IJER
ISSN: 2757-959X

INTERNATIONAL JOURNAL OF ECONOMIC AND ADMINISTRATIVE ACADEMIC RESEARCH

Available online, ISSN: 2757-959X | www.ijerdergisi.com | Economic and Administrative Academic Research

CLUSTER ANALYSIS OF OECD COUNTRIES DURING THE COVID-19 PANDEMIC

Fatma Selen Madenoğlu *[a]

Corresponding Author

ARTICLEINFO

Research Article

Received:13/07/2023

Accepted: 05/10/2023

Keywords:

Cluster Analysis, Hierarchical Cluster, Non-hierarchical Cluster, Health Indicators, COVID-19

ABSTRACT

Health indicators are used to examine society's health status and track its improvement and degradation over time. The health situation of countries can be revealed and examined using health indicators. They help determine what future health policies will be implemented and what advancements in the field of health should be undertaken. The success or failure of the struggle against the COVID-19 pandemic are significantly influenced by the health indicators and COVID-19 pandemic indicators of the OECD countries. The goal of the research is to classify Organization for Economic Cooperation and Development (OECD) member countries in accordance with COVID-19 pandemic indicators and to deeply analyze how this classification influences the countries' responses to the pandemic. The assessments have been made by identifying Türkiye's position within the classification as one of the OECD's founding members. The statistics on health indicators has been compiled using databases from the World Health Organization, OECD, and data.worldbank. A cluster analysis has been used to categorize the countries. The study's findings have demonstrated the effect of the multiple variable evaluation to performance analysis. A hierarchical method and a non-hierarchical method have been applied to cluster the OECD countries considering COVID-19 pandemic indicators. As a result of the two methods, Türkiye was in the same group with France, Germany, Italy, Korea, United Kingdom.

Uluslararasi İktisadi Ve İdari Akademik Araştirmalar Dergisi, 3(2), 2023, 163-180

COVID-19 SALGININDA OECD ÜLKELERİNİN KÜMELEME ANALİZİ

MAKALE BİLGİSİ

Araştırma Makalesi

Geliş :13/07/2023

Kabul : 05/10/2023

Anahtar Kelimeler:

Kümeleme Analizi, Hiyerarşik Kümeleme, Hiyerarşik Olmayan Kümeleme, Sağlık Göstergeleri, COVID-19

ÖZ

Sağlık göstergeleri, toplumun sağlık durumunu incelemek ve zaman içinde iyileşmesini ve kötüye gitmesini izlemek için kullanılmaktadır. Ülkelerdeki sağlık durumu, sağlık göstergeleri kullanılarak ortaya çıkarılabilir ve incelenebilir. Gelecekte hangi sağlık politikalarının uygulanacağını ve sağlık alanında hangi ilerlemelerin gerçekleştirilmesi gerektiğini belirlemeye yardımcı olurlar. COVID-19 pandemisi ile mücadelenin başarısı veya başarısızlığı, OECD ülkelerinin sağlık göstergeleri ve COVID-19 pandemi göstergelerinden önemli ölçüde etkilenmektedir. Çalışmanın amacı, Ekonomik İşbirliği ve Kalkınma Teşkilatı'na üye ülkelerini COVID-19 pandemi göstergelerine göre kategorize etmek ve bu sınıflandırmanın ülkelerin pandemiye verdiği tepkileri nasıl etkilediğini derinlemesine analiz etmektir. Değerlendirmeler, Türkiye'nin OECD'nin kurucu üyelerinden biri olarak sınıflandırmadaki konumu belirlenerek yapılmıştır. Sağlık göstergelerine ilişkin bilgilerin derlenmesinde Dünya Sağlık Örgütü, OECD ve data.worldbank veri tabanları kullanılmaktadır. Ülkeleri kategorize etmek için, küme analizi adı verilen çok değişkenli bir istatistiksel analitik teknik kullanılmıştır. Çalışmanın bulguları, çoklu değişkenle değerlendirmenin performans analizi üzerindeki etkisini ortaya koymuştur. OECD ülkelerini COVID-19 pandemik göstergeleri dikkate alınarak kümelemek için hiyerarşik bir yöntem ve hiyerarşik olmayan bir yöntem uygulanmıştır. İki yöntemin sonucunda Türkiye, Fransa, Almanya, İtalya, Kore, İngiltere ile aynı grupta yer almıştır.

[a] selen.madenoglu@agu.edu.tr https://orcid.org/0000-0002-5577-4471

1. INTRODUCTION

The healthcare sector is one of the most important sectors in the World. The World Health Organization (WHO) defines health as a condition of full physical, mental, and social well-being and not simply the absence of illness or infirmity. As is common knowledge, having access to the best possible health is one of a person's basic rights. Improvement research and novel strategies to be carried out in this context are important for public health. Within the context of establishing a socially equitable community, it is critical that everyone has equal access to healthcare.

A healthy life should be guaranteed for people of all ages and socioeconomic groups, health services should be improved, and population requirements should be satisfied. Understanding the reasons why some communities are healthier than others is essential to increase the number of healthy people (Costa et al., 2019). The essential data about a country's health policies' effectiveness must be provided. Statistical indicators play a significant role in giving this information as well. Indicators are used as monitoring tools that enable setting objectives as well as measurement capabilities.

Health indicators are the measurements of data that can be used in monitoring the health of countries and the factors affecting health (CIHI, 2009). Health indicators are used to reveal the status of countries and societies to have healthy individuals, their development in the field of health, and the effectiveness of their achievements. For this reason, the benefits of performing each data analysis in the health industry with great care are quite high.

Data and data analysis have become the focus of both researchers and practitioners in the health sector, as in other sectors. Widespread use of technology and technological innovations such as medical report, electronic patient records generate large amounts of data (Strang & Sun, 2020). Developments such as the increasing number of healthcare professionals, increasing number of patients, increasing disease diversity, developing treatment methods, and the use of developed technological devices cause a large amount of data to be generated every minute and even every second. The size and complexity of the generated data make it difficult to analyze health data using traditional methods. In practice, a number of research efforts have suggested employing an advanced data analysis technique, to address these data challenges (Gonzalez et al., 2016). Data mining is the process of sorting through enormous amounts of data to find intriguing trends where traditional statistical methods of exploratory data analysis (traditional statistics) were unable to do so (Han et al., 2012). The groundwork of traditional techniques and data mining techniques is mathematics. Data mining uses additional features such as machine learning, visualization. Due to these advantages, data mining is becoming more popular.

