Earthquake & Popular Venues Data Analysis of Turkey

Phys. Zeki ÇIPLAK

November 7, 2020

1. Introduction

1.1. Description & Discussion of the Background

Turkey is an earthquake populous country with **82 million** inhabitants. It consists of **81 provinces**. It is surrounded by the North Anatolia, Western Anatolia and Southeastern Anatolia earthquake zones. The diversity of landforms in Turkey, is a result of earth movement that shaped the terrain in the region for thousands of years.

It has extinct volcanoes and earthquakes still occur frequently. In the north and east of the country, there are large fault lines that cause earthquakes today. The great Marmara earthquake that occurred on the **North Anatolian Fault Line** in 1999 caused the death of thousands of people. [1]

Although Turkey is a country of earthquakes, though it is also a tourist country. It is visited by millions of tourists from all over the world every year. There are even some foreign nationals to buy real estate in Turkey and make the trade. [2]

1.2. Problem

A data analysis based on earthquake statistics and popular locations of each city will be found interesting by investors and consumers. Investors can carry out projects for places with less earthquake risk and invest in jobs where the type of business in those places is less intense. For regular person living in cities, they may want to buy and live in places that are less dangerous and at the same time close to social venues.

Therefore, with a clustering study to be carried out by taking earthquake statistics based on a reasonable past (for example, 100 years) of each city, the regions with high or no earthquake risk can be determined. In addition, by comparing popular venues in these regions, the distinctiveness between regions can be revealed.

2. Data

2.1. Data Sources

The data sources used in this study to solve the problem are as follows:

- From the **AFAD Earthquake Catalog** page, the statistics of the earthquakes that occurred between 1920-2020 and whose magnitudes were between 4-10 according to the Richter scale were taken in csv format with various filtering methods. [3]
- With **OpenStreetMap**, the **GeoJSON data** needed for earthquake distribution and maps showing clustered groups was provided. [4]
- With the **Foursquare API**, the type information of the most popular venues of the cities, the latitude and longitude information of those places were obtained. [5]
- From **Wikipedia** was taken Turkey's cities list. The latitude and longitude information of the cities were obtained by querying them one by one with the help of **Python GeoPy Library**. [6]

2.2. Data Exploring and Cleaning

In the earthquake data scraped from **AFAD**; Date, earthquake latitude, earthquake longitude, earthquake magnitude and depth information was obtained. Data downloaded or scraped from multiple sources were combined into one table. A total of **5707** pieces of data measured between **01.01.1920** and **01.01.2020** and greater than **4.0** Richter were obtained. Some variables with missing information presented in the csv file were not included in the original data set because they were redundant.

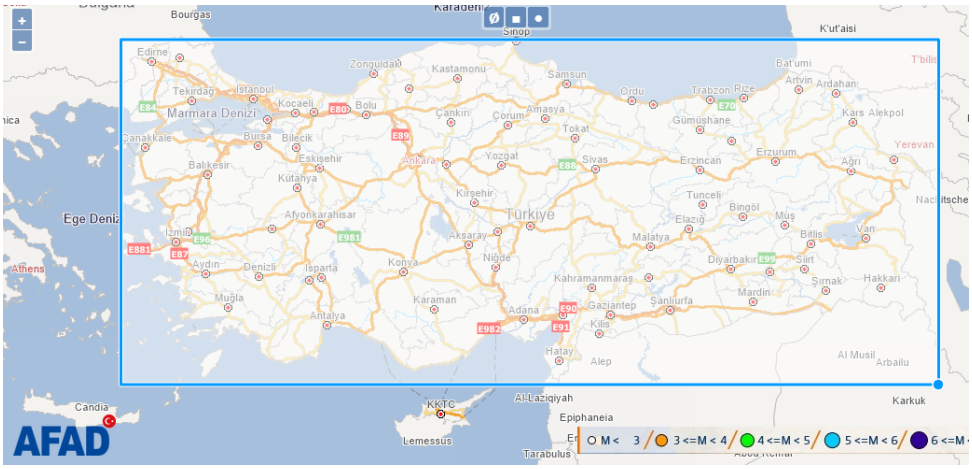

Figure 1. AFAD Earthquake Catalog Page.

While receiving **AFAD** earthquake data, "*Rectangular Search Type*" was implemented. It was realized that some of the data received came from locations outside the borders of Turkey. Since the names of the cities where the earthquakes occurred are not included in the earthquake statistics, the information of the cities where the earthquakes occurred was obtained by using the GeoPy Library and the Reverse Geocoding method for each data. The data of earthquakes have occurred outside of Turkey was removed from the master data set. Thus, a total of **4364** rows of data remained in the master data.

	City	Eq_Latitude	Eq_Longitude	Magnitude	Depth
0	Kars	40.3230	42.1726	4.2	7.00
1	Elazığ	38.3898	39.0158	4.9	11.88
2	Tunceli	39.4648	40.0195	4.2	13.21
3	Balıkesir	39.4411	27.9005	4.0	14.36
4	Balıkesir	39.4435	27.9103	4.3	11.28
5	Balıkesir	39.4430	27.9336	4.6	7.00
6	Kars	40.3255	42.2033	4.0	7.04
7	Kars	40.3106	42.1895	4.0	6.97

Figure 2. The first edited version of earthquake statistics taken from AFAD website.

According to the data set obtained, it can be seen by looking at the **Magnitude Histogram Chart** that the majority of the earthquakes that occurred were between 4 and 4.78 in magnitude. The histogram graph is informative about the distribution of earthquakes and is useful for getting to know the structure of the data set.

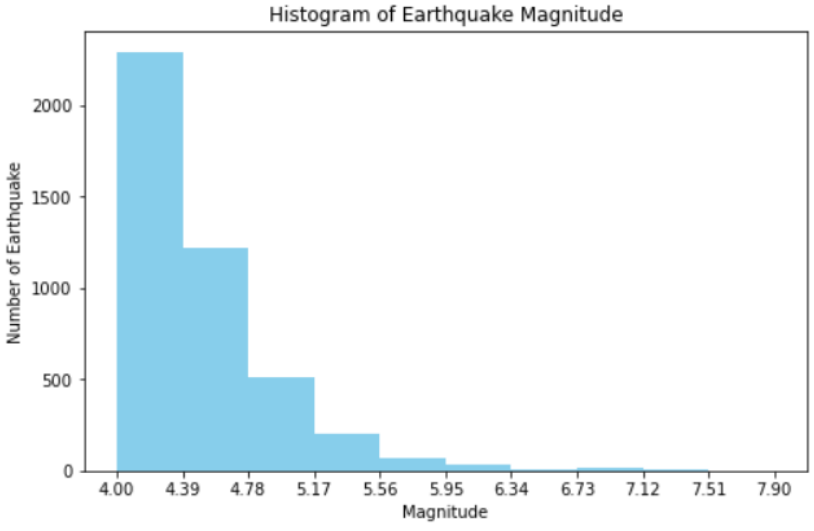

Figure 3. Histogram of Earthquake Magnitude.

