Celal Bayar University Journal of Science
Volume 20, Issue 1, 2024, p 41-48
Doi: 10.18466/cbayarfbe.1416472


**Celal Bayar University Journal of Science**

# Investigation of Earthquakes in Turkey with Cluster Analysis*

Ahmet Murat ŞEN[1]**, Zeki YILDIZ[2]

*This study is generated from the thesis titled "Investigation of Markovian Structure According to the Magnitudes of Eartquakes in Turkey by Cluster Analysis", which is being written in the Department of Statistics, Institute of Science and Technology, Eskişehir Osmangazi University, Eskişehir, Turkey.

[1]Department of Statistics, Institute of Science, Eskişehir Osmangazi University, Eskişehir, Turkey
[2]Department of Statistics, Faculty of Science, Eskişehir Osmangazi University, Eskişehir, Turkey

**501420220002@ogu.edu.tr
** Orcid No: 0000-0003-4359-5000

Received: 8 January 2024
Accepted: 24 March 2024
DOI: 10.18466/cbayarfbe.1416472

**Abstract**

In this study, earthquakes with a magnitude of 3 and above occurring between 36-42 degrees North latitude and 26-45 degrees East longitude, within and around Turkey from 1980 to 2022, were considered. Turkey and its immediate surroundings were divided into 114-unit squares using latitude and longitude coordinates to investigate seismic similarities. The variables including 50130 earthquake depths and magnitudes in our data set were analyzed using the {K-Means Cluster Analysis} method in the SPSS program. Hierarchical clustering method was used to determine the number of clusters and the number of clusters was determined as 5 with the help of the dendrogram obtained. As a result of the cluster analysis, it was observed that 47% of the earthquakes in Turkey were at a depth of 5 to 10 km. It has been observed that 93% of the earthquake intensities occur between 3 and 4 magnitudes. A cluster analysis was conducted to assess the similarities and dissimilarities among earthquakes within these unit squares in terms of their magnitudes and depths. Due to the large sample size and the analysis involving continuous variables, the earthquake data were grouped into 5 clusters using non-hierarchical clustering methods, and the relationships within and between these clusters were observed.

**Keywords:** Cluster Analysis, Earthquake, Depth, Magnitude.

## 1. Introduction

The two severe earthquakes that occurred in our country on February 6, 2023, followed by the earthquake storm that emerged afterwards, as well as the ongoing earthquakes, have been among the biggest disasters our country has faced. These tragic events have deeply saddened us as a whole country and other countries also shared the same pain. It is not possible to predict the earthquake in terms of time and date. However, in terms of probability, we can scientifically predict in what intensity range and in how many years [1-2-3]. When looking at recent studies that have conducted cluster analysis on earthquake-prone area data; earthquakes occurring near the city of Bengkulu in Indonesia were analyzed using the k-means technique for cluster analysis [15]. Similarly, seismic earthquakes in India were analyzed using the k-means technique for cluster analysis [16]. Earthquakes on the Zagros mountains within the borders of Iran were also examined using the k-means technique, as in previous studies. Based on these studies; earthquakes in Turkey have been grouped using the k-means technique for cluster analysis based on their depth and magnitude, yielding scientific results. Although Turkey is located in the Alpine-Himalayan fold belt, a large part of it is located in the earthquake zone and is constantly compressed by the Eurasian plate in the north and the African plate in the south due to the inability of Anatolia to complete its evolution.

Due to this compression, the fault lines formed in Anatolia are located in certain sections and these are called the North Anatolian Fault, the South Anatolian Fault, the Aegean Graben System and the Southeastern Anatolian Thrust Belt. Our country is located between (36º-42º) northern latitudes and (26º-45º) eastern longitudes. It includes 19 longitudes and 6 latitudes. As can be seen in Figure 1, 114 regions are visualized on the map of Turkey by dividing them into latitude and longitude squares.

We refer to each of these allocated sections as a unit

Celal Bayar University Journal of Science
Volume 20, Issue 1, 2024, p 41-48
Doi: 10.18466/cbayarfbe.1416472


square. Earthquake data can be examined through these unit squares and the relationships between these earthquakes can be evaluated.

