Konya Mühendislik Bilimleri Dergisi, c. 10, s. 4, 889-902, 2022
Konya Journal of Engineering Sciences, v. 10, n. 4, 889-902, 2022
ISSN: 2667-8055 (Electronic)
DOI: 10.36306/konjes.1158414

CLUSTERING NEIGHBORHOODS ACCORDING TO URBAN FUNCTIONS AND DEVELOPMENT LEVELS BY DIFFERENT CLUSTERING ALGORITHMS: A CASE IN KONYA

[1]Ali Utku AKAR , [2]Sait Ali UYMAZ

[1]Konya Technical University, Faculty of Engineering and Natural Sciences, Department of Geomatics Engineering, Konya, TÜRKİYE
[2]Konya Technical University, Faculty of Engineering and Natural Sciences, Department of Computer Engineering, Konya, TÜRKİYE
[1]auakar@ktun.edu.tr, [2]sauymaz@ktun.edu.tr

(Geliş/Received: 05.08.2022; Kabul/Accepted in Revised Form: 10.09.2022)

ABSTRACT: Urban functions/activities, which emerged under the influence of the human factor and are in the process of development over time, play a crucial role in the development of neighborhoods. To ensure balanced development status among the neighborhoods, it is necessary to know the development levels of the neighborhoods in advance. This study focuses on the clustering of the 167 central neighborhoods in Konya in terms of urban functions and reveals the similarities or differences in the development status of these neighborhoods. K-means, Hierarchical (agglomerative) and OPTICS clustering analyzes were used to cluster central neighborhoods. 18 features related to urban functions were determined as input parameters in the clustering analyzes. Results showed that cluster analysis can be used in urban studies and determine the development status of cities. It is important to carry out clustering studies to make urban planning by revealing the development differences between the neighborhoods and to provide more appropriate service delivery.

Keywords: *Urban function, K-Means Clustering, Hierarchical Clustering, OPTICS Clustering, Urban Development*

Farklı Kümeleme Algoritmaları ile Kentsel Fonksiyonlara ve Gelişme Düzeylerine Göre Mahallelerin Kümelenmesi: Konya İli Örneği

ÖZ: İnsan faktörünün etkisi altında ortaya çıkan ve zaman içinde gelişim sürecinde olan kentsel fonksiyonlar/faaliyetler, mahallelerin gelişmesinde önemli rol oynamaktadır. Mahalleler arasında dengeli bir gelişme durumu sağlamak için mahallelerin gelişmişlik düzeylerinin önceden bilinmesi gerekmektedir. Bu çalışma, Konya'daki 167 merkez mahallenin kentsel donatılar açısından kümelenmesine odaklanmaktadır ve bu mahallelerin gelişmişlik durumlarındaki benzerlikleri veya farklılıkları ortaya koymaktadır. Merkez mahalleleri kümelemek için K-ortalamalar, Hiyerarşik ve OPTICS kümeleme analizleri kullanılmıştır. Kümeleme analizlerinde kentsel fonksiyonlara ilişkin 18 özellik girdi parametresi olarak belirlenmiştir. Sonuçlar, kümeleme analizinin kentsel çalışmalarda kullanılabileceğini ve kentlerin gelişmişlik durumunu belirleyebileceğini göstermiştir. Mahalleler arasındaki gelişmişlik farklılıklarını ortaya çıkararak kentsel planlama yapmak ve daha uygun hizmet sunumu sağlamak için kümelenme çalışmaları yapılması önemlidir.

Anahtar Kelimeler: *Kentsel Fonksiyon, K-Means Kümeleme, Hiyerarşik Kümeleme, OPTICS Kümeleme, Kentsel Gelişim*

1. INTRODUCTION

Nowadays, cities are residential areas where various life conditions come together with social, demographic, economic, security and political services. Activities in cities are closely related to the diversity, planning and design of settlements (Sisman and Aydinoglu, 2020). Urban functions that are moving, lively, and in constant development and change can be separated into two pieces. The first piece is the factors that meet the needs of people such as housing, education, health, transportation and industry. The second piece is the other factors where social activities that increase the urban comfort and quality of life such as recreation areas and green areas that enrich human life are carried out (Gündüz, 2019). Urban functions, which are indicators of environmental quality, sustainable cities, livable society and level of development, are considered as criteria in the evaluation of the satisfaction, education, culture, consciousness and life conditions of the citizens with their use and regional distribution (Buffel and Phillipson, 2016). The effective and continuous use of these functions is directly related to the cultural level of the society, education and average age of the citizens. It is also an indicator of socio-economic development.

Socio-economic development is synonymous with economic development. The way to improve the quality of life is through economic development (Çetin and Sevüktekin, 2016). However, the quality of life is not always sufficient level everywhere. Because there are differences in development or life quality between provinces/neighborhoods. While some neighborhoods in a province are more developed than others, it is observed that others cannot show the same performance. To ensure balanced socio-economic development among the neighborhoods, first of all, it is necessary to determine the development levels of the neighborhoods with measurable and relatively comparable socio-economic indicators. Determining the level of development provides information on how much and in which area investment should be made in which region or neighborhood (Artmann *et al.*, 2019). In addition, revealing the development in terms of regions and neighborhoods is extremely important in terms of determining the success of the policies implemented so far, identifying and correcting inappropriate policies. There are many studies in the literature on the level of socio-economic development. In the studies carried out, the socio-economic development order of the settlements at different levels is revealed together with the reasons, and it is aimed to develop appropriate regional development policies in this way (Snieška and Šimkūnaitė, 2009; Sakarya and İbişoğlu, 2015; Fuseini and Kemp, 2015).

