Exploratory Analysis of the Last 20 Years of Earthquakes in Turkey

Nilay Çiçekli   
Computer Engineering Department  
FMV Işık Universityİstanbul, Turkey  
nilay.cicekli@isik.edu.tr Faik Boray Tek  
 Computer Engineering Department  
FMV Işık Universityİstanbul, Turkey  
boray.tek@isikun.edu.tr

*Abstract*— Earthquakes are natural disasters that are unpredictable and have sociological and economical costs. There are studies to predict earthquakes using several methods such as statistical and deep learning methods. Yet, there is no truly successful method for such a prediction. Earlier studies show that better results for prediction require better data analysis on bigger historical data and require more parameters (e.g. weather data) are involved in the research process rather than using merely earthquake data. This paper represents a comprehensive analysis of earthquakes in Turkey in the past 20 years and, also associates the earthquake occurrence with climate using data mining techniques.

Depremler ne zaman gerçekleşeceği tahmin edilemeyen doğal felaketlerdir ve toplum üzerinde sosyolojik ve ekonomik sonuçlar doğururlar. Depremleri tahmin etmek üzerine yapılan çalışmalar bulunmaktadır ve bunlar istatistik ve derin öğrenme (yapay zeka) teknikleri kullanırlar. Ancak, depremi tahmin edebilen tam olarak başarılı bir yöntem henüz yoktur. Önceki çalışmalar gösterir ki, deprem tahmininde daha iyi sonuçlar almak için öncelikle daha büyük veriyle, daha iyi veri analizi yapılmalıdır. Ayrıca, araştrma sürecinde yalnızca geçmiş deprem verisi kullanmak yerine, başka parametreler de araştırmaya dahil edilmelidir (örneğin iklim verisi). Bu makale veri madenciliği tekniklerini kullanarak, Türkiye’nin son 20 yıldaki deprem verisinin kapsamlı bir analizini çıkartır ve deprem oluşumu ile iklim arasındaki ilişkiyi inceler.

# Introduction

Earthquakes are the results of the Earth’s crust movement. At the surface, they appear by shaking. They might lead to loss of life and destruction of property. So, they have sociological and economical costs.

The World’s costliest earthquake is recorded as the 2011 Tōhoku Earthquake and Tsunami in Japan. According to the Japanese National Police Agency the earthquake caused over 15,000 deaths, and the property damage valued at about $235 billion making it not only the costliest earthquake ever but also the most costly natural disaster ever.[[1]](#_References)

In Turkey, in 1999 Izmit Earthquake over 17,000 people died, even more than the Tōhoku earthquake. And it resulted in more than 250,000 people becoming homeless[.[2]](#_References)

Veysel Yılmaz Murat Erişoğlu, H. Eray Çelik studied the prediction of the next earthquake in the North Anatolian Fault Zone in Turkey with statistical methods[.[3]](#_References)

Saptarsi Goswami, Sanjay Chakraborty, Sanhita Ghosh, Amlan Chakrabarti and, Basabi Chakraborty studied the application of data mining techniques to combat natural disasters where data mining models are applied to various types of data.[[4]](#_References)

Q. Wang, Y. Guo, L. Yu, and P. Li studied earthquake prediction with a deep learning technique called long short-term memory to learn spatial-temporal relationship among earthquakes in different locations and they resulted that two-dimensional input made better earthquake predictions[.[5]](#_References)

Orhan Polat, Elçin Gök and, Doğuser Yilmaz analyzed the earthquake hazard of the Aegean Region of Turkey. Their analysis covered earthquakes from 1900 to 2002 for the West Anatolian Region. [[6]](#_References)

Lee, Jin A., JongGyu Han, and Kwang Hoon Chi tried to find a relationship between date, location, magnitude, and depth of earthquake using global earthquake data since 1973[.[7]](#_References)

There is no comprehensive study of earthquakes from a data science perspective. In this study, we hope to fulfill this gap and investigate some research questions. Our research questions are as follows:

1. How does the number of earthquake occurrence change over the years in Turkey? Is there a trend?
2. Which region of Turkey has the highest number of earthquakes?
3. Is there an “earthquake season”?
4. How is Marmara Region's and Istanbul's situation in particular?
5. Drawing Turkey's fault-line map by using the 20-years of data.
6. Is there a pattern between the weather and the earthquake occurrence?
7. Can we predict the average daily number of earthquakes monthly for the upcoming months in Istanbul?

# Data

In this study two different data sources are used, which are explained below. There are 159584 rows for the earthquake and 26000 rows for the climate data.

## Kandilli Observation Center Data

A dataset of earthquakes with a magnitude of 2 to 9 in Turkey over the years 2000-2019. The dataset is generated by Kandilli Observation Center’s data engine (see http://www.koeri.boun.edu.tr). The data have 19 columns.

