**COVER PAGE**

**DATA SCIENCE TOOBOX PYTHON PROGRAMMING**

**PROJECT REPORT**

(Project Semester January-April 2025)

***(Exploring Earthquake Magnitude, Depth & Trends )***

Submitted by

Shiwani Agrawal

Registration No.12317789

Programme and Section. K23FA

Course Code INT375

Under the Guidance of

**(Ms. Sandeep Kaur)**

**Discipline of CSE/IT**

**Lovely School of Computer Science**

**Lovely Professional University, Phagwara**

**CERTIFICATE**

This is to certify that **Shiwani Agrawal** (student’s name) bearing Registration no. **12317789** has completed INT375 <Course Code> project titled, **“Exploring Earthquake Magnitude, Depth & Trends”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

**Signature and Name of the Supervisor**

**Designation of the Supervisor**

**School of Computer Science**

Lovely Professional University

Phagwara, Punjab.

Date:

**DECLARATION**

I, Shiwani Agrawal, student of ............................ (Program name) under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: Signature

Registration No. 12317789 Shiwani Agrawal

**1. Introduction**

* **Project Overview**:
* The project aims to analyze earthquake data, focusing on the **magnitude**, **depth**, and **trends** associated with earthquakes.
* By performing exploratory data analysis (EDA), the goal is to uncover patterns and insights that can help in understanding earthquake occurrences globally.
* **Dataset**:
* The dataset consists of records of earthquakes, which include information on **magnitude**, **depth**, **location**, **time**, and **other related parameters**.
* The dataset might be sourced from public repositories like the **USGS Earthquake Database** or **Kaggle**.
* **Key Variables**:
* **Magnitude**: Measures the size of the earthquake, which can help assess the severity.
* **Depth**: Refers to how deep the earthquake occurs, which could impact its effects on the surface.
* **Location**: Geographical coordinates where the earthquake occurred.
* **Time**: Time of occurrence, used for analyzing trends over time.

**2. Source of Dataset**

**url:** <https://earthquake.usgs.gov/earthquakes/map/?extent=-89.95867,-409.21875&extent=89.95765,769.21875&range=search&timeZone=utc&search=%7B%22name%22:%22Search%20Results%22,%22params%22:%7B%22starttime%22:%222025-03-29%2000:00:00%22,%22endtime%22:%222025-04-05%2023:59:59%22,%22minmagnitude%22:2.5,%22orderby%22:%22time%22%7D%7D>

**3. Problem Statement**

Earthquakes are unpredictable natural disasters that can cause significant damage to infrastructure, economies, and human life. With thousands of earthquakes occurring globally every year, understanding their behavior is crucial for disaster preparedness and risk mitigation. Despite advancements in technology, predicting the exact occurrence of earthquakes remains a major scientific challenge.

The goal of this project is to perform **Exploratory Data Analysis (EDA)** on recent global earthquake data to uncover trends and patterns related to the **frequency**, **magnitude**, **depth**, **geographic distribution**, and **temporal occurrence** of earthquakes. The study aims to answer key questions such as:

* Which regions are most prone to earthquakes?
* What is the distribution of earthquake magnitudes and depths?
* How often do earthquakes occur on a daily basis?
* Are there any correlations between features like depth, magnitude, and location?

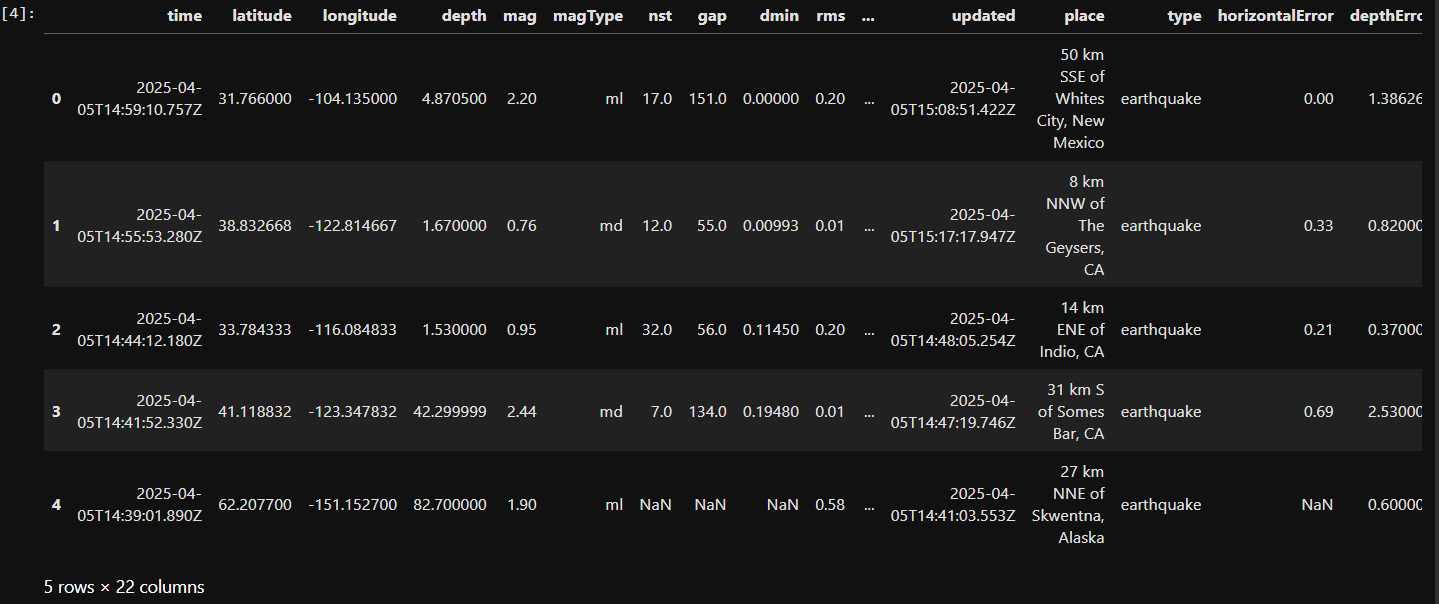
The insights obtained will provide a foundation for potential applications in **predictive modeling**, **clustering**, and **real-time monitoring systems**, contributing to more effective disaster preparedness strategies.

**4. Objectives:**

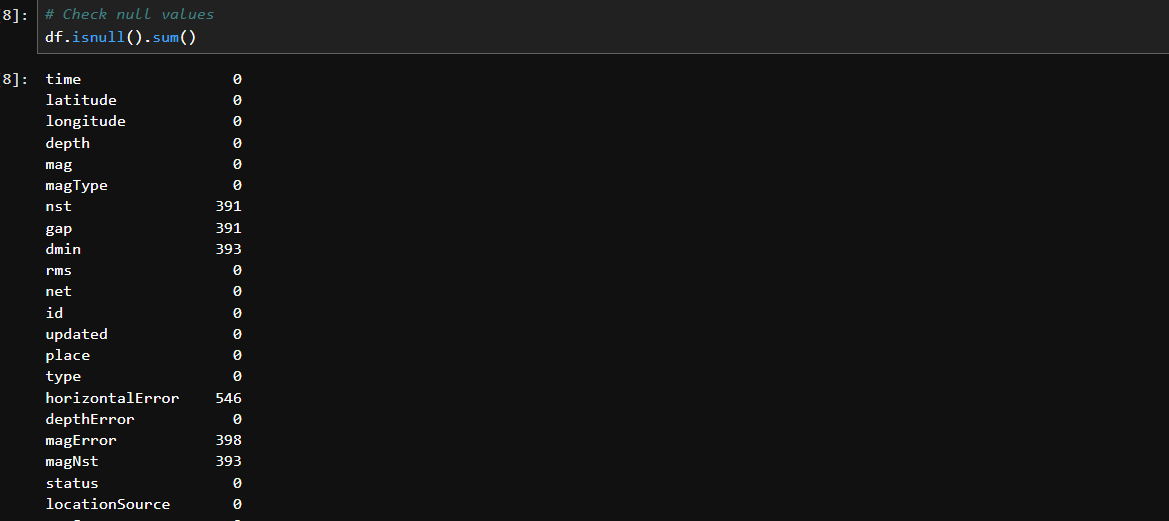
* Perform **Exploratory Data Analysis (EDA)** on earthquake events from the last week.
* Identify the **top regions** with the highest earthquake frequency.
* Understand the **distribution of magnitudes and depths**.
* Analyze **temporal trends**, such as daily frequency patterns.
* Explore **correlations** among numerical features like magnitude, depth, and location.
* Apply basic **visualizations** and **mapping techniques** to uncover geospatial trends.
* Lay the groundwork for **machine learning models** to predict earthquake behavior or cluster regions.