The identification of urban clusters is a topic on which only limited research has been performed (Zhang *et al.*, 2012). Overall, boundaries of urban clusters are approved by setting the thresholds for the determinant criteria based on socio-economic statistical data. Uysal *et al.* (2017) aimed to reveal the provinces that show differences or similarities in terms of life index values in Turkey. Life index values; housing, working life, income and wealth, health, education, environment, security, civic participation, access to infrastructure services, social life and life satisfaction. To achieve this aim, K-means, which is one of the multivariate statistical methods, were examined by cluster analysis and the results obtained were supported by discriminant analysis. Fragkias and Seto (2009) applied the Hoshen-Kopelman algorithm to detect urban clusters in the Pearl River Delta, China, but the objective of that study was to detect the metropolitan area, including the core city and its suburbs, rather than collections of multiple cities. Karabulut *et al.* (2004) tried to determine the position of the provinces among themselves by determining the socio-economic development status of the provinces according to the "Hierarchical Cluster Analysis" method using the demographic, social and economic data of the year 2000 belonging to 81 provinces of Turkey. In other studies, the socio-economic development status of the provinces in Turkey has been examined over economic, social and cultural indicators with principal component analysis, fuzzy clustering and multidimensional scaling analysis (Arı and Hüyüktepe, 2019; Servi and Erişoğlu, 2020). In many of the existing clustering techniques, relevant clusters are detected directly over the various attribute values contained in the datasets. Namely, spatial information is ignored and similarities are only measured according to the attribute values of the data. Except those, the status can be evaluated by producing clustering maps with different data clustering techniques in determining spatial distributions. There are geographical analysis techniques that carry the clustering maps further and examine the value of any criterion simultaneously with its neighbors in the world related to the location (ESRI, 2021). The OPTICS clustering/algorithm is one of the examples that can be given to them. With OPTICS, it is possible to detect similar clusters that are spatially adjacent by considering a set of attribute values and locations of the data (URL1).

The aim of this study is to cluster the central neighborhoods in Konya in terms of urban functions/reinforcement and to reveal the regional similarities or differences of these neighborhoods. The application covers 167 central neighborhoods in Konya. Population density, education level, favorite neighborhood, building density, development potential, geological status, average unit price, green areas, noise pollution, medical, education and public institutions, security units, shopping centers, entertainment centers, industrial zones, city center and transportation datasets were used for this study. After obtaining neighborhood-scale datasets through GIS-based analyzes, K-means, Hierarchical (agglomerative) and OPTICS clustering analyzes were realized to cluster central neighborhoods with similar urban functions.

2. MATERIALS AND METHODS

2.1. Study area

Konya province, which is located in the Central Anatolia Region, which is one of the 7 regions of Turkey, neighbors the provinces of Ankara, Afyonkarahisar, Antalya, Eskişehir, Karaman, Mersin, Isparta, Niğde and Aksaray. Konya, which has 31 districts in total, consists of three central districts, namely Selçuklu, Meram and Karatay. The location where the borders of these districts intersect is described as the city center. When examined in terms of population and area, Konya province is the largest province of Turkey with an area of 38.257 and is the 7th most populous province of Turkey with a total population of 2.277.017 (TÜİK, 2022).

For clustering studies, 167 central neighborhoods of Meram, Selçuklu and Karatay were selected as the study area. In determining the study area, education (university, primary and high school, etc.), medical (health center, hospital, etc.) and public institutions (municipality, service building, etc.), shopping and entertainment/cultural centers, industrial zones, security units, green areas and transportation networks were taken into account (Fig. 1).

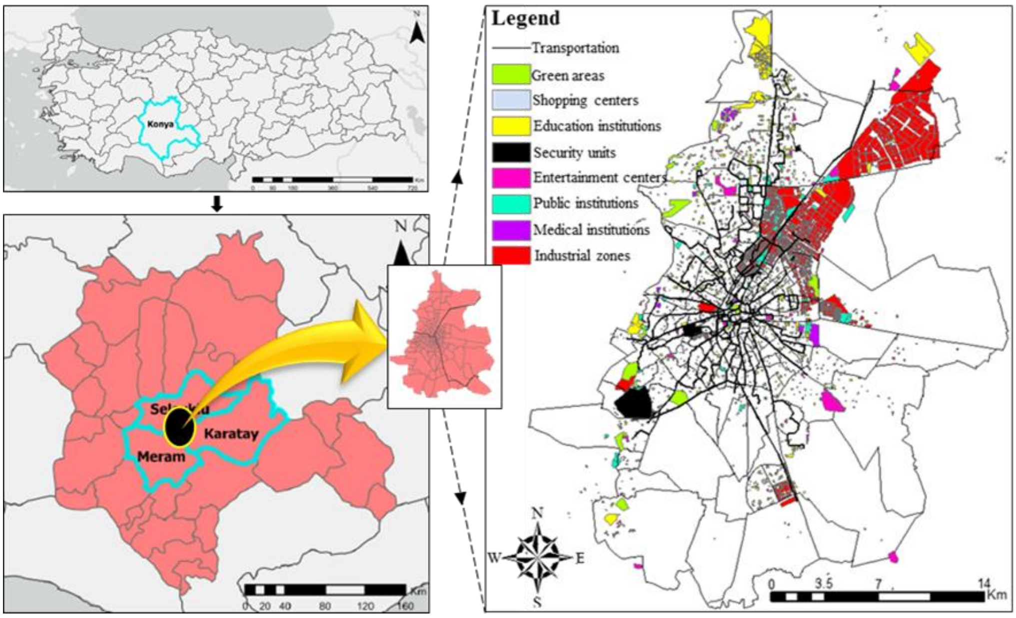

Figure 1. Determining the study area

2.2. Determination of features

Researches in the field of development levels of neighborhoods depending on urban functions are especially important in terms of increasing the speed of socio-economic development and increasing-disseminating social welfare. For regional development to take place, in addition to the macroeconomic policies being made to cover the whole, projections that determine the geographical or spatial distribution of development within the region are also needed. The discussion of which features/factors will be used in evaluating neighborhoods in terms of urban activities is constantly made in the literature. It has brought along the search for features that affected in the development of neighborhoods in line with urban reinforcement.

As a result of research and literature review, criteria affecting the development level of neighborhoods were determined. 18 features (population density, education level, favorite neighborhood, building density, development potential, geological status, average unit price, green areas, noise pollution, medical, education and public institutions, security units, shopping centers, entertainment centers, industrial zones, city center and transportation) were used in this study (Fig. 2). The descriptions of the features, units and the maximum and minimum scale values of the features are given in Table 1.