In addition, the standard deviation, minimum and maximum values of the data set in the study were examined and an opinion was obtained. The table for this review is shown below.

	Eq_Latitude	Eq_Longitude	Magnitude	Depth
count	4364.000000	4364.000000	4364.000000	4364.000000
mean	38.835247	33.601072	4.446677	20.293838
std	1.277391	5.900461	0.470403	19.371195
min	36.087500	25.980000	4.000000	0.000000
25%	37.841325	28.742625	4.100000	10.000000
50%	38.880250	30.800000	4.300000	12.000000
75%	39.750000	39.610500	4.600000	27.000000
max	41.953100	44.800000	7.900000	212.000000

Figure 4. Statistical information of the first data set created.

Looking at **Figure 4**, the Magnitude standard deviation value of **~ 0.47** shows that Magnitude values are in a more uniform distribution and close to each other than other variables. Magnitude Histogram Chart in **Figure 5** also supports this idea. The **Depth Histogram Chart** below, which shows the distribution of Depth values, can also be examined.

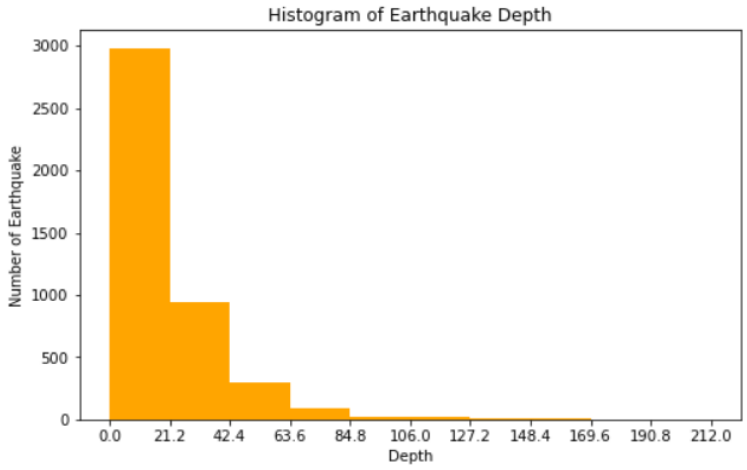

Figure 5. Histogram of Earthquake Depth.

When the maximum and minimum value ranges of the properties are compared, it will be seen that they are values that are generally not too far from each other. In addition, when the maximum values of all properties and minimum values of all properties were compared, it was understood that there was no need for any standardization or normalization.

2.3. Dealing with Missing Data

It was found that there were no statistics for some cities in the first data set. This means that earthquakes with a value greater than 4.0 have not occurred in those cities in 100 years. These cities are ***Kırklareli***, ***Rize***, ***Bartın*** *and* ***Kilis***. Magnitude and Depth variables were filled with reasonable values only for these four cities so that the rows of these provinces did not appear as NaN in the data set.

The fact that there is no earthquake record greater than 4.0 for these cities in the statistics for 100 years does not mean that there is no earthquake. Some earthquakes that are less than 4.0 but we do not have a record may have occurred. Even if there is no earthquake within the boundaries of these cities, earthquakes in neighboring cities can affect these cities. So it makes no sense to fill the Magnitude sections that are NaN with zero.

However, some Depth values were found to be **0.0** in the statistics obtained from AFAD. Considering that this situation may be natural, the values of **Max Depth** and **Min Depth** to be derived below are accepted as **0.0** for these provinces. **Max Magnitude** values were accepted as **3.0** and **Min Magnitude** values as 1.0. Based on this, the **Avg Magnitude** value was accepted as **2.5**.

2.4. Feature Extraction

The average, maximum and minimum features for each city were generated by feature extraction. These attributes were added to the first data set edited. The new attributes are derived based on the Magnitude and Depth variables. The *Eq_Latitude* and *Eq_Longitude* attributes were then removed from the data set. Finally, the **Master data** set was created with the addition of the data table with the latitude and longitude information of **81 cities** to the data set.

	City	Latitude	Longitude	Avg Magnitude	Avg Depth	Max Magnitude	Max Depth	Min Magnitude	Min Depth
0	Adana	36.993617	35.325835	4.389189	23.480270	6.2	86.0	4.0	2.0
1	Adıyaman	37.789360	38.314110	4.466667	16.218462	6.0	64.3	4.0	1.7
2	Afyonkarahisar	38.685273	30.642741	4.490909	17.259091	6.0	44.9	4.0	2.0
3	Ağrı	39.529160	43.383564	4.584783	21.560435	6.0	60.0	4.0	3.2
4	Amasya	40.656945	35.772717	4.269444	10.853056	5.6	33.0	4.0	3.1
5	Ankara	39.716044	32.705995	4.353704	10.589815	5.7	38.0	4.0	1.0
6	Antalya	36.900964	30.695485	4.392553	52.161809	5.4	159.2	4.0	1.6
7	Artvin	41.160506	41.839863	4.300000	15.910000	4.7	61.0	4.0	5.0

Figure 6. Master Data.

3. Methodology

Using the **Python Folium Library**, the geographical details of **4364 earthquake points** were visualized for initial insights. It was a useful visualization, especially in terms of seeing in which regions earthquakes occurred and gaining an idea about the earthquake distribution.

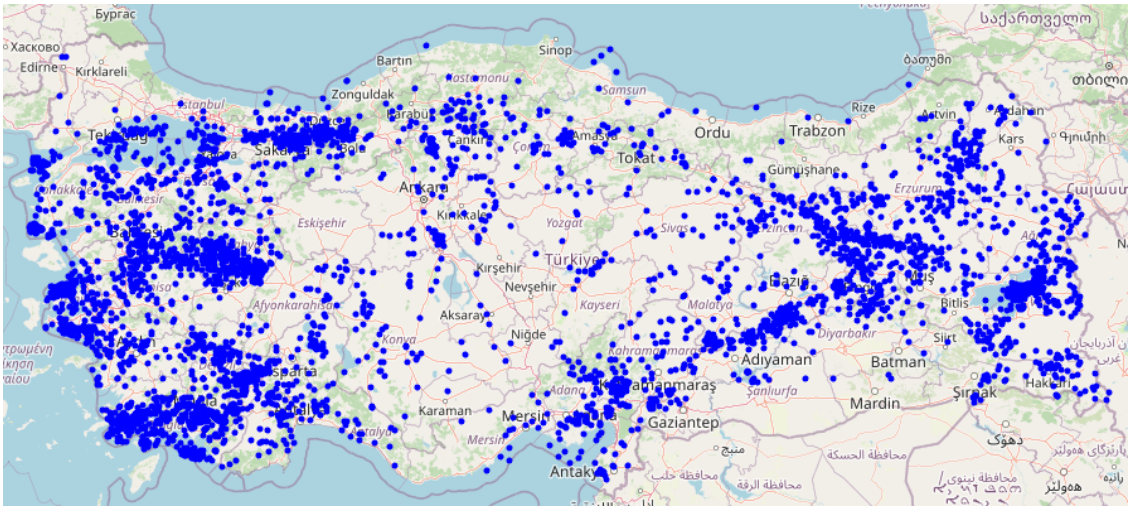

Figure 7. Earthquakes that occurred between the years of 1920-2020 in Turkey.

