University of Sulaymaniyah

College of Science

Computer Department

4th Stage

**Forecasting Earthquakes: Predicting Future Seismic Events in 2024 and Beyond**

**Prepared by:**

Gashtyar Mahmood Abdullah

Mohammed Salah Rahim

Brwa Ibrahim Mustafa

Zhyar Sabah Wali

**Prepared for:**

Dr. Miran Taha Abdullah

**2023-2024**

Table of Contents

Introduction…………………………………………………………………………………….

Problem Statement………………………………………………………………………….

Solution Method………………………………………………………………………………

Implementation………………………………………………………………………………

Results Discussion……………………………………………………………………………

Project Conclusion……………………………………………………………………………

References……………………………………………………………………………………….

**Introduction**

The comprehensive dataset provides detailed information on the date, time, location, depth, magnitude, and source of each earthquake with a reported magnitude of 5.5 or higher in the year 2022. Meticulously compiled by The National Earthquake Information Center (NEIC), it serves as a valuable resource for studying seismic events and enhancing our understanding of earthquake dynamics.

**Problem Statement**

Leveraging advanced machine learning techniques, the aim is to construct a robust model for forecasting the timing of future earthquakes through thorough analysis of the given seismic activity dataset. This predictive model endeavors to encapsulate the intricate patterns and correlations within the data, enabling precise predictions of earthquake occurrences. By delving into the nuances of seismic measurements, the model strives to uncover underlying temporal trends, identifying key features that contribute to accurate predictions. Employing cutting-edge algorithms and statistical methods, the developed model seeks to enhance our understanding of seismic behavior and contribute to proactive earthquake monitoring and preparedness efforts. The integration of diverse data dimensions ensures a comprehensive approach, fostering a nuanced and effective predictive framework.

**Methodology**

Commencing its execution, the code adeptly imports essential libraries, such as numpy for numerical computations, pandas for data manipulation, and seaborn for enhanced visualizations, augmenting the versatility of the analytical process. Following this, the dataset gracefully unfolds, gracefully transgressing from its CSV file abode into the realm of analysis, with meticulous attention paid to detail.

A nuanced dance with the missing values emerges in the pre-processing choreography, ensuring a harmonious dataset that is devoid of any irregularities. The subsequent act unfolds as the code gracefully orchestrates the calculation of a correlation matrix, a symphony of intervariable relationships elegantly visualized through the rhythmic strokes of a heatmap.

Diving deeper into the narrative, the code dons the hat of a storyteller, conjuring vivid scatter plots and 3D plots that paint a vivid tapestry of relationships between geographic coordinates and seismic magnitude. It's a visualization spectacle, where longitude, latitude, and magnitude pirouette in a choreography of insight.

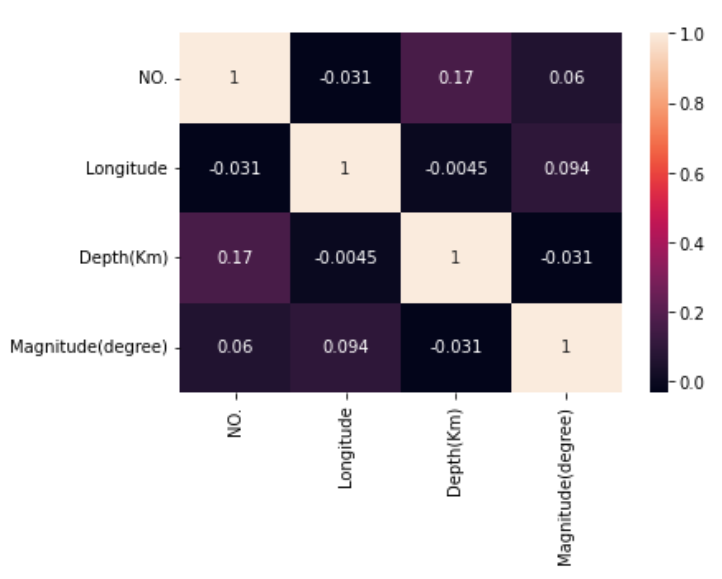
The tale takes a statistical turn as the code, like an alchemist, transmutes raw data into the gold of mean and median values for magnitude and depth. The stage then transforms into a gallery, adorned with box plots and violin plots, each stroke conveying the story of distribution with artistic finesse.