Table 1. Features, description of the features, units and scale value (min-max) of the features

Features	Description	Units	Scale value (min-max)
Population density	*Population/neigh. area*	%	0.35 - 303.08
Education level	*Literacy rate*	%	0.13 - 1.00
Favorite neighborhood	*Development score*	score	1.00 - 5.00
Building density	*Total building/neigh. area*	%	0.14 - 153.74
Development potential	*Advancement potential*	score	22.00 - 91.00
Geological status	*Ground condition*	score	2.50 - 5.00
Average unit price	*Average unit price*	TL/m^2	1.515 - 7.403
Green areas	*Park etc.*	number	0 - 11+
Noise pollution	*Noise level*	score	0.00 - 4.00
Medical institutions	*Healthcare center*	number	0 - 11+
Education institutions	*Education center*	number	0 - 61+
Public institutions	*Government agencies*	number	0 - 27+
Security units	*Security units*	number	0 - 7+
Shopping centers	*Shopping center*	number	0 - 7+
Entertainment centers	*Activity center*	number	0 - 33+
Industrial zones	*Industrial areas*	m^2	0 - 626
City center	*Distance to city center*	meter	459 - 16661
Transportation	*Transportation line (stops)*	number	0 - 43

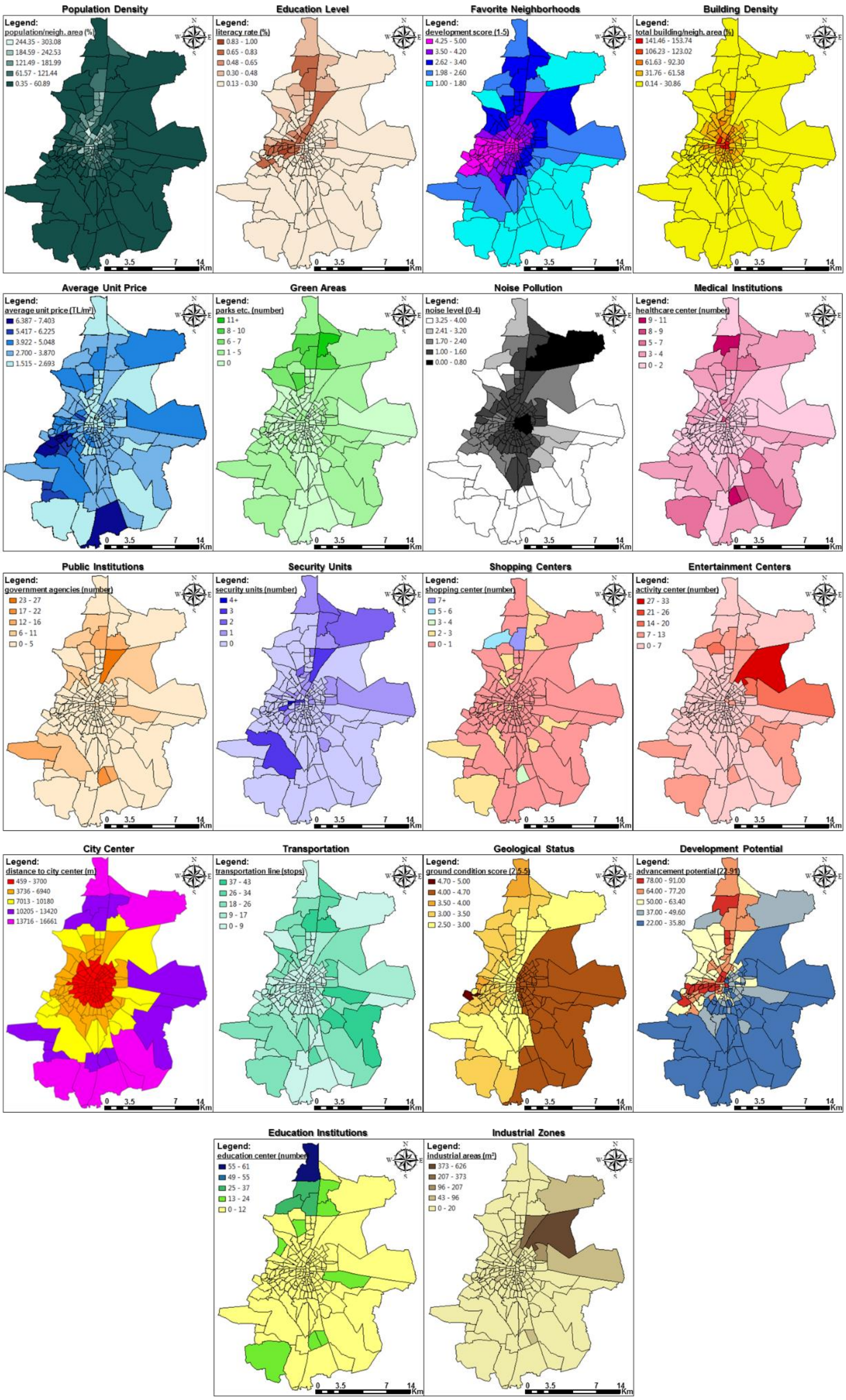

Figure 2. Features used for the study

2.3. Cluster Analysis

Cluster analysis, which is among the multivariate statistical methods, used to group data in a large number of complex datasets and to compare the groups formed, is a method that is frequently used because it is easy to apply and the results are understandable (Everitt *et al.*, 2001). Nowadays, there are many clustering algorithms with different structures. The main purpose of cluster analysis is to classify and make sense of a group of data whose cluster is unknown. Briefly, cluster analysis is used to classify variables according to their basic features. Within the scope of this study, K-Means, Hierarchical and OPTICS clustering are discussed.