A healthcare system involves innovative and developed data storage, administration, analysis, and data mining tools in order to extract knowledge from big data (Pramanik et al., 2020). The technology and methods for converting enormous amounts of data into information that can be utilized for decision-making are provided by data mining (Dash et al., 2019). Classification, clustering, association, and outlier detection are data mining methods (Santos-Pereira et al., 2022). One of the key areas of data mining is cluster analysis, especially with grouping massive amounts of data (Kurosava et al., 2014). This method examines the similar and separate aspects of the data with respect to each other and clusters and analyzes the data according to these characteristics. As a result of the analysis, the components that make up the clusters are similar to each other and differ from the compoments of other clusters (Çelik, 2013).

Cluster analysis is one of the multivariate statistical analysis methods that helps divide units and variables into similar subsets (groups, classes) whose groups are not known exactly.

The main purpose of cluster analysis is to group units based on their characteristic features. Within the scope of the study, it is aimed to cluster OECD countries according to their struggle against the COVID-19 pandemic to determine which countries Türkiye is similar to. The Novel Corona Virus (COVID-19), which emerged in the city of Wuhan (Hubei) in China in December 2019, spread all over the world in a short time and was declared a global pandemic by the World Health Organization-WHO on March 11, 2020 (T.C. Sağlık Bakanlığı). With the outbreak of this pandemic, the vital value of the health sector, which is already a very important sector, has emerged. In the struggle against the pandemic in the field of health, countries tried to adopt good practices to themselves. Health indicators are used to evaluate the performances by revealing the results of good practices. In addition to the incredible efforts of healthcare professionals since the first day of the pandemic, other fields of science (such as social sciences, and natural sciences) have contributed to the process by carrying out various studies to support them. Especially, there is a large amount of unprocessed and processed data in different structures in the health sector.

The main contribution of this study to the literature is presented a clustering analysis carried out based on the COVID data in OECD member countries and the pandemic health indicators of these countries by using two different type of clustering methods. The aim of the study is to carry out a comparative analysis of various preventive measures of countries during the pandemic process, to present a comprehensive view considering the strategies developed by countries in combating the pandemic, and to guide countries in preparing for future pandemics or health crises with the help of health indicators.

In the literature section of the study, studies examining the clustering of OECD countries are included. In the method section, clustering analysis is explained. In the forth section, the results of cluster analysis are presented. The last section presents the results.

2. LITERATURE REVIEW

The studies carried out with cluster analysis in the field of health have been examined. Ersöz (2009) used clustering and discriminant analyzes to compare selected health indicators of OECD countries and similar countries were identified. In the 2004 health indicators of OECD member countries, four variables, namely the ratio of total health expenditures to gross domestic product (GDP), per capita health expenditure, life expectancy at birth, and infant mortality per 1000 births, were used in the study. Comparisons were conducted with the hierarchical clustering method, k-means clustering and medoid clustering methods. According to the hierarchical clustering method, it was determined that the number of clusters should be 3. It has been seen that Türkiye has similar health characteristics with Poland, Slovakia, Czech Republic, Hungary, Mexico, Republic of Korea and is in the same cluster in the clustering method among OECD countries. In the K-Means clustering method; It has been concluded that Portugal is in the same cluster, including the 6 countries given, and in the Medoid clustering method, it is in the same cluster as Mexico and shows similar health indicators. Çelik (2013) identified the province groups showing the same structure with the help of 10 health variables belonging to 81 provinces in Türkiye. The results of the cluster analysis on 81 provinces, which were classified into 7, 10, and 15 clusters, were discussed. Provinces with the worst health conditions were also identified as a result of the investigation. Alptekin (2014) employed fuzzy clustering analysis to group 27 member states of the European Union and Türkiye according to healthcare indicators. This study also examines Türkiye's status in relation to the countries of the European Union in terms of healthcare information. Data from the 2012 World Health Report have been subjected to fuzzy clustering analysis. The countries were divided into two categories based on the Fuzzy clustering analysis. Bulgaria, Cyprus, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, and Slovakia are all grouped together with Türkiye. Sonğur (2016) clustered the OECD countries in terms of health indicators with clustering algorithms. Türkiye was included in the same cluster with Israel, Mexico and Chile. In addition, it was concluded that the common aspects of these countries are that they use the Bismarck financing model. Mut and Akyürek (2017) presented the clustering of countries based on OECD health metrics. They identified which OECD countries Türkiye is comparable to through the study. The hierarchical clustering approach identified three groups as the total number. The clustering technique k-means was used. It was determined that Türkiye is in the same cluster as Mexico and Chile. Proksch et al. (2019) compared 30 OECD countries considering innovative output in healthcare. They employed a hierarcical cluster analysis method using Ward' method with the squared Euclidean Distance. The result of cluster analysis depicted four cluster is the most effective result. According to results, the two countries with the highest innovation output in both knowledge production and knowledge commercialization are the Netherlands and Switzerland. The Scandinavian countries were in the medium rank except Switzerland and Switzerland was assigned to separate, single cluster. Japan and Korea were in the same cluster. Türkiye, Japan, South Korea, Hungary, Poland, Slovak Republic, Chile, Mexico included in same cluster. Reibling et al. (2019) classified OECD countries in terms of thirteen selected health indicators. As a result of the clustering analysis, nine clusters were determined, and Japan and Korea were clustered individually in separate clusters. Switzerland and the United States were in the same cluster. Spain and Italy were in the same cluster. An updated health system classification based on a theoretical framework integrating ideas from comparative institutional welfare state and comparative health policy research was conducted. They combined nine clusters and the clusters were determined again, with the number of clusters being five. Çetintürk and Gençtürk (2020) evaluated Türkiye's position by comparing health expenditure indicators of 36 OECD countries with the help of OECD health statistics data between 2003 and 2017. In this study, 14 different expenditure variables used in health services were analyzed by using the Ward method. As a result of the hierarchical clustering analysis regarding the shares of total health expenditures in GDP, five clusters were generated. The USA was included in a single cluster. The USA differed from other countries with its high spending share compared to other OECD countries. In terms of various health spending types, the study's findings show that Türkiye is comparable to Estonia, Latvia, Mexico, the Czech Republic, Luxembourg, Belgium, and Australia. Demircioğlu and Eşiyok (2020) classified 36 countries which are OECD and EU members based on health indicators by using k-means method. Değirmenci and Yakıcı-Ayan (2020) discussed the position of Türkiye in terms of health indicators compared to OECD countries. They evaluated the rankings of OECD countries using the TOPSIS method as well as classifying them using fuzzy clustering analysis. They determined five health indicators with the data of 2015. As a result of the applied fuzzy cluster analysis, it was seen that Türkiye is in the same cluster as Korea, Mexico, and Poland, and the number of clusters was determined as four. Ünsal and Kasap (2020) investigated Covid data for G20, EU, and OECD countries by using time series analysis and k means cluster analysis. They used the data of April and May with three variables: confirmed, recovered, and deaths. The optimum number of clusters was determined as three and the country that is most affected by the pandemic was the United States. One cluster included the only USA and China, France, Germany, Italy, Spain, Türkiye, and United Kingdom (UK) were in a cluster and the remaining countries were assigned to another cluster in June. In May, China, France, Germany, Italy, Spain, Türkiye, the United Kingdom, Russia, and Brazil were assigned to the same cluster, and China moved to the first cluster. Utilizing the cluster analysis technique, Kartal et al. (2020) examined the COVID data from Türkiye and around the entire world. Fourty nations' COVID and death statistics were examined. The Cross-Industry Standard Process Model for Data Mining was used in this study as a methodology, and the k-means clustering method was used to conduct a clustering analysis.

