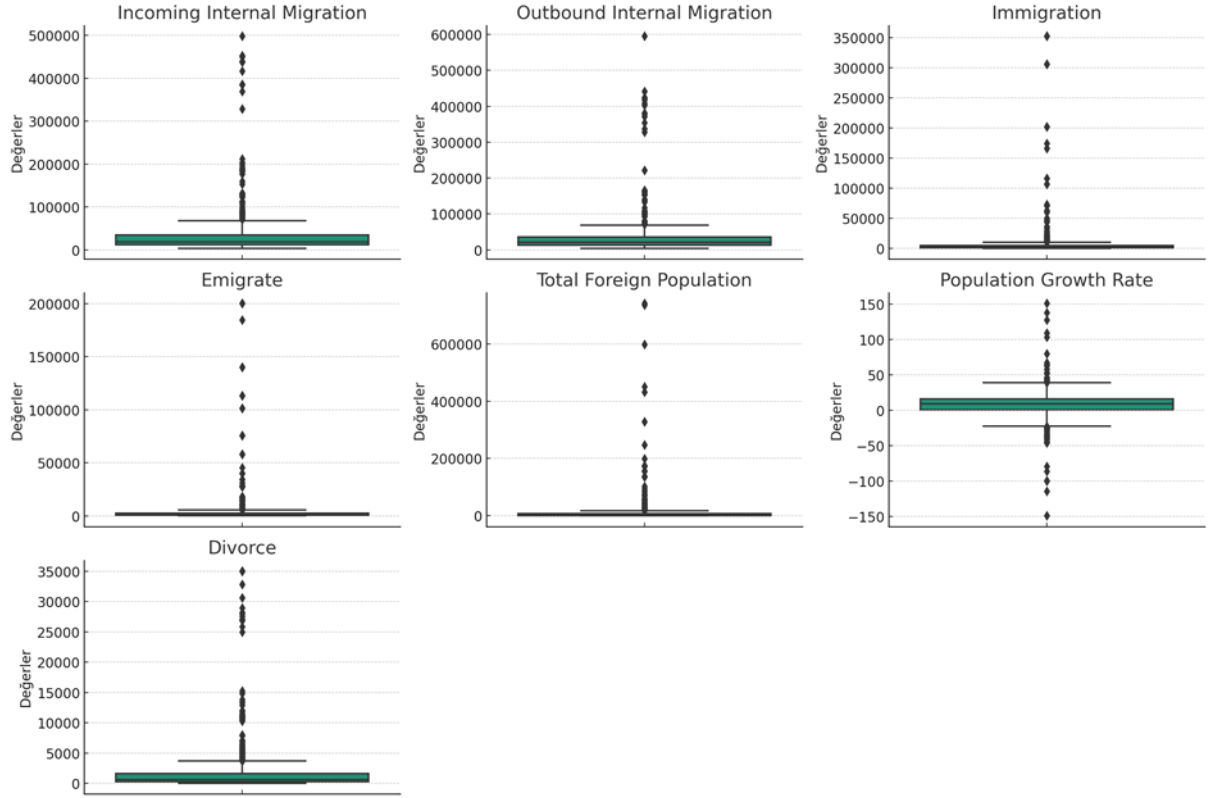

Figure 6. *Box Plots*

Box plots reveal the distribution characteristics of demographic and social variables by visualizing the distribution and outliers of numerical columns in the data set. Observations for each variable are presented below:

Incoming and Outbound Internal Migration

There is a wide distribution and many outliers in the numbers of internal migration arrivals and departures. This indicates that some provinces have much higher internal migration mobility than others. It is understood that migration movements are concentrated around certain centres.

Immigration and Emigration

Similarly, there is a wide dispersion and a significant number of outliers in the numbers of immigrant arrivals and departures. This indicates that certain provinces play an important role in international migrant flows and that these provinces stand out from others in terms of migrant mobility.

Total Foreign Population

A similar trend is observed in the total foreign population; There is a wide distribution and many outliers. This reveals that some provinces are much denser in terms of foreign population than others.

Population Growth Rate

The population growth rate includes both positive and negative values, indicating that the population in some regions of Turkey is increasing, while in others it is decreasing. Additionally, there are outliers in this variable.

Divorce

The number of divorces also has a wide distribution, with outliers indicating that in some provinces, this rate is much higher compared to others. This suggests that social structure and family dynamics vary by province.