* *ID:* Unique row number of the event.
* *Code*: Unique ID for the event that is assigned by Kandilli Observation Center.
* *Date*: Date of the event.
* *Time:* Origin time of the event (UTC).
* *Latitude, Longitude*: Coordination of event (in decimal degrees)
* *Depth*: Depth of the event in kilometers.
* *xM*: Biggest magnitude value in specified magnitude values.
* *MD, ML, Mw, Ms, Mb*: Magnitude types (MD: Duration, ML: Local, Mw: Moment, Ms: Surface wave, Mb: Body-wave).
* *Type*: Ke (Earthquake) or Sm (Suspected Explosion).

While conducting the study 2 more columns added due to classify the earthquakes for a better analysis. The added columns are region and season.

## NASA’s Weather Data of Turkey

A dataset generated using Turkey’s coordinates by NASA’s climate database engine. The dataset represents the climate data of Turkey in between 2000-2019. It is coordinate-based and shows 0.5 x 0.5 Degree Interannual Averages/Sums. The data have 16 columns that represent the values for each month, annual average, latitude, longitude, and parameter. Parameters are as follows:

* Minimum Temperature at 2 Meters (C)
* Maximum Temperature at 2 Meters (C)
* The Temperature at 2 Meters (C)
* Temperature Range at 2 Meters (C)
* Wet Bulb Temperature at 2 Meters (C)
* Relative Humidity at 2 Meters (%)
* Precipitation (mm)
* Maximum Wind Speed at 10 Meters (m/s)
* Minimum Wind Speed at 10 Meters (m/s)
* Maximum Wind Speed at 2 Meters (m/s)
* Minimum Wind Speed at 2 Meters (m/s)
* Dew/Frost Point at 2 Meters (C)
* Surface Pressure (kPa)

# Methods

All data analysis and experiments were done in Python using numpy [[8]](#_References) , pandas [[9]](#_References), matplotlib [[10]](#_References) , sklearn [[11]](#_References) libs.

The analysis was prepared as a Jupyter Notebook and shared on Nilay Çiçekli’s GitHub page. [[12]](#_References)

The weather-earthquake correlation hypothesis was tested both by comparing the patterns for weather parameters and earthquake graphs and using numpy’s correlation method.

Linear and polynomial regressions [[13][14]](#_References) are used for earthquake prediction using 20 years of historical data.

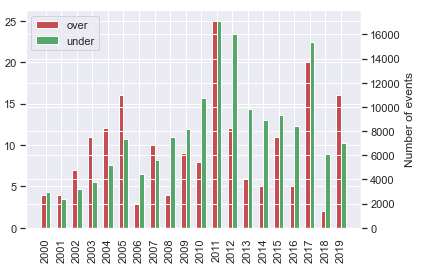


Fig. 3. Yearly Number of Small (xM<5) and Big (xM>5) Earthquakes in Turkey in One Bar Chart

# EXPERIMENTS & ANALYSIS

The data is grouped by day and the daily number of earthquake occurrences is found and shown in Fig. 1. The x-axis shows the order of days (20 years \* 365 days = 7000 days) and the y-axis shows the number of occurrences. The greatest number of occurrences in one day is 505 on the 21st of July 2017. On the same day, there are several earthquakes with a magnitude of over 4. They are close by each other, around Muğla and Gökova Körfezi (Mediterranean). This is called foreshock; small earthquakes trigger larger ones.

Another experiment made was by classifying the quakes according to their magnitudes and finding the yearly number of occurrences. In Fig. 2. distribution of small and big earthquakes is seen over the years. Small earthquakes reach up to 16000 in a year while big ones are about 25 at most, and sometimes even less than 5. That means Turkey does not have many strong earthquakes. Turkey is a country that has smaller but frequently occurring earthquakes.

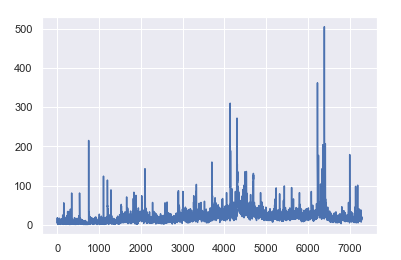


Fig. 1. Daily Number of Quake Occurrences

The next thing is studying the result of the previous experiment in one chart. As seen on this bar chart, sometimes small size earthquakes get the upper hand, sometimes great ones. That means, there is no conclusion such as great magnitude earthquakes trigger small-sized earthquakes or vice versa. Both are possible, and there is no pattern.