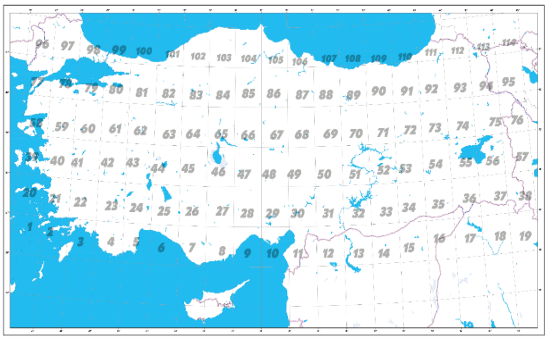

**Figure 1.** Dividing the latitude and longitude coordinates of the Turkey map into unit squares.

## 2. Materials and Methods

### 2.1. Cluster Analysis

If we define cluster analysis, it is a method of dividing the stack of variables into certain groups in terms of their similarities within themselves. It is also a way to find general patterns of distribution in data. It is widely used in statistical analysis [4].

Cluster analysis provides the emergence of certain important features in data analysis. It has the ability to work with unorganized data sets whose dependent variable is uncertain or neglected during the data analysis phase. To measure the distances between units, measurement techniques such as Euclidean distance, Minkowski distance, Manhattan distance, and Jaccard coefficient are used. The main purpose of using these techniques is to ensure the formation of clusters where units exhibit homogeneous properties within themselves. In more general terms, these measurement techniques aim to minimize distances within clusters and maximize distances between clusters [5].

Assumptions that are vital for other statistical methods such as normality, linearity, and constant variance do not have much importance in Cluster Analysis. For this reason, in Cluster Analysis applications, it is recommended to take into account the ability of the data to represent the sample in question and whether it has multiple linked variables. The most important issue to be decided at the stage of choosing clustering analysis will be the degree to which the sample to be examined represents the subgroups in the stack [6].

Cluster analysis is one of the most used techniques as a result of the advancement of technology and is widely used in data mining, especially for large data sets. The cluster analysis method, which was mostly performed manually in the past, can now be applied to large datasets with the assistance of computer technologies, in parallel with technological advancements. Today, it is widely used primarily in market research, customer portfolios of companies, medicine, etc. [7]. Basically, cluster analysis is an analysis technique that aims to collect data with different characteristics in the same group by making use of certain similarities. There are many algorithms available in cluster analysis. In general, these algorithms are grouped into two groups: hierarchical clustering techniques that create dendrograms and non-hierarchical clustering techniques. The main goal of both of these techniques is to maximize the differences between clusters and intra-cluster similarities. In other words, it is to maximize the similarity situation within the sets and to minimize the similarity between the sets. Which of these techniques is preferred depends on the number of clusters, but it is much more useful to use both techniques together? Thus, it helps us to compare the results of both techniques used and to choose the most appropriate technique to be used [8]. Hierarchical clustering method is used in data sets where the number of clusters is not specified. The number of clusters is determined based on the chosen hierarchical clustering technique. In cases where the number of clusters is known, the K-Average technique, which is one of the non-hierarchical clustering methods, is used. These groupings are made according to the specified number of clusters. When determining the clustering analysis technique, the nature of the data set is generally taken into consideration. In this case, some data are collected by determining the first data, while some data are divided from the total and clustered. Figure 1 illustrates the separation of latitude and longitude of the map of Turkey into unit squares.

#### 2.1.1. Hierarchical Clustering Techniques

Hierarchical techniques start with a matrix of distances between units. At the beginning of the analysis phase, each unit is considered as a single set. Afterwards, the groups that are closest to each other are combined with each other and this merging process is repeated one after the other [8]. Among the hierarchical techniques, the most accepted ones are [12];

- Single Linkage Technique (Nearest Neighborhood)
- Complete Linkage Technique (Farthest Neighborhood)
- Average Group Linkage Technique
- Ward Technique
- Median Technique
- Centroid Technique.

##### 2.1.1.1. Single Linkage Technique

In single linkage technique, the existing distance between two clusters begins with the minimum distance between the clusters [9].

#### 2.1.1.2. Complete Linkage Technique

In other words, the "Farthest Neighborhood" technique is similar in theory to the single linkage technique. The difference between this technique and the single linkage technique is that it makes use of the maximum distance between two units, instead of the minimum distance between clusters [9].

#### 2.1.1.3. Average Group Linkage Technique

In the Average Group Linkage Technique, the distance between the two groups must be short in order to unite. In other words, the calculation of the difference between two sets is done by taking the average differences between the pairs of units in one set and the pairs of units in another set [9].

#### 2.1.1.4. Ward Technique

It is a general hierarchical clustering technique created by Ward in 1963 to assist with partial problems. It is also called the minimum variance method. The technique aims to combine two sets with a minimum sum of squares within sets (minimum variance within the group) [13].

