AÜİİBFD

Anadolu Üniversitesi İktisadi ve İdari Bilimler Fakültesi

Anadolu University Journal of Economics and Administrative Sciences

Cilt: 25, Sayı: 2, Yıl: 2024
e-ISSN: 2687-184X
https://dergipark.org.tr/en/pub/anadoluibfd

Makale Türü / Submission Type: Araştırma Makalesi / Research Article
Başvuru Tarihi / Date Submitted: 17/08/2023
Kabul Tarihi / Date Accepted: 28/01/2024

DOI: 10.53443/anadoluibfd.1344618
2024, 25(2), 135-157

EVALUATION OF PROVINCES IN TÜRKİYE WITH HEALTH INDICATORS BY DENSITY-BASED SPATIAL CLUSTERING ANALYSIS

Ahmet Bahadır ŞİMŞEK[1]

Abstract

This study aims to evaluate the health resource distribution of provinces in Turkey using DBSCAN cluster analysis method. The optimum values of DBSCAN parameters (epsilon and minPts) were tested by simulation and the clustering silhouette value was taken as the basis for selecting the appropriate parameter set. The results of the descriptive statistical analysis of the dataset show a high coefficient of variation, indicating inequalities in the distribution of health resources. By dividing provinces into two clusters, the study reveals the similarity of local dynamics in the inequality of resource distribution. The findings provide important insights for relevant stakeholders to address the disparities between provinces in Turkey. The fact that the study adopts a method other than the hierarchical and k-means clustering methods dominant in the literature and that the codes of the algorithm are shared in Python language broadens the horizons of the relevant researchers and increases the transparency and reproducibility of the study.

Keywords: Healthcare Resource Distribution, Density-Based Spatial Clustering, DBSCAN, Regional Disparities

JEL Codes: I11, I18, R23, C88

YOĞUNLUK TABANLI MEKÂNSAL KÜMELEME ANALİZİ İLE TÜRKİYE'DEKİ İLLERİN SAĞLIK GÖSTERGELERİYLE DEĞERLENDİRİLMESİ

Öz

Bu çalışma DBSCAN kümeleme analizi yöntemiyle Türkiye'deki illerin sağlık kaynağı dağılımını değerlendirmeyi amaçlamaktadır. DBSCAN parametrelerinin (epsilon ve minPts) optimum değerleri simülasyon ile test edilmiş uygun parametre setini seçmek için kümeleme siluet değeri baz alınmıştır. Veri setinin tanımlayıcı istatistik analiz sonuçlarında yüksek varyasyon katsayısı göze çarpmakta ve sağlık kaynakları dağılımındaki eşitsizliklere işaret etmektedir. Çalışma, illeri iki kümeye ayırarak kaynak dağılımının eşitsizliğinde yerel dinamiklerin benzerliğini ortaya koymaktadır. Bulgular, Türkiye'de iller arasındaki farklılıkların giderilmesi için ilgili paydaşlara önemli içgörüler sunmaktadır. Çalışmada ilgili literatürde baskın olan hiyerarşik ve k-means kümeleme yöntemlerinin haricinde bir yöntem benimsenmiş olması ve algoritmanın Python dilinde kodlarının paylaşılmış olması ilgili araştırmacıların ufkunu genişletmekte ve çalışmanın şeffaflığı ve tekrar edilebilirliğini artırmaktadır.

Anahtar Kelimeler: Sağlık Hizmeti Kaynak Dağılımı, Yoğunluk Tabanlı Mekansal Kümeleme, DBSCAN, Bölgesel Eşitsizlikler

JEL Kodları: I11, I18, R23, C88

[1] Dr. Öğr. Üyesi, Gümüşhane Üniversitesi Sağlık Bilimleri Fakültesi, abahadirsimsek@gumushane.edu.tr, https://orcid.org/0000-0002-7276-2376

INTRODUCTION

Equitable service delivery in health systems depends on the balanced distribution of resources and the implementation of well-structured policies (Keya, Islam, Pan, Stockwell, & Foulds, 2020). Evaluating health service delivery, especially at the level of sub-service units such as provinces, is critical to manage public health fairly by using resources effectively (Rabarison, Bish, Massoudi, & Giles, 2015). Conducting the evaluation at the provincial level is important due to the existence of local dynamics that have a significant impact on the accessibility and quality of healthcare services (Lam et al., 2020). Variations in many factors, from socioeconomics to infrastructure, result in health inequalities between provinces. An in-depth understanding of these variations is critical for implementing targeted improvements and equitably distributing resources (Khan and Hussain, 2020).

Cluster analysis emerges as a powerful tool that can cluster similar units and identify spatial patterns to overcome the challenge of assessing differences between provinces (Agterberg, Zhong, Crabb, & Rosenberg, 2020). Policymakers widely use this analytical approach due to the important insights provided by cluster analysis (Hassan and Darwish, 2021; Wartelle et al., 2021). This analytical approach not only reveals differences that were unclear at first, but also directs the allocation of health resources, guiding policymakers to customize improvement plans and thus optimize resources (Agterberg et al., 2020; Matthay et al., 2021).

Cluster analysis is the process of dividing data into clusters based on similar characteristics (Everitt et al., 2011). The differences between the well-known and frequently used K-means, Hierarchical, and DBSCAN methods can be discussed as follows (Ahmed et al., 2020; Fuchs & Höpken, 2022; Shahriar et al., 2019). K-means tends to create a certain number of clusters and divide the data between these clusters. Therefore, the number of clusters needs to be known in advance and this may pose a challenge for decision makers. Additionally, since K-means obtains clustering results by focusing on cluster centers, it is limited against noise in the data set and the identification of clusters of different shapes. Hierarchical clustering divides the data into subsets by creating a tree structure. This method offers more flexibility to decision makers because it does not require predetermining the number of clusters. However, clustering results are obtained with a user-specified disconnection threshold. Determining this threshold value may involve trial and error for analysts. The DBSCAN method offers distinct advantages to the decision maker compared to the other two methods. It can recognize dense regions in the data set and neglect outliers without needing to know the number of clusters in advance. This feature allows the data set to be analyzed flexibly. DBSCAN also stands out for its ability to distinguish noise in the data set. Here, “Noise” is a term that represents unwanted or meaningless information in the dataset (Ester et al., 1996). Especially in clustering algorithms, dealing with noise is important because these algorithms need clean data to create accurate and meaningful clusters. DBSCAN can identify noise and identify clusters without taking it into account, allowing it to better deal with uncertainty in data sets. DBSCAN method was used in this study due to its mentioned advantages.