Box plots reveal that Turkey's demographic and social structure exhibits significant diversity, and certain variables significantly differentiate some regions from others. The presence of outliers indicates that some provinces exhibit unusual conditions in terms of migration mobility, population growth rate, and social changes. These observations highlight the need for local governments and policymakers to develop strategies tailored to the specific needs and characteristics of each region(See figure 7: Distribution of Demographic and Social Variables in Turkey).
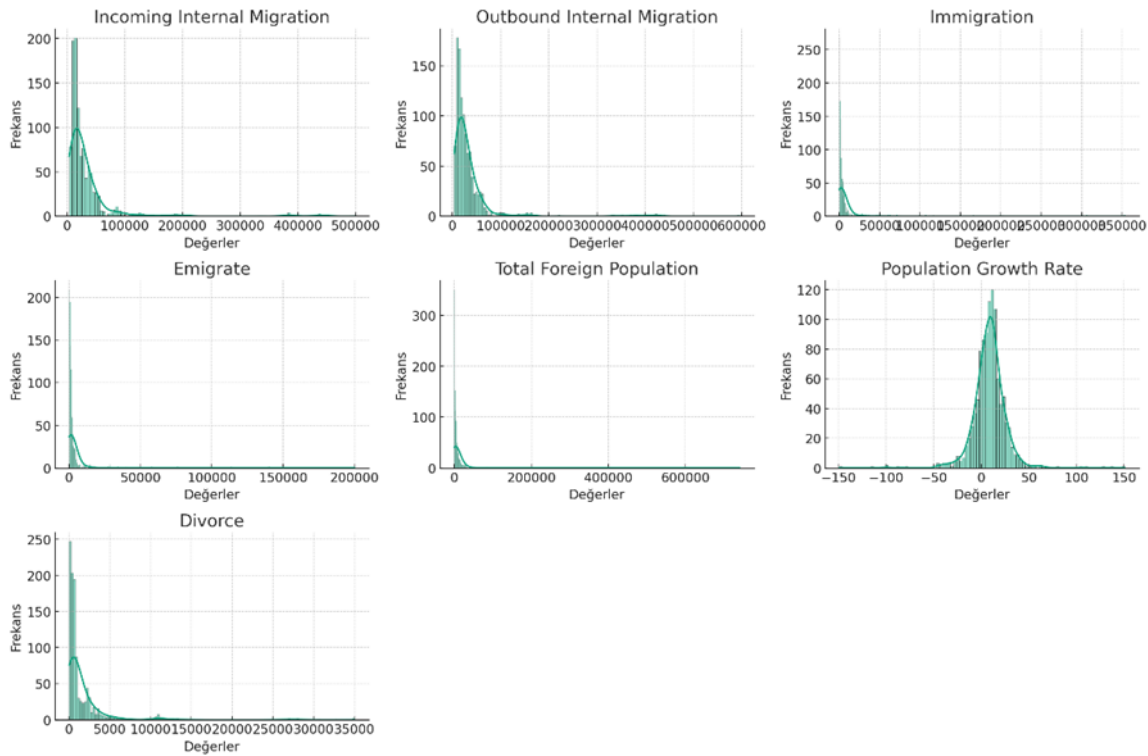
Figure 7. *Distribution of Demographic and Social Variables in Turkey*

Histograms visualize the data distributions of various demographic and social characteristics in Turkey, revealing the frequency distribution and densities of these characteristics. Observations for each variable are presented below:

Incoming and Outbound Internal Migration

The distribution of incoming and outbound internal migration numbers shows a right-skewed structure, meaning that while most provinces have relatively low migration numbers, a few provinces experience very high migration mobility. This indicates that migration movements are concentrated in specific centers.

Immigration and Emigrate

Similarly, the numbers of immigration and emigration exhibit a right-skewed distribution. It is understood that a few provinces receive or send a high amount of immigration, but the majority are at lower levels.

Total Foreign Population

The distribution of the total foreign population also shows a right-skewed structure, indicating that most provinces have a low foreign population, but a few provinces have a significantly high foreign population.

Population Growth Rate

The distribution of the population growth rate includes both positive and negative values, indicating that the population is increasing in some regions while decreasing in others. The distribution typically shows density close to the central tendency. The distribution of the population growth rate shows that the population growth rate of most cities is close to each other, and only certain cities have a high standard deviation compared to the average population growth value.

Divorce

The distribution of divorce numbers also has a right-skewed structure. While most provinces have relatively low numbers of divorces, a few provinces have significantly high numbers. This indicates that social structures and family dynamics vary by province.

These histograms show that the demographic and social characteristics of Turkey generally have right-skewed distributions; that is, while most provinces have relatively low values, a few provinces have very high values. This particularly shows a concentration trend in certain centers for variables such as migration movements, foreign population, and divorce numbers. These distributions reflect the diversity of social and demographic dynamics in Turkey and the unusual high or low mobilities in certain regions. This information can be valuable in social policy and planning processes.

Socio-Economic Characteristics

The clustering analysis has revealed that cities are divided into five different groups based on their socio-economic characteristics. These groups are defined through variables such as migration mobility, foreign population density, population growth rate, and divorce numbers. The PCA analysis has enabled the visualization of these clusters in a two-dimensional space, clearly highlighting the fundamental differences between cities (See figure 8: City Clusters Obtained Through PCA Analysis), (Kemsley, E. K. (1996)).
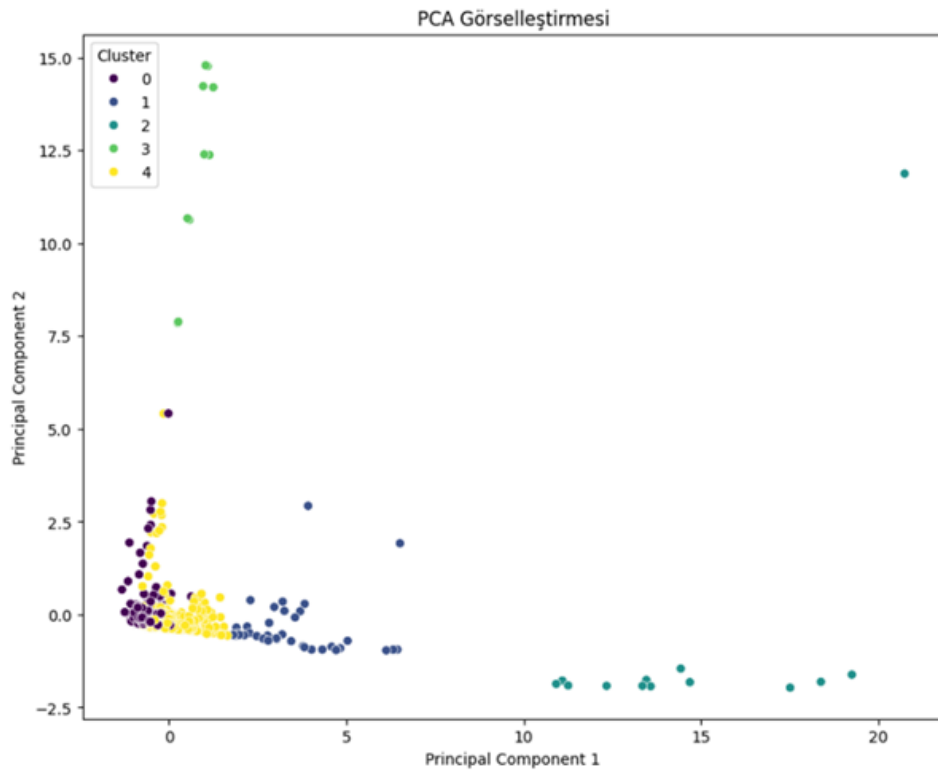
Figure 8. *City Clusters Obtained as a Result of PCA Analysis*