Three groups were given as the number. Yıldırım et al. (2020) provided an insight into the relationship between economic growth and health in OECD member countries. The analysis identified two major clusters as high (level) and low health status (level) countries. It was determined that an increase in the birth life expectancy of countries with better health has no appreciable effect on economic growth. The improvement in the life expectancy at birth of countries with worse health, however, had a positive impact on economic growth. Alkaya and Alkaş (2021) investigated 37 OECD member countries using cluster analysis using 2017 data on health indicators. Köse (2022) evaluated Türkiye's healthcare service in 2019 with hierarchical clustering analysis of statistical regions according to data on demand, production and capacity dimensions. Wulandari and Yogantara (2022) used the k-means and fuzzy c-means methods to classify developed and developing countries based on economic and health factors properly. Paumelle et al. (2023) applied a hierarchical clustering analysis which was applied using Ward’s criterion and k-means clustering algorithm and presented 7 territorial profiles in France.

3. METHOD

Today's rapid technological development and broad use of technological instruments cause massive data accumulation in all fields and an accelerated rate of data flow. These data piles, which have accumulated in every sector of production and consuming, are so massive that processing them is necessary. The question of how to use the available data in detail has become increasingly important as a result of a substantial increase in data availability. In order to support decision-making processes by extracting meaningful and meaningless data from information piles, a new concept called "data mining" has just recently appeared. It is utilized across every possible industry, including marketing, manufacturing, banking, and healthcare etc.. In any situation where there is a large amount of data, it is possible to use data mining techniques to uncover sensitive information and forecast future trends and behavioural patterns (Zhang & Zhou, 2004). Data mining is the process of extracting core information from huge volumes of data (Ganesh, 2002). Data mining is the process of using computer programs to search for relationships and patterns that will help us forecast the future from an immense amount of data. The technology and methods for converting enormous amounts of data into information that can be used for decision-making are provided by data mining (Dash et al., 2019). Data mining makes it possible to maximize the return on investments made in data gathering, which is an expensive and time-consuming process (Kecman, 2001).

Classification, clustering, association, and outlier detection are data analysis method includeds in data mining. Classificiation methods are used to predict categorical attributes. Clustering methods are utilized to construct the subgroups. Association is aimed to find the association rules between attributes. Outlier detection is to identify attributes that aren’t normal (Santos-Pereira et al., 2022). Clustering is the most commonly used method for data processing in data mining.

3.1. Cluster Analysis

Cluster analysis is one of the data mining method and one of the multivariate statistical analysis method (Tüzüntürk, 2010). Cluster analysis can be defined as a collection of methods developed to separate (clustering) the units or variables in the X data matrix into homogeneous subgroups within the framework of their properties (Alpar, 2013). This method divides these variables into subsets that are supposed to explain common features according to values determined in units and construct common factor structures. Its leading principal is that members belong to the same cluster share more similarities than members who belong to different clusters, which share the least similarities (Ji et al., 2013).

In the first step of cluster analysis, after the data matrix is generated. x_{ig} represents the gth attribute value of ith unit in the data set. For every unit in the dataset, the data matrix shows the sum of all of the attribute entries. Determine the absolute deviation's average and the normalized measure.

$$s_g = 1/n(\left|x_{1g} - \mu_g\right| + \left|x_{2g} - \mu_g\right| + \cdots + \left|x_{ng} - g\right| \tag{1}$$

$$m_g = 1/n\ (x_{1g} + x_{2g} + \ldots + x_{ng}\ ) \tag{2}$$

$$z_{ig} = \frac{x_{1g} - \mu_g}{sg} \tag{3}$$

The similarity or distance measure is selected. Some of the distance measures involved in cluster analysis are Euclidean distance (Equation 4), Manhattan distance (Equation 5), Mingkosiji distance (Equation 6) etc.

$$d(x,y) = \sqrt{(x_1 - y_{1)}{}^2 + (x_2 - y_{2)}{}^2 + \ldots + (x_n - y_{n)}{}^2} \tag{4}$$

$$d(x,y) = \left|x_1 - y_{1)}\right| + \left|x_2 - y_{2)}\right| + \ldots + \left|x_n - y_{n)}\right| \tag{5}$$

$$d(x,y) = \left|x_1 - y_{1)}\right|^q + \left|x_2 - y_{2)}\right|^q + \ldots + \left|x_n - y_{n)}\right|^q \tag{6}$$

In the Equation 6, q presents a positive integer. In this equation, if the value of q is one, it demonstrates the Manhattan distance. If the value of q is two, it demonstrates the Euclidean distance (Zou, 2020).

The second step is to decide on the clustering algorithm. Cluster analysis is performed using hierarchical or non-hierarchical clustering techniques. In the last stage, a clustering method suitable for the selected algorithm is selected and the results of the clustering analysis are interpreted. In cluster analysis, assumptions such as normality, linearity, and covariance, which are important in other methods, are not taken into account (Cao et al., 2013). Clustering methods in cluster analysis are hierarchical and classified as non-hierarchical clustering methods. The common goals of these two clustering techniques are to increase the homogeneity within the cluster and to provide heterogeneity between the clusters. The main difference between these two techniques is the differences in determining the number of clusters. In the hierarchical clustering method, the number of clusters is determined with the help of graphical methods such as dendrogram (tree graph), whereas in non-hierarchical clustering, the number of clusters is determined by the researcher at the beginning of the analysis.