This study offers useful policy recommendations by evaluating the health service provision of provinces in Turkey with a cluster analysis approach. To achieve this goal, an analysis was conducted on the ratio of population to healthcare personnel and infrastructure assets in 81 provinces of Turkey for 2021. The aim is to reveal patterns among provinces according to health indicators using DBSCAN cluster analysis.

The rest of the study is organized as follows. Section 2 discusses similar studies on the subject. The methodology of the study is presented in Section 3. Section 4 exhibits the findings of the analysis, while conclusions and recommendations are presented in Section 5.

RELATED RESEARCHES

Cluster analysis is a widely used management approach in evaluating health systems. In this section, studies conducted with cluster analysis in the field of health are discussed.

Revealing the textures within the data set is the main task of cluster analysis. Schaefers et al. (2022) and Manortey et al. (2014) are ideal examples for those who want to observe the ability of cluster analysis to discern patterns in healthcare use. Schaefers et al. While focusing on demographic determinants related to cancer prevention in Indonesia, Manortey et al. It addresses the issue of exploring spatial record variability in insurance records in Ghana. On the other hand, Kumari and Raman (2022) and Sun et al. (2018) addresses health inequalities in specific regions. Both examined health development inequalities at the regional level using k-means and hierarchical clustering, respectively. Additionally, Ullah et al. (2020) and Korkhmazov and Perkhov (2023) used cluster analysis to understand disease dynamics. Ullah et al. While focusing on tuberculosis clusters in Pakistan, Korkhmazov and Perkhov have addressed disparities in Covid-19 death rates in Russia. Kurji et al. (2020) and Jamtsho, Corner, and Dewan (2015) are concerned with inequalities in access to healthcare. Kurji et al. examined maternal health services in Ethiopia, while Jamtsho et al. analyzed access to primary healthcare in Bhutan. The mentioned studies emphasize the versatile application practicality of cluster analysis in health research. Key characteristics about the studies are presented in Table 1.

Table 1: Clustering methods in related studies

Study	Clustering Analysis Method	Unique Characteristics
Schaefers et al., (2022)	Multivariate Statistical Approach	Identifying demographic determinants in cancer care
Manortey et al., (2014)	Kulldorff's Spatial Scan Statistic	Detecting geographic clusters in insurance enrollment
Kumari & Raman, (2022)	K-means Cluster Analysis	Grouping districts by healthcare disparities
Sun et al., (2018)	Hierarchical Cluster Analysis	Classifying provinces based on healthcare development
Ullah et al., (2020)	Space-time Scan Statistics (SaTScan)	Identifying spatial and space-time clusters in TB
Korkhmazov & Perkhov, (2023)	Hierarchical Clustering	Dividing regions by COVID-19 mortality rates
Kurji et al., (2020)	Getis Ord Gi* and Kulldorf's Method	Unveiling clusters of maternal healthcare use
Jamtsho et al., (2015)	Nearest-Neighbour Modified 2SFCA Model	Mapping spatial accessibility of primary healthcare

The studies listed above highlight the effectiveness of cluster analysis in addressing research problems ranging from identifying regional inequalities to deciphering patterns of disease spread. Cluster analysis studies that can be associated with health in Turkey are presented in Table 2.

Table 2: Clustering methods in related studies on Türkiye

Study	Scope & Objective	Motivation for Clustering	Method Used
Çınaroğlu, (2021)	Examine health personnel distribution across Türkiye’s provinces and identify regional disparities	Identify patterns and group provinces based on similarities	K-means clustering
Güleç & Yılmaz Işıkhan, (2016)	Compare social media use by health units in Türkiye and WHO region	Group countries based on social media usage patterns	Hierarchical & K-means clustering
Köse, (2022a)	Evaluate OECD countries based on health indicators	Identify hidden patterns and relationships among indicators	K-means clustering
Çelik, (2013)	Classify Türkiye’s provinces based on health indicators	Group provinces based on similarities	Hierarchical clustering

Alkaya, (2022)	Evaluate Türkiye's health indicators among EU countries	Group countries based on similarities in health indicators	K-means clustering
Köse, (2022b)	Analyze Türkiye's statistical regions based on health dimensions	Classify regions based on health service demand, production, and capacity	Hierarchical clustering
Şahin, (2017)	Classify Türkiye and EU countries based on life index values	Group countries based on life index values	Hierarchical clustering
Çınaroğlu & Avcı, (2019)	Determine health provinces based on various variables	Emphasize economic, health, and demographic factors in planning	Two-step cluster analysis
Suner & Çelikoğlu, (2010)	Examine factors affecting health institution selection	Support multiple correspondence analysis interpretations	Kohonen method
Uysal et al., (2017)	Identify differences/similarities in Türkiye's provinces based on life index	Group provinces based on life index values	Ward linkage hierarchical clustering
Tekin, (2015)	Classify Türkiye's provinces based on basic health indicators	Identify differences and group provinces	Ward hierarchical clustering
Eren & Ömürbek, (2019)	Cluster Türkiye's provinces based on health indicators	Rank clusters using Multi-Moora method	Multi-Moora method
Yıldız, (2021)	Evaluate health indicators of Türkiye's provinces	Classify provinces based on health status and healthcare indicators	Hierarchical clustering (Ward method)
Gençoğlu, (2018)	Assess health services and development level of Türkiye's provinces	Identify similarities/differences based on health services	Ward hierarchical clustering

Grouping the studies based on common themes could provide a comprehensive perspective. First, these studies tend to use cluster analysis as a means of grouping and categorizing entities according to their common characteristics, whether provinces, countries or health units. Cluster analysis is constantly motivated by the desire to reveal patterns, similarities and differences within the data set and thereby shed light on important aspects of the subject under study. It is worth noting that cluster analysis has been adopted in studies due to its ability to handle complex, multidimensional data and facilitate the identification of previously hidden relationships.