PCA allows for the reduction of high-dimensional datasets to lower dimensions by maximizing the variance. In this analysis, the first two principal components capture the most significant variance of the dataset, and visualization is conducted on these components. Metropolises like Istanbul, Ankara, İzmir, and tourism centers like Antalya constitute a large part of the variance captured by PCA, indicating the

diversity and complexity in the demographic and social structures of these cities (İslamoğlu, A. H., & Alnıaçık, Ü. (2014)).

The clusters represented by different colors in the visualization clearly show how provinces are grouped according to their similarities in demographic and social variables. This helps in discovering structural relationships and patterns in the dataset (Gündem, F., 2023). For instance, Istanbul, forming Cluster 2 on its own, indicates its unique position by separating from all other provinces in terms of demographic and social characteristics.

PCA results show how the clustered provinces differentiate on two principal components. This differentiation implies that demographic and social characteristics are related and these characteristics divide the provinces into specific groups. For example, provinces in Cluster 1, such as Ankara, Antalya, Bursa, İzmir, come together by exhibiting similar socio-economic and cultural dynamics.

In conclusion, PCA visualization is a powerful tool for understanding the complex nature of Turkey's demographic and social structure. This method uncovers significant patterns and relationships in the dataset, paving the way for more informed policy and strategic planning. Specifically, how provinces differentiate and are divided into similar groups based on demographic and social variables allows policymakers to develop strategies tailored to regional needs and potentials.

Policy Implications and Regional Strategy Recommendations

The results of the Cluster and Layered Development Analysis reveal that demographic and socioeconomic disparities among provinces in Turkey necessitate the development of differentiated policy frameworks at the regional level. In this context, because each cluster has a unique socioeconomic profile, policymakers and local governments must develop targeted strategies responsive to regional conditions. Policies for metropolitan clusters should prioritize initiatives that increase housing affordability, access to childcare, and social cohesion to mitigate migration-related pressures (OECD, 2020). Transitional clusters will benefit from targeted infrastructure investments and localized education and employment programs that support labor market integration (UNDP, 2021). Rural and peripheral clusters, on the other hand, require inclusive development policies that combine agricultural modernization, digital inclusion, and social protection programs, in line with the "people-centered development" approach of the UN Development Programme (UNDP) and DESA (2022).

In parallel, international policy recommendations for each provincial group can be summarized as follows:

for Cluster 0 (Traditional provinces), supporting rural revitalization through education and small-scale industrial development (OECD, 2020);

for Clusters 1 and 2 (Metropolitan cities and Istanbul), prioritizing the integration of the foreign population into the labor market and affordable housing policies (World Bank, 2022);

and for Clusters 3-4 (Transition regions), strengthening digital infrastructure and encouraging interregional mobility to reduce inequalities (UNDP, 2021).

In conclusion, the current policy approach we present offers a holistic development perspective that considers the demographic diversity of Türkiye's provinces, socioeconomic inequalities, and various factors, including migration. As emphasized in the literature (Rodríguez-Pose, 2018; Barca, McCann, & Rodríguez-Pose, 2012), policy designs that address local potential and account for spatial differences are crucial for promoting sustainable growth and regional development. In this context, multilayered policy frameworks that advocate for integrating regional differences into policymaking will also encourage diverse governance approaches.

Discussion, Conclusion, and Recommendations

This study has comprehensively examined the demographic and socio-economic differences among Turkey's cities through an extensive data analysis process. The research has not only revealed the complex demographic structure of Turkey and the diversity among its cities but also demonstrated the applicability and effectiveness of various statistical and machine learning techniques used in the analysis of large datasets. The study, incorporating methods such as correlation analysis, clustering, principal component analysis (PCA), and various visualization techniques, has presented significant findings towards a better understanding of the demographic and socio-economic profiles of Turkey's cities.

Our analyses have distinctly highlighted Istanbul's unique position in demographic and social structure by forming its own cluster. This clearly indicates that Istanbul significantly differs from other provinces in terms of foreign population density and migration mobility. On the other hand, provinces such as Ankara, İzmir, Bursa, and Antalya have shown similar characteristics in terms of economic, social, and cultural dynamics and have come together. These findings emphasize how certain provinces contribute to Turkey's demographic and social structure and enrich its diversity.