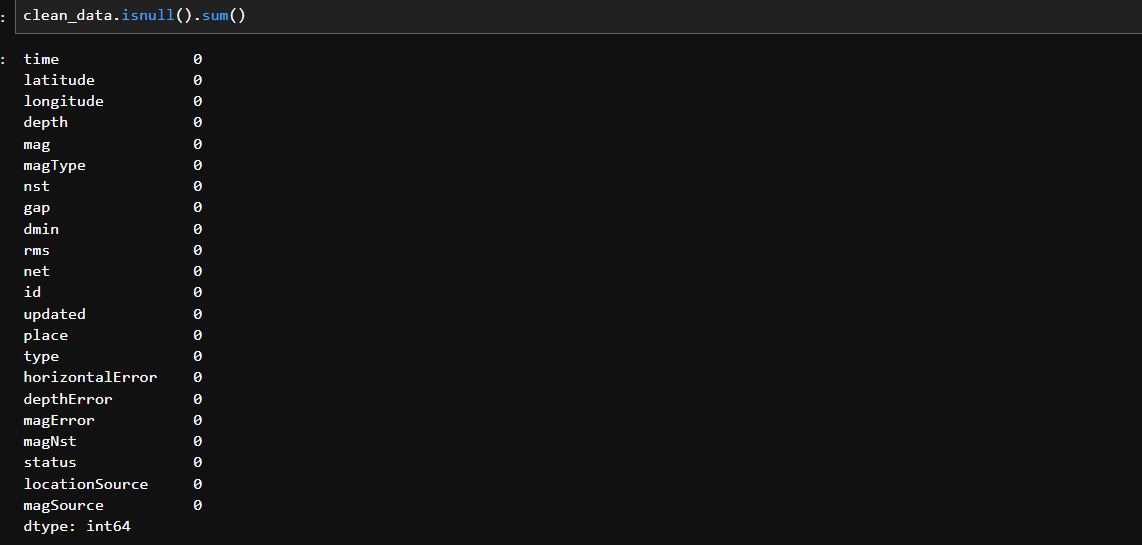
**5. EDA Process (Exploratory Data Analysis)**

1. **Import the Data**
   * 1. Load the dataset into your environment (e.g., using Pandas in Python).

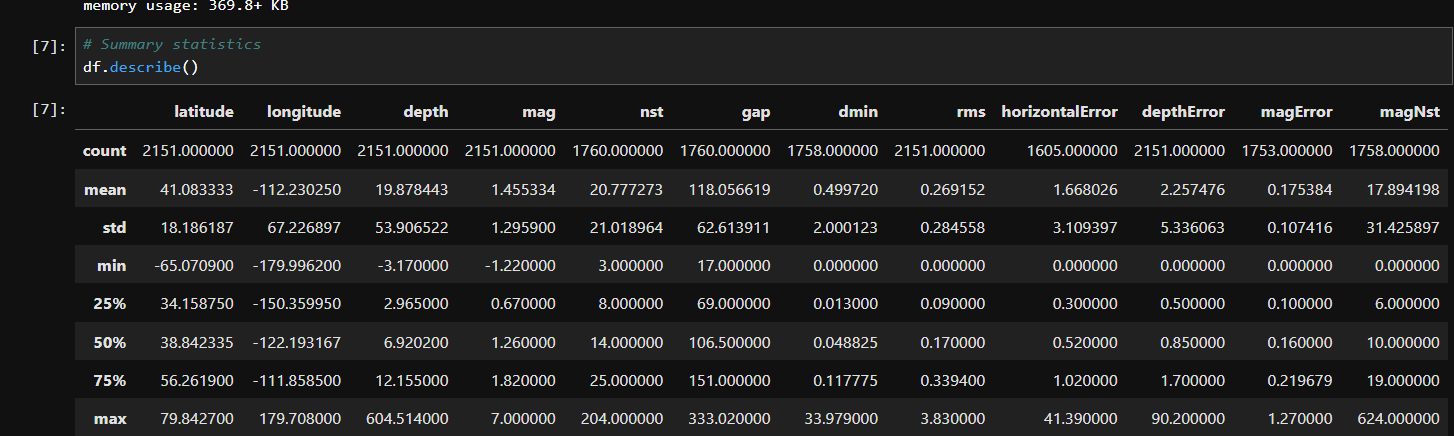


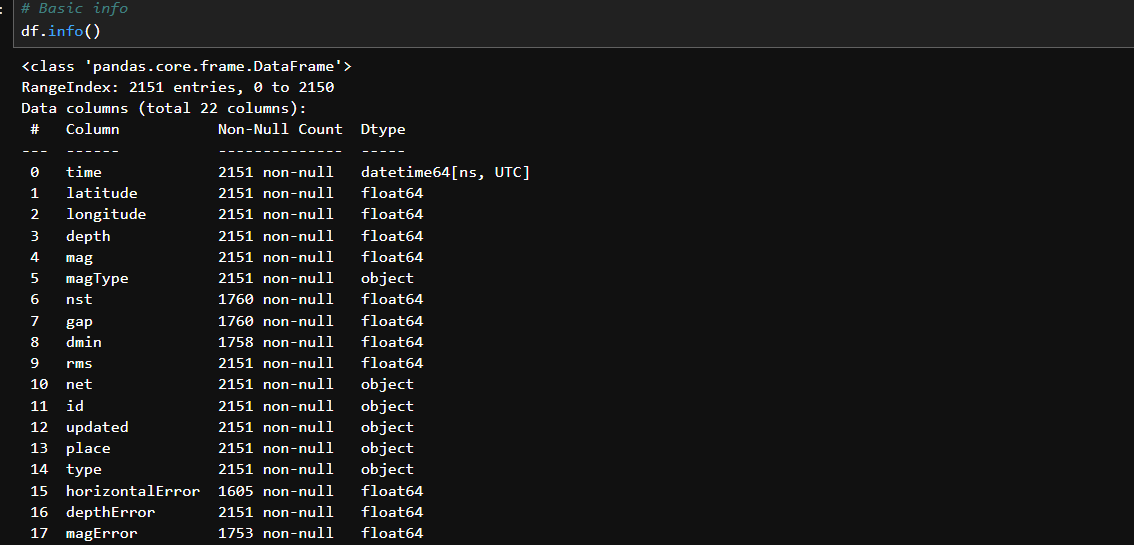
1. **Understand the data**
   * 1. Look at the first few rows (head ()).
     2. Check column names, data types, and shape (rows × columns).
2. **Clean the Data**
   * 1. Handle missing values (remove or fill them).
     2. Remove duplicate rows (if any).





1. **Summary Statistics**
   * 1. Use functions like describe () and info()to see mean, median, min, max, etc.
     2. Helps understand the range and spread of data.





1. **Check for Missing Values**
2. Use isnull().sum() to find missing data in each column.
3. **Data Visualization**
   * 1. Create charts to see patterns:
        1. **Line plots** to show trends over time.
        2. **Bar charts** to compare categories.
        3. **Heatmaps** to check correlations.
        4. Histogram to see distribution
4. **Check Correlations**
   * 1. Find relationships between columns using corr() and heatmaps.