When the reviewed studies are evaluated on the basis of clustering results, some important insights come to the fore. The most prominent of these is the effectiveness of clustering in revealing the differences within the data set. For example, Çınaroğlu (2021) analyzed the province-based distribution of health personnel and was able to reveal regional inequalities that shape health resource planning. Similarly, Güleç and Yılmaz Işıkhan (2016) tried to identify countries with similar digital engagement strategies based on social media usage patterns of official health units. Another important point is that clustering results can provide input for policy recommendations. For example, Köse (2022a) clustered OECD countries based on health indicators and provided practical insights aimed at strengthening health systems. Similarly, Alkaya (2022) clustered European Union member and candidate countries in terms of health indicators and positioned Turkey at the point of health policy development. The reviewed studies emphasize the utility of clustering as a tool and provide concrete insights that enable actionable decisions. While some studies use K-means clustering (Çınaroğlu, 2021; Köse, 2022a), others use hierarchical clustering (Köse, 2022b; Güleç and Yılmaz Işıkhan, 2016; Şahin, 2017) and even special techniques such as the Multi-Moora method (Eren and Ömürbek, 2019). When the reviewed studies are evaluated on a subject basis, the diversity of the application area of cluster analysis draws attention. Some of these subjects are as follows. Clustering of regions according to health metrics (Tekin, 2015; Yıldız, 2021), examining the factors affecting health institution preference (Suner and Çelikoğlu, 2010), evaluating the demand for health services (Köse, 2022b) and defining regions according to demographic and economic variables (Çınaroğlu and Avcı, 2019).

METHODOLOGY

This study employs a quantitative research design to comprehensively evaluate health service levels within Turkish provinces, leveraging clustering analysis.

Data

The data utilized for this study emanated from the Health Statistical Yearbooks for the year 2021, by the General Directorate of the Ministry of Health of the Republic of Türkiye. In this context, the population per health human and infrastructure resources of the provinces are included in the analysis.

Clustering Analysis Techniques:

The literature review highlights a preference for Hierarchical and K-means clustering methods. However, this study places a distinct focus on density-based algorithms among the fundamental clustering types. When considering local factors affecting healthcare delivery, density-based algorithms provide a robust approach to uncover spatial patterns. Leveraging data point density, these algorithms effectively identify similar clusters while also managing noise and outliers. Hierarchical, K-means and DBSCAN clustering methods are compared in the introduction section and the reason for choosing the DBSCAN method is explained.

Density-based spatial clustering of applications with noise (DBSCAN)

The DBSCAN algorithm was introduced by Ester et al. (1996), and is effective in identifying clusters of different shapes and sizes in the data set. The underlying reason for this is that it is based on the concept of density. So much so that it defines clusters as dense regions of data points separated by low-density regions (Xie et al., 2021). The algorithm is based on two basic parameters. The first parameter, epsilon (ε), represents the maximum distance between two data points, while the second parameter, minPts, is the minimum number of points needed to form a cluster. DBSCAN starts from a random data point and creates clusters by detecting minPts points within the epsilon (ε) distance.

Data points are labeled in three ways. i) Core Points are points with at least minPts data points within radius ε. ii) Border Points lie within radius ε of a seed point but have fewer neighbors than minPts in radius ε. iii) Noise Points are data points that are not labeled as Core or Border. These tags allow the data structure to be distinguished. The performance of DBSCAN is sensitive to the ε and minPts parameters (Oyelade et al., 2019).

The stages of the DBSCAN algorithm are as follows. 1. Determination of Core Points: points within radius ε of each data point are examined. If there are at least minPts points within radius ε, it is labeled as a core point. 2. Cluster Formation: Each core point and its neighbors within radius ε form a cluster. Core points located at radius ε are included in the same cluster. 3. Noise and Border Points: Points that are not included in a cluster are considered noise. Additionally, if the points within the radius ε do not meet the minPts criterion, these points are considered border points.

Clustering Process

The ideal values of ε and minPts parameters were determined by simulation. The DBSCAN algorithm was simulated 535 times and the clustering silhouette coefficient of each iteration was recorded. The ε and minPts values that provide the highest silhouette coefficient were used in the clustering analysis. The silhouette coefficient is a metric that measures how well cluster units are separated (Shahapure and Nicholas, 2020). The silhouette coefficient measures the similarity of each data point to its assigned clusters relative to other clusters and reflects the quality of the clustering result. The higher the silhouette coefficient, the better the cluster separation.

All computations in this study were performed on a computer platform equipped with Intel Core i5-8250U 1.60GHz CPU and 12 GB system memory, using the Scikit-learn machine learning framework in Python programming language. The code developed in this study is presented in the appendix. This piece of code provides direct access to the algorithmic procedures, increasing transparency and reproducibility.

FINDINGS

This section presents the key findings of the study, including the descriptive statistics of the dataset, the results of the DBSCAN parameter simulation, and the corresponding silhouette coefficients, followed by the clustering outcomes.

Descriptive Statistics of the Dataset

The descriptive statistics of dataset (Table 3) are related to the population's access to different healthcare resources within the provinces. These indicators encompass the availability of specialized medical professionals, healthcare facilities, and emergency services.

Table 3: Descriptive statistics of the dataset

Code	Indicator	Mean	Standart Daviation	Coefficient of Variation	Min	Max
H1	Pop. per Specialist Physician	1233	328	27%	546	1995
H2	Pop. per General Physician	1415	250	18%	729	1984
H3	Pop. per Dentist	2955	1043	35%	1366	6332
H4	Pop. per Pharmacist	2637	676	26%	1758	5350
H5	Pop. per Nurse	386	89	23%	229	704
H6	Pop. per Midwife	1246	342	27%	581	2523
H7	Pop. per Other Health Personnel	373	86	23%	208	667
I1	Pop. per Hospital	48011	17084	36%	13941	107151
I2	Pop. per Bed	365	107	29%	186	751
I3	Pop. per Qualified Bed	546	177	32%	308	1310
I4	Pop. per Intensive Care Bed	2197	844	38%	953	4881
I5	Pop. per Family Physician	3070	181	6%	2637	3549
I6	Pop. per 112 Station	21001	8274	39%	6435	43164
I7	Pop. per 112 Ambulance	11169	5877	53%	2261	29066

Pop.: Population

In light of the aforementioned data, several pivotal observations emerge, each shedding light on the intricate landscape of healthcare resource distribution across Türkiye’s provinces.