Foursquare API was used to discover popular venues in cities. As a limit for each city, **100 popular venues** and **20 kilometers in diameter** were measured. In some cities, it has been observed that when the diameter measure is shorter, no popular space data is returned. For this reason, a 20-kilometer-wide search was conducted in order to obtain **at least 10 rows** from each city. A total of **6581** rows of data were obtained.

	City	Venue	Venue Category
0	Adana	Adana Divan	Hotel
1	Adana	Fitness Akademi Reşatbey	Gym
2	Adana	İzmir Hair Design	Salon / Barbershop
3	Adana	Beta Tea House	Café
4	Adana	Frame Adana	Restaurant
5	Adana	Kazım Büfe	Snack Place
6	Adana	Burn Latin Dans Okulu	Dance Studio
7	Adana	Geko	Café
8	Adana	Golden Deluxe Hotel	Hotel

Figure 8. Location data obtained by Foursquare API query.

A summary table was created for the popular places identified in the cities. The pivot table shows the total number of venues returned by the Foursquare API for each city. The chart of this table is as follows.

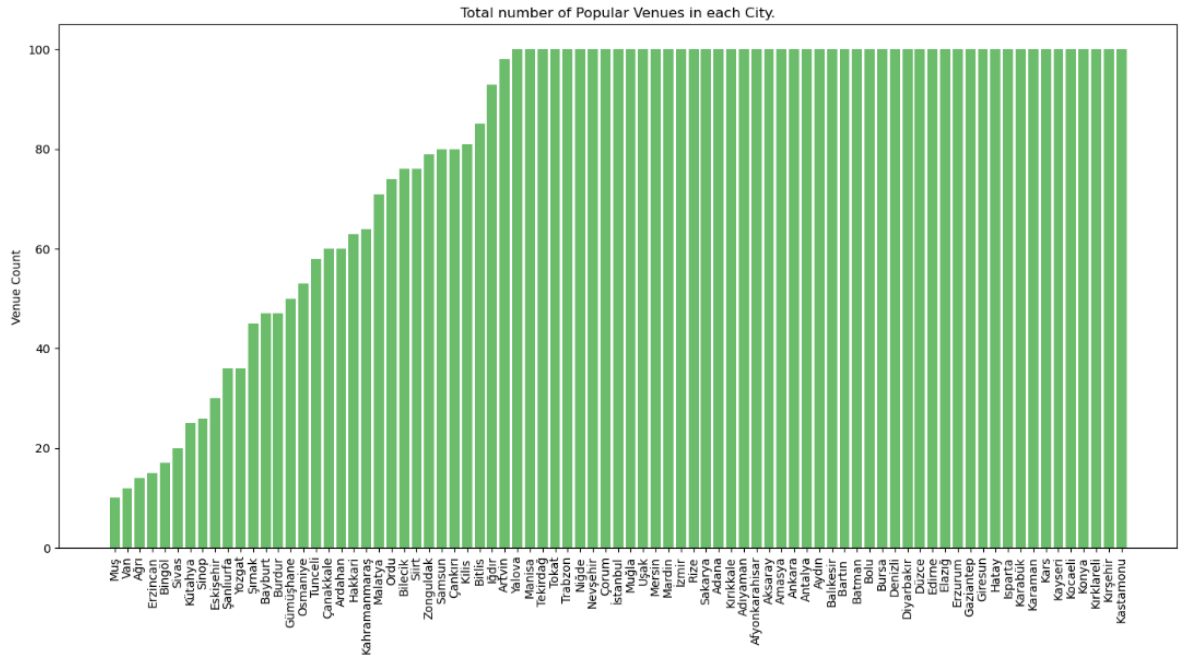

Figure 9. Total number of Popular Venues in each City.

When **Figure 9** was examined, it was seen that **100 results** were returned for many large cities. While provinces such as *Istanbul*, *Yalova*, *Izmir*, *Bursa*, *Trabzon*, *Adana*, *Mersin*, *Diyarbakır* and *Antalya* are seen as rich in popular places; Provinces such as *Ağrı*, *Tunceli*, *Hakkâri*, *Siirt*, and *Erzurum* remained below **20 results**.

Of course, this graphic does not include all popular places in the provinces. Because a search was made at a distance of 20 kilometers for each city and only one latitude-longitude pair was used to represent each city. Considering its area, this search may be considered a narrow search for some cities.

There may also be many locations that were not detected or considered popular by Foursquare. More popular location information can be obtained by performing more detailed searches with more latitude and longitude information about a city.

When the data obtained with the Foursquare API was summarized, it was seen that a total of **347** types of popular venues belonging to different categories were identified. A new data table has been created showing the **10 most common venues** in each province.

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adana	Hotel	Coffee Shop	Steakhouse	Turkish Restaurant	Candy Store	Kebab Restaurant	Café	Clothing Store	Dessert Shop	Gym
1	Adıyaman	Café	Hotel	Dessert Shop	Coffee Shop	Turkish Restaurant	Restaurant	Gym / Fitness Center	Park	Campground	Tea Room
2	Afyonkarahisar	Café	Dessert Shop	Gym / Fitness Center	Bakery	Coffee Shop	Candy Store	Garden	Park	Shopping Mall	Kebab Restaurant
3	Aksaray	Café	Hotel	Gym / Fitness Center	Park	Restaurant	Dessert Shop	Turkish Restaurant	Bakery	Doner Restaurant	Clothing Store
4	Amasya	Café	Hotel	Turkish Restaurant	Coffee Shop	Restaurant	History Museum	Plaza	Comfort Food Restaurant	Pide Place	Gym
5	Ankara	Farm	Restaurant	Café	Seafood Restaurant	Turkish Restaurant	Bakery	Scenic Lookout	Gym / Fitness Center	Coffee Shop	Other Great Outdoors
6	Antalya	Coffee Shop	Seafood Restaurant	Bar	Park	Hotel	Gym / Fitness Center	Restaurant	Pizza Place	Café	Scenic Lookout
7	Ardahan	Café	Farm	Electronics Store	Turkish Restaurant	Convenience Store	Park	Diner	Coffee Shop	Mobile Phone Shop	Athletics & Sports

Figure 10. Top Ten Most Common Venues.