2.3.1. K-Means Clustering

K-means algorithm is one of the unsupervised learning and data-clustering model developed by J. B. MacQueen (1967). The purpose of the algorithm is to ensure that the clusters obtained at the end of the clustering process have maximum similarities within clusters and minimum similarities between clusters. The application process begins with grouping the features according to a predetermined number of clusters. K-means treats the mean vectors of the features as the cluster center and the clustering process is shaped around it (Aldino *et al.*, 2021). It aims to divide the features into the dataset into k clusters in a way that minimizes the sum of squares within the cluster. A distinctive observation is then selected to represent each cluster itself, and similar features are clustered around the distinctive observation. The process of placing features into clusters is done iteratively. Variables are assigned to different clusters at each iteration, and the optimal solution is determined in a way similar to the permutation approach. Given a series of variables (x_1, x_2, …, x_n), where each variable is an a-dimensional actual vector, k-means algorithm purposes to division the n variables into k (≤ n) series S = {S_1, S_2, …, S_k} to reduce the Within-Cluster Sum of Squares (WCSS). Consequently, the purpose is to calculate (Kriegel *et al.*, 2017) (1):

$$argmin \sum_{i=1}^{k} \sum_{x \in S_i} \| x - \mu_i \|^2 = argmin \sum_{i=1}^{k} |S_i| Var S_i \qquad (1)$$

Where μ_i is the average of points in S_i. This is equation to decreasing the pairwise squared deviations of points in the same cluster (2):

$$\arg min \sum_{i=1}^{k} \frac{1}{|S_i|} \sum_{x,y \in S_i} \| x - y \|^2 \qquad (2)$$

In addition to the easy use and evaluation of the k-means, it is more preferred in datasets with continuous structure. However, the algorithm has some restrictions of its own. If the observations in the dataset to exhibit an asymmetric structure or contain many outliers, a suitable mean position parameter cannot be found. Therefore, cluster centers cannot accurately represent their members, and also a homogeneous structure may not be formed even if clusters due to the smallest differences are formed.

2.3.2. Hierarchical Clustering

The main idea of the hierarchical clustering algorithms is based on the combination of similar objects or vice versa. According to this case, there are two basic structures in application: agglomerative and divisive approaches. Agglomerative clustering algorithm start the analysis by assuming that all objects in the data set form a different cluster. In this method, n objects in the dataset are hierarchically, respectively; it aims to place x clusters, x-2 clusters, ..., x-r clusters, ..., 3 clusters, 2 clusters, and 1 cluster. In practice, each object is initially considered as a separate set. Two objects that are highly similar to each other form a cluster. Then, other features with different similarity levels are added to this cluster and they are connected with each other in such a way that all the objects are collected in a cluster (Davidson and Ravi, 2005). Contrary to the agglomerative approach, a discriminatory strategy is dominant in the divisive approach. In this approach, there is only one cluster at the beginning. The objects are separated from the top-cluster according to the euclidean distance, and different sub-clusters are formed. As a result of the process, each data becomes a cluster. Hierarchical clustering algorithms are a clustering approach that is widely used in shaping the clusters formed by the objects, in terms of showing at what stage and at what level of similarity they form clusters with common objects (Fernández and Gómez, 2008).

2.3.3. OPTICS Clustering

OPTICS (Ordering Points to Identify the Clustering Structure) is a clustering algorithm for identifying clusters of varying density, including spatial data. It was proposed by Ankerst *et al.* (1999) at the SIGMOD'99 conference. OPTICS can be considered as an enhancement of the DBSCAN algorithm rather than an algorithm on its own. OPTICS can save the ε (Eps) value, which is the biggest weakness of DBSCAN, from being entered by the user, and can change this value dynamically while the algorithm is running. In this way, the algorithm can find and display one extra distance parameter of the DBSCAN algorithm in a single run. OPTICS searches two parameters: The first of these is the parameter ε, which defines the maximum radius. The other is the *MinPts* parameter, which defines the number of points required to create a cluster (Campello *et al.*, 2020). A point p is a core point if at least *MinPts* points are found within its ε-neighborhood $N_\varepsilon(p)$. Since OPTICS also considers points that are part of a denser set, it takes two more terms into account for each point. These are core distance and reachability distance. Each point is assigned a core-distance that defines the distance to the $MinPts^{th}$ closest point (URL2) (3):

$$\text{Core\_dist}_{\boldsymbol{\varepsilon,MinPts}}(p) = \begin{cases} \text{undefined} & \text{if } |N_\varepsilon(p)| < MinPts \\ MinPts^{th} \text{ smallest distance in } N_\varepsilon(p) & \text{otherwise} \end{cases} \quad (3)$$

The reachability-distance of another point o from a point p is either the distance between o and p, or the core distance of p, whichever is bigger (4):

$$\text{Reachability\_dist}_{\boldsymbol{\varepsilon,MinPts}}(o,p) = \begin{cases} \text{undefined} & \text{if } |N_\varepsilon(p)| < MinPts \\ \max\left(\text{core}_{\text{dist}_{\boldsymbol{\varepsilon,MinPts}}}(p), \text{dist}(p,o)\right) & \text{otherwise} \end{cases} \quad (4)$$

If o and p are nearest neighbors, this is the ε-prime < ε we need to assume to have p and o belong to the same cluster. Both core-distance and reachability-distance are undefined if no sufficiently dense cluster is available (Breunig *et al.*, 1999). In other words, while the algorithm will find all the clusters with the density determined by the chosen values for ε and *MinPts*, it may miss higher-density clusters that are contained in these clusters. These higher-density clusters become visible only at some ε-prime < ε. The problem is that there is no way of knowing these ε-prime values in advance, so all one can do is run DBSCAN for as many ε-prime values as feasible. OPTICS solves this problem by ordering the points in the dataset and by associating with each point two values (URL3).

3. RESULTS

In this study, clustering algorithms belonging to different structures were used to group central neighborhoods (167 neighborhoods) in Konya in terms of urban functions/reinforcement and to reveal the regional similarity or differences of these neighborhoods (Fig. 3). After obtaining neighborhood-scale datasets through GIS-based analyzes, K-means, Hierarchical (agglomerative) and OPTICS clustering analyzes were realized to cluster central neighborhoods with similar urban functions. 18 features related to urban functions were considered as input parameters in the clustering analysis. Clustering processes were applied for the study area, taking into account various features according to the urban development status of the neighborhoods.