Firstly, a substantial degree of variability characterizes the allocation of healthcare resources. This heterogeneity is conspicuously evident in indicators such as "Pop. per Specialist Physician" (Coefficient of Variation: 27%), "Pop. per Dentist" (Coefficient of Variation: 35%), and "Pop. per 112 Ambulance" (Coefficient of Variation: 53%). These disparities underscore the imperative to meticulously assess and optimize the accessibility and availability of specialized medical services and facilities. Particularly in regions marked by higher coefficients of variation, strategic interventions are required to rectify the imbalances and augment the equitable provision of healthcare resources.

Secondly, the statistics accentuate discernible fluctuations in indicators associated with hospital bed capacity. The "Pop. per Bed" and "Pop. per Qualified Bed" indicators bear witness to notable coefficients of variation (29% and 32% respectively), indicating uneven distribution of bed capacity across provinces. Furthermore, the "Pop. per Intensive Care Bed" indicator demonstrates a substantial spread (Coefficient of Variation: 38%), indicative of potential discrepancies in critical care infrastructure. Remedying these variations assumes pivotal significance in bolstering healthcare readiness and responsiveness, especially during periods of escalated demand.

Thirdly, the "Pop. per Family Physician" indicator presents a notable exception with a low coefficient of variation (6%), implying a relatively consistent accessibility to primary healthcare services throughout provinces. While this suggests equitable distribution of family physicians catering to primary healthcare, it warrants a comprehensive assessment to ascertain alignment with the genuine healthcare requisites of the populace and the potential influence of urbanization and demographic factors.

DBSCAN Parameter Simulation and Silhouette Values

Table 4 encapsulates the outcomes of a systematic simulation approach conducted within the scope of the research. The simulation revolves around the DBSCAN algorithm's parameterization, specifically the values of Epsilon (ε) and minPts. These parameters are pivotal in determining the algorithm's clustering efficacy.

Table 4: DBSCAN parameter simulation and silhouette values

Iteration	Epsilon ε	minPts	Silhouette	# of Clusters
1	1.16	2	-0.148891	2
⋮	⋮	⋮	⋮	⋮
87	2.02	2	0.013505	6
⋮	⋮	⋮	⋮	⋮
232	3.47	2	0.333197	2
233	3.48	2	0.357227	2
234	3.49	2	0.340013	2
⋮	⋮	⋮	⋮	⋮
535	2.16	14	-0.146281	2

Table 4 summarizes the Silhouette coefficients corresponding to the ε and minPts values tested in a total of 535 iterations performed, as well as the number of clusters formed. The Silhouette coefficient serves as a metric for assessing the quality of clusters generated by the algorithm. It quantifies how similar an object is to its own cluster compared to other clusters, aiding in the identification of well-defined clusters.

Upon analysis of the Table 2, several observations come to light. The Silhouette values fluctuate notably, ranging from negative values (e.g., -0.148891) to positive values (e.g., 0.357227), indicative of differing degrees of cluster cohesion and separation. Iteration 233, with ε of 3.48 and minPts of 2, attains a relatively high Silhouette value of 0.357227, suggesting a well-defined cluster configuration for that parameter combination.

Additionally, it is discernible that the number of clusters resulting from the algorithm varies across iterations, with instances of two and six clusters being prominent in this simulation. Notably, the parameters ε and minPts do not always exhibit a straightforward relationship with the Silhouette values and cluster counts. For instance, while a larger ε (e.g., 3.49) might be anticipated to yield improved clustering (as evidenced by its higher Silhouette value), this is not consistently the case, as evidenced by iteration 535.

Clustering Results

The DBSCAN cluster analysis provides valuable insights into the differences in health resources across Turkey's provinces. The results of the DBSCAN cluster analysis are presented in Table 5.

Table 5: Clustering results

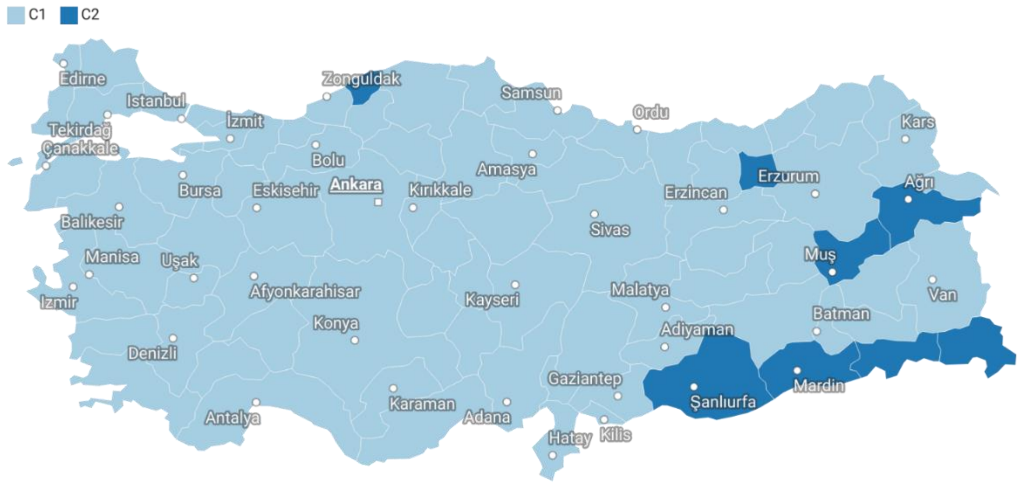
Provinces	Cluster	Core Points	Border Points	Noise Points	Provinces	Cluster	Core Points	Border Points	Noise Points
Agri	C2	True	False	False	Kars	C1	True	False	False
Mus	C2	True	False	False	Kastamonu	C1	True	False	False
Hakkari	C2	False	True	False	Kayseri	C1	True	False	False
Mardin	C2	False	True	False	Kirklareli	C1	True	False	False
Sanliurfa	C2	False	True	False	Kirsehir	C1	True	False	False
Bayburt	C2	False	True	False	Kocaeli	C1	True	False	False
Sirnak	C2	False	True	False	Konya	C1	True	False	False
Bartin	C2	False	True	False	Kutahya	C1	True	False	False
Adana	C1	True	False	False	Malatya	C1	True	False	False
Adiyaman	C1	True	False	False	Manisa	C1	True	False	False
Afyonkarahisar	C1	True	False	False	Kahramanmaras	C1	True	False	False
Amasya	C1	True	False	False	Mugla	C1	True	False	False
Ankara	C1	True	False	False	Nevsehir	C1	True	False	False
Antalya	C1	True	False	False	Nigde	C1	True	False	False
Artvin	C1	True	False	False	Ordu	C1	True	False	False
Aydin	C1	True	False	False	Rize	C1	True	False	False
Balikesir	C1	True	False	False	Sakarya	C1	True	False	False
Bilecik	C1	True	False	False	Samsun	C1	True	False	False
Bingol	C1	True	False	False	Siirt	C1	True	False	False
Bitlis	C1	True	False	False	Sinop	C1	True	False	False
Bolu	C1	True	False	False	Sivas	C1	True	False	False
Burdur	C1	True	False	False	Tekirdag	C1	True	False	False
Bursa	C1	True	False	False	Tokat	C1	True	False	False
Canakkale	C1	True	False	False	Trabzon	C1	True	False	False
Cankiri	C1	True	False	False	Tunceli	C1	True	False	False
Corum	C1	True	False	False	Usak	C1	True	False	False
Denizli	C1	True	False	False	Van	C1	True	False	False
Diyarbakir	C1	True	False	False	Yozgat	C1	True	False	False
Edirne	C1	True	False	False	Zonguldak	C1	True	False	False
Elazig	C1	True	False	False	Aksaray	C1	True	False	False
Erzincan	C1	True	False	False	Karaman	C1	True	False	False