According to the cluster results based on migration and urbanization indicators, certain provinces stand out as migration hubs exhibiting distinctive demographic transformations. The rise in foreign populations—particularly following the post-2011 Syrian migration—has reshaped Turkey's urban and social dynamics (İçduygu & Nimer, 2020). Migration hotspots such as İstanbul and Gaziantep show growing demographic diversification and labor-absorption effects similar to those observed in major European metropolitan migration corridors. These population dynamics accelerate urban concentration processes and contribute to the emergence of new socio-economic configurations in western provinces while simultaneously altering the demographic balance in the country's southeastern border regions. This coexistence of modernized and traditional demographic zones within the same national context reflects the "dual demographic zones" pattern, as described by Bloom and Canning (2003), where advanced and transitional demographic regimes operate side by side.

The clustering analysis has revealed that Turkey's cities are distinctly divided into groups based on their demographic and socio-economic characteristics. This grouping serves as a significant resource for policymakers and local governments in developing policies and strategies tailored to the specific needs and characteristics of each region. In particular, groups such as metropolises and tourism centers have come together by exhibiting similar economic and social dynamics. This can assist local governments in determining priorities for focus, especially in areas such as strengthening social services, improving access to education and health services, and enhancing economic opportunities.

Another significant contribution of this study is the time series analyses that examine the changes in demographic and socio-economic trends and migration movements over time. These analyses have revealed significant trends and changes in the demographic structure and social dynamics among Turkey's cities. For instance, the increase in the foreign population over time and the dynamics of migration movements indicate that Turkey plays a significant role in immigration flows and that these flows have profound effects on the social and economic structure.

Furthermore, the insights provided by this study for policymakers and local governments include crucial points that should be considered in regional policy and planning processes. In particular, regions with intense migrant flows require the strengthening of social services, enhancement of economic opportunities, and development of integration policies. Additionally, in regions with low or negative population growth rates, measures to encourage population growth and strategies to support economic development are essential.

In conclusion, PCA and k-means clustering analysis offer valuable tools for understanding the demographic and social structure of Turkey. These methods enable us to better understand the demographic and social challenges faced by the country and to develop more informed and effective strategies to address these challenges. However, this study is just the beginning of analyses that can be conducted on the subject. More detailed and comprehensive studies are crucial to better understand Turkey's demographic and social structure and to cope with the challenges it faces. Future research, by examining different demographic and social variables in more detail, can provide deeper insights into Turkey's socio-economic and cultural structure. Such studies will play a critical role in developing policies and strategies that contribute to our country's social and economic development.

The research has comprehensively revealed the complexity of Turkey's demographic structure and the diversity among its cities, including an analysis of various factors such as migration movements, population growth, foreign population density, and divorce rates. In this section, the significance of the findings has been evaluated in the context of social sciences and policy-making, and limitations of this study along with suggestions for future research have been presented. The findings of this study have shown that there are significant differences in migration and demographic dynamics among Turkey's cities. Especially the high rates of internal and external migration observed in major cities reflect Turkey's rapid urbanization process and economic changes (Göçmen and Ekmekçi, 2015). The positive correlations found indicate that internal migration has a significant impact on population growth rates,

which has important implications for urban planning and infrastructure development (Jussibaliyeva et al., 2023).

The results of the research represent an important step in understanding Turkey's demographic and socio-economic structure. The significant differences shown by cities in terms of migration movements, population growth rates, and socio-economic characteristics underline the necessity for local governments and policymakers to develop specific policies and strategies tailored to the specific needs of each region. Particularly, a detailed analysis of internal and immigration movements is critical in shaping migration policies and steps towards the social integration of migrants. The clustering analysis results have shown that cities can be divided into different groups based on their socio-economic characteristics. These groups indicate differentiation among cities in terms of economic opportunities, quality of life, and demographic structures. Especially, the distinct characteristics of Istanbul (Cluster 2) highlight the city's prominence as Turkey's economic and cultural center (Akçapar and Şimşek, 2016), (Ak, M. A., 2023).

The findings obtained through methods such as clustering analysis and PCA have revealed the complex structure of cities' demographic and socio-economic profiles and the unique characteristics within this structure. These results show that Turkey's cities are not homogeneous in terms of demographic and socio-economic aspects; rather, each city has its own unique characteristics. This situation necessitates the development of policies sensitive to local needs, especially in areas such as urban planning, education, health, and social services. The outcomes of clustering and PCA analyses have demonstrated that the demographic and socio-economic profiles of cities are complex and multidimensional (Pereira, Vale, & Santana, 2023). These analyses have clearly identified the fundamental trends and groups in the dataset, thus providing significant insights for policymakers and urban planners (Jolliffe, 2002; MacQueen, 1967).

The research also includes time series analyses that examine the changes in demographic and socio-economic trends over time. These analyses have shown that the demographic transformation and socio-economic developments experienced in recent years have created significant impacts on the structure and dynamics of cities. For example, the increase in the foreign population and the dynamics of migration movements necessitate the development of new strategies in terms of social policy and migration management.

The analysis results have revealed that provinces such as Ankara, İzmir, Bursa, and Antalya exhibit similar socio-economic and cultural dynamics and come together. These findings indicate that certain provinces display similar trends in terms of economic development, migration movements, and social change. On the other hand, it is observed that smaller and rural areas face different demographic and social challenges. These differences underline the need for policymakers and local governments to develop policies tailored to the specific needs of each region.