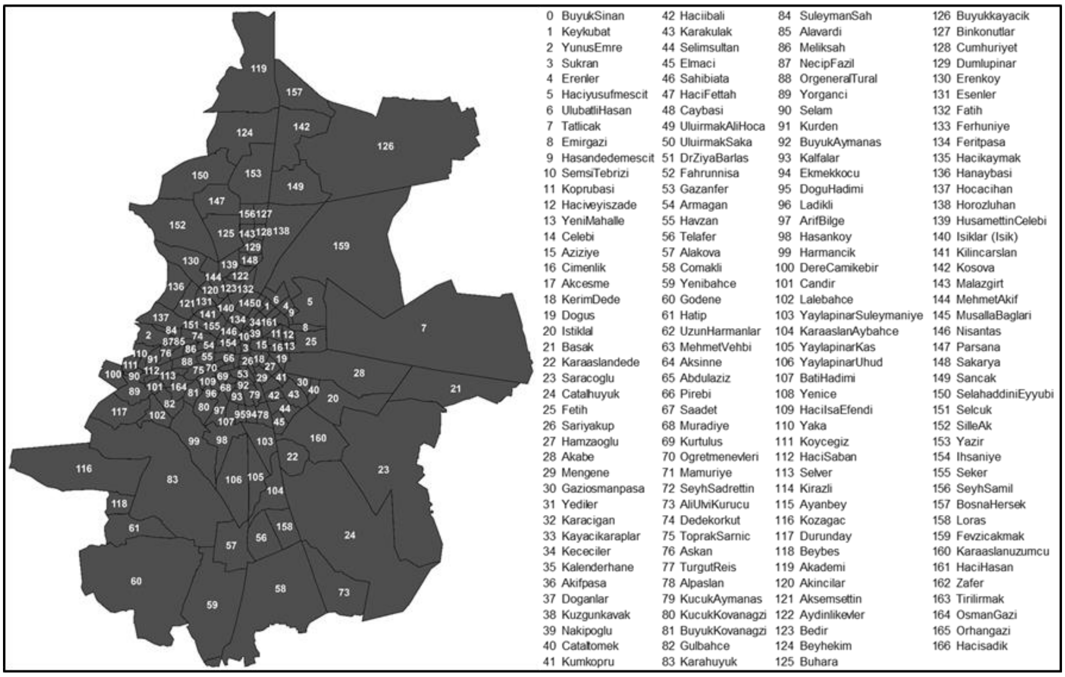

Figure 3. The numbering of 167 central neighborhoods in Konya

3.1. Clustering of neighborhoods with the K-Means algorithm

The K-Means clustering was performed for the determination of central neighborhoods with similar development functions and urban features. A clustering process was carried out with the features showing the development level of neighborhoods and the results of the analysis were interpreted. An open-source software, Python v3.8 program language, was preferred to perform the clustering analysis. Firstly, python libraries were imported for K-means clustering and datasets (18 features) belonging to 167 neighborhoods were included in the application. To cluster the data, the optimal K value must be found. Here, the Elbow Method was used to find the optimal K value. WCSS value was calculated for each K value. Then WCSS was shown on the graph together with the K value (Fig. 4).

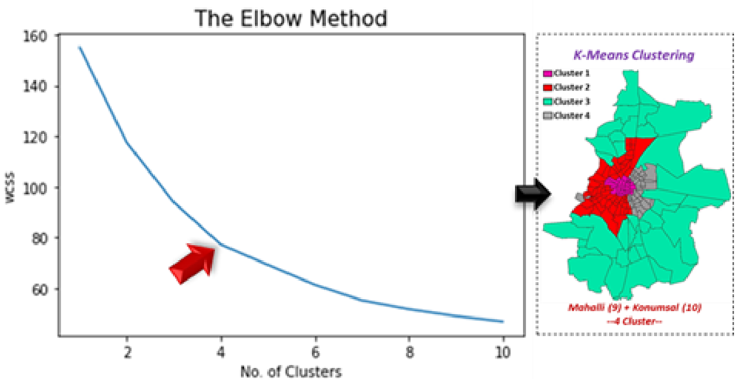

Figure 4. The optimal number of clusters according to the Elbow Method

The point at which the elbow shape is created is 4, that is, our K value or an optimal number of clusters is 4. As a result of the analysis, the number of 4 clusters determined by the elbow method was not used, since the optimal number of clusters could not reflect the diversity in the urban structure of Konya. In the study, it is purposed to determine a common number of clusters to be able to compare the clusters that will be formed with different clustering algorithms. Therefore, the number of 5 clusters that could reveal different socio-development for the study area was decided. In addition, it was concluded that this number of clusters (5) could be used as a standard in other clustering analyses in the study. The purpose is also to display the effect of different clustering analyzes. Neighborhood clusters formed as a result of K-Means clustering are shown in Fig. 5.

It was observed that there were 25 neighborhoods in Cluster-1 (red), 21 neighborhoods in Cluster-2 (yellow), 50 neighborhoods in Cluster-3 (grey), 52 neighborhoods in Cluster-4 (green), and 19 neighborhoods in Cluster-5 (blue). In the regions close to the city center in Konya, three different clusters have been observed and these are divided as Cluster-1, Cluster-3 and Cluster-4. Cluster-2 and Cluster-5 are located in the north and south of the study area, respectively.

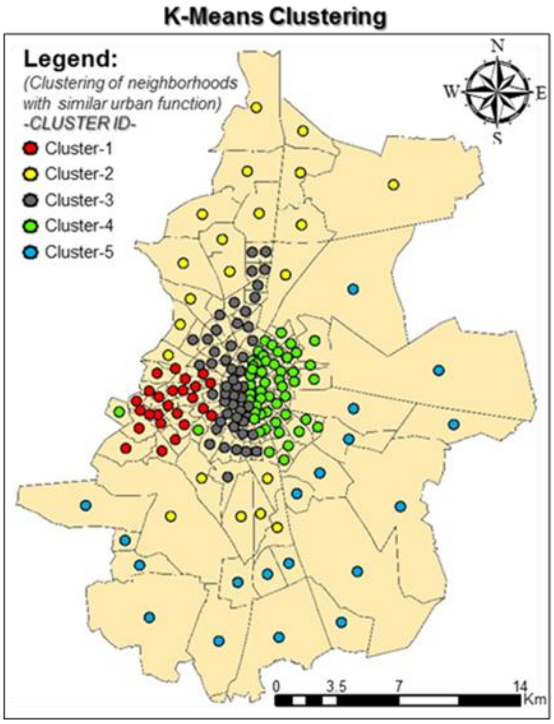

Figure 5. Clustering of neighborhoods with similar urban function (K-means)

3.2. Clustering of neighborhoods with the Hierarchical method

Within the scope of the study, the clustering was performed using the hierarchical (agglomerative) algorithm. At first, NumPy, Pandas and Matplotlib libraries in Python v3.8 were imported for the application. In addition, the Scikit-Learn library was used for agglomerative hierarchical clustering analysis, and the SciPy library was used for Dendrogram. In hierarchical clustering analyses, dendrograms were preferred to determine the number of clusters or to easily interpret the results obtained.