Erzurum	C1	True	False	False	Kirikkale	C1	True	False	False
Eskisehir	C1	True	False	False	Batman	C1	True	False	False
Gaziantep	C1	True	False	False	Ardahan	C1	True	False	False
Giresun	C1	True	False	False	Igdir	C1	True	False	False
Gumushane	C1	True	False	False	Yalova	C1	True	False	False
Hatay	C1	True	False	False	Karabuk	C1	True	False	False
Isparta	C1	True	False	False	Kilis	C1	True	False	False
Mersin	C1	True	False	False	Osmaniye	C1	True	False	False
Istanbul	C1	True	False	False	Duzce	C1	True	False	False
Izmir	C1	True	False	False					

The provinces have been categorized into two distinct clusters, denoted as Cluster C1 and Cluster C2. Cluster C1 is characterized by the presence of core points, implying a denser concentration of data points that share mutual proximity. These provinces exhibit substantial internal cohesion, aligning with the criteria set forth by the DBSCAN algorithm for forming dense clusters. Conversely, Cluster C2 comprises provinces denoted as border points, indicating a less dense configuration compared to core points. These provinces exhibit a moderate level of connectivity to the central body of their respective clusters. It is noteworthy that neither cluster includes noise points, indicating that all provinces have been successfully classified into meaningful clusters without residual or outlier data points.

The structuring of the clusters reflects the patterns of uniqueness and differences within the data set through the association of provinces with different clusters. The C1 Cluster, which has core points, symbolizes provinces with a high degree of homogeneity in terms of relevant health resource indicators. This suggests that these provinces have similar profiles in health resource distribution, possibly reflecting consistent health policies or demographics. On the other hand, the C2 Cluster of border points includes provinces that exhibit a certain difference in health resource distribution compared to the C1 Cluster. The presence of border points indicates a transitional or Decur Decal situation between the more densely clustered provinces.

Figure 1: Clusters on the map



The clustering results shown in Figure 1 can be interpreted geographically as follows. Cluster C1, which is dominated by core points, includes 73 provinces from different regions such as Central Anatolia Region, Aegean Region and Marmara Region. The fact that a large number of provinces from different regions come together indicates homogeneity in the distribution of health resources. Accordingly, despite regional differences, provinces in cluster C1 may have similar health resource profiles, possibly influenced by common socio-economic factors or policy interventions. On the other hand, cluster C2, which mostly includes border points, includes a total of 8 provinces, mostly located in Eastern Anatolia and Southeastern Anatolia. The concentration of provinces in cluster C2 in these two regions suggests that certain geographical and socio-economic dynamics influence the distribution of health resources across provinces, differentiating them from other provinces. This may reflect challenges in ensuring equitable access to health resources in economically disadvantaged regions.

Table 6: Cluster-based descriptive statistics

Cluster C1	**Mean**	**St. Dv.**	**CoV.**	**Min**	**Max**
Pop. per Specialist Physician	1175	281	24%	546	1782
Pop. per General Physician	1427	251	18%	729	1984
Pop. per Dentist	2753	816	30%	1366	5110
Pop. per Pharmacist	2510	480	19%	1758	3959
Pop. per Nurse	368	64	17%	229	536
Pop. per Midwife	1208	318	26%	581	2523
Pop. per Other Health Personnel	358	66	18%	208	509

Pop. per Hospital	45311	14866	33%	13941	76068
Pop. per Bed	345	82	24%	186	650
Pop. per Qualified Bed	513	126	25%	308	853
Pop. per Intensive Care Bed	2086	718	34%	953	4147
Pop. per Family Physician	3069	177	6%	2637	3549
Pop. per 112 Station	20922	8258	39%	6435	43164
Pop. per 112 Ambulance	11291	5933	53%	2261	29066
Cluster C2	**Mean**	**St. Dv.**	**CoV.**	**Min**	**Max**
Pop. per Specialist Physician	1753	257	15%	1251	1995
Pop. per General Physician	1313	209	16%	997	1626
Pop. per Dentist	4799	1079	22%	3271	6332
Pop. per Pharmacist	3800	1008	27%	2241	5350
Pop. per Nurse	548	117	21%	327	704
Pop. per Midwife	1594	359	22%	1000	2079
Pop. per Other Health Personnel	505	129	26%	258	667
Pop. per Hospital	72655	16265	22%	55644	#####
Pop. per Bed	545	134	25%	266	751
Pop. per Qualified Bed	845	270	32%	333	1310
Pop. per Intensive Care Bed	3216	1164	36%	1751	4881
Pop. per Family Physician	3074	219	7%	2726	3393
Pop. per 112 Station	21723	8377	39%	10631	36949
Pop. per 112 Ambulance	10048	5212	52%	2502	18756
# of Provinces in C1	73				
# of Provinces in C2	8				