After the determination of popular venues, **Onehot encoding** was done for 347 categorical variables and *City* attribute was extracted from Master Data and clustering study was performed with **K-Means** algorithm. The K-Means algorithm clusters data by trying to separate samples in n groups of equal variance, minimizing a criterion known as the inertia or within-cluster sum-of-squares. K-Means algorithm is one of the most common cluster method of unsupervised learning.

	Latitude	Longitude	Avg Magnitude	Avg Depth	Max Magnitude	Max Depth	Min Magnitude	Min Depth	Afghan Restaurant	African Restaurant	Airport	Airport Food Court	Airport Lounge	Airport Service	A Re
0	36.993617	35.325835	4.389189	23.480270	6.2	86.0	4.0	2.00	0.000000	0.0	0.00	0.0	0.00	0.0	
1	37.789360	38.314110	4.466667	16.218462	6.0	64.3	4.0	1.70	0.000000	0.0	0.01	0.0	0.01	0.0	
2	38.685273	30.642741	4.490909	17.259091	6.0	44.9	4.0	2.00	0.000000	0.0	0.00	0.0	0.00	0.0	
3	39.529160	43.383564	4.584783	21.560435	6.0	60.0	4.0	3.20	0.000000	0.0	0.00	0.0	0.00	0.0	
4	40.656945	35.772717	4.269444	10.853056	5.6	33.0	4.0	3.10	0.000000	0.0	0.00	0.0	0.00	0.0	
5	39.716044	32.705995	4.353704	10.589815	5.7	38.0	4.0	1.00	0.000000	0.0	0.00	0.0	0.00	0.0	
6	36.900964	30.695485	4.392553	52.161809	5.4	159.2	4.0	1.60	0.000000	0.0	0.00	0.0	0.00	0.0	
7	41.160506	41.839863	4.300000	15.910000	4.7	61.0	4.0	5.00	0.000000	0.0	0.00	0.0	0.00	0.0	
8	37.841301	27.832837	4.478431	18.662353	6.8	101.0	4.0	2.80	0.010204	0.0	0.00	0.0	0.00	0.0	
9	39.540080	28.022879	4.461333	15.355267	6.2	80.0	4.0	1.10	0.000000	0.0	0.00	0.0	0.00	0.0	
10	40.143510	29.975291	4.350000	22.060000	5.1	60.0	4.0	7.50	0.000000	0.0	0.00	0.0	0.00	0.0	

Figure 11. A summary of the dataset used for the K-Means algorithm.

Figure 11 contains a summary of the data set (81 rows x 355 columns) used in the K-Means clustering algorithm. There are different metric distance function for spatial distance. **Sqeuclidean** metric was chosen in the study. Because it made the elbow break point to be seen more clearly. In order to determine the optimum number of clusters, clustering study was conducted with different trials. The results were analyzed by increasing the cluster constant (k

value) of the K-Means algorithm. When analyzed by **The Elbow Method**, it was decided that the optimum value was **K = 3**. This situation can also be seen in the graphic below.

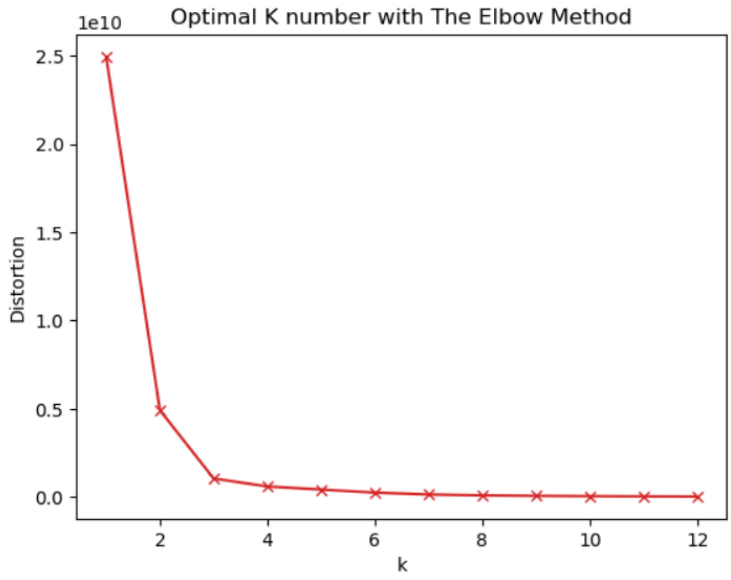

Figure 12. Optimal K number with The Elbow Method.

After the clustering study, a new data table was created giving the clusters found, city names, avg. magnitude of earthquakes occurring in cities, and the 10 most common venues in each city.

	Cluster Numbers	City	Avg Magnitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	0	Adana	4.389189	Hotel	Coffee Shop	Steakhouse	Turkish Restaurant	Candy Store	Kebab Restaurant	Café	Clothing Store	Dessert Shop	Gym
1	0	Adıyaman	4.466667	Café	Hotel	Dessert Shop	Coffee Shop	Turkish Restaurant	Restaurant	Gym / Fitness Center	Park	Campground	Tea Room
2	1	Afyonkarahisar	4.490909	Café	Dessert Shop	Gym / Fitness Center	Bakery	Coffee Shop	Candy Store	Garden	Park	Shopping Mall	Kebab Restaurant
3	0	Ağrı	4.584783	Café	Hotel	Gym / Fitness Center	Park	Restaurant	Dessert Shop	Turkish Restaurant	Bakery	Doner Restaurant	Clothing Store
4	1	Amasya	4.269444	Café	Hotel	Turkish Restaurant	Coffee Shop	Restaurant	History Museum	Plaza	Comfort Food Restaurant	Pide Place	Gym
5	1	Ankara	4.353704	Farm	Restaurant	Café	Seafood Restaurant	Turkish Restaurant	Bakery	Scenic Lookout	Gym / Fitness Center	Coffee Shop	Other Great Outdoors
6	2	Antalya	4.392553	Coffee Shop	Seafood Restaurant	Bar	Park	Hotel	Gym / Fitness Center	Restaurant	Pizza Place	Café	Scenic Lookout

Figure 13. Most Common Venues and Cluster Numbers.

When the **3 clusters detected** were examined in terms of earthquake risk, they were labeled as **LOW** – **MEDIUM** – **HIGH**. When the average earthquake magnitude of all cities in each cluster was calculated, the risk ratio by clusters was also determined.