#### 2.1.1.5. Median Technique

It was developed by Gower in 1967 to determine the calculation point of cluster distances as the midpoint. Thus, for n clusters, each of which contains a single unit at the beginning, the distance between each cluster will be equal and equal to its median value [13].

#### 2.1.1.6. Centroid Technique

In the centroid technique, the distance between clusters is expressed as the Euclidean distance between cluster averages. Thus, the weighted average of the cluster means is considered to be the cluster center [12].

### 2.1.2 Non-Hierarchical Clustering Techniques

If the researcher has a preliminary knowledge about the number of clusters, in other words, if the number of clusters is decided by the researcher, non-hierarchical techniques are used instead of hierarchical techniques [8]. Non-hierarchical clustering techniques have the advantage of being faster than hierarchical clustering techniques. In addition, the higher the sample size, the more meaningful the results will be. In many non-hierarchical clustering techniques, the number of clusters is predetermined by the researcher as ‘k’ according to the sample size. In general, it is recommended to use hierarchical techniques and non-hierarchical techniques together [8]. The k-mean technique, which is a hierarchical clustering technique, was developed by J.A. Hartigon and M.A. Wong, and the purpose of this technique is to create a small number of clusters among a large number of units. Although this technique is used both for units that contain discrete variables, it can also be applied to continuous variables that do not contain extreme values. The primary objective in using this technique is to minimize the sum of squares within clusters, thereby minimizing cluster variability and dividing a p-dimensional unit into 'k' clusters. [14].

## 3. Cluster Analysis in Earthquake Studies

In the study conducted by [15], earthquakes occurring near the city of Bengkulu in Indonesia were grouped according to seismic classifications, based on location and magnitudes, using cluster analysis to categorize earthquake characteristics. While performing the cluster analysis of the earthquake source regions, weightings were made with the k-means technique. In this study, 19 optimal earthquake clusters were obtained. The data consists of earthquakes of magnitude 5 Ms (surface wave) and above, which occurred from January 1970 to December 2015 in and around Bengkulu, West Sumatra, Lampung, and the Indian Ocean regions. As a result, earthquakes in and around the city of Bengkulu mostly occur in the Indian Ocean , with only minor earthquakes occurring on the mainland. Most earthquakes with a magnitude of 5 ms to 6 ms are tectonic earthquakes that occur at a depth of less than 100 km. Among the earthquake clusters created, the most striking is the earthquake that occurred around the Mentawai Islands of West Sumatra province, and another in the seas of Lampung Province.

In the study conducted by [16], cluster analysis was performed using the k-means method based on the spatial and positional magnitudes of seismic earthquakes that occurred in India over the last 10 years. The performance of the study is calculated using the sum of the squares of the intra-cluster error coefficients. In the current study, 6 earthquake clusters were obtained. The applied dataset consists of 1657 seismic events that occurred in India between January 1, 2005 and December 31, 2015. Each seismic event includes the year, month, day, hour, latitude, longitude, and magnitude. The focus of the study has been on the magnitude of the earthquake rather than its time. As a result, other variables have been excluded from the research, and only earthquake magnitudes have been included in the cluster analysis. Euclidean distances were used to calculate the distance between the two clusters. As a result, the k-means method has a strong potential to provide a superior tool in earthquake cluster analysis. In the study conducted by [17], earthquakes occurring on the Zagros Mountains within the borders of Iran were investigated and categorized into seismic zones based on their respective faults. The k-means technique was used for clustering these regions, with the dataset containing 554 earthquakes of magnitude 4 and above that occurred between 2006 and 2019. By including outlier values, a total of 13 clusters have emerged, and it is observed that the clusters are more homogeneous as a result of the weighting process performed using the k-means technique. Examples of widespread studies of cluster analyses in the field of earthquakes are [15-16-17-18-19-20-22]. In the data sets consisting of earthquake data, aftershocks and precursor shocks were determined as dependent variables, and these variables formed distributions according to probability laws and parameters. These distributions are estimated by statistical methods [21].

Since non-hierarchical clustering methods generally contain large data sets and the earthquake data in question consists of continuous variables, it would be more appropriate to group them with the non-hierarchical clustering method in accordance with the data set we have. Although it is expected for us that the intra-group homogeneity is maximum and the similarity between the groups is minimum in order to create ideal clusters, it is aimed to create optimal groups by using the K-Mean technique for the number of clusters determined by us and then to examine the heterogeneity of the said groups with the Euclidean Distance. In the stages of cluster analysis, while determining certain objectives and after the selection of variables, the specific questions that the researcher should answer are as follows [24];

- Is the sample size large enough?
-Are there outliers in the data and can these outliers be removed from the research?
-How can the similarities between observations be determined?
-Is there a need for standardization of the data?