St. Dv.: Standart Deviation
CoV.: Coefficient of Variation

The descriptive statistics in Table 6 reveal distinct patterns in healthcare resource distribution across Turkish provinces in Cluster C1 and Cluster C2. The difference in cluster sizes, 73 provinces in C1 and 8 in C2, should be noted when interpreting the results. The higher coefficient of variation in Cluster C1 can be attributed to its larger representation, introducing greater resource variability due to diverse regional characteristics. In contrast, the smaller size of C2 leads to a more uniform allocation, yielding a lower coefficient of variation. This underscores the significance of considering both cluster size and indicator variability in healthcare resource distribution analysis.

When focusing on the average values of the indicators within each cluster, a clear differentiation emerges. Cluster C1 represents a scenario in which health resources are more evenly and effectively distributed among the population, with lower mean values across multiple indicators. Low values in indicators such as "Population per Specialist Doctor," "Population per General Practitioner," "Population per Dentist" indicate that there are more health professionals in the head, in order to provide better access to medical services in the field of specialization. Low values in indicators such as "Population per Hospital" and "Population per Bed" indicate a more accessible health infrastructure in the provinces in the C1 cluster. On the other hand, cluster C2 has higher average values on the same indicators. High values in indicators such as "Population per Specialist Doctor", "Population per Dentist", "Population per Pharmacist", "Population per Hospital" and "Population per Bed" indicate that there may be a shortage of healthcare professionals in the provinces in the C2 cluster. It indicates that access to services is limited and the health infrastructure is less accessible or denser.

Policy Making

The following are the three critical insights that can be communicated to policymakers:

1.Border Region Resource Dynamics: The findings reveal that the provinces in the C2 cluster, which consists mainly of border points, are located in the Eastern Anatolia and Southeastern Anatolia Regions. This finding indicates that the two regions mentioned differ from the others in terms of health resource distribution. This difference can be resolved by developing policies appropriate to the local dynamics of these regions and allocating resources. Decision makers should primarily prioritize special improvement strategies that tolerate this difference.

2.Urban Health Infrastructure Optimization: C1 Cluster includes a total of 73 provinces that are close to each other in terms of health resource distribution in various regions. Despite this relative homogeneity, local dynamics between provinces may differ. Decision makers may prefer to work on clusters containing a smaller number of provinces with derivative clustering analyzes in order to achieve pinpoint improvements. This should provide a nuanced approach to resource allocation to identify specific points across smaller clusters that may need additional health infrastructure or services.

3.Synergistic Resource Collaboration: Findings reveal the potential for collaborative resource sharing within each cluster. Provinces within the same cluster should explore collaboration options to optimize resource use and develop better healthcare delivery. Promoting regional cooperation, knowledge exchange and sharing of best practices can ensure effective resource allocation. This understanding increases overall healthcare efficiency by minimizing duplication, especially in neighboring provinces that face similar challenges or have complementary resources.

Discussion

The benefits obtained by cluster analysis in this study are parallel to the studies in the literature. The insights that cluster analysis helps uncover can be briefly listed as follows. Using the k-means clustering approach, Yıldırım (2018) was able to reveal notable inequalities in health accessibility across Turkey's regions. Çınaroğlu (2021) put forward the policy formulation regarding the distribution of health personnel in the provinces in Turkey with the k-means clustering method. Gençoğlu (2018) was able to reveal a positive correlation by examining the link between socioeconomic progress and health opportunities with hierarchical cluster analysis.The findings obtained in this study reveal that there are differences in the allocation of health resources among the provinces of Turkey. Additionally, the findings, combined with supporting findings from previous studies, strengthen advocacy for targeted policies and resource allocation strategies to increase health accessibility and equity in Turkey.

CONCLUSION

In this study, we tried to determine the differences at the provincial level regarding the distribution of health resources in Turkey with the DBSCAN clustering algorithm. By running the algorithm many times with the simulation method, the algorithm parameters that would maximize the clustering silhouette value were determined. Findings of the DBSCAN clustering algorithm run with the most appropriate parameter values are reported.

According to the findings, it was determined that the provinces in Turkey are not homogeneous in terms of health resource distribution and there are noticeable inequalities. While the provinces are divided into two clusters, the cluster, which includes a small number of provinces and provinces with relatively low resource distribution, consists of provinces in the eastern region of the country. The fact that cluster findings can be linked to geographic regions highlights the impact of local dynamics on health disparities.

The study demonstrates the benefits offered by cluster analysis and encourages policy makers, healthcare professionals and researchers to use the DBSCAN method in the decision-making process. This effort has the potential to play a significant role in developing concrete policy improvements and shaping a comprehensive and effective health ecosystem. Proactive integration of health management and clustering methodologies is important for creating an inclusive and efficient health system and equitable distribution of resources.

AUTHOR STATEMENT / YAZAR BEYANI

Researchers have jointly contributed to the article. Researchers have not declared any conflict of interest.

Araştırmacı(lar) makaleye ortak olarak katkıda bulunduğunu bildirmiştir. Araştırmacılar herhangi bir çıkar çatışması bildirmemiştir.

REFERENCES

Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, *9*(8), 1-12. doi: 10.3390/electronics9081295

Agterberg, J., Zhong, F., Crabb, R., & Rosenberg, M. (2020). Cluster analysis application to identify groups of individuals with high health expenditures. *Health Services and Outcomes Research Methodology*, *20*(2–3), 140–182. doi: 10.1007/S10742-020-00214-8/FIGURES/9

Alkaya, A. (2022). Sağlık göstergeleri açısından avrupa birliğine üye ve aday ülkelerin çok boyutlu ölçekleme ve kümeleme analiziyle sınıflanması. *Hacettepe Sağlık İdaresi Dergisi*, *25*(1), 29–46.