The cluster of cities with the highest earthquake risk was determined as HIGH, and the cluster of the lowest cities was determined as LOW. A new data table was created consisting of the total number of cities owned by the clusters, the representation colors of the clusters on shapes and maps, and cluster labels.

Cluster Numbers	Avg Magnitude	Color	Labels	City Count
1	4.236	green	LOW	33.0
2	4.484	orange	MEDIUM	10.0
0	4.493	red	HIGH	38.0

Figure 14. Label, color and count information of the detected clusters.

When the count of cities between clusters is compared, it is seen that **Cluster HIGH** is in the first place with **46.9%**. This result shows that most of the cities in the country are at high risk of earthquakes.

On the other hand, the total number of cities in **Cluster LOW**, where the earthquake risk is low, is too high to underestimate. For all kinds of investments to be made to protect against earthquake risk, attention can be drawn to the provinces in this cluster.

Cluster MEDIUM, where the earthquake risk is at a medium level, appears to be the smallest cluster in terms of the total number of cities. Below you can examine the pie chart containing the city numbers of the clusters.

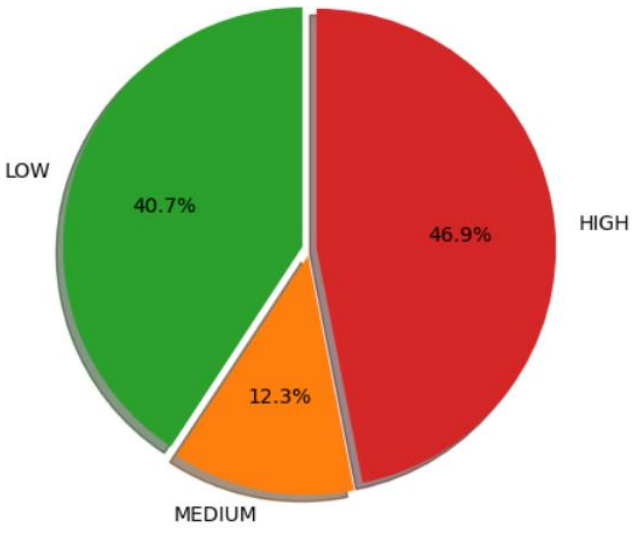

Figure 15. Count of the cities in clusters.

4. Results Analysis

The results we have achieved so far have been a guide for regular person who want to live in a safe city or make their real estate buy decisions in this direction. It is necessary to examine clusters in more detail and separately in order to make comments to guide investors.

4.1. Cluster: "HIGH"

Cluster Numbers	City	Avg Magnitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adana	4.389189	Hotel	Coffee Shop	Steakhouse	Turkish Restaurant	Candy Store	Kebab Restaurant	Café	Clothing Store	Dessert Shop	Gym
0	Adıyaman	4.466667	Café	Hotel	Dessert Shop	Coffee Shop	Turkish Restaurant	Restaurant	Gym / Fitness Center	Park	Campground	Tea Room
0	Ağrı	4.584783	Café	Hotel	Gym / Fitness Center	Park	Restaurant	Dessert Shop	Turkish Restaurant	Bakery	Doner Restaurant	Clothing Store
0	Artvin	4.300000	Café	Farm	Electronics Store	Turkish Restaurant	Convenience Store	Park	Diner	Coffee Shop	Mobile Phone Shop	Athletics & Sports
0	Aydın	4.478431	Café	Restaurant	Mountain	Forest	Bar	Park	Turkish Restaurant	Hotel	Campground	Steakhouse
0	Balıkesir	4.461333	Café	Bar	Coffee Shop	Gym / Fitness Center	Bakery	Dessert Shop	Gym	Ice Cream Shop	Turkish Restaurant	Kuruyemişçi

Figure 16. Pivot Table of Cluster HIGH.

Cluster HIGH is the cluster of cities with the greatest earthquake risk. **38** of the all cities in Turkey are in this cluster. Considering all clusters, it is the first cluster in terms of the count of cities. When the characteristic structure of the cities is examined, it is seen that they are generally located far from the sea and in the inner parts, in places mountainous and some of them have a coast to the sea.

Most of the cities that have a coast in the cluster are neighbors of the Aegean Sea. There are also two cities in the cluster, neighboring Marmara, Black Sea and Mediterranean.

		1st Most Common Venue	Total
HIGH	**0**	Café	24
	1	Hotel	4
	2	Turkish Restaurant	2
	3	Farm	2
	4	Steakhouse	1

Figure 17. Pivot Table of 1st Common Venues in Cluster HIGH

The table given in **Figure 17** was obtained when a comparison was made in terms of "**the first most common places**" among the cities in the cluster.

According to this table, Café appears to be the first most common location in 24 of 38 cities in this cluster. Then, "**Hotel**" and "**Turkish Restaurant**" come respectively. It is anticipated that there will be a great competition in these sectors for the investors. It would be wiser to turn attention to less competitive sectors.

Almost all of the cities in this cluster are located on a fault line. In this respect, especially those who want to buy real estate, can choose a city from Cluster MEDIUM or Cluster LOW that are farther from the fault lines.

In addition, since the cluster with the highest earthquake risk, it may be preferable that the investment to be made is a less costly business type that is not affected by the earthquake. For high-cost investments, cities in clusters with less earthquake risk may be preferred.

When a summation is made among all common places from 1 to 10 of all cities in this cluster, the top 20 venues at the top of the list are shown in the graph below.