Many different approaches can be used to answer the above questions. But none of these questions is sufficient to be a definitive answer. For this reason, each different approach will produce different results for the same data. For this reason, just like factor analysis, cluster analysis will be more affected by the design of the research and the method selection process compared to other multivariate methods [25]. The main purpose of cluster analysis is the similarity and distances between units. For this reason, the first step of the analysis in question is the creation of a similarity or distance matrix. However, one of the most important decision points is whether standardization of the data is necessary before calculating similarity measurements. The reason for this is that these measurements are sensitive to size differences between different scales and variables [25]. Various units of distance measurement are proposed to calculate the distances of units to variables and between each other. The measurement units in question also vary according to the measurement units of the variables present in the data matrix. If the variables are obtained by proportional or intermittent scales, distances or relationship type measures are used. If the measurements are made according to binary observations, the existing similarity and difference measurements between the units are beneficial. [11]. In order to obtain reliable results in cluster analysis, the number of clusters must be accurately determined after the variables are included in the model. The researcher's initiative is important in deciding the number of clusters, but certain methods have been developed to minimize bias. The most practical formula for determining the number of clusters is as follows [8-10].

$$\boldsymbol{k = (\tfrac{n}{2})^{1/2}} \qquad \textbf{(3.1)}$$

k: Number of clusters
n: Sample size
The above formula is mostly used in small volume data sets. Its use in large samples creates difficulties for the researcher in reaching healthy results.
The second method for determining the number of clusters is the one proposed by Marriot;

$$\boldsymbol{M = k^2|W|} \qquad \textbf{(3.2)}$$

W = Within -Groups Sum of squares and Cross Products Matrix
M = Number of clusters
The third method of determining the number of clusters is the method developed by Calinsky and Harabasz;

$$\boldsymbol{C = [Iz(B)/(k-1)]\ /\ [Iz(W)/(n-k)]} \quad \textbf{(3.3)}$$

W = Within -Groups Sum of squares and Cross Products Matrix
B = Between-Groups Sum of squares and Cross Products Matrix

The fourth method of determining the number of clusters was proposed by Lewis and Thomas. In this method, two criteria are taken into account when deciding on the number of clusters. The first of these criteria is the explanatory power of the total variance, and the second is that the addition of a new set increases the set variance. Apart from these methods, there are also some informative and advisory rules when deciding on the number of clusters [8].

-Theoretical and practical considerations can suggest a certain number of sets.
- In hierarchical clustering, the distances at which the clusters converge can be used critically. This information can be obtained from the dendrogram.
-In non-hierarchical clustering, the ratio of total Between-Groups variance to Within-Groups variance can be graphed against the number of clusters. The point at which a certain break occurs indicates the appropriate number of clusters. After this point, it is not beneficial to increase the number of clusters.
-Considering the total number of units and the number of variables, the number of clusters obtained should be significant.

Celal Bayar University Journal of Science
Volume 20, Issue 1, 2024, p 41-48
Doi: 10.18466/cbayarfbe.1416472


## 4. Data

This study includes earthquake data covering the year 1900 and after, which is published on the website of Boğaziçi University Kandilli Observatory Regional Earthquake-Tsunami Monitoring and Evaluation Center [23]. In this data set, since the earthquake data measured between 1900 and 1980 did not contain significant results, a study was carried out on earthquakes of 3 and above that occurred between 01.01.1980 and 31.12.2022. The dataset contains a total of 50130 earthquake records. Specific concepts related to our data set are as follows;

**Magnitude**: "xM" The concept of Magnitude, which is used as the unit of measurement of earthquakes, is used instead of the Richter scale.
**Depth:** It is the shortest distance from the point where the energy is released in the earthquake to the earth.
**Latitude:** The angular distance of any point north or south of the Equator to the Equator is called Latitude.
**Longitude:** The angular distance of any point east or west of the initial meridian is called Longitude.

When we examine the earthquake data from 1980 to 2022, we reach the following conclusions:

- **5 ≤ M** "There have been 306 earthquakes with a magnitude of 5 or higher."
- **4 ≤ M < 5** "There have been 3355 earthquakes between magnitude 4 and 5."
- **3 ≤ M < 4** "There have been 46469 earthquakes between magnitude 3 and 4."