3.1.1 Hierarchical Clustering Methods

Clustering in hierarchical clustering methods gradually constructs subgroups of clusters in the next phase. Grouping is carried out according to the similarities or dissimilarities of the units. By using similarity and dissimilarity measures, homogeneous and heterogeneous groups (clusters) are formed (Alkaya & Alkaş, 2021). There are two types of hierarchical clustering methods agglomerative and divisive hierarchical clustering methods. Agglomerative hierarchical clustering methods start with individual units and initially, there are as many clusters as there are units. In the first stage, the most similar units are grouped and these first groups are combined according to their similarities. When the similarity decreases, all subgroups are combined into a single cluster. The agglomerative clustering methods are median clustering, centroid clustering, furthest neighbor, nearest neighbor, within-groups linkage, between-groups likage, Ward's method. In divisive hierarchical clustering methods, there is only one cluster containing all the observations. In the first stage, a single group is divided into two subgroups. Here, the units in the subgroup are far away from the units in the other subgroup.

These subgroups are subdivided into dissimilar subgroups. The process continues until there are as many subgroups as the number of units, with each unit forming a group (Alpar, 2013; Johnson & Wichern, 1998). Ward technique, which is one of the hierarchical techniques, was used in the study. This technique focuses on minimizing the variance within the cluster. For this purpose, the sum of squares error formula (ESS) which is given in Equation 7 is used. Here x_i is the *i*th unit value. In the first step, each unit is a cluster, so the sum of squares of the error is equal to zero. The procedure then proceeds on to identifying the clusters or units that have the smallest sum of squares error rise (Çakmak et al., 2005).

$$\mathrm{ESS}=\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 \tag{7}$$

3.1.2 Non-Hierarchical Clustering Methods

The non-hierarchical clustering methods are the methods in which the number of clusters is determined before the clustering process, the units are grouped into *k* clusters determined according to the variables, and they are divided into clusters with a homogeneous and heterogeneous structure according to the variables. The methods start from the initial division of units into groups, or from an initial set of center points that will form the centers of the clusters. One way to get started is to randomly select starting points from among items or randomly sort items into starting groups (Alpar, 2013). When constructing clusters, non-hierarchical clustering techniques aim to maximize a similarity criterion, which is frequently specified locally or globally. Non-hierarchical methods can be applied to much larger data sets than hierarchical techniques. When we compare the methods in terms of time, non-hierarchical clustering methods take less time than hierarchical methods. If the number of clusters can be determined by another method, non-hierarchical clustering methods are employed. By applying non-hierarchical methods according to different cluster numbers without predetermining the number of clusters, the number of clusters that should be can be determined later as a result of comparative analysis. Non-hierarchical techniques ensure that data are gathered in the proper groups. Examples of non-hierarchical methods are the k-means clustering method and the maximum likelihood method. In general, the steps of the algorithm are as follows. First, the *k* value, which is the number of clusters, is determined. All points is assigned to the cluster whose cluster centroid is closest to it. When all data are assigned, the centroid of each cluster are recalculated. The process is repeated until the centroids do not change (Karypis et al., 2000).

In this study, k-means method, which is one of the non-hierarchicak clustering method, has been used in this study. This effective algorithm is used in a lot of low-dimensional and big data set applications (Singhal & Shukla, 2018). This particular grouping technique used by Mac Queen is one of the most liked ones. The observations are repeatedly reassigned to the clusters until a set of numerical criteria are met. The data is divided into a user-specified number of clusters. As compared to different clustering analysis techniques, the goal of the k-means method is to optimize both in-group homogeneity and heterogeneity between clusters (Tzortzis & Likas, 2014). The k-means criterion aims to maximize distance between clusters and minimize observation distance within clusters (Hair et al., 2014). The k-means technique is frequently assessed using the square error. The clustering results with the lowest squared error is considered to be the best one (Tan et al., 2016).

4. APPLICATION AND RESULTS

In this study, a cluster analysis has been conducted to cluster OECD countries. The aim here is to ensure that data are combined at certain levels, taking into account certain characteristics, and to determine which OECD countries are in the same clusters in struggle against COVID-19 pandemic. The 38 countries are the members of OECD have been considered in this study. The health indicators and COVID-19 indicator have been used. The analyses have been performed using hierarchical clustering method and non-hierarchical clustering method. The SPSS program has been implemented to evaluate the clustering algorithms for the data analysis. In the first stage, the number of clusters and clustering results have been obtained by applying the Ward method, which is a hierarchical clustering approach. Afterwards, the k-means method, which is a non-hierarchical clustering approach, has been applied with the information on the number of clusters determined. Finally, the results of the two methods have been compared. The first stage of cluster analysis is to determine the variables that form the data matrix. In the study, the results have been obtained by considering the data taken from different sources, since all the current data for the statistics to be made could not be found together. Research data have been taken from the official websites of OECD, World Health Organization, data.worldbank on 16 August 2022. When the health indicators in the WHO's world health data platform have been examined, there are many indicators. The following ten variables commonly used in the literature have been used within the scope of the study.

Case/Population Ratio: For each country, it is the ratio of the number of cases announced from the first case to the study date to the country's population. Values realized per thousand people are taken as basis.

Death/Population Ratio: For each country, it is the ratio of the number of registered deaths reported from the day the first case occurred to the study date, to the country's population. Values realized per thousand people are taken as basis.

Total Recovered: For each country, it is the ratio of the number of registered recoveries announced from the date of the first case to the date of the study.

Number of Doctors: The number of doctors per thousand people was taken for each country.

Number of Nurses: The number of nurses per thousand people was taken for each country.

Number of Beds: The number of beds per thousand people was taken.

Health spending: It is the measurement of the final consumption of health care goods and service. This measurement is presented as a share of total health spending and in USD per capita.

Pharmaceutical spending: It covers in most countries to “net” spending. This measure is a share of total health spending.

Vaccine: For each country, it is the number of vaccine doses administered per one hundred population due the study date.

Intensive care beds capacity: For each country, it’s the capacity of intencisve care beds per 100000 population.

In this stage of the study, outlier evaluation has been carried out in the data. Outliers have been determined by the Mahalanobis distance values (m_i^2). Mahalanobis distance is the distance of any observation in the multivariate data matrix X from the center of the data (Alpar, 2013). When compared to a chi-square distribution with the same degrees of freedom, these Mahalanobis distances are evaluated. The amount of variables you have combined to determine the Mahalanobis Distances will be indicated in the degrees of freedom. The p-value for the chi-square distribution's right tail is computed. Table 1 presents the Mohalanobis distance values and probability values. The values of the new probability variable are less than 0.001, multivariate anomalies are detected. From the calculated distance values, it’s shown that United states is an outlier.