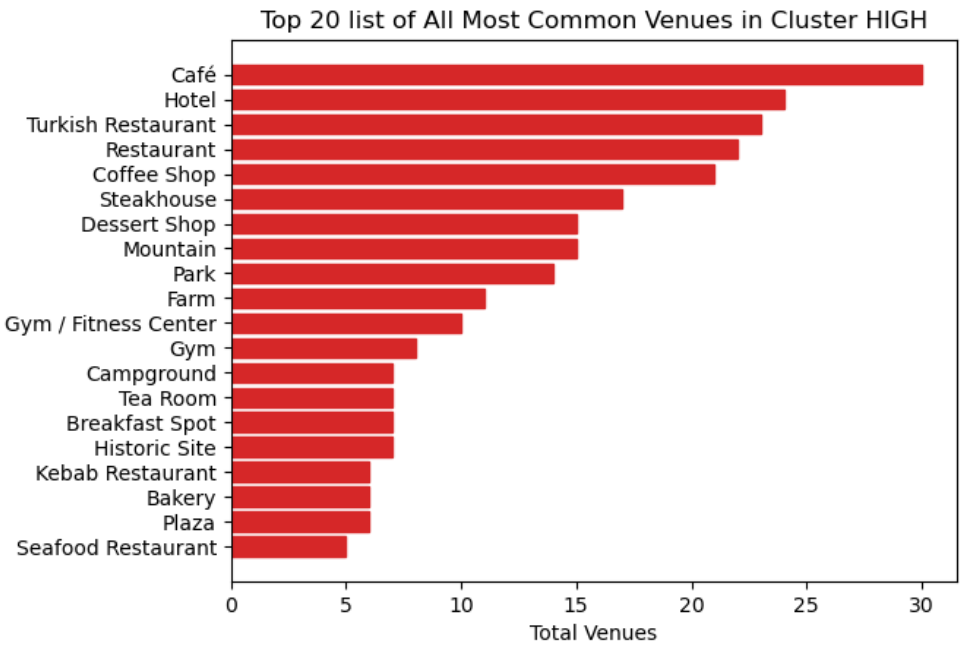

Figure 18. Top 20 Chart of All Most Common Venues in Cluster HIGH

According to the graphic in **Figure 18**, "**Seafood Restaurant**" appears as a less competitive place than other sectors. An investment can be made in this sector, which can be utilized especially in the seaside cities of this cluster.

In addition, as an alternative to the seventh baked food business "**Dessert Shop**", investing in the "**Bakery**" sector, which serves a similar area and has less competition than the graph, can be considered as a good opportunity.

Some of the cities in this cluster are located in the inner parts of the country where the continental climate is experienced and in the mountainous regions with very cold winters. On the other hand, looking at the chart, it is seen that there are many businesses related to beverages such as tea and coffee.

Rather than heading into this highly competitive area, "**Breakfast Spot**", where you can have both tea and coffee, have breakfast on the one hand, and has less competition, can be a good investment option.

4.2. Cluster: "MEDIUM"

Cluster Numbers	City	Avg Magnitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	Antalya	4.392553	Coffee Shop	Seafood Restaurant	Bar	Park	Hotel	Gym / Fitness Center	Restaurant	Pizza Place	Café	Scenic Lookout
2	Bingöl	4.440000	Café	Coffee Shop	Kebab Restaurant	Convenience Store	Gym / Fitness Center	Dessert Shop	Kokoreç Restaurant	Hotel	Kofte Place	Restaurant
2	Denizli	4.483226	Mountain	Steakhouse	Convenience Store	Restaurant	Bay	Beach	Farm	Campground	Lake	Café
2	Diyarbakır	4.525490	Café	Historic Site	Hotel	Mosque	Turkish Restaurant	Restaurant	History Museum	Kebab Restaurant	Dance Studio	Seafood Restaurant
2	Erzurum	4.531481	Café	Restaurant	Turkish Restaurant	Coffee Shop	Gym / Fitness Center	History Museum	Mosque	Hotel	Gym	Historic Site
2	Mersin	4.582353	Café	Kebab Restaurant	Hotel	Steakhouse	Mobile Phone Shop	Coffee Shop	Bagel Shop	Restaurant	Mountain	Bar
2	Konya	4.485185	Café	Coffee Shop	Vineyard	Steakhouse	Burger Joint	Scenic Lookout	Dessert Shop	Bakery	Gym / Fitness Center	Other Great Outdoors

Figure 19. Pivot Table of Cluster MEDIUM.

The cluster, which is in the second degree in terms of earthquake hazard, consists of 10 cities. Neighbors of cities are usually cities from the Cluster HIGH. In this respect, landforms can be similar to Cluster HIGH in terms of fault lines passing through cities and earthquake risk. Except for Muğla, Antalya and Mersin, among the cities in the cluster, none of the cities have a seafront. Cities in Eastern Anatolia are located in mountainous and high altitude regions.

When a comparison is made between the cities in the cluster in terms of "**first most common places**", the following table is obtained.

		1st Most Common Venue	Total
MEDIUM	**0**	Café	8
	1	Mountain	1
	2	Coffee Shop	1

Figure 20. Pivot Table of 1st Common Venues in Cluster MEDIUM

As can be seen from this table, the **Café** is a sector that is widely preferred by investors among the first two clusters. Therefore, the competition is very high. Especially when the clusters in Cluster MEDIUM are examined, the "**Mountain**" sector, which is common in some provinces but has less competition, may be suitable for investment.

For further comment on this issue, it would be useful to look at the top 20 venues list, which includes all common venues in cities. Below you can see the graph that gives the sum of all common places from 1 to 10.

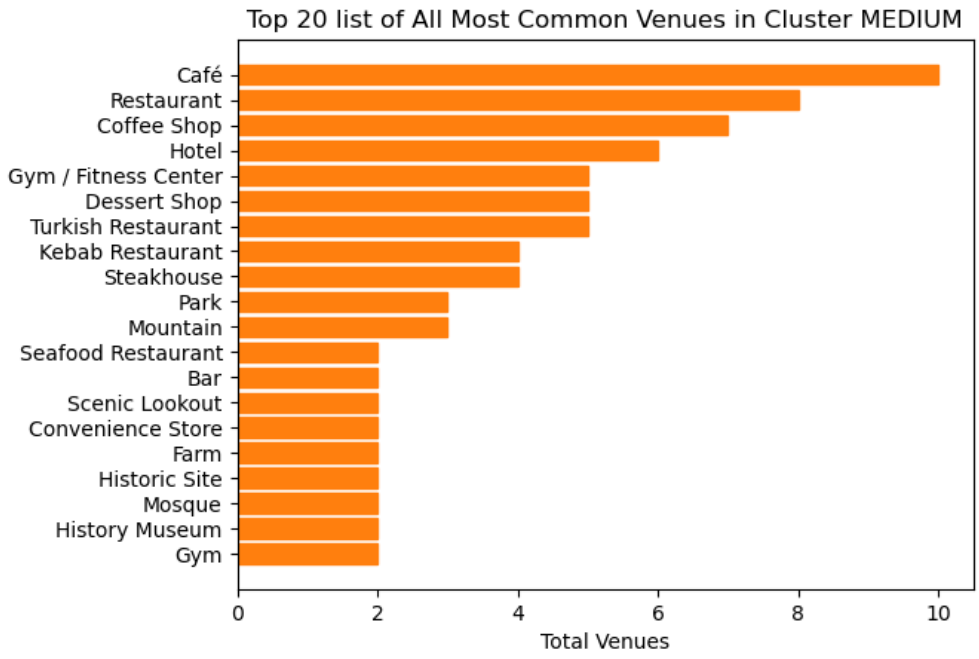

Figure 21. Top 20 Chart of All Most Common Venues in Cluster MEDIUM

Cities with a coast to the sea in the cluster can be considered as cities that can be invested especially in terms of tourism. As with Cluster HIGH, we can examine industries that are less competitive in terms of investment and are lower on the list.