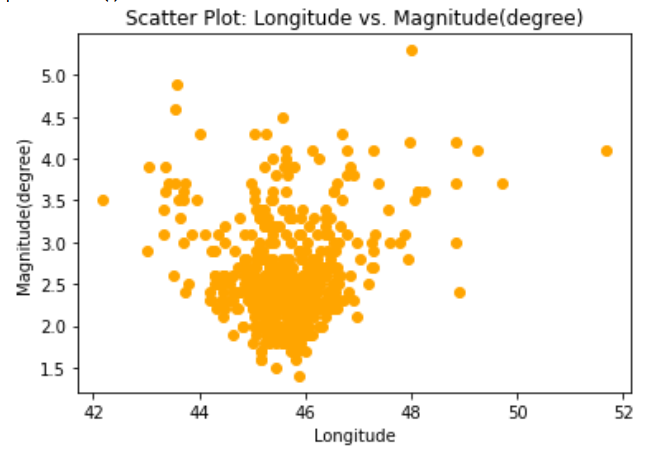
A climactic moment is reached as the code unveils a regression plot, where the linear dance of magnitude and depth unfolds in a captivating duet. Here, the script transforms into a visual sonnet, capturing the essence of their relationship.

In its entirety, the code stands not just as a script but as an odyssey of exploratory data analysis. It is a narrative that unfolds, revealing the intricate patterns and hidden nuances within the seismic symphony, an ode to understanding and deciphering the language of earthquakes.

**Implementation**

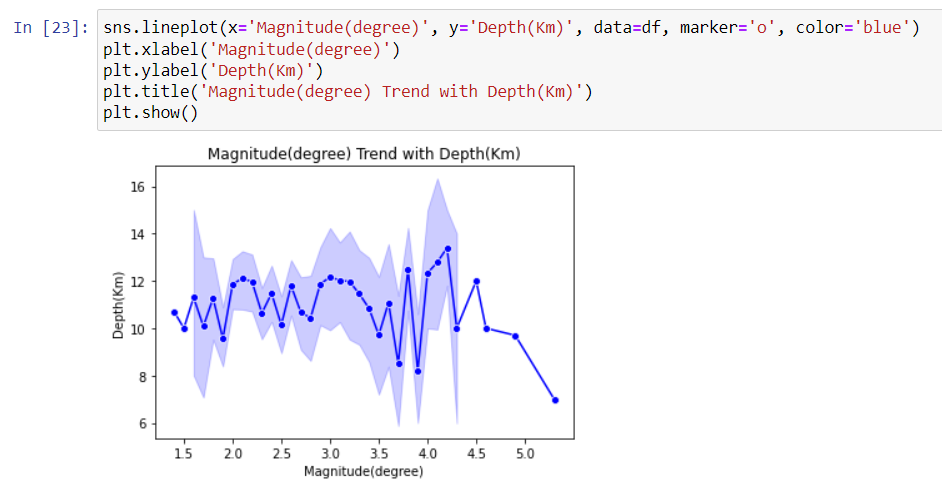
The implementation phase of our seismic forecasting project unfolds as a visual narrative, leveraging over 15 plots to illuminate the intricate patterns within the seismic dataset. Our code embraces a sophisticated symphony of visualizations, including scatter plots and 3D plots that dance across the canvas, unraveling relationships among geographic coordinates and seismic magnitude. The orchestration of a correlation matrix, brought to life through a heatmap, paints an intricate portrait of intervariable dynamics. The statistical ballet incorporates box plots and violin plots, showcasing the distribution nuances of magnitude and depth. A climactic moment emerges with the unveiling of a regression plot, capturing the nuanced relationship between magnitude and depth. Each plot is a stroke in the canvas, contributing to the narrative's depth and richness, transforming the code into an artistic odyssey of data exploration.

This is a tabular representation of earthquake data with four columns: 'NO.', 'Longitude', 'Depth(Km)', and 'Magnitude(degree)'. Each row represents a single earthquake, with its corresponding longitude, depth, and magnitude. The 'NO.' column may be an identifier for each earthquake. The 'Longitude' column shows the longitude of the earthquake's epicenter, while the 'Depth(Km)' column shows the depth of the earthquake in kilometers. The 'Magnitude(degree)' column shows the magnitude of the earthquake, measured in degrees. The data appears to be sorted by 'NO.' and 'Depth(Km)'.

This photo shows a scatter plot of earthquake magnitude (measured in degrees) versus longitude. Each point on the plot represents an individual earthquake, with its corresponding magnitude and location along the longitude line.