## 5. Results and Discussions

50130 earthquake depths and earthquake intensities in our data set were considered as variables and the "K-Means Cluster Analysis" method was used in the SPSS program. First of all, the Hierarchical Clustering Method was used to calculate the optimal number of clusters, and as a result of the dendrogram obtained, it is seen that in figure 2 we need to continue the analysis with 5 clusters.

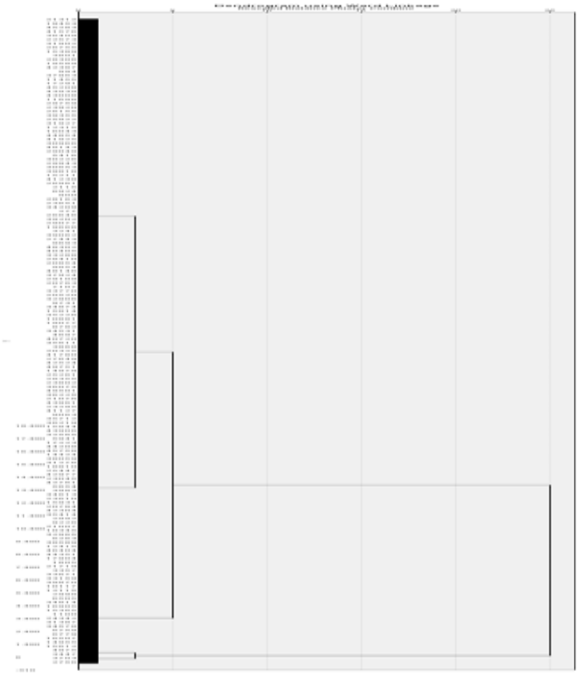

**Figure 2.** Dendrogram representation of depth and magnitudes

Different cluster element membership in dendrograms emerges as a result of differentiation in the sub-dendrogram and is determined by calculating the maximum number of edges [26].
Due to the large volume of our dataset, we will be using the Non-Hierarchical Clustering Method. When we analyze the results obtained by incorporating the 5 clusters identified in the dendrogram, the frequency values of each cluster are presented in Table 1 below;

**Table 1**. Frequency Distribution of 5 Formed Clusters

| | | |
|---|---|---|
| **Clusters** | 1 | 487 |
| | 2 | 164 |
| | 3 | 340 |
| | 4 | 42070 |
| | 5 | 7069 |
| **Total Frequency** | | 50130 |

As can be seen Table 1; the most agglomeration occurs in the 4th cluster. When the center depth and magnitude values of each cluster are examined, the following results are obtained.

**Table 2**. 5 Average depth and magnitude of cluster centers

| **Clusters** | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| **Depth(km)** | 63.7 | 136.2 | 97.7 | 6.3 | 20.5 |
| **xM** | 3.6 | 3.7 | 3.5 | 3.3 | 3.3 |

Celal Bayar University Journal of Science
Volume 20, Issue 1, 2024, p 41-48
Doi: 10.18466/cbayarfbe.1416472


47% of the earthquakes in Turkey occurred at a depth of 5 to 10 km and 19% of the earthquakes occurred at a depth of 0 to 5 km as seen in Table 2. When we look at the earthquake intensities, it was observed that 93% of them occurred between 3 and 4 magnitudes. Based on these data, it seems that the most earthquakes are concentrated in the center of the 4th cluster as seen in Table 2.

**Table 3**. 5 Distances of cluster centers relative to each other

| Clusters | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| **1** | | 72.49 | 34.06 | 57.33 | 43.21 |
| **2** | 72.49 | | 38.43 | 129.82 | 115.70 |
| **3** | 34.06 | 38.43 | | 91.39 | 77.27 |
| **4** | 57.33 | 129.82 | 91,39 | | 14.13 |
| **5** | 43.21 | 115.70 | 77,27 | 14.13 | |

As can be seen from Table 3; it is seen that the earthquakes observed in the 2nd cluster differ from the earthquakes observed in the other clusters at the distance of the concluded earthquake centers from each other. The 1st, 2nd, 3rd, 6th, 24th and 25th regions included in the 2nd cluster are the regions where earthquake faults and earthquake intensities are low and occur deeply. Therefore, there are 302 earthquakes that occur at a depth of 100 km or more. 285 of them occur in these regions. Clusters 4 and 5 have close average magnitudes of 3.3 at their centers, while depths are 6.3 km in cluster 4 and 20.5 km in cluster 5. In this respect, it has been observed that they differ from other clusters.