Table 1. The Outlier Results

Countries	m_i^2	p-value	Countries	m_i^2	p-value
Australia	3,4761	,9679	Austria	16,0338	,0987
Belgium	10,01114	,4395	Canada	16,4918	,0864
Chile	5,7850	,8330	Colombia	3,5431	,9656
Costa Rico	15,0443	,1305	Czech Republic	4,0183	,9465
Denmark	26,2750	,0034	Estonia	5,6926	,8404
Finland	7,5256	,6751	France	9,3660	,4978
Germany	9,2191	,5114	Greece	4,0086	,9470
Hungary	11,5221	,3183	Iceland	9,6947	,4677
Ireland	6,1709	,8007	Israel	9,5843	,4777
Italy	10,3295	,4121	Japan	5,3550	,8662
Korea	4,0006	,9473	Latvia	8,0351	,6254
Lithuania	12,5724	,2486	Luxemburg	6,4053	,7801
Mexico	11,8203	,2973	Netherlands	7,3101	,6959
New Zealand	19,8129	,0311	Norway	3,5298	,9661
Poland	9,3290	,5012	Portugal	4,9964	,8914
Slovak	7,3738	,6897	Slovenia	8,6285	,5677
Spain	2,7838	,9861	Sweden	20,2875	,0266
Switzerland	3,5376	,9658	Türkiye	8,5087	,5793
UnitedKingdom	5,8965	,8239	United states	36,0243	,0001

In the third phase of the study, it is investigated whether there is a multicollinearity problem in the data. If there is a multi-correlation problem in cluster analysis, one of the suggested methods is to choose between multi-correlated variables and continue working with the selected variable. The correlation coefficients are examined in determining the multicollinearity situation. The Shapiro-Wilk test is used to determine which factors to use following establishing whether or not the data is normally distributed. The result are given in Table 2. From Table 2, it is seen that the data for the Pharmaceutical spending, Doctor, Nurse, and Death rate variables are normally distributed, and for the other variables, the data are not normally distributed ($p<0.05$). Since the assumption of normal distribution for all variables in the data could not be ensured, Spearman correlation coefficient values given in Table 3 are examined to investigate the relationships between variables. According to the Spearman correlation coefficients, the same variables are used because there isn’t a very strong relationship between the variables.

Table 2. Test of Normality

Variables	Statistics	Sd	p
Health Spending	0,472	38	<0,001
Pharmaceutical spending	0,976	38	0,581
Doctor	0,963	38	0,232
Nurse	0,979	38	0,684
Bed	0,833	38	<0,001
Recovered	0,546	38	<0,001
Case Rate	0,159	38	<0,001
Death Rate	0,967	38	0,316
Test	0,935	38	0,030
Intensive care bed	0,837	38	<0,001

Table 3. Spearman correlation coefficient values

	Health-spending	Pharmet-spending	doctor	nurse	bed	recovered	case	death	test	Intensive care
Health-spending	1,0	-,565**	,386*	,784**	,05	,161	,372*	-,498**	,375*	,247
Pharmet-spending		1,000	-,090	-,479**	,46**	,039	-,124	,559**	-,375*	,037
doctor			1,000	,238	,16	-,172	,408*	,065	,107	,124
nurse				1,000	,04	-,083	,239	-,479**	,438**	,036
bed					1,0	,086	,250	,224	-,244	,519**
recovered						1,000	-,062	-,115	,235	,096
case							1,000	-,033	-,034	,325*
death								1,000	-,535**	,141
test									1,000	-,248
intensive care										1,000

** Correlation is significant at the 0.01 level (2-tailed).
*Correlation is significant at the 0.05 level (2-tailed).

In clustering analysis, observations close to each other and clusters consisting of different observation groups are determined. Distance measures are used to determine the clusters. The Euclidean distance is used in quantitative data and is one of the most widely used distance measures (Alpar, 2013). The Euclidean distance is the most widely recognized measure of distance, often referred to as the straight line distance. The study data is in a quantitative structure, the Euclidean distance measure, which is one of the most used distance measures in the quantitative data structure, has been used in the study. The analyzes have been performed using Ward method, which provides a reliable analysis by minimizing the variance difference between the variables, in the hierarchical clustering method and dendrogram analysis and agglomeration are explained with the help of charts. The tree graph of the Ward method is presented in Figure 1. According to the tree graph, it is seen that United States is a separate cluster, Australia, Netherlands, Japan, Spain, Türkiye, France, Germany, Italy, Korea, and United Kingdom are a separate cluster, and the other countries are classified in a single cluster. It is evident that there are three separate clusters within the parameters of the research.