While generating the dendrogram for the related dataset, "the linkage method of Ward" was used to cluster. The euclidean distance between the clusters is seen on the y-axis of the produced dendrogram. On the x-axis, there are the ID (label) numbers of the data points that form the clusters (Fig. 6). When the y-axis is examined, there are 4 dashed lines. Threshold was applied because we wanted the distance between our final clusters to be at least 4. As can be seen, this lower border crosses 4 vertical columns. This actually says that there are 4 clusters representing our dataset.

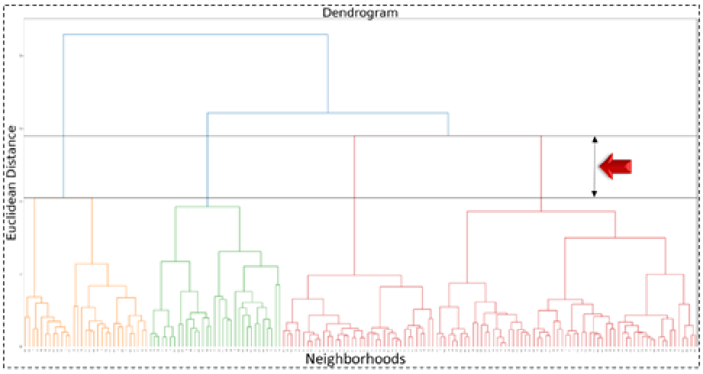

Figure 6. Dendrogram representation

As in K-Means algorithm, the optimum number of clusters was determined as 5 in this clustering analysis. Already at least 4 clusters were found through dendrograms. For this reason, the neighborhoods are divided into 5 clusters (Fig. 7). It was observed that there were 25 neighborhoods in Cluster-1 (red), 18 neighborhoods in Cluster-2 (yellow), 51 neighborhoods in Cluster-3 (grey), 53 neighborhoods in Cluster-4 (green), and 20 neighborhoods in Cluster-5 (blue). In the regions close to the city center in Konya, three different clusters have been observed and these are divided as Cluster-1, Cluster-3 and Cluster-4. Cluster-2 and Cluster-5 are located in the north and south of the study area, respectively.

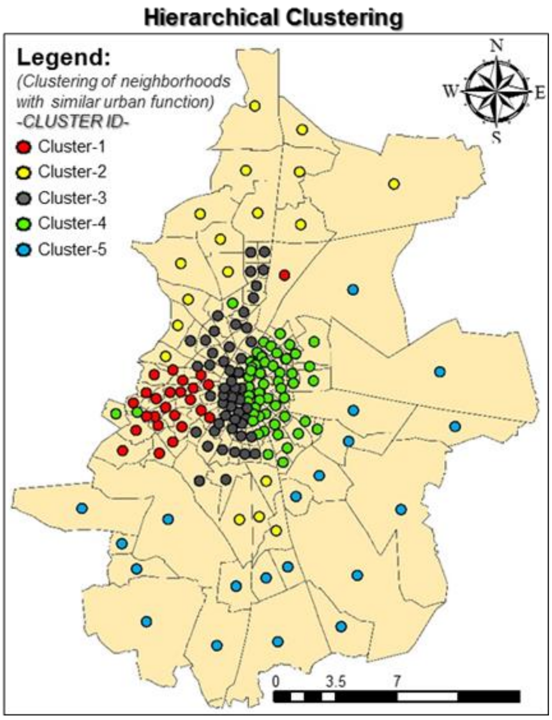

Figure 7. Clustering of neighborhoods with similar urban function (Hierarchical)

Similar results were obtained from K-Means and Hierarchical clustering. In both clusters (Cluster-2 and Cluster-4), there are neighborhoods with similar features in different regions. Therefore, neighborhoods that are not spatially dependent on each other can coexist as a result of similar features (depending on attributes).

3.3. Clustering of neighborhoods with the OPTICS algorithm

In this study, ArcGIS Pro 2.8 was used to implement the OPTICS algorithm. The distance between neighbors and reachability values were determined to separate clusters of varying densities from noise. OPTICS offers the most flexibility in fine-tuning the clusters that are detected, though it is computationally intensive, particularly with a large search distance. The max_search_dist parameter was preferred as the search_dist for the OPTICS option in this study. Min_features_cluster to be considered as a cluster was chosen as 10 according to the distinguishability of neighborhood features. As selected in K-means and Hierarchical clustering, the optimum number of clusters was determined as 5 and, the OPTICS clustering results were given in Fig. 8.

OPTICS (multi-scale) Clustering

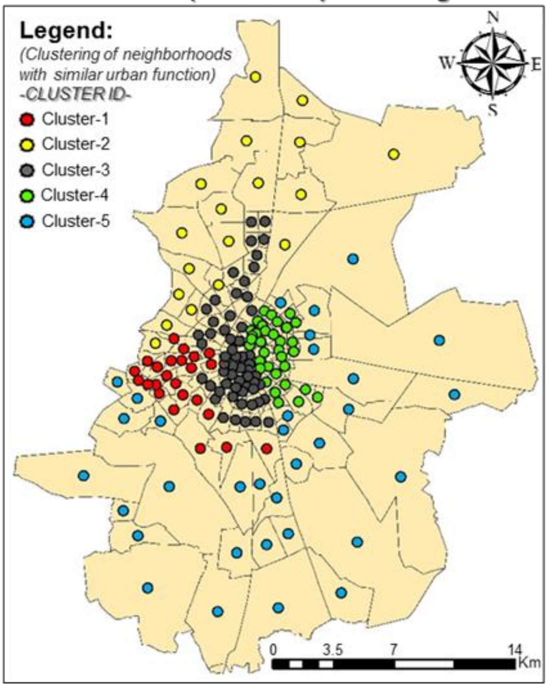

Figure 8. Clustering of neighborhoods with similar urban function (OPTICS)

It was shown that there were 24 neighborhoods in Cluster-1 (red), 18 neighborhoods in Cluster-2 (yellow), 58 neighborhoods in Cluster-3 (grey), 34 neighborhoods in Cluster-4 (green), and 33 neighborhoods in Cluster-5 (blue). Compared to other clustering methods, different neighborhoods have come together in OPTICS. The reason is that: while there are similarities between spatial and non-spatial (such as K-means and Agglomerative method) statistics in terms of concepts and objectives, spatial statistics are unique in that they were developed specifically for use with geographic data. Unlike non-spatial statistical methods, they incorporate space (proximity, area, connection or other spatial relationships) directly into their mathematics.

