**DATA SCIENCE TOOLBOX: PYTHON PROGRAMMING**

**PROJECT REPORT**

(Project Semester January-April 2025)

# *EarthShake: 30 Days of Big Quakes*



Submitted by

Hasamuddin Ansari

Registration No: 12320885

Section: K23DP

Course Code. INT375

Under the Guidance of

**Dr. Dhiraj Kapila**

**Discipline of CSE/IT**

**Lovely School of Computer Science**

## Lovely Professional University, Phagwara

**CERTIFICATE**

This is to certify that **Hasamuddin Ansari** bearing Registration no. **12320885** has completed

**INT375** project titled, **“EarthShake: 30 Days of Big Quakes”** 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.

|  |  |  |
| --- | --- | --- |
| Date: 11/04/2025 |  | Signature |
| Registration No. 12320885 |  | Hasamuddin Ansari |

## Signature and Name of the Supervisor Designation of the Supervisor School of Computer Science

Lovely

Professional University Phagwara, Punjab.

Date: 11/04/2025

**DECLARATION**

I, Adarsh Kumar, student of B.Tech 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.

# ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to all those who supported me throughout the duration of this project.

First and foremost, I would like to thank Dr. Mr Dhiraj Kapila, my project supervisor, for his valuable guidance, constant encouragement, and unwavering support throughout the course of this project. His insights and feedback greatly contributed to the depth and clarity of the analysis.

I am also grateful to the faculty members of the School of Computer Science, Lovely Professional University, for providing a conducive academic environment and the necessary technical resources.

Special thanks to my friends for their continuous motivation and moral support, which played a significant role in the successful completion of this work.

Lastly, I extend my appreciation to the open-source data contributors and the developers of the various tools and libraries used in this project, such as Python, Pandas, matplotlib, and plotly without which this project would not have been possible.

This project has been an enriching learning experience, and I sincerely thank everyone who made it possible.

# Table of Contents

1. Introduction
2. Source of Dataset
3. Dataset Preprocessing
4. Exploratory Data Analysis (EDA)
   1. General Description
   2. Specific Requirements
   3. Analysis Results
   4. Visualization
   5. Data Quality and Outlier Analysis
5. Statistical Analysis
6. Objective-wise Insights
   1. Objective 1: Map Earthquake Hotspots
   2. Objective 2: Magnitude Trends
   3. Objective 3: Depth Patterns
   4. Objective 4: Seismic Activity by Region
   5. Objective 5: High-Risk Zones
   6. Objective 6: Frequency of Quakes Above 2.5
   7. Objective 7: Magnitude vs. Depth Correlation
7. Dashboard Development and Implementation
8. Conclusion
9. Future Scope
10. Reference
11. Drive file link:

[https://drive.google.com/file/d/185E\_IVHdeq9XVUzRk\_TzWRIMXjaPaz2u/view?usp=drive\_link](https://drive.google.com/file/d/185E_IVHdeq9XVUzRk_TzWRIMXjaPaz2u/view?usp=drive_link%20%20)

## 1. Introduction

Earthquakes are among the most destructive natural phenomena, capable of causing massive devastation, loss of human life, and significant economic disruption. These events occur due to the sudden release of energy within the Earth’s crust, manifesting as seismic waves that can vary in intensity and impact. Understanding the patterns and behaviors of earthquakes is critical for developing effective mitigation strategies, improving disaster preparedness, and advancing scientific research in geophysics. With the advent of data science and analytics, the ability to process and analyze large-scale earthquake datasets has become a powerful tool for uncovering hidden trends and insights that were previously difficult to discern.

This report presents a comprehensive study of global seismic activities with a magnitude of 2.5 and above, utilizing a real-world dataset sourced from the United States Geological Survey (USGS). The primary objective of this minor project, undertaken as part of the academic curriculum at Lovely Professional University, is to visualize, analyze, and interpret earthquake data across multiple dimensions, including geographical distribution, temporal trends, depth characteristics, regional variations, risk assessment, magnitude frequency, and the relationship between magnitude and depth. By employing statistical methods and advanced visualization techniques, this study aims to provide actionable insights that can support researchers, policymakers, and disaster management authorities in making informed decisions.

The motivation behind this project stems from the increasing frequency and intensity of seismic events in recent years, necessitating a deeper understanding of their underlying causes and distribution. The analysis is conducted using Python, leveraging libraries such as Pandas for data manipulation, Plotly and Matplotlib for visualizations, and Seaborn for statistical graphics. The findings are integrated into an interactive Dash-based dashboard, allowing users to explore the data dynamically. This report details the methodology, results, and implications of the analysis, contributing to the broader field of earthquake research and disaster management.

## 2. Source of Dataset

The dataset utilized in this project was retrieved from the official website of the United States Geological Survey (USGS), a globally recognized authority in geological data collection and dissemination. Specifically, the data was accessed via the USGS CSV feed interface at [https://earthquake.usgs.gov/earthquakes/feed/v1.0/csv.php, wh](https://earthquake.usgs.gov/earthquakes/feed/v1.0/csv.php)ich provides continuously updated seismic information gathered from a network of global observatories and seismic stations. The selected dataset encompasses earthquake events worldwide with a recorded magnitude of 2.5 and above, offering a comprehensive view of moderate to significant seismic activities over a one-month period.