**6.** **Analysis on Dataset (for each analysis)**

**i. Introduction**

* Earthquakes are sudden and potentially devastating natural disasters.
* Understanding seismic activity is essential for safety and preparedness.
* Data analysis helps in identifying patterns and understanding geographical impacts.
* This project analyzes a real-world dataset containing global earthquake data.
* It focuses on magnitude, depth, time, and regional distribution of earthquakes.
* Visualizations and statistics are used to extract meaningful insights.

**ii. General Description**

This project focuses on analyzing how vaccination has impacted the outcomes of COVID-19. The main aim is to compare the number of cases, hospitalizations, and deaths among vaccinated and unvaccinated people. By studying real-world data over time, we can understand the importance and effectiveness of vaccines in reducing the severity of the disease. The dataset used contains historical information and is organized by vaccination status. Python programming is used for data cleaning, analysis, and creating visualizations. This helps to clearly see trends, patterns, and differences between groups, making the data easy to understand and interpret.

1. **Specific Requirement, Functions and Formulas**

**Specific Requirement**

**Python libraries used:**

* **Pandas**: For loading and handling data.
* **NumPy**: For numerical operations.
* **Matplotlib & Seaborn**: For creating various graphs and plots.

**Functions Used**

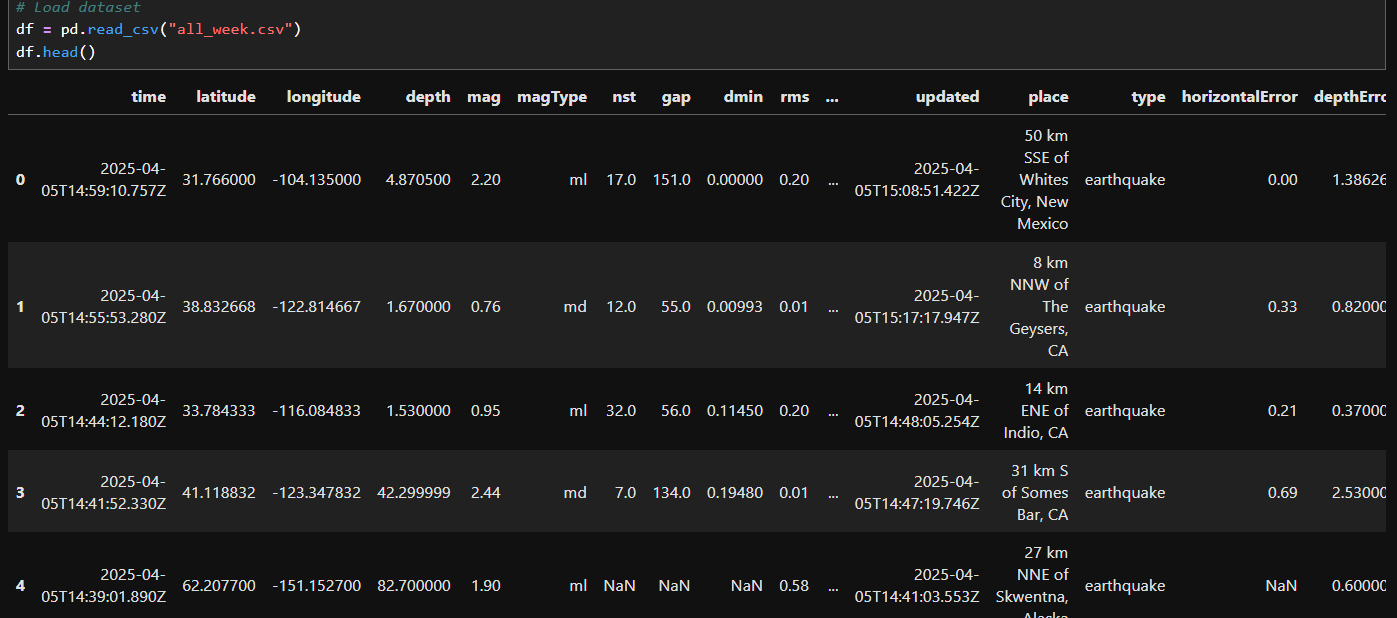
1. pd.read\_csv() – To load the dataset into a DataFrame.
2. df.head() / df.tail() – To preview the first or last few records.
3. df.info() – To understand data structure, column types, and missing values.
4. df.describe() – To get summary statistics for numerical columns.
5. df.isnull().sum() – To check for missing values in each column.
6. df.dropna() – To remove rows with missing values.
7. df.duplicated() – To detect and remove duplicate entries.
8. pd.to\_datetime() – To convert string dates to datetime objects.
9. df['column'].value\_counts() – To count frequency of categorical values.
10. df.groupby() – To group data based on categories like region or date.
11. df.corr() – To compute correlation matrix between numerical columns.

**Formulas Used**

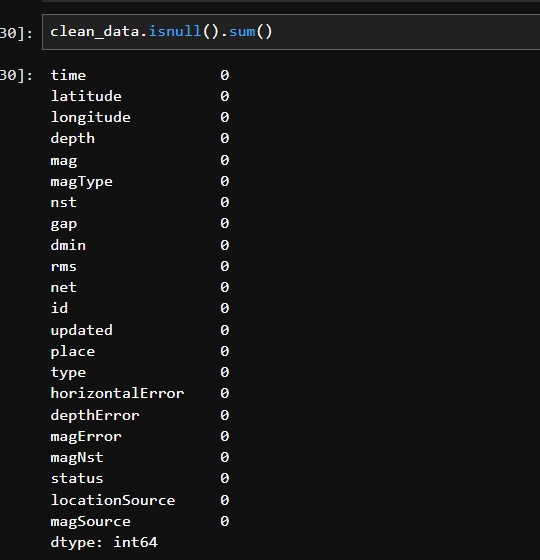
1. **Percentage Formula:** Used to compare case/death rates among vaccinated and unvaccinated**.**
2. **Correlation Coefficient (from df.corr()):** Measures how two variables are related (value between -1 and 1