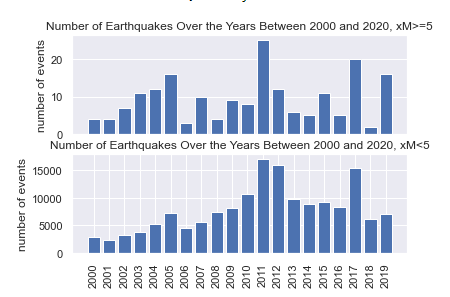


Fig. 2. Yearly Number of Small (xM<5) and Big (xM>5) Earthquakes in Turkey

Another experiment made was classifying the earthquakes according to their regions and observing the regional distribution of earthquakes of Turkey. The experiment made by 159584 data points and the results showed that over 20 years, the earthquakes occurred in the Aegean region the most, with a number 54658 in total, that is over 34% of all the data. The following regions are Eastern Anatolia, Mediterranean, Marmara, Central Anatolia, Black Sea, and Southeastern Anatolia in decreasing order. A pie chart representing this result is shown in Fig. 4. with percentages. There is also a grey slice in the pie chart that shows the borders of Turkey with 1.22%.

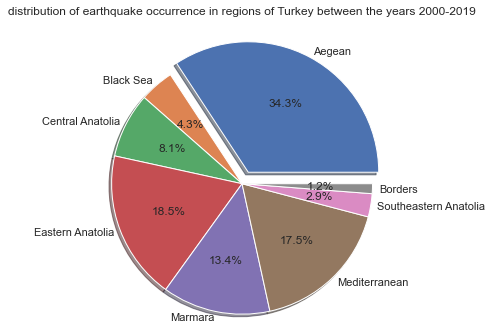


Fig. 4. Distribution of Earthquake Occurrence in Regions of Turkey between 2000-2019

The distribution of the earthquakes according to the regions of Turkey is also made for each year between 2000-2019. Over the years, it is observed that the Aegean region generally dominates other regions, it reaches a peak in 2002 and 2005 with almost 50%. However, there are times that the Marmara region increases unexpectedly. It even dominates the Aegean region by 43% in 2001 while it is 10% in 2005 and 2007; and 7% in 2011 and 2012. This situation proves that the Marmara region does not have a specific interval and it is unstable. This instability of the Marmara and regional distribution of earthquakes over 20 years is shown in figures Fig. 5.a – Fig. 5.c. You can see the rest of the plots of regional distribution of earthquakes in the Jupyter Notebook on GitHub. [[12]](#_References)

Due to observing the changes better, line charts are drawn for each region separately. Some of those charts are represented in Fig. 6.a - Fig. 6.b. You can see the charts on GitHub. [[12]](#_References) It is observed that the Aegean region is moving around 35% and the Marmara region is around 15% mostly, but the Marmara region has a dramatic difference between its highest and lowest points.