3.4. Evaluation of different clustering algorithms

In this study, the clustering results were obtained with three different methods. It has been examined whether the central districts are in the same cluster (according to different methods) in terms of urban functions. When each clustering technique was evaluated individually, the optimal number of clusters found differed (4 for k-means clustering, 4 for hierarchical clustering, and 5 for OPTICS clustering). However, it was noticed that the number of clusters found decreased the cluster score. Therefore, optimal cluster numbers determined as a result of clustering were not used as they could not reflect the diversity in Konya's urban structure. So, the number of 5 clusters that could reveal different socio-development for the study area was decided. Clustering results of neighborhoods whose number order was determined before (as seen in Fig. 3) are shown symbolically in Fig. 9.

In the hierarchical clustering, a total of 6 central neighborhoods (81, 83, 90, 99, 138 and 139) are located in different clusters according to k-means, while other neighborhoods are located in the same clusters. In these methods, neighborhood relations (spatial relations of neighborhoods with each other) were not taken into account, and clustering was carried out only on attribute data. For the OPTICS, more different neighborhood clusters appeared in the study area. The reason is that, as expressed earlier, spatial clustering methods incorporate proximity, area, connection, and/or other spatial relationships directly into their mathematics. Therefore, neighborhood relations of neighborhoods were also taken into account in clustering analyses. The most important criteria that distinguish OPTICS from k-means and hierarchical methods are that it is involved in its spatial dimension.

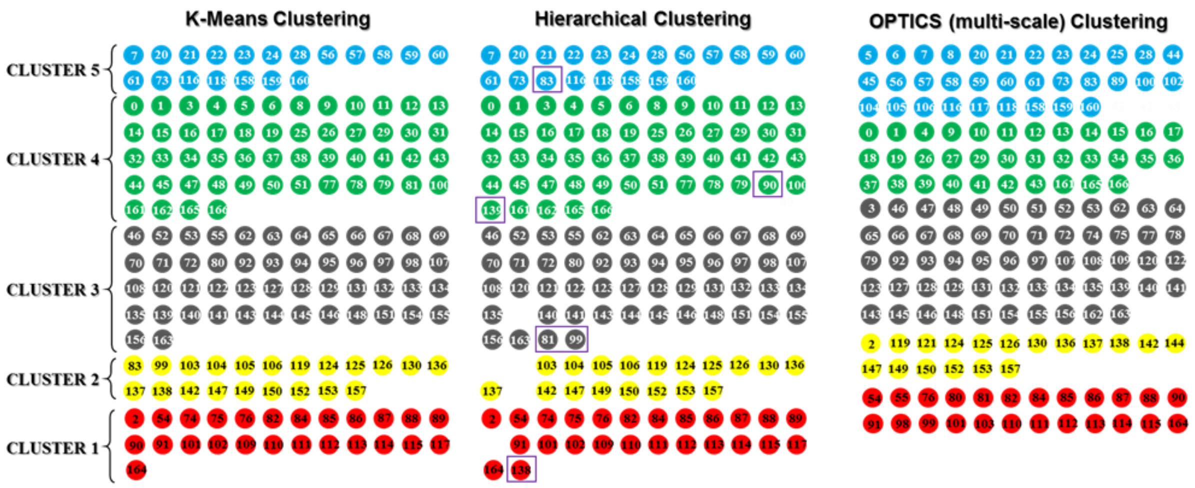

Figure 9. Clustering results (Each number corresponds to a Neighborhoods-ID)

Finally, the cluster score was calculated for each cluster to interpret the degree of discrimination of the clusters and the extent to which the neighborhoods were separated according to urban functions. The following procedure was followed while calculating the cluster scores. In the first stage, each of the 18 features was scored between 0 and 100 according to the degree of influence in the neighborhood. In other words, if a feature has a sufficient level in a neighborhood, it is scored as 100 or 0 if it does not exist. There is, but if it is not at a sufficient level, it has taken the corresponding value in the range of 0-100. The total score for each neighborhood was obtained by adding the scores from each feature. Then, the total scores of the neighborhoods within the same cluster were added and averaged, and an average score was found for Clusters 1, 2, 3, 4 and 5 (Table 2).

Table 2. Average score calculated for each cluster

Methods	C-1 Score (Avg.)	C-2 Score (Avg.)	C-3 Score (Avg.)	C-4 Score (Avg.)	C-5 Score (Avg.)
K-Means	80.50	76.64	74.54	73.36	70.22
Hierarchical	81.52	76.26	73.68	73.28	69.66
OPTICS	84.50	82.72	74.40	73.36	68.48

Note: *C: Cluster; Avg: Average*

It was concluded that the most distinctive clustering was obtained with OPTICS. While deciding on this result, it was examined whether each neighborhood was in the right cluster. In addition, the fact that the highest and lowest cluster scores and differences are in OPTICS clustering indicates that neighborhoods are better grouped according to urban functions.

4. CONCLUSION

In cities that grow and develop with population growth, there has been a development level that differs in settlement areas depending on the proximity to existing urban function and socio-economic structure. This process has developed in line with the changing policies and investment trends over time, not depending on direct planning. However, it is not correct to move the population in one direction only towards regions with high development levels. Because, depending on the population, the current needs may not be met and segregation among citizens in regions with different levels of development has a negative effect. Therefore, it is necessary to determine the adequacy of urban functions by making the distribution within the framework of population needs. The situation of the regions/neighborhoods relative to each other in terms of urban functions should be known beforehand.