This study shows that there is a need to focus on areas such as social services, economic opportunities, and integration policies, especially in regions with intense migrant flows. Moreover, the importance of measures to encourage population growth and strategies to support economic development in regions with low population growth rates also emerges. These findings indicate that regional policy and planning processes should be shaped by taking into consideration the unique demographic and social characteristics of the provinces.

Regarding the limitations of this study, the analyses are primarily based on existing datasets and statistical methods. This situation may lead to some limitations in terms of the scope and quality of the datasets and the applicability of the methods. Additionally, demographic and socio-economic analyses involve complex processes that emerge as a result of the interaction of numerous factors, making it challenging to model and understand these interactions in detail comprehensively. In this context, future research could provide a more thorough analysis of Turkey's demographic and socio-economic structure by utilizing different data sources and analysis methods.

In conclusion, methods such as PCA and k-means clustering analysis offer valuable tools for better understanding Turkey's demographic and social structure. These methods enable us to better comprehend the demographic and social challenges faced by the country and to develop more informed and effective strategies to address these challenges. This study is just the beginning of the analyses that can be conducted to understand our country's demographic and social structure and to cope with future challenges. Moreover, this article has contributed to understanding the demographic and socio-economic

differences among Turkey's cities and presented significant findings on this topic. The results obtained can be used as a valuable resource in the field of social sciences and in the process of policy-making. However, it should not be forgotten that the findings of this study should be viewed as a starting point for a deeper understanding of Turkey's demographic and socio-economic structure, and future research should be built upon this foundation.

Ethical Declaration

Scientific rules, ethics, and citation rules were followed during the writing process of the study titled "*Analysis of Turkey's Demographic And Social Structure Using Pca And K-Means Clustering.*" There was no tampering with the data collected, and this study was not sent to any other academic publication environment for evaluation.

Etik Beyan

"Türkiye'nin Demografik ve Sosyal Yapısının PCA ve K-Means Kümeleme Yöntemiyle Analizi" başlıklı çalışmanın yazım sürecinde bilimsel kurallara, etik ve alıntı kurallarına uyulmuş; toplanan veriler üzerinde herhangi bir tahrifat yapılmamış ve bu çalışma herhangi başka bir akademik yayın ortamına değerlendirme için gönderilmemiştir.

Declaration of Researchers' Contribution Rate

The authors have equal contributions.

Araştırmacıların Katkı Oranı Beyanı

Yazarların çalışmadaki katkı oranları eşittir.

Declaration of Conflict

There is no potential conflict of interest in the study.

Çatışma Beyanı

Çalışmada herhangi bir potansiyel çıkar çatışması söz konusu değildir.

Note

Cities According to Clusters

Cluster 0: Adıyaman, Afyonkarahisar, Ağri, Amasya, Ardahan, Artvin, Bartin, Bayburt, Bilecik, Bingöl, Bitlis, Bolu, Burdur, Çanakkale, Çankiri, Çorum, Edirne, Elaziğ, Erzincan, Erzurum, Giresun, Gümüşhane, Hakkari, Iğdir, Isparta, Karaman, Kars, Kastamonu, Kirikkale, Kirklareli, Kirşehir, Kütahya, Malatya, Muş, Nevşehir, Niğde, Ordu, Rize, Samsun, Siirt, Sivas, Tokat, Trabzon, Tunceli, Uşak, Van, Yozgat, Zonguldak, Kahramanmaraş, Sinop, Denizli, Düzce, Eskişehir, Karabük, Kilis, Balikesir, Mardin, Şirnak, Aksaray, Hatay, Manisa, Yalova

Cluster 1: Ankara, Antalya, Bursa, İzmir

Cluster 2: İstanbul

Cluster 3: Yalova, Yozgat

Cluster 4: Adana, Aksaray, Aydin, Balikesir, Batman, Denizli, Diyarbakir, Düzce, Eskişehir, Gaziantep, Hatay, Kahramanmaraş, Karabük, Kayseri, Kilis, Kocaeli, Konya, Manisa, Mardin, Mersin, Muğla, Osmaniye, Sakarya, Sinop, Şanliurfa, Şirnak, Tekirdağ, Yalova, Afyonkarahisar, Bartin, Bayburt, Bilecik, Bolu, Burdur, Çanakkale, Çankiri, Edirne, Erzincan, Gümüşhane, Isparta, Karaman, Kirklareli, Nevşehir, Samsun, Tokat, Trabzon, Uşak, Adiyaman, Bursa, Malatya, Ordu, Van, Amasya, Ardahan, Artvin, Bingöl, Bitlis, Çorum, Elaziğ, Erzurum, Giresun, Hakkari, Iğdir, Kastamonu, Kirikkale, Kirşehir, Kütahya, Niğde, Rize, Siirt, Sivas, Tunceli, Yozgat, Antalya, Ağri, Kars, Muş

Not

Kümelere Göre Şehirler

Küme 0: Adıyaman, Afyonkarahisar, Ağri, Amasya, Ardahan, Artvin, Bartin, Bayburt, Bilecik, Bingöl, Bitlis, Bolu, Burdur, Çanakkale, Çankiri, Çorum, Edirne, Elaziğ, Erzincan, Erzurum, Giresun,

Gümüşhane, Hakkari, Iğdir, Isparta, Karaman, Kars, Kastamonu, Kirikkale, Kirklareli, Kirşehir, Kütahya, Malatya, Muş, Nevşehir, Niğde, Ordu, Rize, Samsun, Siirt, Sivas, Tokat, Trabzon, Tunceli, Uşak, Van, Yozgat, Zonguldak, Kahramanmaraş, Sinop, Denizli, Düzce, Eskişehir, Karabük, Kilis, Balikesir, Mardin, Şirnak, Aksaray, Hatay, Manisa, Yalova