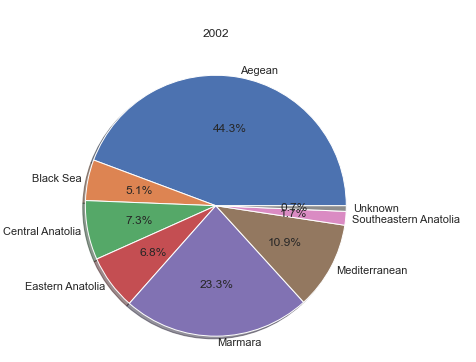


Fig. 5.a. Regional Distribution of Earthquakes of Turkey in 2002

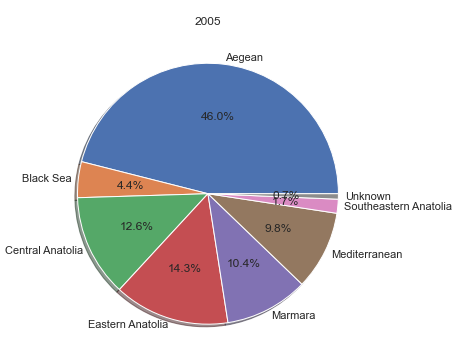


Fig. 5.b. Regional Distribution of Earthquakes of Turkey in 2005

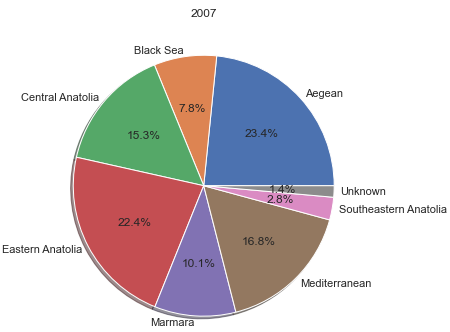


Fig. 5.c. Regional Distribution of Earthquakes of Turkey in 2007

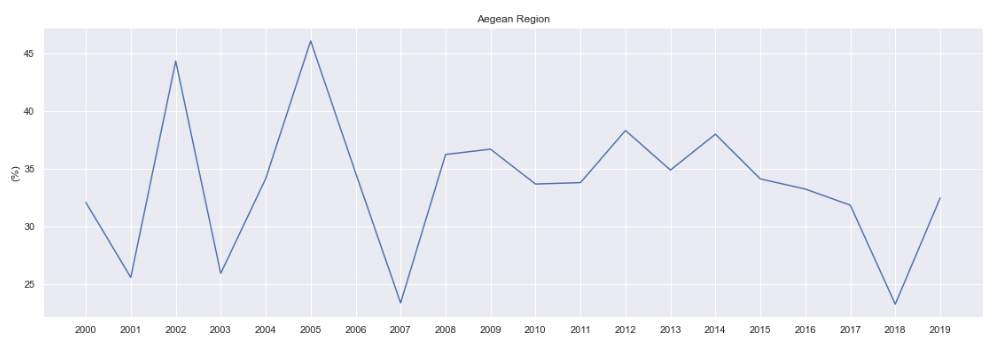


Fig. 6.a. Percentages of Earthquakes Over 20 Years for Aegean Region

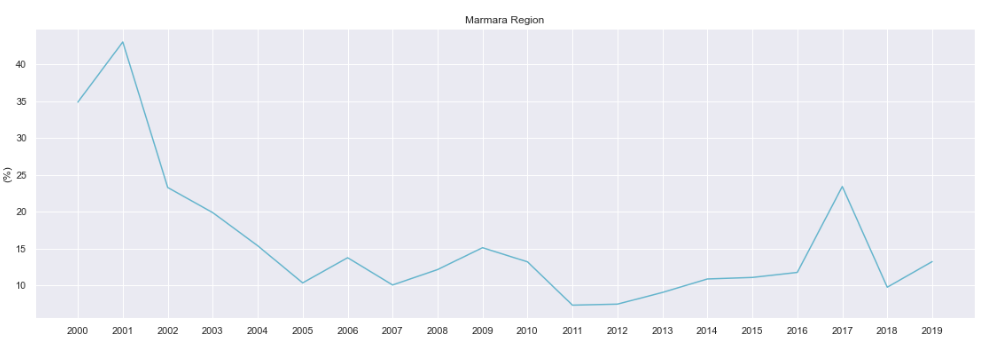
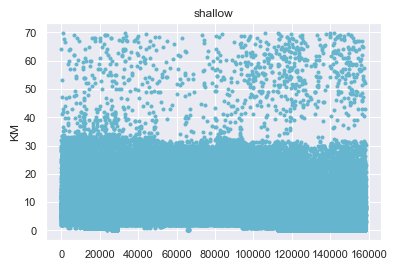


Fig. 6.b. Percentages of Earthquakes Over 20 Years for Marmara Region

Another research is to find if there is any specific “earthquake season”, meaning “is there a season where earthquakes occur more often than other seasons?”. Due to find an answer, data is classified according to months. December, January, February are winter; March, April, May are spring; June, July, August are summer and September, October, November are fall in Turkey. In the experiment, it is found that there is no specific season where earthquakes occur more than other seasons. Distribution of the

Fig. 10. Shallow Earthquakes of Turkey

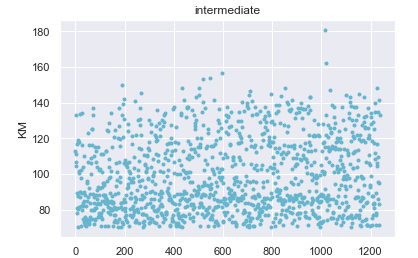


Fig. 11. Intermediate Earthquakes of Turkey

earthquakes according to seasons is shown in Fig. 7.

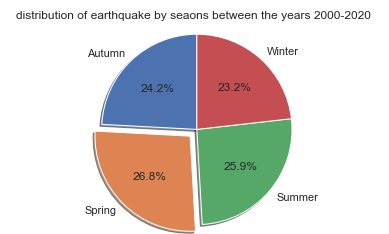


Fig. 7. Distribution of Earthquakes by Seasons in Turkey

After analyzing the Istanbul data, a graph is drawn that shows the number of earthquakes that occurred in Istanbul between 2000 and 2019. (Fig. 8.). The result shows that Istanbul was stable between the years 2004-2013, but around 2016 and 2017, earthquakes increased dramatically.