Çelik, Ş. (2013). Kümeleme analizi ile sağlık göstergelerine göre türkiye'deki illerin sınıflandırılması.*Doğuş Üniversitesi Dergisi, 14*(2), 175-194.

Çınaroğlu, S. (2021). Türkiye'de iller düzeyinde sağlık personeli dağılımı ve daha etkin politika ihtiyacı. *Hacettepe Sağlık İdaresi Dergisi*, *24*(2), 235–254.

Çınaroğlu, S., & Avcı, K. (2019). Sağlıkta bölgesel planlama çalışmalarında verimliliğin artırılması için alternatif bir yaklaşım: iki aşamalı kümeleme uygulaması. *Verimlilik Dergisi*, *2*, 7–25.

Eren, H., & Ömürbek, N. (2019). Türkiye'nin sağlık göstergeleri açısından kümelenmesi ve performans analizi. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *11*(29), 421–452.

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 226–231.

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis*. John Wiley & Sons.

Fuchs, M., & Höpken, W. (2022). Clustering. In R. Egger (eds), *Applied data science in tourism* (p. 129-149). Cham: Springer.

Fuchs, M., & Höpken, W. (2022). *Clustering* ,129–149. doi: 10.1007/978-3-030-88389-8_8

Gençoğlu, P. (2018). Türkiye'de illerin gelişmişlik düzeyi dikkate alınarak sağlık hizmetlerinin kümeleme analizi aracılığıyla değerlendirilmesi. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, *52*, 301–324.

Güleç, D., & Yılmaz Işıkhan, S. (2016). Kümeleme analizi ile türkiye ve dsö bölgesi ülkeleri resmi sağlık birimlerinin sosyal medya kullanımı açısından karşılaştırılması. *Uluslararası Sağlık Yönetimi ve Stratejileri Araştırma Dergisi* , *1*(3), 36–48.

Hassan, A. M., & Darwish, S. M. (2021). Dynamic distributed clustering approach directed to patient-centric healthcare system. *Advances in Intelligent Systems and Computing*, *1339*, 355–365. doi: 10.1007/978-3-030-69717-4_35/COVER

Jamtsho, S., Corner, R., & Dewan, A. (2015). Spatio-Temporal Analysis of Spatial Accessibility to Primary Health Care in Bhutan. *ISPRS International Journal of Geo-Information, 4*(3), 1584–1604. doi: 10.3390/IJGI4031584

Keya, K. N., Islam, R., Pan, S., Stockwell, I., & Foulds, J. (2020). *Equitable allocation of healthcare resources with fair cox models*. doi: 10.13016/M2LINH-RXCX https://www.semanticscholar.org/reader/6f5da8f435a115d41ce04ab743174906c07fe2e7 adresinden erişildi.

Khan, S. U., & Hussain, I. (2020). Inequalities in health and health-related indicators: a spatial geographic analysis of Pakistan. *BMC Public Health*, *20*(1), 1–9. doi: 10.1186/S12889-020-09870-4/FIGURES/4

Korkhmazov, V. T., & Perkhov, V. I. (2023). Cluster analysis results for assessment of COVID-19-Related mortality differences between Russian regions. *Innovative Medicine of Kuban*, *0*(1), 65–71. doi: 10.35401/2541-9897-2023-26-1-65-71

Köse, A. (2022a). OECD ülkelerinin sağlık göstergeleri açısından kümeleme analizi yöntemine göre değerlendirilmesi. *Nevşehir Hacı Bektaş Veli Üniversitesi SBE Dergisi*, *12*(4), 2010–2021.

Köse, A. (2022b). Türkiye'de sağlık göstergelerine göre istatistik bölgelerin kümeleme analizi ile sınıflandırılması. *Alanya Akademik Bakış*, *6*(2), 2167–2189. doi: 10.29023/ALANYAAKADEMIK.1021019

Kumari, R., & Raman, R. (2022). Regional disparities in healthcare services in Uttar Pradesh, India: a principal component analysis. *GeoJournal*, *87*(6), 5027–5050. doi: 10.1007/S10708-021-10542-Y/TABLES/13

Kurji, J., Talbot, B., Bulcha, G., Bedru, K. H., Morankar, S., Gebretsadik, L. A., Wordofa, M. A., Welch, V., Labonte, R., & Kulkarni, M. A. (2020). Uncovering spatial variation in maternal healthcare service use at subnational level in Jimma Zone, Ethiopia. *BMC Health Services Research*, *20*(1), 1–14. doi: 10.1186/S12913-020-05572-0/TABLES/3

Lam, H. Y., Zarsuelo, M. A. M., Capeding, T. P. J. Z., Silva, M. E. C., Mendoza, M. A. F., & Padilla, C. D. (2020). Policy analysis on province-level integration of healthcare system in light of the universal health care act. *Acta Medica Philippina*, *54*(6), 650–658. doi: 10.47895/AMP.V54I6.2608

Manortey, S., Vanderslice, J., Alder, S., Henry, K. A., Crookston, B., Dickerson, T., & Benson, S. (2014). Spatial analysis of factors associated with household subscription to the National Health Insurance Scheme in rural Ghana. *Journal of Public Health in Africa*, *5*(1), 1–8. doi: 10.4081/jphia.2014.353

Matthay, E. C., Hagan, E., Joshi, S., Tan, M. L., Vlahov, D., Adler, N., & Glymour, M. M. (2021). The revolution will be hard to evaluate: How co-occurring policy changes affect research on the health effects of social policies. *Epidemiol Rev*, *44*(1), 19-32. doi: 10.1101/2020.10.02.20205971

Oyelade, J., Isewon, I., Oladipupo, O., Emebo, O., Omogbadegun, Z., Aromolaran, O., Uwoghiren, E., Olaniyan, D., & Olawole, O. (2019). Data clustering: Algorithms and its applications. *2019 19th International Conference on Computational Science and Its Applications (ICCSA)*, 71–81. doi: 10.1109/ICCSA.2019.000-1

Rabarison, K. M., Bish, C. L., Massoudi, M. S., & Giles, W. H. (2015). Economic evaluation enhances public health decision making. *Frontiers in Public Health*, *3*, 1-5. doi: 10.3389/FPUBH.2015.00164/BIBTEX

Şahin, D. (2017). Sağlık göstergeleri bakımından Türkiye'nin Avrupa Birliği ülkeleri arasındaki yeri: İstatistiksel bir analiz. *Çankırı Karatekin Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *8*(2), 55–77.