Küme 1: Ankara, Antalya, Bursa, İzmir

Küme 2: İstanbul

Küme 3: Yalova, Yozgat

Küme 4: Adana, Aksaray, Aydin, Balikesir, Batman, Denizli, Diyarbakir, Düzce, Eskişehir, Gaziantep, Hatay, Kahramanmaraş, Karabük, Kayseri, Kilis, Kocaeli, Konya, Manisa, Mardin, Mersin, Muğla, Osmaniye, Sakarya, Sinop, Şanliurfa, Şirnak, Tekirdağ, Yalova, Afyonkarahisar, Bartin, Bayburt, Bilecik, Bolu, Burdur, Çanakkale, Çankiri, Edirne, Erzincan, Gümüşhane, Isparta, Karaman, Kirklareli, Nevşehir, Samsun, Tokat, Trabzon, Uşak, Adiyaman, Bursa, Malatya, Ordu, Van, Amasya, Ardahan, Artvin, Bingöl, Bitlis, Çorum, Elaziğ, Erzurum, Giresun, Hakkari, Iğdır, Kastamonu, Kirikkale, Kirşehir, Kütahya, Niğde, Rize, Siirt, Sivas, Tunceli, Yozgat, Antalya, Ağrı, Kars, Muş

Kaynakça

Ak, M. A. (2023). Comparative analysis of Turkey and Russia's public diplomacy on the Balkans (Example of the Russian World Foundation and Yunus Emre Institute). Marmara Üniversitesi Siyasal Bilimler Dergisi, 11(1), 1-22. https://doi.org/10.14782/marmarasbd.1159265

Akçapar, S. K., & Şimşek, D. (2016). Türkiye'de uluslararası göçün sosyo-ekonomik etkileri. Ankara Üniversitesi SBF Dergisi, 71(4), 1019-1042.

Barca, F., McCann, P., & Rodríguez-Pose, A. (2012). The case for regional development intervention: Place-based versus place-neutral approaches. Journal of Regional Science, 52(1), 134–152.

Behrens, J. T. (1997). Exploratory data analysis. In (Editör Adı, Ed.), Research methods in psychology (pp. 221-236). Springer.

Bloom, D. E., & Canning, D. (2003). Contraception and the Celtic Tiger. The Economic and Social Review, 34(3), 229–247.

Da Poian, V., Theiling, B., Clough, L., McKinney, B., Major, J., Chen, J., & Hörst, S. (2023). Exploratory data analysis (EDA) machine learning approaches for ocean world analog mass spectrometry. Frontiers in Astronomy and Space Sciences, 10, 1134141. https://doi.org/10.3389/fspas.2023.1134141

Docquier, F., Machado, J., & Sekkat, K. (2015). Efficiency gains from liberalizing labor mobility. Scandinavian Journal of Economics, 117(2), 303-346. https://doi.org/10.1111/sjoe.12097

Eken, S. (2020). Büyük verinin interaktif görselleştirilmesi: Tableau üzerine öğrenci deneyimleri. Avrupa Bilim Ve Teknoloji Dergisi, (18), 262-271. https://doi.org/10.31590/ejosat.659823

Emekdaş, E. F. (2010). İmge arama sonuçlarının baskın kümeler kullanılarak gruplandırılması [Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü].

Fielding, T., & Ishikawa, Y. (2021). COVID-19 and migration: A research note on the effects of COVID-19 on internal migration rates and patterns in Japan. Population, Space and Place, 27(6), e2499. https://doi.org/10.1002/psp.2499

Filiztekin, A. (2020). Regional economic disparities and convergence in Turkey: A spatial approach. Growth and Change, 51(3), 1338–1359.

Göçmen, İ., & Ekmekçi, P. E. (2015). Türkiye'de iç göç ve ekonomik etkileri. Ege Akademik Bakış Dergisi, 15(2), 231-243.

Gündem, F. (2023). Beliefs, economics, and spatial regimes in voting behavior: The Turkish case, 2007–2018. Humanities and Social Sciences Communications, 10(1), 1-15.

İçduygu, A., & Nimer, M. (2020). The politics of an emerging "immigration country": Turkey and its migrants. International Migration, 58(4), 270–289.

İslamoğlu, A. H., & Alnıaçık, Ü. (2014). Sosyal bilimlerde araştırma yöntemleri (511). Beta Yayınevi.

Jolliffe, I. T. (2002). Principal component analysis. Springer-Verlag.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. Philosophical Transactions of the Royal Society A, 374(2065), 20150202.

Jussibaliyeva, A., Iskakova, D., Kurmanalina, A., Khassenova, K., & Amerkhanova, I. (2023). Internal mıgration and its impact on population income in different sectors of economy. Science and Innovation, 2(A3), 125-137.

Karataş, A. (2023). Global migration governance: An analysis of Turkey on the basis of migration governance indicators. İ. Dursunoğlu (Ed.), International Research in Social, Human and Administrative Sciences X, (ss. 47–90). Eğitim Yayınevi.

Kaya, Y., & Çolak, M. (2020). Türkiye'de boşanma oranları ve sosyo-ekonomik faktörler. Türk Sosyal Bilimler Dergisi, 25(3), 456-472.

Kemsley, E. K. (1996). Discriminant analysis of high-dimensional data: A comparison of principal components analysis and partial least squares data reduction methods. Chemometrics and Intelligent Laboratory Systems, 33(1), 47-61.