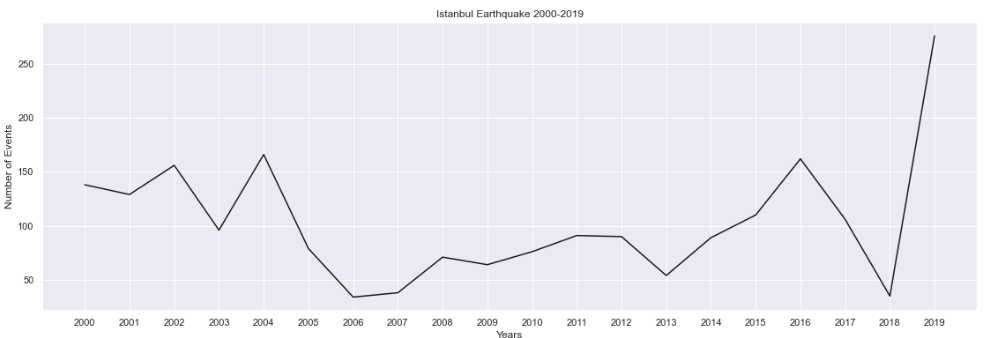


Fig. 8 Earthquake Occurrence in Istanbul between 2000-2019

And, drawing the graph again daily, not only with the number of occurrences but with their magnitude results as in Fig. 9. This figure tells that magnitudes of earthquakes in Istanbul are smaller after 2010 compared to the period 2000-2010.

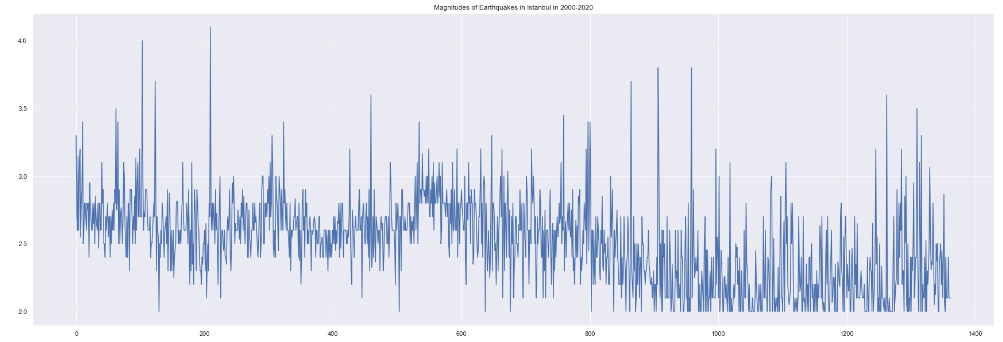


Fig. 9. Magnitudes of Earthquakes in Istanbul in 2000-2020, Daily

Another experiment made was about the depth of the earthquakes. There are three depth zones. Shallow, intermediate, and deep earthquakes. Deep earthquakes are 300-700 km deep in the ground and they are not seen in Turkey. Intermediate is 70-300 km deep, and shallow ones are 0-70 km deep. The experiment shows that the earthquakes in Turkey are mostly shallow quakes. And, sadly they cause the most damage. As seen in Fig. 10 and Fig. 11, intermediate earthquakes are less frequent and range is 80-180 km, shallows are dense in the interval 0-30 km.

There is also an experiment made to state the distribution of shallow and intermediate earthquakes according to regions. Since shallow ones dominate and the Aegean region is already found as the most earthquake occurring region before in another experiment in this paper (Fig.4), the distribution of shallow earthquakes is not so different from the previous experiment’s result. In Fig. 12.1 this result is shown and while this figure is compared with Fig. 4., similarity can be observed.

However, the result for intermediate earthquakes is more interesting. The result shows that most intermediate ones, 89.5%, happened in the Mediterranean region in the past 20 years. This result proves that Turkey is a shallow earthquake area except for the Mediterranean. (Fig. 12.2)

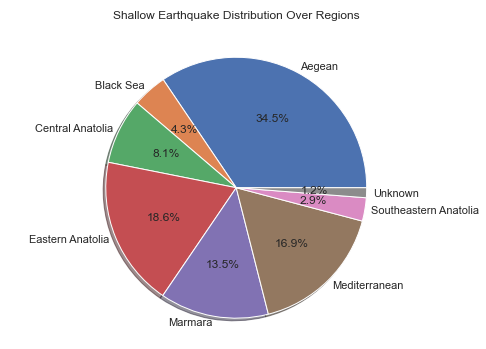


Fig. 12.a. Shallow Earthquake Distribution for Regions of Turkey

The final experiment made about depth is investigating the shallow distributions of the Marmara region. The result is represented in Fig. 13. The result suggests that 2017 was a year where the number of earthquakes increased drastically in the Marmara. The same result was found in previous experiments for Istanbul in this paper. So, it can be concluded that Istanbul (and its close area) dominates the earthquakes of the Marmara region.