**7.Analysis Results (EDA)**

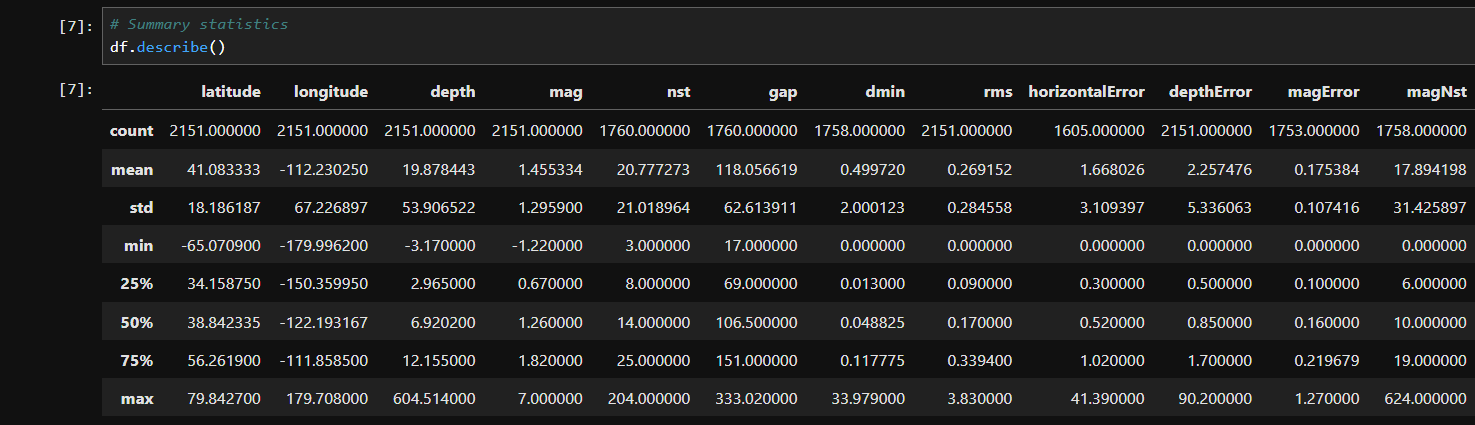
* **Import Data**

****

* **Clean the Data**

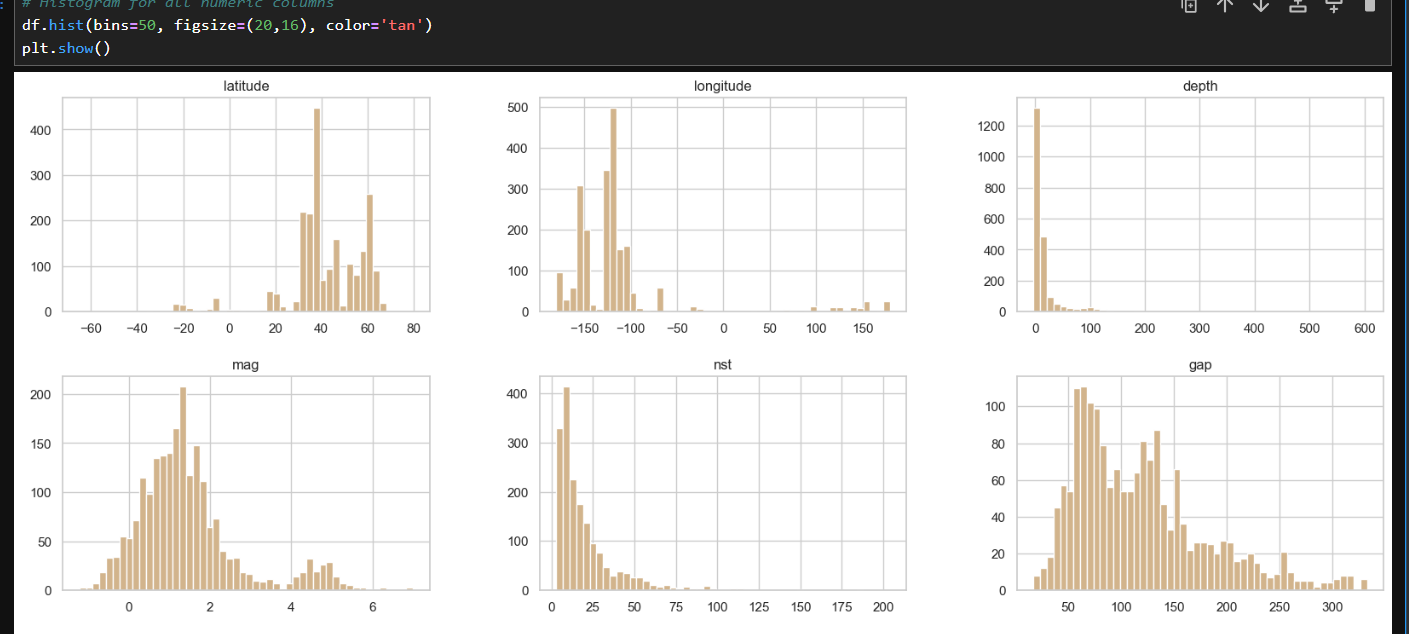
****

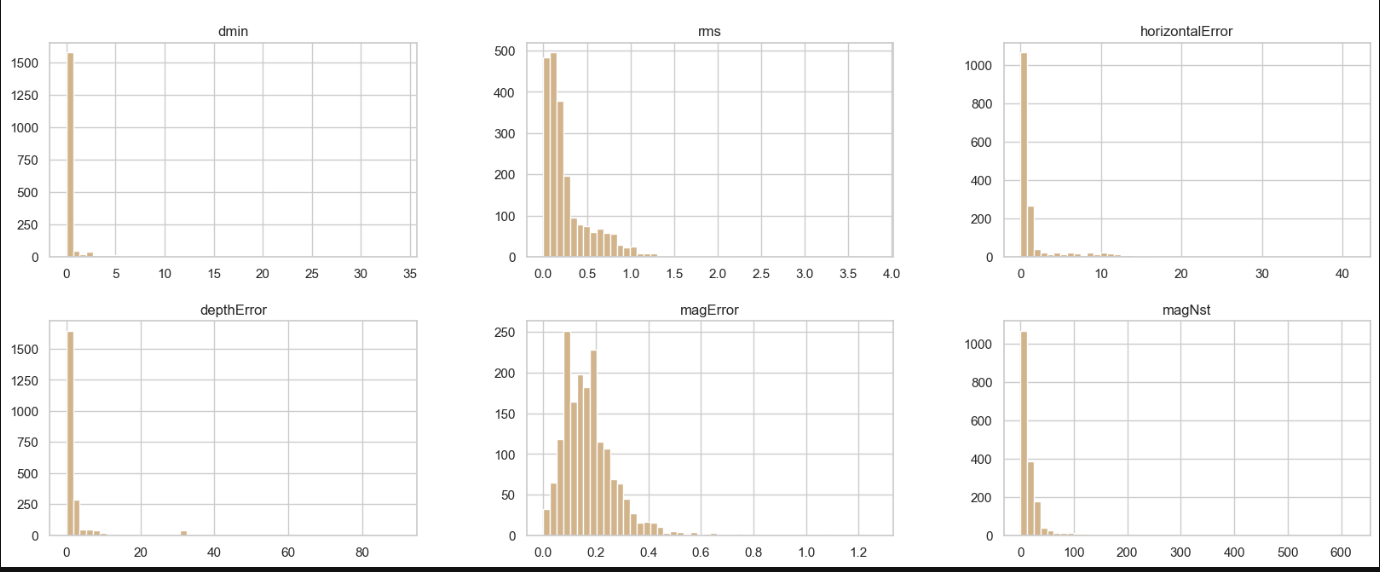
* **Summary Statistics**