Depending on the development of tourism in the country, the "**Hotel**" sector, which ranks 4th in the list of top 20 places, and the "**Scenic Lookout**", which seems less competitive in cities with high altitude, may be a good option for investment.

Cities by the sea can be attractive in terms of real estate. However, for this, it would be more appropriate to choose the Cluster LOW, which has the least earthquake risk and has a seafront.

4.3. Cluster: "LOW"

Cluster Numbers	City	Avg Magnitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Afyonkarahisar	4.490909	Café	Dessert Shop	Gym / Fitness Center	Bakery	Coffee Shop	Candy Store	Garden	Park	Shopping Mall	Kebab Restaurant
1	Amasya	4.269444	Café	Hotel	Turkish Restaurant	Coffee Shop	Restaurant	History Museum	Plaza	Comfort Food Restaurant	Pide Place	Gym
1	Ankara	4.353704	Farm	Restaurant	Café	Seafood Restaurant	Turkish Restaurant	Bakery	Scenic Lookout	Gym / Fitness Center	Coffee Shop	Other Great Outdoors
1	Bolu	4.415625	Café	Coffee Shop	Steakhouse	Turkish Restaurant	Restaurant	Hotel	Dessert Shop	Gym / Fitness Center	Clothing Store	Kebab Restaurant
1	Gaziantep	4.527778	Park	Convenience Store	Moving Target	Turkish Restaurant	Café	Historic Site	Pier	Mountain	Lake	Food Court
1	Giresun	4.683333	Café	Historic Site	Turkish Restaurant	Hotel	Gym / Fitness Center	Kebab Restaurant	Ski Area	Mountain	Breakfast Spot	Coffee Shop
1	Gümüşhane	4.309091	Farm	Recreation Center	Campground	Light Rail Station	Rest Area	Bus Stop	Lake	Steakhouse	Garden	Turkish Home Cooking Restaurant

Figure 22. Pivot Table of Cluster LOW.

It is the cluster with the least earthquake risk. It consists of **33** cities in total and ranks second among other clusters in terms of number of cities. Almost all of the cities with a coastline are adjacent to the **Black Sea**. Except Kocaeli and Sakarya, Turkey's neighbor to the Black Sea all the cities that are located in this cluster.

Cities in this cluster may be preferred especially for real estate investment. Those who want to buy a house by the sea and from a place with the least earthquake risk should evaluate the cities in this cluster and the real estate opportunities in these cities in detail.

When evaluated in terms of population density, it consists of the most populous cities in the country such as Istanbul and Ankara. The fact that the population is dense and the earthquake risk is lower than other provinces has paved the way for investment.

The "first most common places" in cities are shown in the table below.

		1st Most Common Venue	Total
LOW	0	Café	22
	1	Farm	3
	2	Park	2
	3	Restaurant	1
	4	Mosque	1

Figure 23. Pivot Table of 1st Common Venues in Cluster MEDIUM

In the cities in this cluster, investment opportunities are higher than in the provinces in other clusters. The abundance of cities on the seaside, fewer earthquakes, and a wide variety of sectors to invest in make the cities in Cluster LOW more attractive.

Therefore, it would be a better decision to choose a sector that is different and open to development, instead of a sector that is very common and competitive in every cluster like the Café sector. In the image below, you can see top 20 location list of the cities in Cluster LOW.

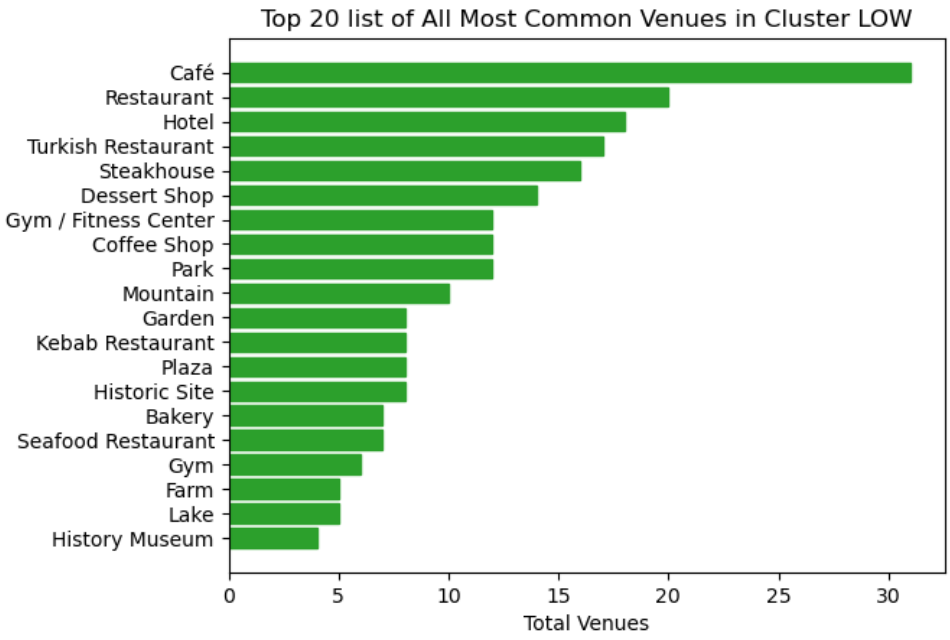

Figure 24. Top 20 Chart of All Most Common Venues in Cluster LOW

The "**Dessert Shop**" and "**Bakery**" investment idea we mentioned for Cluster HIGH before is also valid in this cluster. When **Figure 24** is examined, it will be seen that these sectors can be alternatives to each other since they are in similar areas.

Fishing has developed in cities with a coastline to the Black Sea. Therefore, “**Seafood Restaurant**”, which is less competitive, can be a good investment. It is also a good business idea to supply fish for Seafood Restaurants.

In “**Garden**” and “**Farm**”, they can be evaluated as similar and less competitive business types, as joint or separate investments.

When the tables and graphs of all clusters are examined so far, it will be seen that the “**Mountain**” sector is in a remarkable order in each cluster. For this purpose, it can be considered as a good investment opportunity in cities with high altitude in all clusters.

Again, when all of the clusters are considered together, it is possible to see that the “**Farm**”, “**Park**” and “**Plaza**” sectors, whose competition is at medium levels, have a large percentage in total.

4.4. Cluster Maps

All detected clusters were visualized with the point in the map of Turkey. **Python Folium Library** was used for this. You can see the visual version of all the explanations we made above about the clusters on the map below.