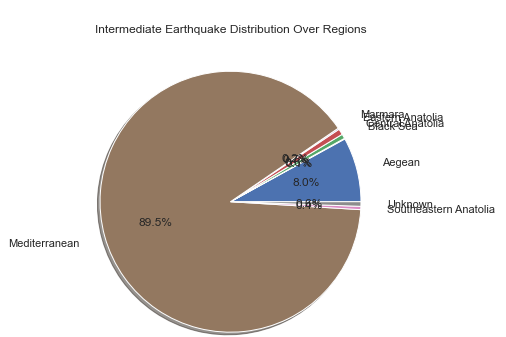


Fig. 12.b. Intermediate Earthquake Distribution for Regions of Turkey

Using all the data it can be drawn Turkey’s fault line map by marking the coordinates for each earthquake. The result is convenient and represents a roughly drawn map. The map is given in Fig. 14.

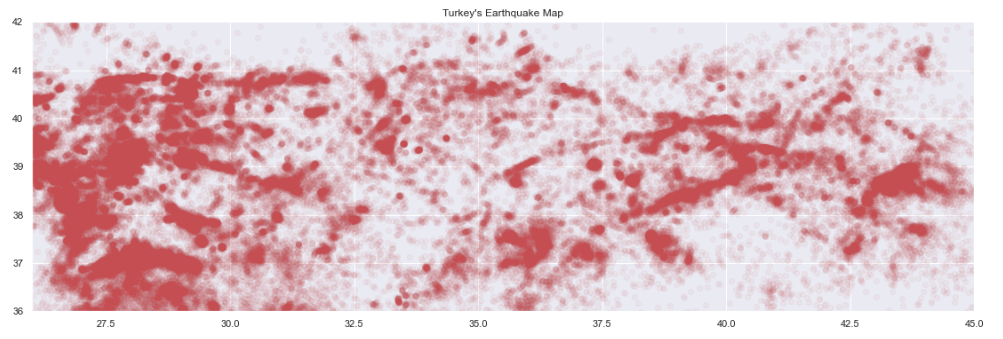


Fig. 14. Turkey Fault Line Map

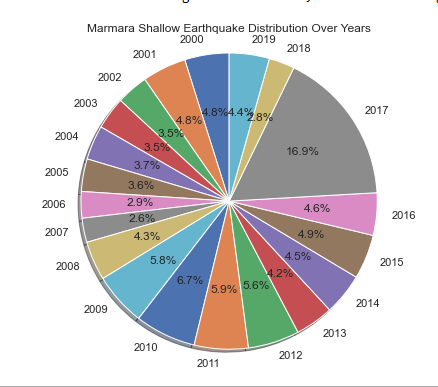


Fig. 13. Marmara Region’s Shallow Earthquakes

In the next experiment, earthquake data and temperature data are studied together to see if there is a correlation on yearly basis. In Fig. 15 max, min, average temperatures, and the number of earthquakes are seen in the same graph. The number of earthquakes is downscaled to fit into the graph, the main idea was to see if there is a matching pattern.

The correlation coefficient is also investigated for min temp-number of earthquakes and max temp-number of earthquakes. Coefficients were -0.045 and -0.015 which are very low to call the relation a non-coincidental relation.

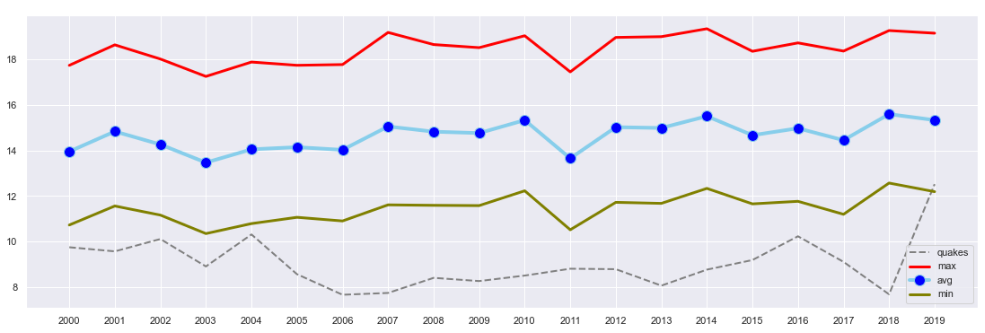


Fig. 15. Temperature and Earthquake

The same experiment is made with precipitation and humidity data combined with earthquake data separately. Resulting graphs are shown in Fig. 16 and Fig. 17. Either Fig. 15, Fig. 16 or Fig. 17 do not suggest any pattern between earthquake and weather data. The correlation scores for humidity and precipitation were 0.059 and -0.18 respectfully. Even though the correlation for precipitation level appears to be higher than the others, the score is not high enough to worth investigating. However, when the experiment is repeated for magnitude and depth of the earthquakes with weather parameters, some interesting results are found. A correlation score of 0.309 over 1 for humidity-magnitude relation, a score of -0.330 for max temperature-magnitude, and a score of -0.362 for min temperature-magnitude are found. These scores suggest that there might be a relation between weather and earthquakes somehow. Due to a lack of data, this paper has not taken further analysis on the topic, so it does not claim to have exhaustive conclusions but it does urge future research on the topic.