Schaefers, J., Wenang, S., Afdal, A., Mukti, A. G., Sundari, S., & Haier, J. (2022). Population-based study on coverage and healthcare processes for cancer during implementation of national healthcare insurance in Indonesia. *The Lancet Regional Health - Southeast Asia*, *6*, 100045. doi: 10.1016/J.LANSEA.2022.100045

Shahapure, K. R., & Nicholas, C. (2020). Cluster quality analysis using silhouette score. *Proceedings - 2020 IEEE 7th International Conference on Data Science and Advanced Analytics, DSAA 2020* (p. 747–748). doi: 10.1109/DSAA49011.2020.00096

Sun, Y., Liu, X., Jiang, J., & Wang, P. (2018). Comprehensive evaluation research on healthcare development in China from 2002 to 2014. *Social Indicators Research*, *138*(3), 1259–1278. doi: 10.1007/S11205-017-1685-2/TABLES/8

Suner, A., & Çelikoğlu, C. C. (2010). Toplum tabanlı bir çalışmada çoklu uygunluk analizi ve kümeleme analizi ile sağlık kurumu seçimi. *Dokuz Eylül Üniversitesi İktisadi İdari Bilimler Fakültesi Dergisi*, *25*(2), 43–55.

Tekin, B. (2015). Temel sağlık göstergeleri açısından Türkiye'deki illerin gruplandırılması: bir kümeleme analizi uygulaması. *Journal of the Faculty of Economics and Administrative Sciences*, *5*(2), 389–416.

Ullah, S., Daud, H., Dass, S. C., Fanaee-t, H., Kausarian, H., & Khalil, A. (2020). Space-time clustering characteristics of tuberculosis in Khyber Pakhtunkhwa Province, Pakistan, 2015–2019. *International Journal of Environmental Research and Public Health*, *17*(4), 1-10. doi: 10.3390/IJERPH17041413

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

Wartelle, A., Mourad-Chehade, F., Yalaoui, F., Chrusciel, J., Laplanche, D., & Sanchez, S. (2021). Clustering of a health dataset using diagnosis co-occurrences. *Applied Sciences*, *11*(5), 1-19. doi: 10.3390/APP11052373

Xie, Y., Jia, X., Shekhar, S., Bao, H., & Zhou, X. (2021). Significant DBSCAN+: Statistically robust density-based clustering. *ACM Transactions on Intelligent Systems and Technology*, *12*(5). doi: 10.1145/3474842

Yıldırım, H. (2018). Comparison of provinces of Turkey ın terms of accessing health care services by using different clustering algorithms. *Anadolu University Journal of Science and Technology-A Applied Sciences and Engineering*, *19*(4), 907–925. doi: 10.18038/aubtda.413890

Yıldız, A. (2021). Türkiye'de illerin sağlık göstergeleri açısından kümeleme analizi ile değerlendirilmesi. *OPUS International Journal of Society Researches*, *17*(35), 1607–1625.

Appendix 1. DBSCAN clustering with silhouette-based parameter search Python script

```
import numpy as np

import pandas as pd

from sklearn.cluster import DBSCAN

from sklearn.preprocessing import StandardScaler

from sklearn.metrics import silhouette_score

import matplotlib.pyplot as plt



pd.set_option('display.max_columns', None)

pd.set_option('display.max_rows', None)



# Read data and preprocess

data = pd.read_csv('veri.csv')  # Update with file path

X = data.drop('City', axis=1)

X_scaled = StandardScaler().fit_transform(X)



# Parameter ranges

epsilon_values = np.arange(1, 10, 0.01)

minPts_values = np.arange(2, 81, 1)
```

```
# Find best parameters using silhouette score
best_silhouette = -1
best_params = {}
results = []

for minPts in minPts_values:
    for epsilon in epsilon_values:
        dbscan = DBSCAN(eps=epsilon, min_samples=minPts)
        labels = dbscan.fit_predict(X_scaled)
        core_samples_mask = np.zeros_like(labels, dtype=bool)
        core_samples_mask[dbscan.core_sample_indices_] = True
        num_clusters = len(set(labels)) - (1 if -1 in labels else 0)
        if num_clusters > 1:
            silhouette = silhouette_score(X_scaled, labels)
            results.append({'Epsilon': epsilon, 'MinPts': minPts, 'Silhouette Score': silhouette, 'Num Clusters': num_clusters})
            if silhouette > best_silhouette:
                best_silhouette = silhouette
                best_params['epsilon'] = epsilon
                best_params['minPts'] = minPts
```

```
# Apply DBSCAN with best parameters
best_dbscan = DBSCAN(eps=best_params['epsilon'], min_samples=best_params['minPts'])
best_labels = best_dbscan.fit_predict(X_scaled)

# Visualize clusters
plt.scatter(X_scaled[:, 0], X_scaled[:, 1], c=best_labels, cmap='viridis')
plt.title("DBSCAN Clustering with Best Parameters")
plt.xlabel("Feature 1")
plt.ylabel("Feature 2")
plt.show()

# Convert results to DataFrame
results_df = pd.DataFrame(results)

# Print best parameters and results
print("Best minPts:", best_params['minPts'])
print("Best epsilon:", best_params['epsilon'])
print("Best silhouette score:", best_silhouette)
print(results_df)

# Core, Border, and Noise points
```

```
core_samples_mask = np.zeros_like(best_labels, dtype=bool)

core_samples_mask[best_dbscan.core_sample_indices_] = True

border_samples_mask = (best_labels == -1) & (~core_samples_mask)

noise_samples_mask = (best_labels == -1) & (~border_samples_mask)


# Convert results to DataFrame

results_df = pd.DataFrame({

    'City': data['City'],

    'Cluster Label': best_labels,

    'Core Points': core_samples_mask,

    'Border Points': border_samples_mask,

    'Noise Points': noise_samples_mask

})


# Print results

print(results_df)
```