In this study, clustering analyzes belonging to different structures were used to group 167 central neighborhoods in Konya in terms of urban functions and to reveal the regional similarity or differences of these neighborhoods. After obtaining neighborhood-scale datasets through GIS-based analyzes, K-means, Hierarchical (agglomerative) and OPTICS clustering analyzes were realized to cluster central neighborhoods with similar urban functions. 18 features related to urban functions were considered as input parameters in the clustering analysis. Clustering processes were applied for the study area, taking into account various features according to the urban development status of the neighborhoods. Results showed that cluster analysis can be used in such studies. It is recommended to determine the economic, social and technological changes experienced in the society and the service demands that develop / may develop in parallel with this. It is important to carry out such clustering studies to eliminate the socio-economic development differences in the neighborhoods and to target more fair-egalitarian service delivery.

REFERENCES

Arı, E., Hüyüktepe, B., 2019, "Sosyo-ekonomik göstergeler için çok değişkenli veri analizi: Türkiye için ampirik bir uygulama", *Aksaray Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi,* 11, 7-20.

Aldino, A. A., Darwis, D., Prastowo, A. T., Sujana, C., 2021, Implementation of K-means algorithm for clustering corn planting feasibility area in south lampung regency, *In Journal of Physics: Conference Series*, 1751, 1, 012038, IOP Publishing.

Ankerst, M., Breunig, M. M., Kriegel, H. P., Sander, J., 1999, OPTICS: Ordering points to identify the clustering structure, *ACM SIGMOD record*, 28(2), 49-60.

Artmann, M., Inostroza, L., Fan, P. 2019, Urban sprawl, compact urban development and green cities, How much do we know, how much do we agree?, *Ecological indicators*, 96, 3-9.

Breunig, M. M., Kriegel, H. P., Ng, R. T., Sander, J. 1999, Optics-of: Identifying local outliers, *In European Conference on Principles of Data Mining and Knowledge Discovery*, 262-270, Springer, Berlin, Heidelberg.

Buffel, T., Phillipson, C., 2016, Can global cities be 'age-friendly cities'? Urban development and ageing populations, *Cities*, 55, 94-100.

Campello, R. J., Kröger, P., Sander, J., Zimek, A., 2020, Density-based clustering, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(2), e1343.

Çetin, İ., Sevüktekin, M., 2016, Türkiye'de gelişmişlik düzeyi farklılıklarının analizi, *Uluslararası Ekonomik Araştırmalar Dergisi*, 2(2), 39-61.

Davidson, I., Ravi, S. S., 2005, Agglomerative hierarchical clustering with constraints: Theoretical and empirical results, *In European Conference on Principles of Data Mining and Knowledge Discovery*, 59-70. Springer, Berlin, Heidelberg.

ESRI, 2021, Data classification methods (https://pro.arcgis.com/en/pro-app/latest/help/mapping/layer-properties/data-classification-methods.html).

Everitt, B. S., Landau, S., Leese, M., 2001, Cluster analysis arnold, *A member of the Hodder Headline Group*, London, 429-438.

Fernández, A., Gómez, S., 2008, Solving non-uniqueness in agglomerative hierarchical clustering using multidendrograms, *Journal of Classification*, 25(1), 43-65.

Fragkias, M., Seto, K. C., 2009, Evolving rank-size distributions of intra-metropolitan urban clusters in South China. *Computers, Environment and Urban Systems*, 33(3), 189-199.

Fuseini, I., Kemp, J., 2015, A review of spatial planning in Ghana's socio-economic development trajectory: A sustainable development perspective, *Land use policy*, 47, 309-320.

Gündüz, M. Y., 2019, Ataköy Ve Şirinevler Mahallelerinin Kentsel Donatı Elemanlarının Kent Kimliği Açısından Değerlendirilmesi, *The Journal of Turk & Islam World Social Studies*, 23, 255-281.

Karabulut, M., Gürbüz, M., Sandal, E. K., 2004, Hiyerarşik Kluster (Küme) Tekniği Kullanılarak Türkiye'de İllerin Sosyo-Ekonomik Benzerliklerinin Analizi. *Coğrafi Bilimler Dergisi*, 2(2), 65-78.

Kriegel, H. P., Schubert, E., Zimek, A., 2017, The (black) art of runtime evaluation: Are we comparing algorithms or implementations?, *Knowledge and Information Systems*, 52(2), 341-378.

MacQueen J. B., 1967, Some Methods for classification and Analysis of Multivariate Observations, *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, *Berkeley*, University of California Press, 1:281-297.

Sakarya, Â., İbişoğlu, Ç., 2015, Türkiye'de İllerin Sosyo-Ekonomik Gelişmişlik Endeksinin Coğrafi Ağırlıklı Regresyon Modeli İle Analizi, *Marmara Coğrafya Dergisi*, 32, 211-238.

Servi, T., Erişoğlu, Ü., 2020, Türkiye'deki şehirlerin sosyo-ekonomik gelişmişlik düzeylerinin istatistiksel analizi. *Al Farabi Uluslararası Sosyal Bilimler Dergisi*, 5(2), 174-186.

Sisman, S., Aydinoglu, A. C., 2020, Using GIS-based multi-criteria decision analysis techniques in the smart cities, *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.

Snieška, V., Šimkūnaitė, I., 2009, Socio-economic impact of infrastructure investments, *Inžinerinė ekonomika*, 3, 16-25.

TÜİK, 2022, Türkiye İstatistik Kurumu, Nüfus ve Demografi İstatistikleri, *2022 Yılı Adrese Dayalı Nüfus Kayıt Sistemi Sonuçları*, https://cip.tuik.gov.tr/#.

Uysal, F. N., Ersöz, T., Ersöz, F., 2017, Türkiye'deki illerin yaşam endeksinin çok değişkenli istatistik yöntemlerle incelenmesi. *Ekonomi Bilimleri Dergisi*, 9(1), 49-65.

Zhang, Q., Hu, Y., Liu, J., Liu, Y., Ren, W., Li, J., 2012, A quantitative assessment of the distribution and extent of urban clusters in China. *Journal of Geographical Sciences*, 22(1), 137-151.

https://pro.arcgis.com/en/pro-app/2.8/tool-reference/spatial-statistics/, (09.09.2022), URL1.

https://en.wikipedia.org/wiki/OPTICS_algorithm#cite_note-1 , (27.04.2022), URL2.

https://nodepit.com/node/org.knime.base.node.mine.optics.compute.OPTICSComputeNodeFactory , (26.04.2022), URL3.