**Figure 3.** The 1st distribution of cluster between northern latitudes (36°-42°) and eastern longitudes (26°-45°)

The center of the 1st cluster was 3.6 magnitude and 63.7 km deep. As can be seen from Figure 3; the 2nd, 3rd, 4th, 5th regions of the 1st cluster are observed intensively and some cluster data are observed in the 39th region.

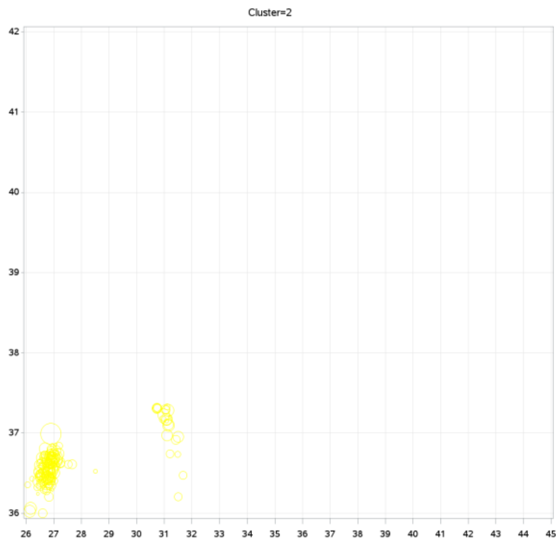


**Figure 4.** The 2nd distribution of the cluster between northern latitudes (36°-42°) and eastern longitudes (26°-45°)

The 2nd cluster formed was centered at a magnitude of 3.7 and a depth 136.2 km. As can be seen from Figure 4; the 2nd cluster, which is clustered around this center, is densely populated by 1st, 2nd, 3rd, 6th, 24th and 25th regions. It has been observed that it consists of data in regions. 2.3% of the earthquakes that occurred in Turkey occurred at depths of more than 35 km.

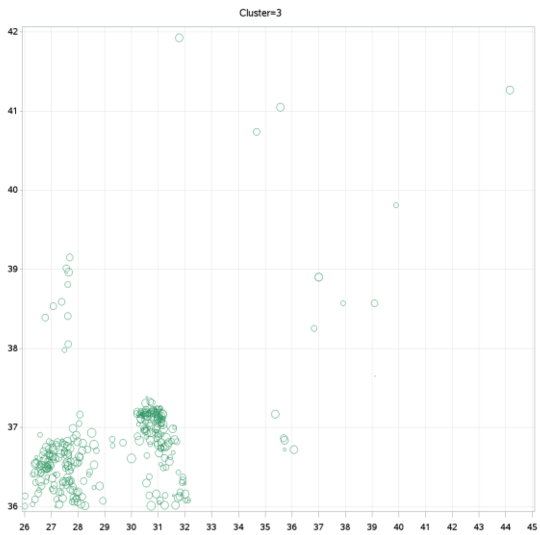


**Figure 5.** The 3rd distribution of the cluster between northern latitudes (36°-42°) and eastern longitudes (26°-45°)

The 3rd cluster formed was centered at 3.5 magnitude and 97.7 km deep. The 3rd cluster, clustered around this center, is densely populated by 1st, 2nd, 3rd, 5th, 6th and 24th regions. It has been observed that it consists of data in regions. As can be seen from Figure 5 and as can be understood from these 3 clusters we have observed, the potential for earthquakes at every depth and magnitude is seen in the 1st, 2nd, 3rd, 6th and 24th regions. It is seen that these regions we examined on the map are in the Mediterranean, off the coast of Antalya and Datça, as well as in Datça, Marmaris and Burdur. We can say that the Burdur Grabeni in these regions and the Selimiye and Taşlıca Faults in Marmaris produce earthquakes at all depths [27].