****

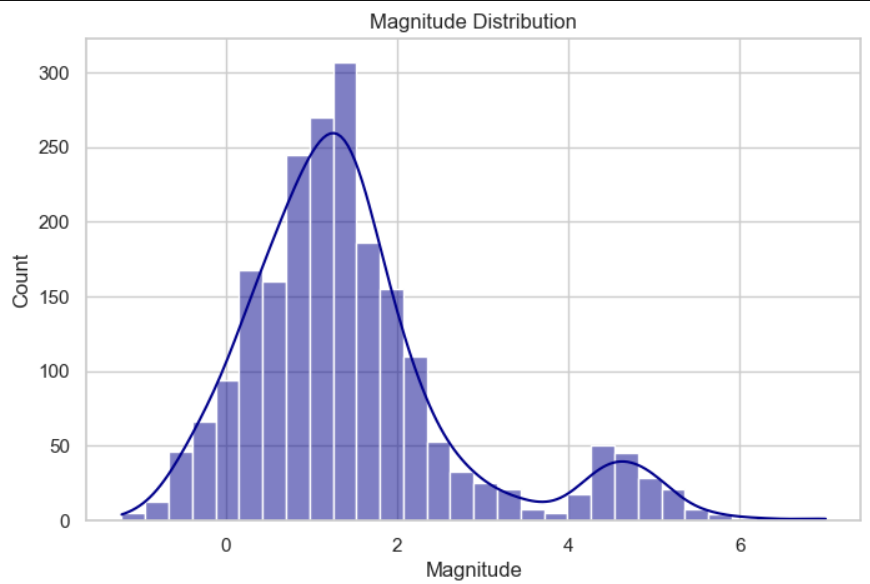
**8.Visualization**

* **Histogram for all numeric column**

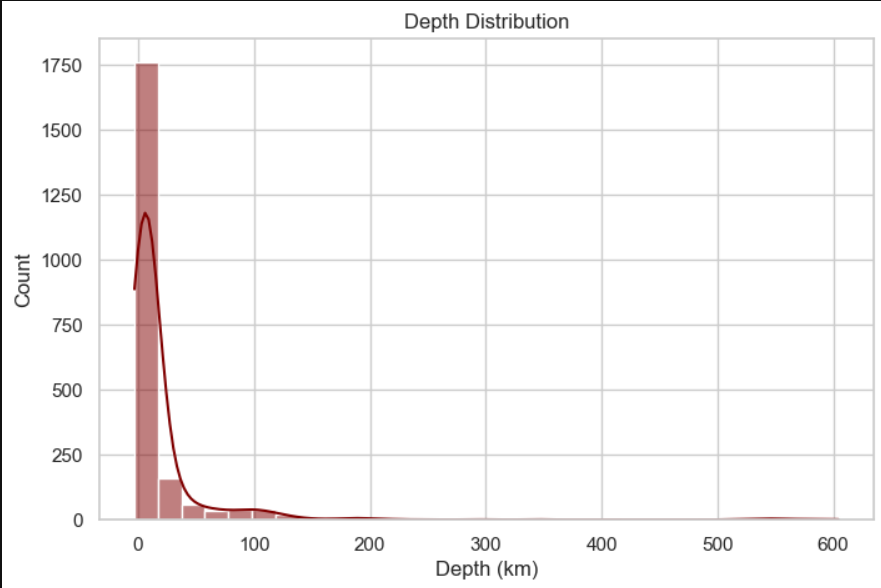
****

****

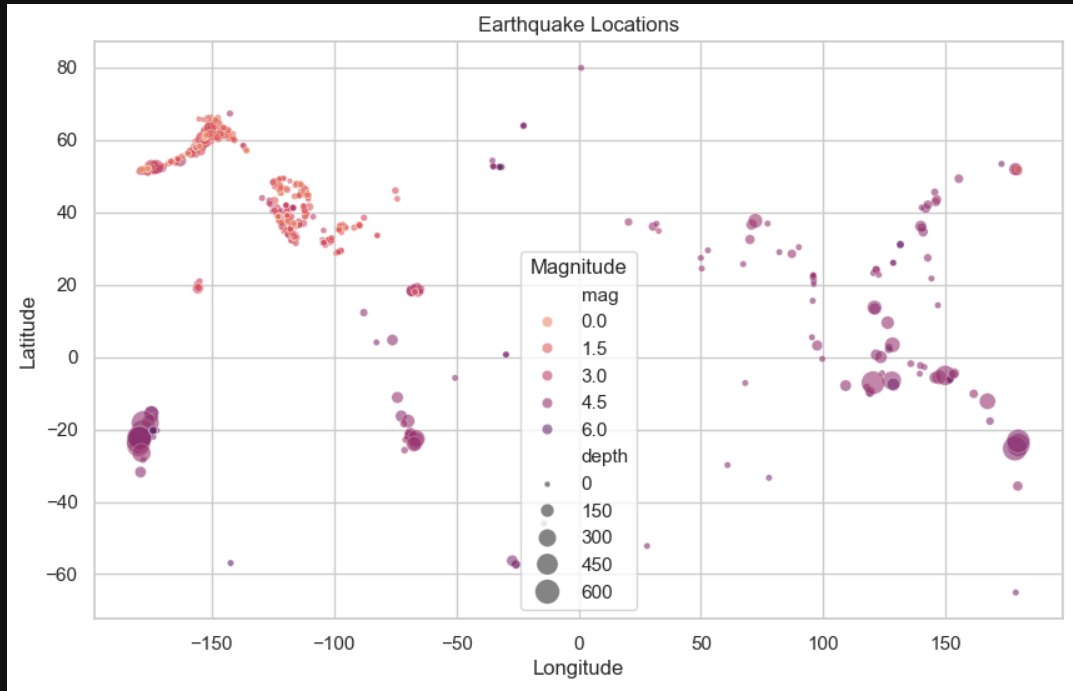
* **Magnitude Distribution (Histogram)**