The very final experiment was on making a prediction of future earthquakes for Istanbul. Linear and polynomial regression techniques are used. Linear regression did not provide a good fit, but polynomial did. But towards the ends, an increase in the year 2017 affected the line wrongly. Linear and polynomial results are given on the same graph in Fig. 18. In conclusion, the prediction should be improved with other methods.

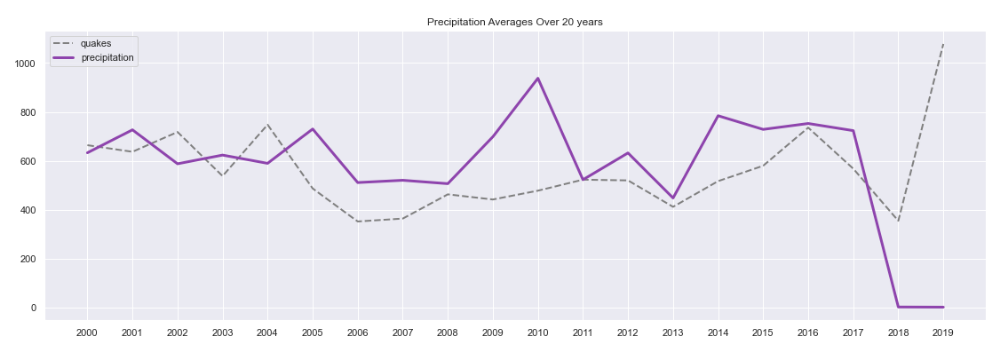


Fig. 16. Precipitation Levels and Earthquake

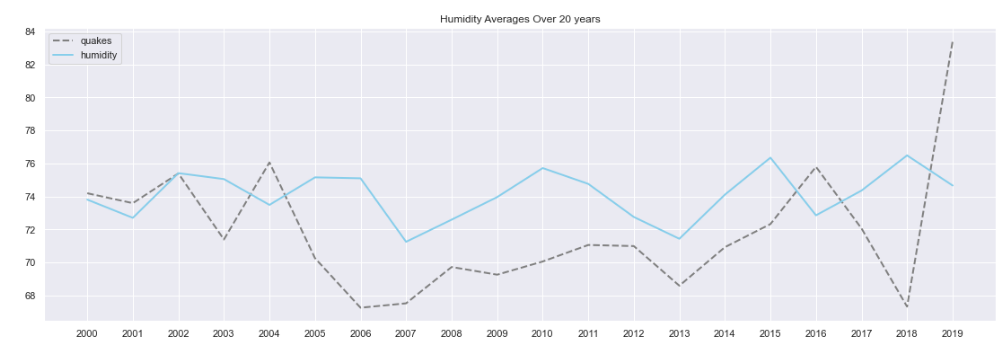


Fig. 17. Humidity Level and Earthquake

The linear prediction suggested that the average daily number would be around 10 for upcoming months while polynomial prediction suggested that from January to June in 2020, the daily number of earthquakes in Istanbul would be around 25. Normally, real values are in a range from 0 to 25. So, predicted numbers look viable but there are still question marks since 25 is the highest value normally.

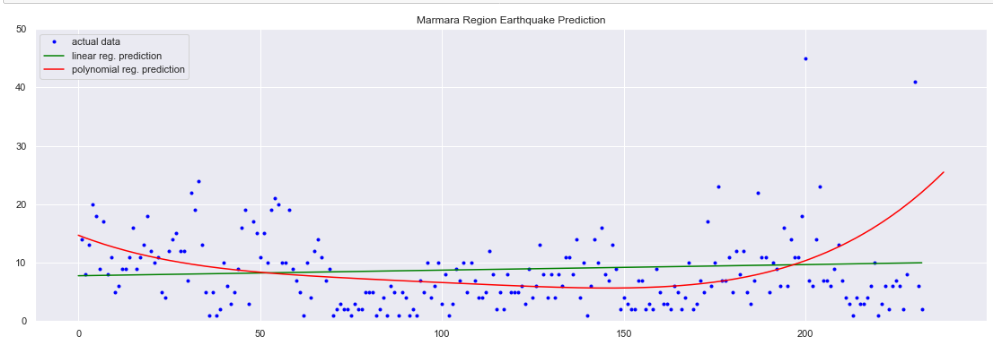


Fig. 18. Prediction Using Linear and Polynomial Regression

# Dıscussıon

In this paper, an extensive earthquake analysis of Turkey is conducted. Due to carrying out the experiments data mining techniques are used. Earthquakes over the past 20 years are in Turkey are investigated and represented in graphs. The regional analysis of earthquakes is also made. It is found that the Aegean region is the region that has the highest number of earthquakes in Turkey. But the Marmara is also a region at risk, because of the unstable nature of the region. It is shown that Istanbul and its nearby area dominate the earthquakes in the Marmara region. Also, the experiments represented that Turkey is a shallow earthquake area, and shallow earthquakes are the most damaging ones among all types.