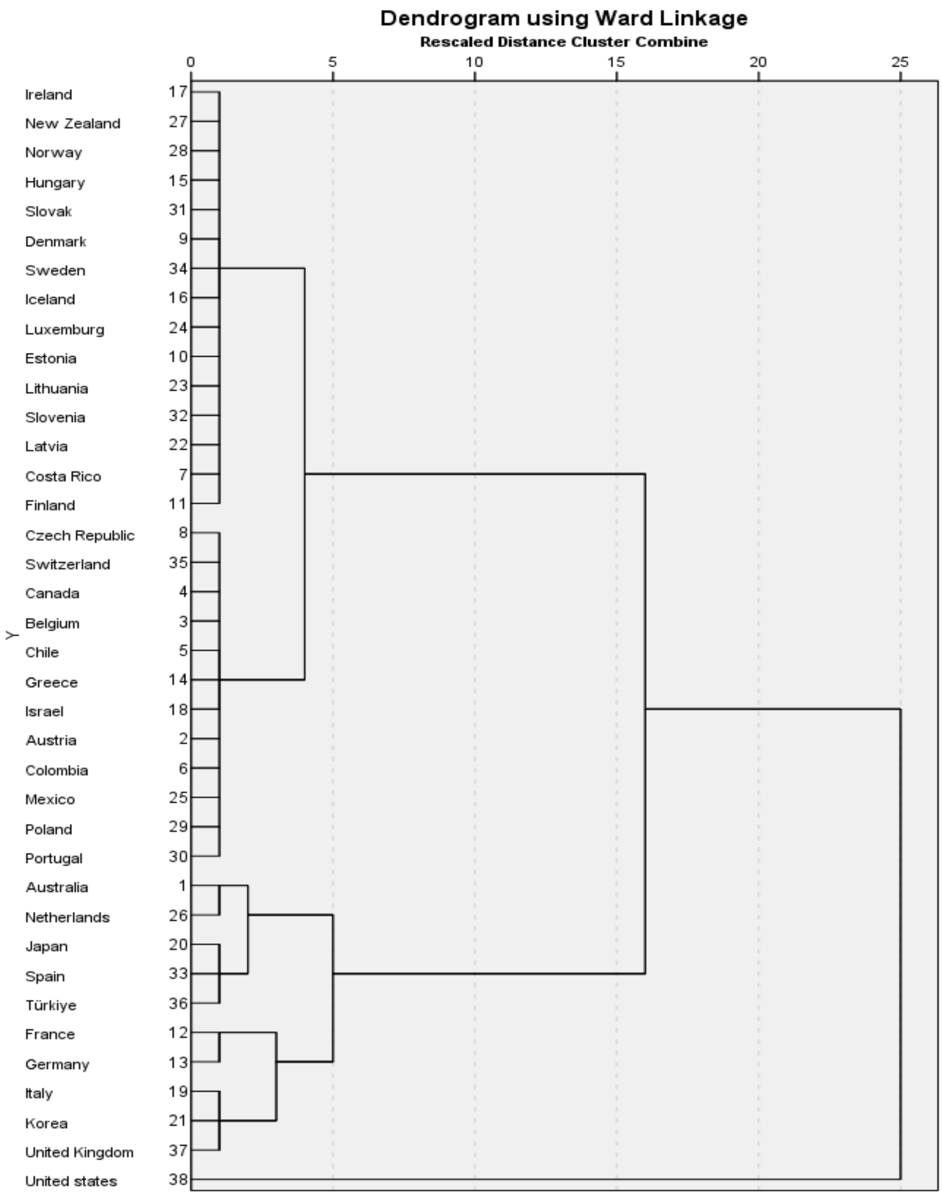

Figure 1. The tree graph of the Ward method
Figure by Author

The agglomeration chart is given in Table 4. The column of coefficients in the accumulation chart shows the Euclidean distance, and the Euclidean distance measures the distance between observations. The smallest coefficient is 17: Ireland and 27: New Zealand. In the first phase of clustering, the two closest countries, Ireland and New Zealand are combined into a cluster. In the second stage of clustering from Table 2, 6: Colombia and 25: Mexico are combined in a cluster as the two closest countries in the second place. In the third stage, 14: Greece and 18: Israel merge into one cluster. In the thirty-seventh stage, which is the last stage of clustering, all countries are united in a cluster.

Table 4. Agglomeration Schedule

Stage	Cluster1	Cluster2	Coefficients	Stage	Cluster1	Cluster2	Coefficients
1	17	27	0,0000	20	19	21	0,0201
2	6	25	0,0001	21	20	33	0,0233

3	14	18	0,0004	22	1	26	0,0284
4	23	32	0,0007	23	2	4	0,0338
5	8	35	0,0010	24	6	29	0,0396
6	3	5	0,0015	25	9	15	0,0488
7	16	24	0,0020	26	7	10	0,0584
8	15	31	0,0025	27	12	13	0,0703
9	10	16	0,0032	28	20	36	0,0842
10	2	14	0,0040	29	19	37	0,0983
11	22	23	0,0048	30	2	6	0,1190
12	17	28	0,0056	31	7	9	0,1412
13	7	11	0,0068	32	1	20	0,1807
14	29	30	0,0080	33	12	19	0,2662
15	4	8	0,0093	34	2	7	0,3964
16	7	22	0,0108	35	1	12	0,5604
17	15	17	0,0125	36	1	2	1,1414
18	9	34	0,0146	37	1	38	2,0512
19	2	3	0,0169				

The k-mean clustering method, which is one of the non-hierarchical methods, is applied. Considering the number of clusters determined by the hierarchical clustering method, the k-mean clustering method is performed. Since three clusters were determined with the Ward method, the analyses for 3 clusters and 10 iterations have been discussed with this method. The obtained clusters are given in Table 5. France, Germany, Türkiye, Italy, Korea, and United Kingdom are in cluster 1, United States is in cluster 2, the remained 31 countries are in cluster 3. When the number of clusters is taken as three, the cluster centers are given in Table 6. The highest health spending is in cluster 3, the highest pharmaceutical spending is in cluster 3, the highest number of doctors is in cluster 3, the highest number of nurses is in cluster 2, the highest total recovered number is in cluster 2, the lowest case rate is in cluster 3, the lowest death rate is in cluster 2, the highest test rate is in cluster 1, the highest number of intensive care bed is in cluster 2.

Table 5. The Results of k=3 Cluster Analysis

Countries	Cluster	Countries	Cluster
Australi	3	Israel	3
Austria	3	Italy	1
Belgium	3	Japan	3
Brazil	3	Korea	1
Bulgaria	3	Latvia	3
Canada	3	Lithuania	3
Chile	3	Luxemburg	3
China	3	Mexico	3
Colombia	3	Netherlands	3
Costa Rica	3	New Zealand	3
Croatia	3	Norway	3
Cypruz	3	Poland	3
Czech Republic	3	Portugal	3
Denmark	3	Romania	3
Estonia	3	Slovak	3

Finland	3	Slovenia	3
France	1	Spain	3
Germany	1	Sweden	3
Greece	3	Switzerland	3
Hungary	3	Türkiye	1
Iceland	3	United Kingdom	1
India	3	United states	2
Ireland	3		

Table 6. Clustering Center

Variable	Cluster 1	2	3
Health Spending	4582,40	12318,10	5287,80
Pharmaceutical spending	13713,66	11044,00	14699,29
Doctor	3293,33	2640,00	3765,90
Nurse	8291,67	11980,00	9332,26
Bed	5,81	2,80	4,09
Recovered	23873156,00	90035472,00	3910676,16
Case Rate	367491,00	95070423,00	345883,58
Death Rate	1887,50	1064,00	2104,39
Test	221,83	181,00	201,90
Intensive care bed	19,95	25,80	11,88

Table 7 presents the distance between clusters. The distance between clusters 2 and 3 is the largest. Cluster 1 and cluster 3 are the closest to one another. More similarities exist between cluster 1 and cluster 3 than between cluster 2 and cluster 3. When the results of the two methods are compared, the Unites Stated is assigned to a single cluster in both methods, Australia, Netherlands, Japan and Spain, which are in the second cluster in the ward method, are assigned to the third cluster in the k-mean clustering method. In this study, the United States is the country that is an outlier. Due to the effect of outlier data, it seems likely that many countries are classified in one cluster and one country is classified in other clusters.