Kirk, D. (1996). Demographic transition theory. Population Studies, 50(3), 361–387.

Lee, R., & Mason, A. (2014). Is low fertility really a problem? Science, 346(6206), 229–234.

Likas, A., Vlassis, N., & Verbeek, J. J. (2003). The global k-means clustering algorithm. Pattern Recognition, 36(2), 451-461.

Little, R. J. A., & Rubin, D. B. (2019). Statistical analysis with missing data (3rd ed.). Wiley. https://doi.org/10.1002/9781119482260

Lutz, W., Butz, W. P., & KC, S. (Eds.). (2014). World population and human capital in the twenty-first century. Oxford University Press.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press.

Marques, A. M., Domingos, T., & Costa, P. (2020). Socio-economic typologies of EU regions using PCA and clustering. Regional Science Policy & Practice, 12(6), 1073–1093.

Notestein, F. W. (1945). Population—the long view. In (Editör Adı, Ed.), Food for the world (pp. 36–57).

OECD. (2020). Rural well-being: Geography of opportunities. OECD Publishing.

Öncel, H., & Levend, S. (2023). The effects of urban growth on natural areas: The three metropolitan areas in Türkiye. Environmental Monitoring and Assessment, 195(7), 816. https://doi.org/10.1007/s10661-023-11539-7

Pearson, K. (1895). Notes on regression and inheritance in the case of two parents. Proceedings of the Royal Society of London, 58(347-352), 240-242.

Pereira, M. F., Vale, D. S., & Santana, P. (2023). Is walkability equitably distributed across socio-economic groups?– A spatial analysis for Lisbon metropolitan area. Journal of Transport Geography, 106, 103491.

Reher, D. (1998). Family ties in Western Europe: Persistent contrasts. Population and Development Review, 24(2), 203–234.

Rodríguez-Pose, A. (2018). The revenge of the places that don't matter. Cambridge Journal of Regions, Economy and Society, 11(1), 189–209.

Saccenti, E. (2024). A gentle introduction to principal component analysis using tea-pots, dinosaurs, and pizza. Teaching Statistics, 46(1), 38-52. https://doi.org/10.1111/test.12363

Thornes, B., & Collard, J. (2023). Who divorces? Taylor & Francis.

Todaro, M. P. (1969). A model of labor migration and urban unemployment in less developed countries. American Economic Review, 59(1), 138-148.

Tukey, J. W. (1977). Exploratory data analysis. Addison-Wesley.

UN DESA. (2022). World social report.

UNDP. (2021). Regional human development report.

GENİŞ ÖZET

Yaptığımız çalışma, Türkiye'nin benzersiz konumundan ve karmaşık toplum yapısından yararlanarak, Türkiye'deki illerin demografik ve sosyoekonomik özelliklerinin kapsamlı bir analizini yapmaktadır. Çalışmada kullanılan veriler korelasyon çalışmaları, kümeleme, Temel Bileşen Analizi (PCA) ve yenilikçi görselleştirmeler gibi farklı ve yenilikçi istatistiksel ve makine öğrenimi teknikleriyle incelenmiştir. İncelem sonucunda iç göç, nüfus artışı, yabancı nüfus yoğunluğu ve boşanma oranları gibi faktörler hakkında bulgulara ulaşılmıştır. Sonuçlar, zaman serisi analizleri gelecek değerlerini değişen trendlere ışık tutmasıyla oluşan sonuç verileri, göç kalıpları ve sosyo-ekonomik faktörler arasında şehirler arasında önemli farklıklar olduğunu ortaya koymuştur. Çalışma bu yönüyle Türkiye'nin bölgesel ve küresel göçteki kritik rolünü ve toplumsal yapı üzerindeki etkisini vurgulayarak politika oluşturma ve sosyal bilim araştırmalarını ilerletme konusunda içgörüler sunmaktadır. Aynı zamanda, etkili politikaların geliştirilmesi ve sosyal uyumun güçlendirilmesi açısından hayati önem taşıyan Türkiye'nin nüfus dinamikleri ve sosyal dokusunun ayrıntılı bir şekilde anlaşılmasını sağlamayı da amaçlamaktadır. PCA ve kümeleme gibi teknikleri kullandığımız araştırma deseni, karmaşık veri gruplarını ortaya çıkarmış, demografik ve sosyal özelliklerin nasıl etkileşimde bulunduğunu belirlemiş ve Türkiye'nin nüfus yapısına ilişkin ayrıntılı bilgiler sunmamıza imkan vermiştir. Yaptığımız analizlerin, çeşitli istatistiksel ve makine öğrenimi tekniklerinin büyük veri kümelerini analiz etmede uygulanabilirliğini ve etkinliğini göstererek şehirlerin karşılaştığı zorlukların ve fırsatların daha iyi anlaşılmasına ve daha etkili yerel ve ulusal politikalar geliştirilmesine yardımcı olacağı düşünülmektedir. Çalışmamız, çok boyutlu bir veri seti ve ileri veri analiz teknikleri kullanarak Türkiye'nin demografik ve sosyo-ekonomik yapısının kapsamlı bir analizini sunmayı amaçlamaktadır. Bu çalışmada kullanılan veri seti, Türkiye'nin 81 ilinin 2010-2022 yıllarını kapsayan demografik ve sosyo-ekonomik verilerini içermektedir. Veriler Türkiye İstatistik Kurumu (TÜİK) tarafından sağlanan veri setleri kullanılmıştır. Analiz edilen temel değişkenler iç göç girişi ve çıkışı, göç girişi ve çıkışı, toplam yabancı