****

* **Depth Distribution (Histogram)**

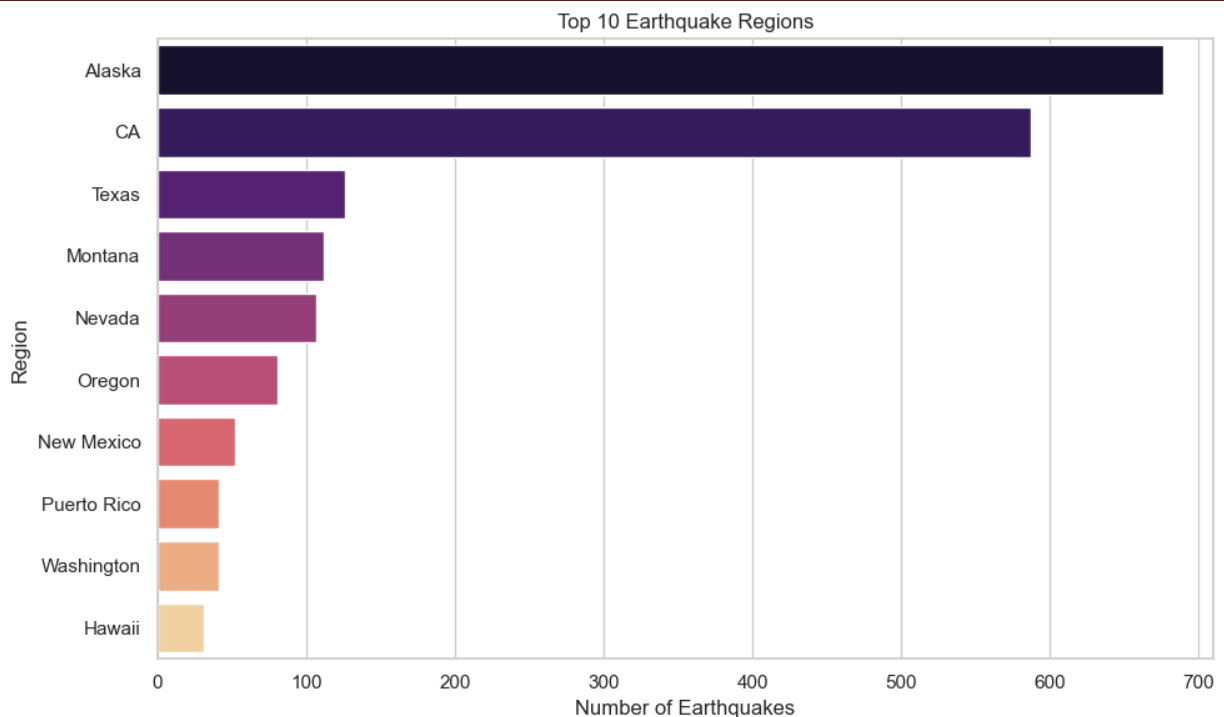
****

* **Earthquake Locations**

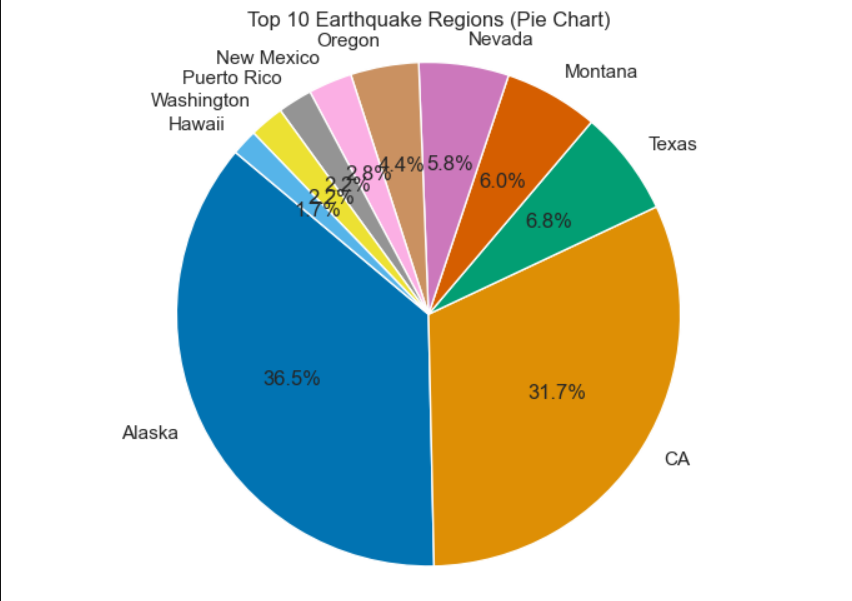
****

* **Top 10 Earthquake Region**

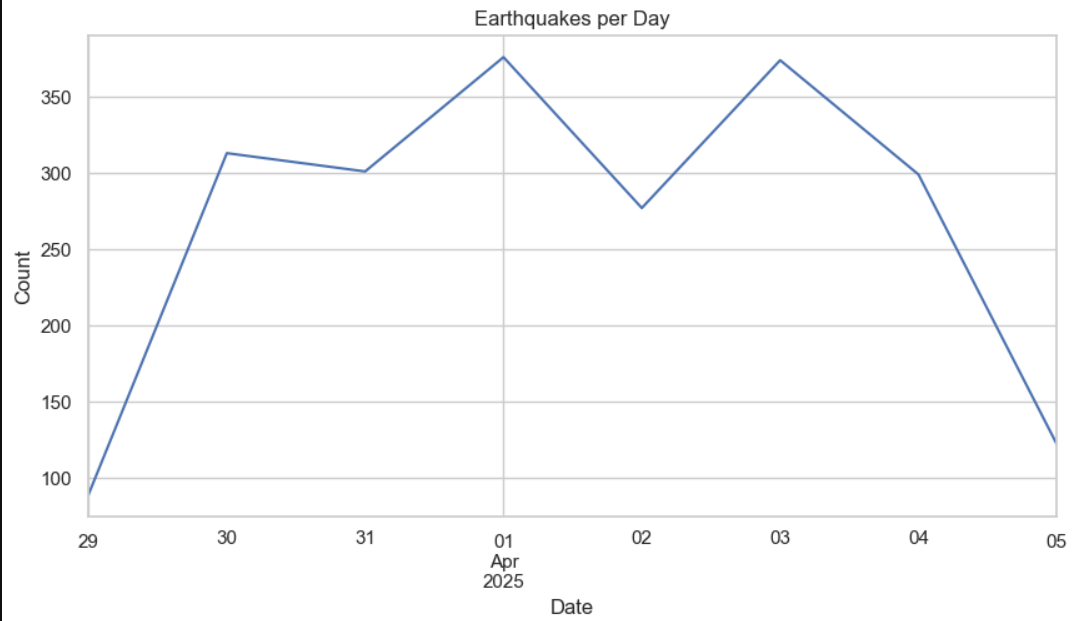
**Bar graph**

****

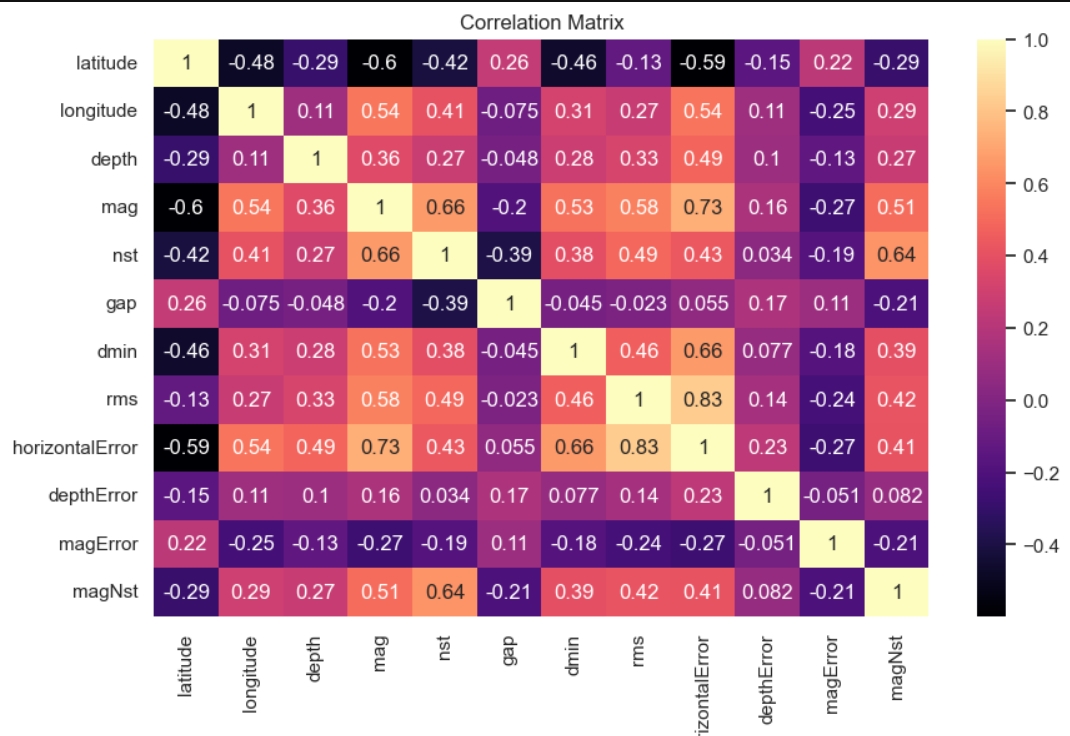
**Pie chart**

****

**Earthquake per day (line plot)**

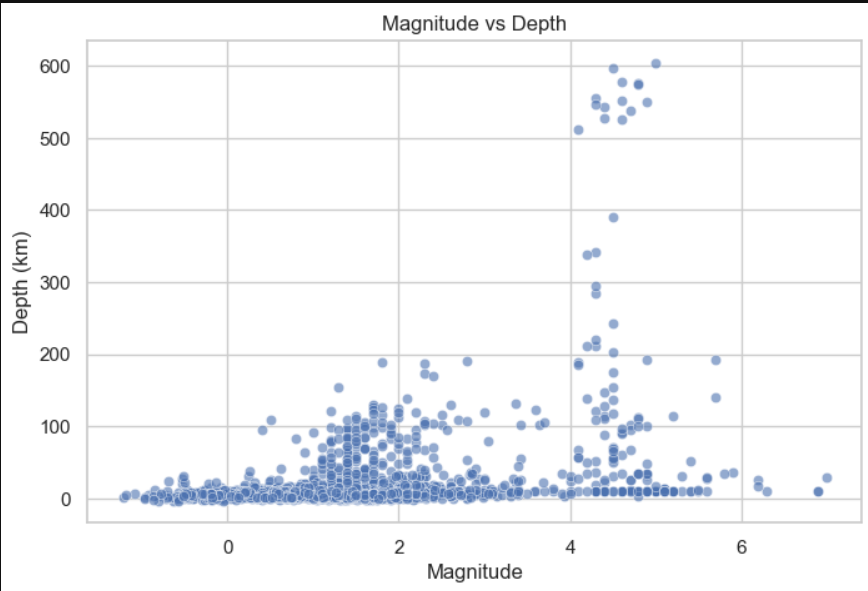
****

* **Correlation Matrix**

****

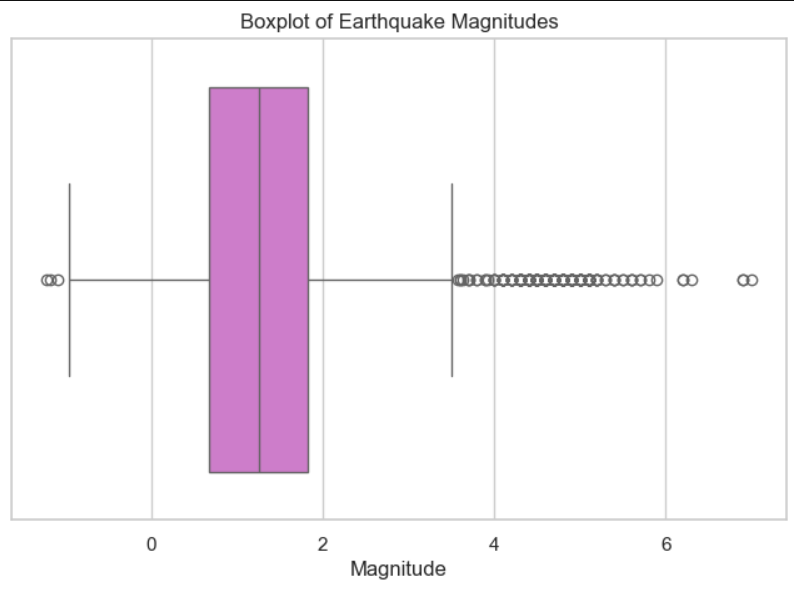
This matrix is calculating correlation between all numeric columns.