Weather data of Turkey is studied combined with earthquake data to find a correlation and there found possible relations between humidity-magnitude and temperature-magnitude. The paper does not have a certain conclusion, but the results are worth working on more.

Due to make earthquake prediction 2 different methods are used. Linear regression could not fit the data well. Polynomial regression performed a better fit but still, because of the lack of the second-dimensional data, the prediction was not successful. But it was a good step to start with. This paper can encourage other studies to work relational analysis on earthquakes.

##### References

1. Sawe, B. (2018, July 12). The World's Costliest Earthquakes. Retrieved June 24, 2020, from https://www.worldatlas.com/articles/the-world-s-costliest-earthquakes.html
2. Gurenko, Eugene; Lester, Rodney; Mahul, Olivier; Gonulal, Serap Oguz (2006). Earthquake Insurance in Turkey: History of the Turkish Catastrophe Insurance Pool. World Bank Publications. p. 1. ISBN 9780821365847.
3. YILMAZ, Veysel; ERİŞOĞLU, Murat; ÇELİK, H. Eray. Probabilistic Prediction of the Next Earthquake in the NAFZ (North Anatolian Fault Zone), Turkey. Doğuş Üniversitesi Dergisi, [S.l.], v. 5, n. 2, p. 243-250, mart 2011. ISSN 1308-6979.
4. Saptarsi Goswami, Sanjay Chakraborty, Sanhita Ghosh, Amlan Chakrabarti, Basabi Chakraborty, A review on application of data mining techniques to combat natural disasters, Ain Shams Engineering Journal, Volume 9, Issue 3, 2018, Pages 365-378, ISSN 2090-4479.
5. Q. Wang, Y. Guo, L. Yu and P. Li, "Earthquake Prediction Based on Spatio-Temporal Data Mining: An LSTM Network Approach," in IEEE Transactions on Emerging Topics in Computing, vol. 8, no. 1, pp. 148-158, 1 Jan.-March 2020, doi: 10.1109/TETC.2017.2699169.
6. Polat, Orhan, Elcin Gok, and Doğuser Yilmaz. "Earthquake hazard of the Aegean extension region (West Turkey)." Turkish Journal of Earth Sciences 17.3 (2008): 593-614.
7. Lee, Jin A., JongGyu Han, and Kwang Hoon Chi. "Mining quantitative association rule of earthquake data." Proceedings of the 2009 International Conference on Hybrid Information Technology. 2009.

[8] Stéfan van der Walt, S. Chris Colbert and Gaël Varoquaux. **The NumPy Array: A Structure for Efficient Numerical Computation**, Computing in Science & Engineering, **13**, 22-30 (2011), [DOI:10.1109/MCSE.2011.37](http://dx.doi.org/10.1109/MCSE.2011.37) ([publisher link](http://scitation.aip.org/content/aip/journal/cise/13/2/10.1109/MCSE.2011.37))

[9] Wes McKinney. **Data Structures for Statistical Computing in Python**, Proceedings of the 9th Python in Science Conference, 51-56 (2010) ([publisher link](http://conference.scipy.org/proceedings/scipy2010/mckinney.html))

[10] John D. Hunter. **Matplotlib: A 2D Graphics Environment**, Computing in Science & Engineering, **9**, 90-95 (2007), [DOI:10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55) ([publisher link](http://scitation.aip.org/content/aip/journal/cise/9/3/10.1109/MCSE.2007.55))

[11] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Édouard Duchesnay. **Scikit-learn: Machine Learning in Python**, Journal of Machine Learning Research, **12**, 2825-2830 (2011) ([publisher link](http://jmlr.org/papers/v12/pedregosa11a.html))

[12] Çiçekli Nilay, Earthquake, (2020), GitHub repository, <https://github.com/nilaycicekli/earthquake>

[13] Duda, R. O., Hart, P. E.,, Stork, D. G. (2000). *Pattern Classification (2nd Edition)*. Wiley-Interscience. ISBN: 0471056693

[14] VanderPlas, J. (2017). *Python data science handbook: Essential tools for working with data*. Sebastopol, CA: O'Reilly.