Table 7. Distances between Cluster Centers

Cluster	1	2	3
1	-	115525310,95	19962491,603
2	115525310,95	-	128024290,27
3	19962491,603	128024290,27	-

5. CONCLUSION

Healthy individuals and healthy societies composed of these healthy individuals are an indicator of a country's degree of development in terms of health. At the level of society, the establishment and accessibility to health facilities is crucial. As a result, both the quality and quantity of health services provided to people should be adequate. The current COVID-19 pandemic, technological advancements, rising population demand, and high standards for service expectations have an impact on every industry, including health care. It is crucial to keep track of the health services provided in order to analyze how effectively they were performed for this reason. The World Health Organization (WHO) is one of the examples of international groups that decide on health indicators. Health indicators are indicators that are used to monitor and assess the health status of the community they are used for, as well as for planning and managing health services, producing and supervising policies related to those services, and planning and managing health services. Health indicators offer benchmarks for tracking changes in population health state and comparing populations around the world.

Starting in China in December 2019, the pandemic of Covid-19 has been spreading across the world since March 2020. Within the scope of this study, a cluster analysis of 38 OECD member countries was carried out using health indicators and the latest data to provide a comparative analysis of various preventive measures adopted by different countries. Understanding the strategies adopted by each country can provide a comprehensive perspective on the worldwide reactions to the pandemic. In additon, these health indicators can be useful in showing how countries can be better prepared for future pandemics or health crises. Additionally, the study may provide some suggestions to policy makers in future research. Since the test numbers of China have been not reported, China isn't included in the analysis of the study. As indicators, a total of ten variables consisting of the number of doctors, the number of nurses, the number of beds, the number of totals recovered, intensive care beds capacity, health spending, pharmaceutical spending, the number of vaccines, case/population ratio, death/population ratio are considered. For the analysis of the data, the ward method, which is one of the hierarchical clustering algorithm, and the k-means method, which is one of the non-hierarchical clustering algorithm, are employed.

OECD countries are divided into three clusters according to the results of Ward's method, which is a hierarchical method. In 38 countries, 3% (1 country) are in the first cluster, 26% (10 countries) are in the second cluster, and 71% (27 countries) are in the third cluster. The United states is in Cluster 1. Countries where the pandemic is intense, such as Australia, Netherlands, Japan, Spain, Türkiye, France, Germany, Italy, Korea, and United Kingdom, are included in Cluster 2. The remained countries are in Cluster 3. The success of the countries in the third cluster against COVID-19 isn't as successful as in the second countries, and the number of cases and deaths in these countries is higher than the countries are assigned to the second cluster.

OECD countries are classified into three clusters according to the ward method. In 38 countries, 16% (6 countries) are in the first cluster, 3% (1 country) are in the second cluster, and 82% (31 countries) are in the third cluster. Cluster 1 includes France, Germany, Türkiye, Italy, Korea, and United Kingdom. Cluster 2 includes The United states. The remained countries are assigned to Cluster 3. In this method, the countries that have performed successfully during the Covid-19 pandemic process and are in a better position than the indicators are in the first cluster. The United States are assigned to a separate cluster as the only country in both clusters, in proportion to the result of being found to be outliers in the analysis.

The results demostrated that the USA differs from other OECD member countries according to the health indicators examined. The United States has a highest expenditure on healthcare than any other country. Similarly, it is the country with the highest value in the number of cases and the number of recovered compared to other countries. The USA has been found to be the outlier country in the statistical analysis. In the results of the cluster analysis, the USA is included as the only country in a cluster. This shows that the statistical analyzes performed have obtained parallel results. In the two methods, Türkiye has been grouped together with the following countries: France, Germany, Italy, Korea, United Kingdom. They have carried out the process successfully with the policies they put forward during these crisis periods. One of the countries that has been used as an example in the struggle against the epidemic is Korea, one of the countries with a high number of coronavirus cases. The country has taken steps such as testing a large population, sanitizing the streets, implementing treatment plans, carefully observing the citizens, and administering tests to those who are thought to be at risk. Following the steps, the country's daily case count dropped and the rise in deaths came to an end. It is an indicator that effective administration can make a difference in the struggle against epidemics. Germany is one of the countries that has been successful in the struggle against the epidemic. The country took measures to control the spread of the epidemic and to control it, protected the people in the risk group and increased the capacity of the health system. In terms of healthcare spending, pharmaceutical spending, recovered, and test numbers, France and Italy are leading among the other countries. Türkiye is among the competent countries in the prevention of epidemics, the number of recovered patients, the intensive care bed, and the number of tests. It is the second-highest capacity among the countries that are taken into consideration for the number of intensive care beds, per country statistics. By increasing the number of recovered patients, this has raised the whole country to the status of a leader in its struggle against the epidemic. In accordance to their similarity in terms of health indicators during the epidemic time period, OECD countries are grouped in this study. According to the performance indicators and clustering results, Türkiye has shown a successful effort in general during the pandemic period and still.

REFERENCES

Alkaya, A., & Alkaş, C. (2021). OECD Ülkelerinin Sağlık Göstergelerine Göre Kümeleme Analizi Sınıflaması. *Sosyal Güvence*, (19), 427-474.

Alpar R. (2013). *Çok Değişkenli İstatistiksel Yöntemler*. Detay Yayıncılık.

Alptekin, N. (2014). "Comparison of Türkiye and European Union Countries' Health Indicators by Using Fuzzy Clustering Analysis." *International Journal of Business and Social Research,* 4-10.

Cao, H., Jia, L., Si, G., & Zhang, Y. (2013). A clustering-analysis-based membership functions formation method for fuzzy controller of ball mill pulverizing system. *Journal of process Control*, 23(1), 34-43.

CIHI (2009). Canadian Institute for Health Information. *Report from the Third Health Indicators Consensus Conference*.

Costa, C., Freitas, Â., Stefanik, I., Krafft, T., Pilot, E., Morrison, J. & Santana, P. (2019). Evaluation of Data Availability on Population Health Indicators at The Regional Level Across The European Union. *Population Health Metrics,* 17(11), 1-15.

Çakmak, Z., Uzgören, N., & Keçek, G. (2005). Kümeleme analizi teknikleri ile illerin kültürel yapılarına göre sınıflandırılması ve değişimlerin incelenmesi. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, (12).

Çelik, Ş. (2013). "Kümeleme Analizi ile Sağlık Göstergelerine Göre Türkiye'deki İllerin Sınıflandırılması", *Doğuş Üniversitesi Dergisi,* 14(2), 175-194.