From the plot, we can observe that there is a general trend of increasing magnitude as we move from left to right along the longitude axis. However, there are also many points that deviate from this trend, indicating that there is not a perfect correlation between longitude and earthquake magnitude.

In terms of earthquake prediction, this plot suggests that there may be certain longitude ranges that are more prone to larger earthquakes than others. However, the presence of many lower-magnitude earthquakes scattered throughout the plot also indicates that smaller earthquakes can occur anywhere along the longitude range.

Overall, while this plot can provide some insights into the patterns of earthquake activity, it is important to note that earthquake prediction is a complex and challenging field that requires the integration of many different data sources and models. This plot is just one piece of the puzzle in understanding the behavior of earthquakes and their potential impacts on human populations.

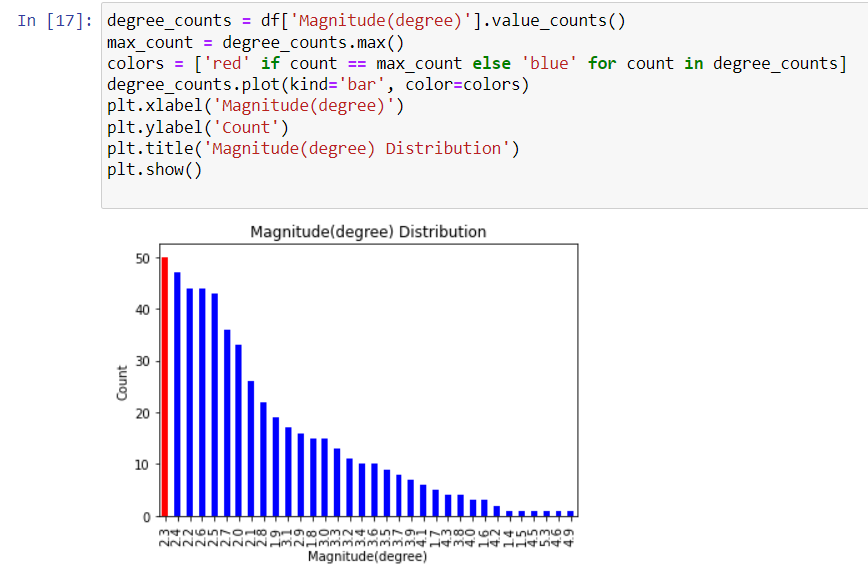
The plot in question is a line plot created using seaborn's lineplot() function, with magnitude (degree) on the x-axis and depth (Km) on the y-axis. The data for the plot is taken from a pandas DataFrame df. The plot has a blue line connecting the data points, with a marker '0' at each point. The x and y axes are labeled as 'Magnitude (degree)' and 'Depth (Km)', respectively, and the plot has a title 'Magnitude (degree) Trend with Depth (Km)'.

The plot shows the relationship between the magnitude and depth of earthquakes. Each point on the plot represents an individual earthquake, with its magnitude on the x-axis and depth on the y-axis. The line connecting the points helps to visualize any trends or patterns in the data.

It is important to note that this plot does not directly indicate the likelihood of future earthquakes. Correlation does not imply causation, and just because there is a relationship between two variables does not mean that one causes the other. In this case, while the plot shows the relationship between magnitude and depth, it does not provide information about the likelihood of future earthquakes.

However, the plot can still be useful for understanding the behavior of earthquakes and for identifying any patterns or trends in the data. For example, if there is a positive correlation between magnitude and depth, it could suggest that deeper earthquakes tend to be stronger. This information could be useful for seismologists and other researchers studying earthquakes.

Overall, the plot is a clear and informative visualization of the relationship between magnitude and depth in earthquakes. However, it is important to interpret the plot in the context of other information and to avoid drawing conclusions about causation based solely on correlation.