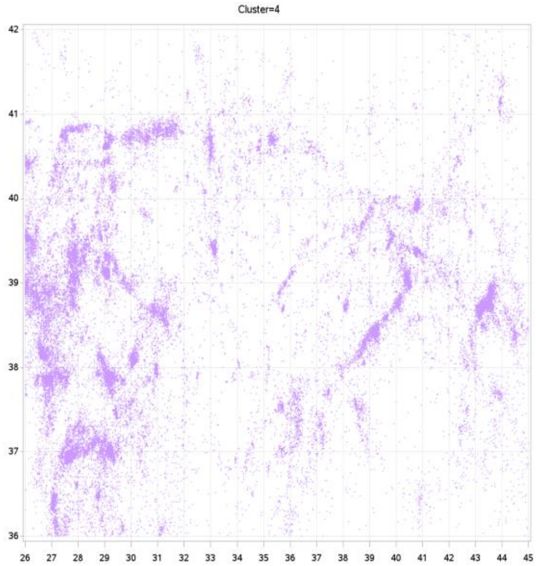


**Figure 6.** The 4th distribution of cluster (36º-42º) between northern latitudes and (26º-45º) eastern longitudes

The 4th cluster formed was centered at 3.3 magnitude and 6.3 km deep. As can be seen from Figure 6; the 4th cluster, which is clustered around this center, generally contains observations from every region. However, some regions where the density is evident stand out. It can be said that the Aegean region, Southern Marmara, Malatya and Lake Van are observed more intensely. The geological structure of the fault lines of the regions in these clusters is also a separate research topic.

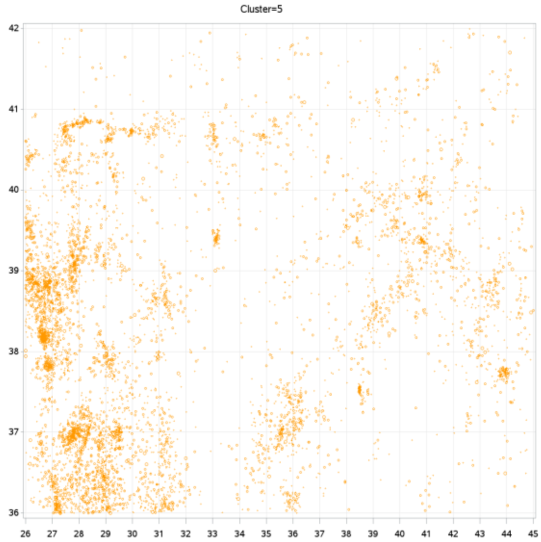


**Figure 7.** The 5th distribution of cluster (36º-42º) between northern latitudes and (26º-45º) eastern longitudes

The 5th cluster formed was centered at 3.3 magnitude and 20.5 km deep. As can be seen from Figure 7; the 5th cluster, which is clustered around this center, is similar to the 4th cluster, and the depth centers are different. In this study, it was observed that certain fault lines became evident in the clustering of earthquakes with magnitude 3 and above that occurred between 1980 and 2022 depending on their depth and magnitude.

## 6. Conclusions

Although this study does not have any limitations, unlike previous studies, cluster analysis was performed on depth and magnitude values and consistent information was obtained. On similar subjects, it has been contributed that clustering analysis can obtain consistent and strong results for numerical variables such as depth, magnitude, etc.

This study was conducted on earthquakes of 3 and above that occurred in Turkey between 01.01.1980 and 31.12.2022. The results of the cluster analysis performed by grouping 5 clusters with non-hierarchical cluster analysis of earthquake data show that intra-group and inter-group relationships were detected. In addition, the density varies between regions, and at this point, examining the geological structures of the fault lines of the regions considered in the clusters constitutes a separate research topic.

Clustering analysis studies in earthquakes can be detailed by diversifying the techniques and methods or certain region used in following studies.

## Author's Contributions

**Ahmet Murat ŞEN:** Conducted the literature review, performed analysis, evaluated the results, and wrote the manuscript.

**Zeki YILDIZ:** Contributed to the analysis advancement, manuscript preparation, interpreted findings.

## Ethics

There are no ethical issues after the publication of this manuscript.

## References

**[1].** Özel, G., Solmaz, A. (2012). Estimation of earthquake recurrence time in Turkey and investigation of seismicity according to neotectonic regions with Markov chain. *Çankaya University Journal of Science and Engineering*, 9(2), 125-138.

**[2].** Doğaner, A., Çalık S. (2012). Investigation of the relationship between seismic activities in and around Elazig and earthquake parameters. *Fırat University Journal of Science,* 26(1), 73-77.

**[3].** Kasap, R., Gürlen, Ü. (2003). Obtaining recurrence years for earthquake magnitudes: The case of Marmara Region. *Journal of Doğuş University*, 4(2), 157-166.

**[4].** Orman, F. (2021), Investigation of Eurasian Countries in terms of Economic Freedom with Cluster Analysis. *International Conference on Eurasian Economies:* 24-25 August 2021, Istanbul: p.79-86.