Çetintürk, İ., & Gençtürk, M. (2020). OECD Ülkelerinin Sağlık Harcama Göstergelerinin Kümeleme Analizi ile Sınıflandırılması. *Süleyman Demirel Üniversitesi Vizyoner Dergisi*, 11(26), 228-244.

Dash, S., Shakyawar, S.K., Sharma, M., Kaushik, S., 2019, "Big data in healthcare: management, analysis and future prospects," J. *Big Data* 6, vol. 54.

Değirmenci, N. & Yakıcı-Ayan, T. (2020). OECD Ülkelerinin Sağlık Göstergeleri Açısından Bulanık Kümeleme Analizi ve TOPSIS Yöntemine Göre Değerlendirilmesi. *Hacettepe Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 38(2), 229-241.

Demircioğlu, M., & Eşiyok, S. (2020). Covıd–19 Salgını İle Mücadelede Kümeleme Analizi İle Ülkelerin Sınıflandırılması. *İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi*, 19(37), 369-389.

Ersöz, F. (2009). "OECD'ye Üye Ülkelerin Seçilmiş Sağlık Göstergelerinin Kümeleme ve Ayırma Analizi ile Karşılaştırılması.", *Türkiye Klinikleri Journal of Medical Sciences*, 29(6), 1650-1659.

Ganesh, S. (2002). Data mining: Should it be included in the statistics curriculum?. *In The 6th international conference on teaching statistics (ICOTS 6), Cape Town, South Africa.*

Gonzalez, G. H., Tahsin, T., Goodale, B. C., Greene, A. C., & Greene, C. S. (2016). Recent advances and emerging applications in text and data mining for biomedical discovery. *Briefings in bioinformatics*, 17(1), 33-42.

Hair, J. F. Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L.(2014), *Multivariate Data Analysis*. USA: Pearson.

Han, J., Kamber, M., & Pei, J., 2012, Data mining: concepts and techniques. *Elsevier/Morgan Kaufmann*, 2012.

Ji, J., Bai, T., Zhou, C., Ma, C., & Wang, Z. (2013). An improved k-prototypes clustering algorithm for mixed numeric and categorical data. *Neurocomputing*, 120, 590-596.

Kartal, E., Balaban, M. E., & Bayraktar, B. (2021). Küresel Covıd-19 Salgınının Dünyada ve Türkiye'de Değişen Durumu ve Kümeleme Analizi. *Journal Of Istanbul Faculty Of Medicine*, 84(1), 9-19.

Karypis, M. S. G., Kumar, V., & Steinbach, M. (2000, August). A comparison of document clustering techniques. *In TextMining Workshop*.

Kecman, V. (2001). *Learning and soft computing: support vector machines, neural networks, and fuzzy logic models*. MIT press.

Köse, A. (2022). Türkiye'de sağlık göstergelerine göre istatistik bölgelerin kümeleme analizi ile sınıflandırılması. *Alanya Akademik Bakış*, 6(2), 2167-2189.

Kurasova, O., Marcinkevicius, V., Medvedev, V., Rapecka, A., & Stefanovic, P. (2014) "Strategies for Big Data Clustering". *IEEE 26th International Conference on Tools with Artificial Intelligence.*

Mut, S., Akyürek, Ç. E. (2017). "OECD Ülkelerinin Sağlık Göstergelerine Göre Kümeleme Analizi ile Sınıflandırılması". *International Journal of Academic Value Studies*, 3(12), 411-422.

Pramanik, M. I., Lau, R. Y., Azad, M. A. K., Hossain, M. S., Chowdhury, M. K. H., & Karmaker, B. K. (2020). Healthcare informatics and analytics in big data. *Expert Systems with Applications*, 152, 113388.

Paumelle, M., Occelli, F., Wakim, L. M., Brousmiche, D., Bouhadj, L., Ternynck, C., ... & Deram, A. (2023). Description of the multi-dimensional environment at the territorial scale: A holistic framework using cluster analysis and open data in France. *Ecological Indicators*, 154, 110562.

Proksch, D., Busch-Casler, J., Haberstroh, M. M., & Pinkwart, A. (2019). National health innovation systems: Clustering the OECD countries by innovative output in healthcare using a multi indicator approach. *Research Policy*, 48(1), 169-179.

Reibling, N. Ariaans, M. & Wendt, C. (2019). Worlds of Healthcare: A Healthcare System Typology of OECD Countries. *Health Policy*, 123(7), 611- 620.

Santos-Pereira, J., Gruenwald, L., & Bernardino, J. (2022). Top data mining tools for the healthcare industry. *Journal of King Saud University-Computer and Information Sciences*, 34(8), 4968-4982.

Singhal, M., & Shukla, S. (2018, February). Centroid selection in kernel extreme learning machine using K-Means. *In 2018 5th International Conference on Signal Processing and Integrated Networks*.

Sonğur, C. (2016). "Sağlık Göstergelerine Göre Ekonomik Kalkınma ve İş birliği Örgütü Ülkelerinin Kümeleme Analizi". *Sosyal Güvenlik Dergisi*, 6(1).

Strang, K.D., & Sun, Z., 2020. Hidden big data analytics issues in the healthcare industry *Health Informatics Journal* 26(2) 981–998.

Tüzüntürk, S. (2010). Veri madenciliği ve istatistik. *Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi,* 29(1), 65-90.

Tzortzis, G., & Likas, A. (2014). The MinMax k-Means clustering algorithm. *Pattern recognition*, 47(7), 2505-2516.

Ünsal, M. G., & Kasap, R. (2020). Investigating Covid 19 data for G20, EU, and OECD countries via using time series analysis & cluster analysis. *Journal of Turkish Operations Management*, 4(2), 424-432.

WHO (1998). Health Promotion Glossary https://www.who.int/healthpromotion/about/HPR%20Glossary%201998.pdf

Wulandari, L., & Yogantara, B. O. (2022). Algorithm analysis of K-means and fuzzy C-means for clustering countries based on economy and health. *Faktor Exacta*, 15(2).

Yıldırım, S., Yildirim, D. C., & Calıskan, H. (2020). The influence of health on economic growth from the perspective of sustainable development: a case of OECD countries.

World Journal of Entrepreneurship, *Management and Sustainable Development,* 16(3), 181-194.

Zhang, D., & Zhou, L. (2004). Discovering golden nuggets: data mining in financial application. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 34(4), 513-522.

Zou, H. (2020). Clustering algorithm and its application in data mining. *Wireless Personal Communications*, 110(1), 21-30.