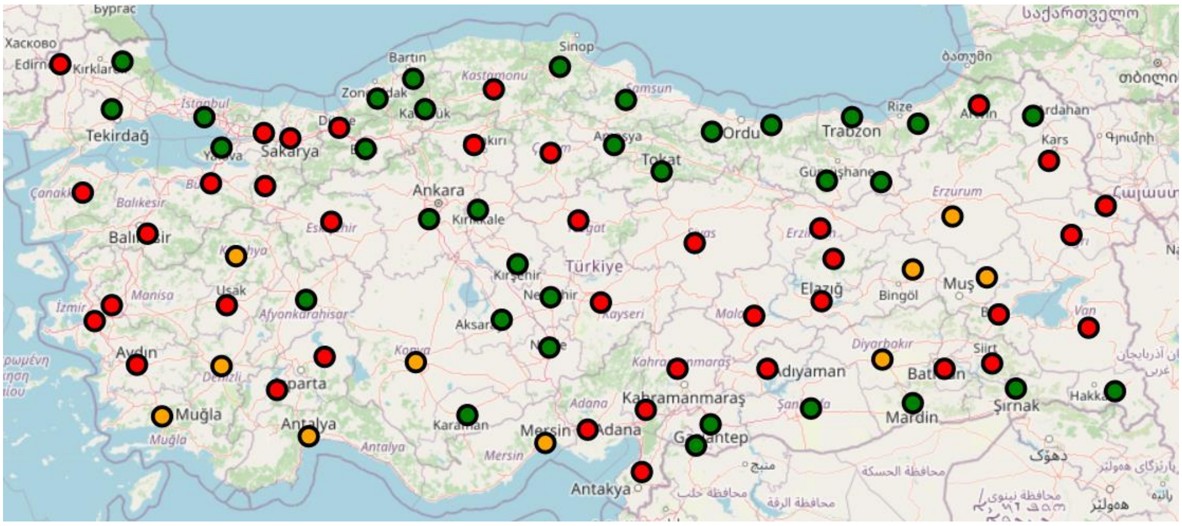

Figure 25. The visualization of cluster point in the map of Turkey.

The dots shown in Red on the map above belong to Cluster HIGH, which has the highest earthquake risk. Orange dots represent MEDIUM and Green dots represent Cluster LOW.

In addition to this map, **Choropleth Map** with color distribution according to the average earthquake magnitude was also created. Both average earthquake magnitude and clusters are shown together in the Choropleth map.

In addition, when clicking on each point, the name of the relevant city, the cluster information and the 100-year average earthquake magnitude information are given as **Popup**.

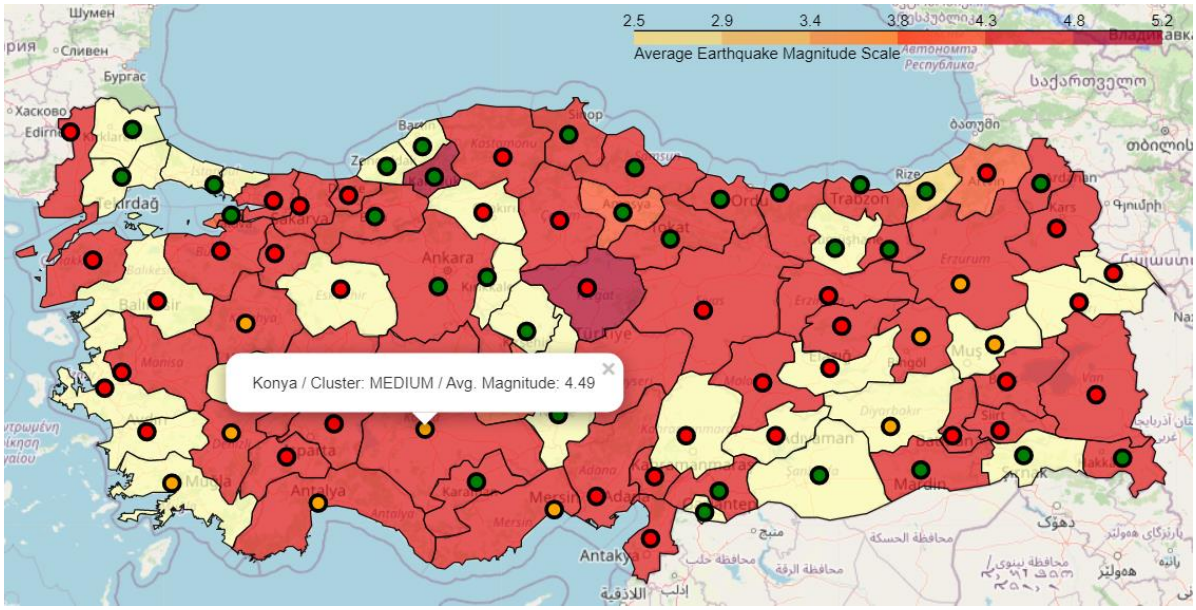

Figure 26. The visualization of Cluster Points and Average Earthquake Magnitude distribution in the map of Turkey

5. Discussion

The first aim of this study is to help business people and ordinary citizens in their investment decisions and social lives for the future by analyzing popular venues in cities with earthquake data.

Although the statistical data used in the project represented a part of the city, it was seen as representing the whole. For example, an earthquake that occurred in only one part of the city was considered to have occurred in the whole. Therefore, the city-based study can be carried out on a district or neighborhood basis by digging deeper.

In addition, more data can be obtained with different data sources other than Foursquare API data. Thus, it will be possible to obtain sharper and different results.

Another aim of the project is to make inferences by examining old earthquakes and to produce information that we can benefit from.

For example, if there have been too many earthquakes in a city and the magnitude of the earthquake is high, it is likely that an earthquake will occur in that city again due to fault activity. However, this acceptance does not lead to the conclusion that places where there have been no earthquakes for many years are safe. The approaches used in this study can be handled with different data and different methods in the future, and different results can be reached.

When the red points in Figure 25 are examined, it will be seen that they are compatible with The Seismic Hazard Map of Turkey. [7]

6. Conclusion

As a result, this project has helped to better understand the characteristics of cities in terms of earthquakes and social venues. The project has supported not only the investors but also the city managers or planners in their decisions. It also played a guiding role for all kinds of researchers using data analysis types in areas similar to the one in this study.

All data sets, Python codes and image files used in the study are stored in the relevant GitHub account. [8] Thus, it is made ready to be used for different projects in the future.

Stay with the Science.

Zeki ÇIPLAK
Physicist, Data & Computer Science Enthusiast.

7. References

- 1. Wikipedia "Turkey" topic. (Turkish)
- 2. GYODER - Why Invest in Turkey? (Turkish)
- 3. AFAD - Earthquake Catalog
- 4. OpenStreetMap ©
- 5. Foursquare API
- 6. Cities of Turkey (Turkish)
- 7. GitHub Repository of Project
- 8. The Seismic Hazard Map of Turkey. (dmc.metu.edu.tr)