**[5].** Boehmke, B., & Greenwell, B. (2020). Hands-on machine learning with R. Boca Raton: *CRC Press, Taylor and Francis Group.*

**[6].** Alpar, R. (2022). Applied Statistics and Validity-Reliability with Examples from Sports. *Health and Educational Sciences 7th Edition. Detay Publishing*, 2022.

**[7].** Demircioğlu M., Eşiyok, S. (2019). *Classification of countries* by cluster analysis in the fight against the Covid-19 pandemic. *Istanbul Commerce University Journal of Social Sciences*, 37, 369-389.

**[8].** Koldere Akın, Y. (2008). Clustering Algorithms and Clustering Analysis in Data Mining. *Marmara University Doctoral Thesis*, 8, 10, 16, 26-29.

**[9].** Lean, L., Ricardo ve Powers L, Thomas. (1997). A Taxonomy of Countries Based on Inventive Activity. *International Marketing Rewiev,* Volume 14, No.6, 1997, ss.445-460.

**[10].** Tatlıdil, Hüseyin. Applied Multivariate Statistical Analysis. *Ankara: Ziraat Printing House*, 2002, ss 330-341.

**[11]**. Özdamar, Kazım. Statistical Data Analysis with Package Programs. *Eskisehir: Kaan Bookstore*, 1999.

**[12].** Ravinda, Khattree ve Naik N. Dayanand. Multivariate Data Reduction and Discrimination. *First Edition. North Caroline: Cary*, 2002.

**[13].** Latin, M. James, Carroll J. Douglas ve Paul E. Green. Analyzing Multivariate Data. *United States: Thomson Brooks-Cole,* 2003.

**[14].** NCSS. User Guide, Kaysville, *National Council for the Social Studies Incorporated*, 2001

**[15**]. Novanti P., Setyorini D., Rafflesia U., 2017. K-Means Cluster Analysis in Eartquake Epicenter Clustering. *International Journal of Advances in Intelligent Informatics*, 3(2). 2017.

**[16].** Kamat R.K. ve Kamat R.S., 2017. Earthquake Cluster Analysis: K-Means Approach, 2017. *Journal of Chemical and Pharmaceutical Sciences,* 10(1). 2017.

**[17].** Samadi et al. (2020), Robust Earthquake Cluster Analysis Based On K-Nearest Neighbor Search, 2020. *Pure and Applied Geophysics, ISSN* 0033-4553. 2020.

**[18].** Scitovski, R., Scitovski, S. (2013). A fast partitioning algorithm and its application to earthquake investigation. *Computers & Geosciences Volume* 59, 124-131.

**[19].** Morales-Esteban, A., Martínez-Alvarez, F., Scitovski, S., Scitovski, R., (2014). A fast partitioning algorithm using adaptive Mahalanobis clustering with application to seismic zoning. *Computers & Geosciences* Volume *73,* 132-141.

**[20].** Prozorov, A. G., Dziewonski, A. M. (1982). A Method of Studying Variations in the Clustering Property of Earthquakes: Application to the Analysis of Global Seismicity. *Journal of Geophysical Research,* Volume 87, No. B4, pp. 2829-2839.

**[21].** Yan Y. Kagan, David D. Jackson (1991). Long-term earthquake clustering. *Geophysical Journal International*, Volume 104, 117-133.

**[22].** Scitovski, S. (2018). A density-based clustering algorithm for earthquake zoning. *Computers & Geosciences* Vol.110, 90-95.

**[23].** Kandilli observatory and earthquake research institute website active fault maps accessed on 17.11.2023. http://www.koeri.boun.edu.tr/sismo/zeqdb/

**[24].** Hair, J.F., Anderson, R. E., Tatham, R. L., & Black, W. C., (1995). Multivariate Data Analysis with Readings. *New Jersy: Prentice Hall (1995)*.

**[25].** Yalçın, N., (2013). Clustering Analysis and Application. *Fırat University Institute of Natural and Applied Sciences Master Thesis*, 21,

**[26].** Jafarzadegan, M., Safi-Esfahani, F.ve Beheshti, Z. (2019). Combining hierarchical clustering approaches using the PCA method, *Expert Systems with Applications,* 137(15), 1-10. https://doi.org/10.1016/j.eswa.2019.06.064.

**[27]**. General Directorate of Mining Technical Exploration website active fault maps accessed on 04.12.2023. https://www.mta.gov.tr/v3.0/sayfalar/hizmetler/diri-fay-haritalari/