nüfus, nüfus yoğunluğu, nüfus artış hızı ve boşanma sayısıdır. Çalışmada, Keşif Amaçlı Veri Analizi (EDA), Temel Bileşen Analizi (PCA), K-ortalamalı kümeleme, korelasyon analizi ve istatistiksel görselleştirme teknikleri dahil olmak üzere çeşitli veri analizi yöntemleri kullanılmıştır. EDA süreci, veri kümesinin temel istatistiksel özelliklerinin incelenmesini, eksik verilerin ve aykırı değerlerin ele alınmasını ve veri dağılımının görselleştirilmesini için kullanılmıştır. PCA, çok boyutlu verileri iki boyuta indirerek daha erişilebilir analiz ve yorumlamaya amacıyla kullanılmıştır. K-Means algoritması, illerin demografik ve sosyo-ekonomik özelliklere göre gruplandırılarak kümeleme analizi için yapılması amacıyla uygulanmıştır. Analiz süreci, farklı ölçeklerde ölçülen değişkenlerin etkilerini normalleştirmek için özellik ölçeklendirmeyle başlatılmıştır. Sonuçlar, dağılım grafikleri, çubuk grafikler ve ısı haritaları gibi çeşitli görselleştirme araçları kullanılarak sunulmuştur. Bulgular, göç hareketleri, nüfus artış oranları ve sosyo-ekonomik özellikler açısından şehirler arasında önemli farklılıklar olduğunu ortaya koymaktadır. Çalışma, Türkiye'deki şehirlerin demografik ve sosyo-ekonomik profillerine ilişkin bilgiler sunarak, bu faktörler arasındaki karmaşık etkileşimi ve bunların şehirlerin ve bölgelerin geleceği üzerindeki potansiyel etkilerini ortaya çıkarılmasını sağlamıştır. Çalışmamız, çeşitli demografik değişkenler ve bunların korelasyonlarının bir analizinin yanı sıra, Türkiye'deki toplam yabancı nüfusun zaman serisi analizini de sunmaktadır. Korelasyon analizi, iç göç (giriş ve çıkış), nüfus yoğunluğu, toplam yabancı nüfus ve boşanma oranları arasında güçlü pozitif ilişkiler olduğunu ortaya koymuştur. Bu bulgular göç hareketleri, nüfus yoğunluğu ve boşanma oranları gibi sosyal dinamiklerin birbiriyle yakından ilişkili olduğunu göstermektedir. Zaman serisi analizi, Türkiye'nin toplam yabancı nüfusunda 2010'dan 2022'ye kadar kademeli büyüme (2010-2013) ve hızlı büyüme (2013-2018 ve 2018-2022) dönemlerinde önemli bir artış olduğunu göstermektedir. Yabancı nüfus 2013 ile 2018 arasında iki kattan fazla artış göstermiş; 2021 yılına gelindiğinde ise yabancı nüfus sayısında önemli bir sıçrama olduğu gözlenmiştir. Bu eğilim, Türkiye'nin muhtemelen coğrafi konumu, ekonomik fırsatları, bölgesel çatışmalar ve genel göç eğilimleri nedeniyle göçmenler için önemli bir ev sahibi ülke haline geldiğini göstermektedir. Türkiye'de artan yabancı nüfus, hizmet sunumu, sosyal entegrasyon, işgücü piyasası ve konut talebi gibi alanlarda zorluklar oluşturmanın yanısıra fırsatlar da sunmaktadır. 2018'den sonraki sürekli artış, gelecekteki sosyo-ekonomik planlama ve politika oluşturmada bu eğilimin dikkate alınması ihtiyacını vurgulamaktadır. İstanbul, demografik ve sosyal yapı açısından kendi kümesini oluşturarak diğer illerden belirgin bir şekilde ayrışmaktadır. Bu, İstanbul'un yabancı nüfus yoğunluğu ve göç hareketliliği açısından diğer illerden önemli ölçüde farklı olduğunu göstermektedir. Ankara, İzmir, Bursa ve Antalya gibi iller, ekonomik, sosyal ve kültürel dinamikler açısından benzer özellikler göstermekte ve bir araya gelmektedir. Bu bulgular, belirli illerin Türkiye'nin demografik ve sosyal yapısına nasıl katkıda bulunduğunu ve çeşitliliğini zenginleştirdiğini vurgulamaktadır. 2010-2022 yılları arasında Türkiye'deki çeşitli demografik ve sosyal göstergeler zaman serisi grafikleri ve kümeleme teknikleri kullanılarak analiz edildiğinde ortaya çıkan temel gözlemler şunlardır: İç göç (gelen ve giden) genel bir artış eğilimi göstermektedir; 2018'de gözle görülür bir artış, göç hareketlerinin yoğunlaştığını göstermektedir. Göçmen girişlerinde (göç) 2017, 2015 ve 2021'de görülen keskin artış rakamları; göç akışında önemli sonuçlara işaret etmektedir. Sonuç olarak; yoğun göç akışına sahip bölgelerde sosyal hizmetlerin güçlendirilmesi, ekonomik fırsatların artırılması ve entegrasyon politikalarının geliştirilmesi gerektiği ortaya çıkmıştır. Ayrıca, düşük veya negatif nüfus artış oranlarına sahip bölgelerde nüfus artışını teşvik edici önlemler ve ekonomik kalkınmayı destekleyici stratejiler belirlenmesi de önem taşımaktadır.