* **Magnitude vs Depth (scatter plot)**

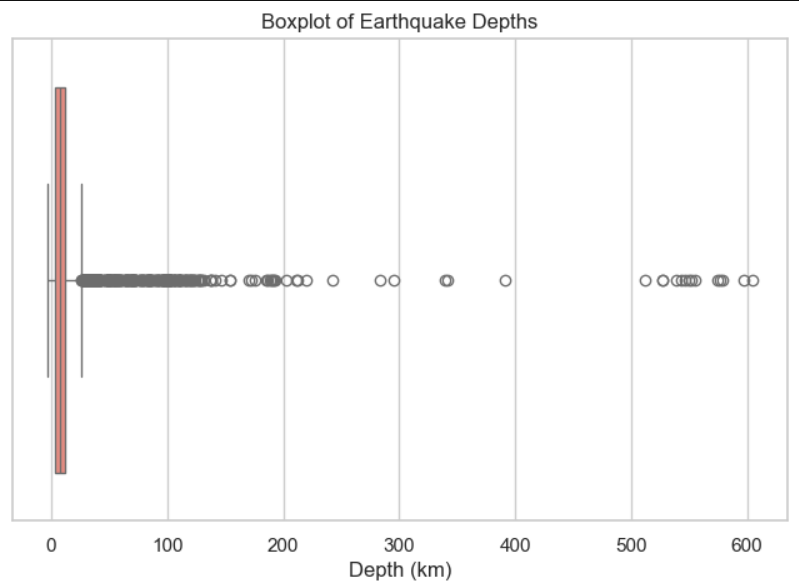
****

* **Box plot for outliers**

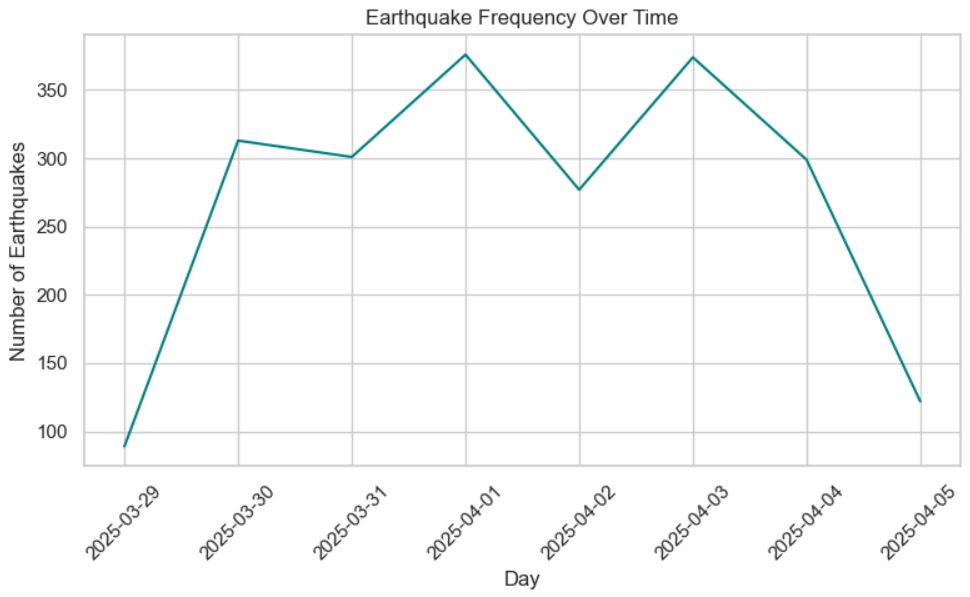
Magnitudes

****

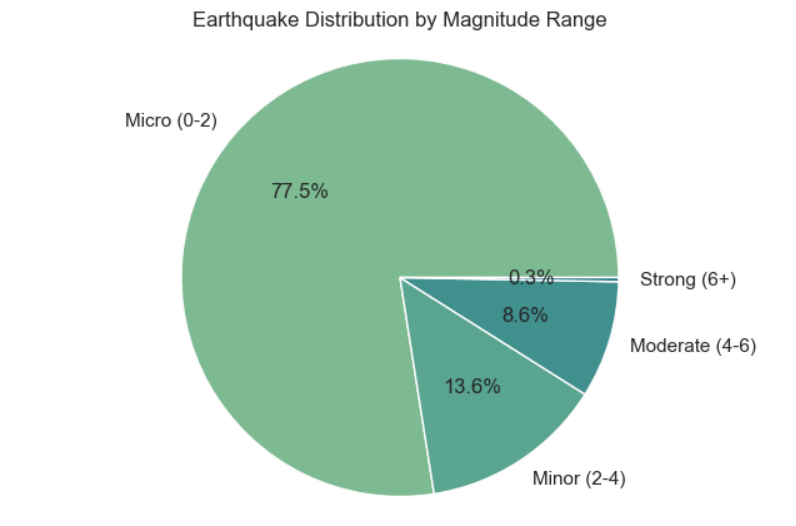
Depths

****

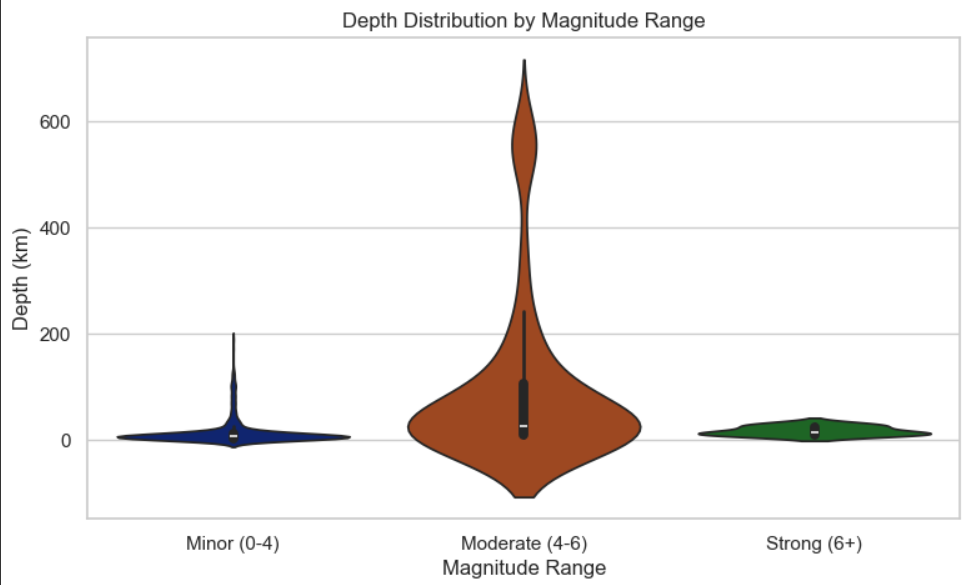
* **Earthquake frequency over time**

****

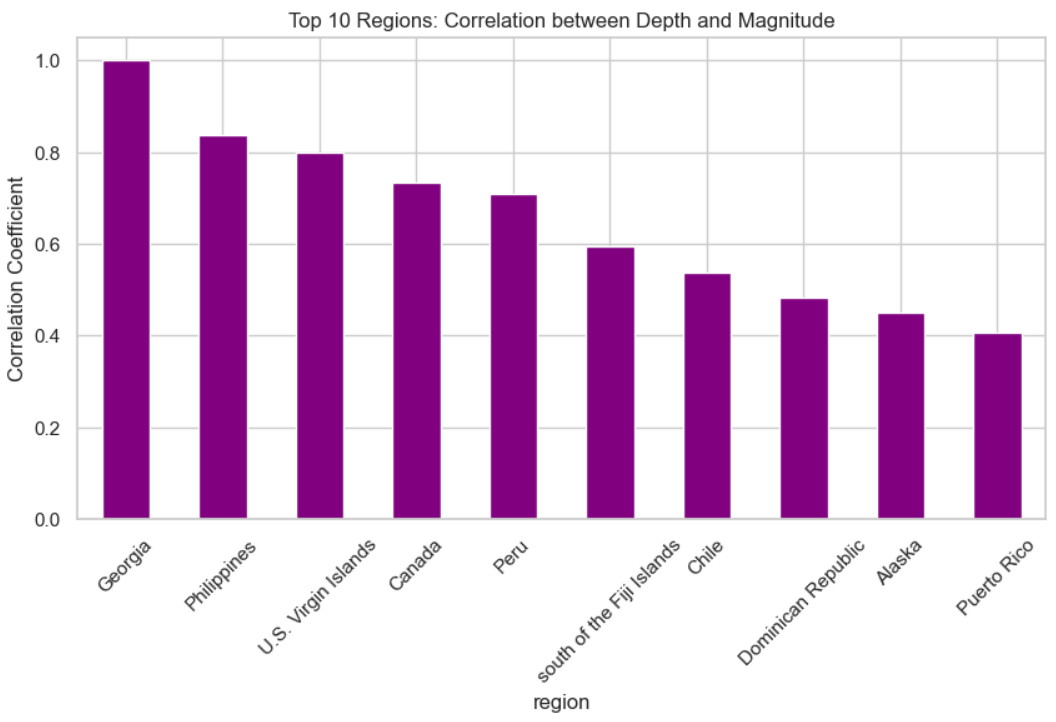
* **Earthquake Distribution by Magnitude Range**

****

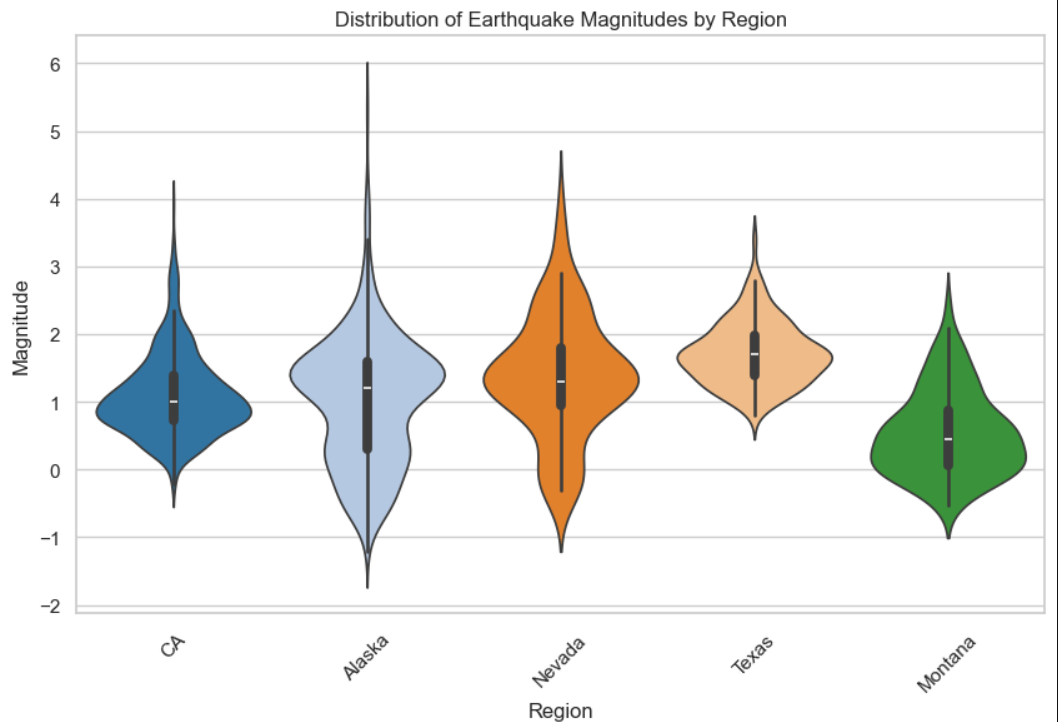
* **Depth Distribution by Magnitude Range**

****

* **Correlation between Depth and Magnitude**

****

* **Distribution of Earthquake Magnitude by region**

****

* 1. **Conclusion**
* The analysis revealed that **earthquakes are most frequent in geologically active regions** like Alaska, California, and around the Pacific Ring of Fire.
* **Most earthquakes are shallow and low in magnitude**, indicating minor seismic activities dominate globally.
* **Outliers** were found in both depth and magnitude, representing **potentially hazardous quakes** that need further attention.
* **Daily trends** showed a fairly consistent number of events each day, without any drastic spikes.
* The **magnitude and depth distributions** followed natural patterns, with more frequent events at lower ranges.
* **Correlation analysis** showed **no strong linear relationship** between features like magnitude, depth, latitude, and longitude.