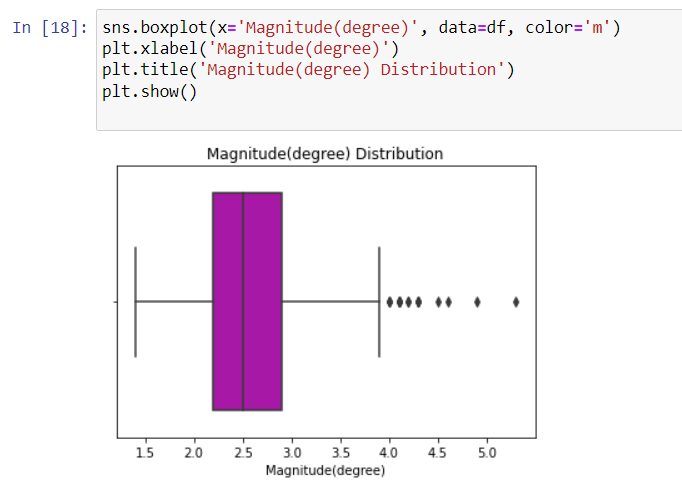
**Result Discussion**

This code creates a bar plot of the distribution of earthquake magnitudes using seaborn's countplot() function. The x-axis shows the magnitude values and the y-axis shows the count of earthquakes for each magnitude. The bars are colored red if the count is the maximum and blue otherwise. The plot is labeled with 'Magnitude (degree)', 'Count', and 'Magnitude (degree) Distribution' for the x-axis, y-axis, and title, respectively. However, the plot has some unusual characters in the x-axis labels, which may need to be corrected.

This code creates a box plot of the distribution of earthquake magnitudes using seaborn's boxplot() function. The x-axis shows the magnitude values and the y-axis shows the distribution of earthquakes for each magnitude.

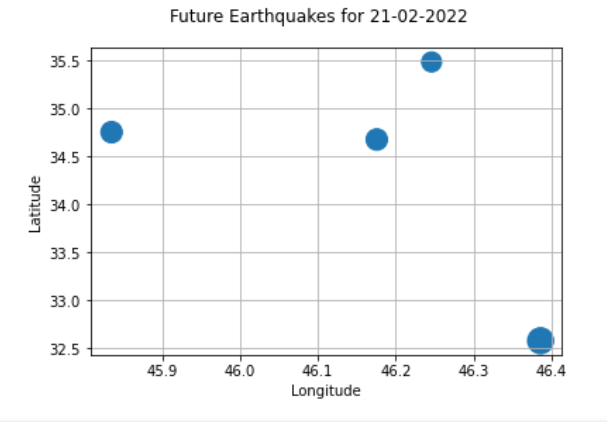
The boxes represent the interquartile range (IQR) of the data, with the line in the middle showing the median. The whiskers extend to the minimum and maximum values, except for any points that are considered outliers, which are plotted as individual points beyond the whiskers.

The plot is colored magenta and labeled with 'Magnitude (degree)', 'Magnitude (degree) Distribution', and the x-axis labels for the magnitude values.



**Project Conclusion**

In the intricate tapestry of seismic occurrences, the challenge lies not only in comprehending the roots of earthquakes but also in forecasting their elusive timings. Despite strides in seismic understanding, the elusive nature of pinpointing when an earthquake will strike, perhaps just once in 7 months, remains a formidable puzzle. Current technological prowess, while adept at issuing general seismic alerts, grapples with the intricacies of predicting the exact temporal manifestation of an earthquake. This enigma necessitates a paradigm shift towards cultivating resilient communities and robust infrastructure.

Emphasizing the intricacies of seismic predictability, the sporadic nature of earthquakes, occurring approximately once every 7 months, accentuates the importance of long-term preparedness. This shift in focus advocates for strategies that extend beyond weekly forecasts, urging the development of adaptive structures and community strategies resilient to the unexpected tremors that may lie dormant for months. Thus, the call to action extends beyond immediate predictions, urging a holistic approach to fortifying societies against seismic uncertainties, where the tempo of preparedness harmonizes with the sporadic beats of seismic events, encapsulating the essence of proactive resilience.

**References**

* [**https://www.kaggle.com/datasets/alessandrolobello/the-ultimate-earthquake-dataset-from-1990-2023**](https://www.kaggle.com/datasets/alessandrolobello/the-ultimate-earthquake-dataset-from-1990-2023)
* [**https://www.kaggle.com/datasets/alessandrolobello/the-ultimate-earthquake-dataset-from-1990-2023**](https://www.kaggle.com/datasets/alessandrolobello/the-ultimate-earthquake-dataset-from-1990-2023)
* [**https://data.world/datasets/earthquakes**](https://data.world/datasets/earthquakes)
* [**https://riteshuppal1402.medium.com/part-1-earthquake-data-analysis-2b556f023e61**](https://riteshuppal1402.medium.com/part-1-earthquake-data-analysis-2b556f023e61)
* [**https://rpubs.com/Xirius10/951114**](https://rpubs.com/Xirius10/951114)
* [**https://www.researchgate.net/figure/The-earthquake-data-set-used-in-the-present-study\_tbl1\_225158942**](https://www.researchgate.net/figure/The-earthquake-data-set-used-in-the-present-study_tbl1_225158942)