* **Geospatial patterns** matched global tectonic boundaries, confirming the reliability of the dataset and analysis.
* The **EDA helped uncover patterns and anomalies** that form a strong base for further machine learning and prediction models.

**10.Future Scope**

* **Real-time Monitoring:** Integrate with APIs (e.g., USGS) to **track and visualize real-time earthquakes** on interactive maps.
* **Clustering and Pattern Detection:** Use **K-Means or DBSCAN** to find hidden clusters and tectonic groupings based on location and depth.
* **Machine Learning Predictions:** Build models to **predict magnitude or likelihood** of earthquakes based on historical trends, time, and location.
* **Time-Series Forecasting:** Explore models like **ARIMA, Prophet** to forecast future earthquake frequency or intensity.
* **Interactive Dashboards:** Develop **web-based dashboards** using Streamlit, Dash, or Power BI for live visualization and reporting.
* **Disaster Preparedness:** Share findings with geologists and disaster management teams for **risk assessment and early warning systems**.

**11. References**

* <https://earthquake.usgs.gov/earthquakes/feed/v1.0/csv.php>
* <https://earthquake.usgs.gov/>
* <https://seaborn.pydata.org/>
* <https://chatgpt.com/>

LINKEDIN LINK : <https://www.linkedin.com/posts/shiwaniag_datascience-eda-python-activity-7318312262375559169-20H_?utm_source=share&utm_medium=member_desktop&rcm=ACoAAEbwoOMBqUL12VXxiHR77_4OOS2kNy8wPOc>