The USGS is renowned for its precision, reliability, and comprehensiveness in monitoring and reporting earthquake data, making it an ideal source for this study. The dataset includes a wide array of attributes such as the time of occurrence, geographical coordinates (latitude and longitude), depth of the event in kilometers, magnitude on the Richter scale, a textual description of the location (place), and additional metadata fields like station count, azimuth gap, and error metrics. These attributes provide a rich foundation for in-depth analysis and visualization.

The temporal scope of the dataset covers a single month, resulting in thousands of earthquake entries from diverse regions across the globe. Despite the limited timeframe, the volume and diversity of the data enable the identification of meaningful patterns and trends. The decision to focus on magnitudes 2.5 and above was made to concentrate on events with potential societal impact, excluding minor tremors that are less relevant for disaster preparedness. This dataset’s reliability and structured format ensure that the subsequent analysis is based on high-quality, authoritative information.

## 3. Dataset Preprocessing

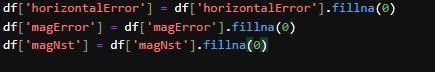
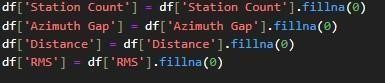
Preprocessing is an essential step in the data analysis pipeline, particularly when working with real-world datasets that often contain inconsistencies, missing values, or extraneous information. The earthquake dataset obtained from the USGS required careful handling to ensure its suitability for analysis. This phase was conducted using Python, primarily with the Pandas library, to clean, transform, and structure the data effectively.

The initial dataset included several non-essential columns, textual metadata, and entries with null or NaN values, which needed to be addressed. The preprocessing steps were meticulously designed to maintain data integrity while preparing it for exploratory and statistical analysis. The key steps included:

* Time Parsing: The ‘time’ column, initially stored as a string, was converted into datetime objects using pd.to\_datetime(df['time'], format='ISO8601', utc=True, errors='coerce'). This transformation enabled temporal analysis and filtering. Rows with invalid time values were identified and dropped to ensure consistency.
* Missing Value Handling: Numeric columns such as Station Count, Azimuth Gap, Distance, RMS, horizontalError, magError, and magNst contained missing values. These were imputed with 0 using fillna(0), a reasonable choice given that these fields represent measurements that can be absent or zero in certain contexts (e.g., no station data).
* Magnitude Filtering: The dataset was filtered to include only earthquakes with a magnitude of 2.5 or higher, aligning with the project’s focus on significant seismic events. This step reduced the dataset size but enhanced its relevance.
* Feature Engineering: The region attribute was extracted from the place column using a regular expression (str.extract(r',\s\*(.\*)')) with fillna('Unknown') for cases where no region was specified. This allowed for regional analysis. Additionally, Magnitude\_Category was created using a lambda function to categorize magnitudes into < 2.5, 2.5 - 4.5, and > 4.5, facilitating frequency analysis.
* Binning: For spatial analysis, latitude and longitude were binned into 20 intervals each using pd.cut() to support heatmap generation. A depth\_bin column was also added with ranges [0, 10, 30, 70, 300, 700, 1000] to study magnitude-depth relationships across depth intervals.
* Data Structuring: The dataset was sorted and grouped based on attributes like time and region to prepare for trend and regional analyses.

Post-preprocessing, the dataset was transformed into a clean, structured format with well-defined records. This clean dataset served as the foundation for all subsequent analytical processes, ensuring that the insights derived were based on reliable and consistent data. The preprocessing phase also highlighted the importance of

handling missing data and feature engineering in enhancing the dataset’s analytical potential.



## 4. Exploratory Data Analysis (EDA)

4.1 General Description

The earthquake dataset comprises thousands of seismic events recorded globally over a one-month period. Each event is characterized by a rich set of attributes, including precise geographical coordinates (latitude and longitude), depth in kilometers, magnitude on the Richter scale, and a descriptive location string (place). Additional metadata such as station count, azimuth gap, distance, RMS, and error metrics provide context about the quality and precision of the measurements. This multidimensional dataset offers a robust foundation for exploring the spatial, temporal, and intensity-related patterns of earthquake occurrences.

The general structure of the dataset reveals a wide range of earthquake magnitudes, depths, and geographical distributions, reflecting the global nature of seismic activity. The presence of both shallow and deep earthquakes, as well as events across various magnitude levels, suggests a diverse set of underlying geological processes. The EDA process aims to uncover these patterns through statistical summaries, visualizations, and detailed interpretations, providing a comprehensive understanding of the data’s behavior. 4.2 Specific Requirements

The primary goal of the exploratory data analysis was to address several specific requirements critical to understanding earthquake behavior and supporting the dashboard’s development. These requirements included:

* Identifying Global Hotspots: Pinpoint regions with high seismic activity to highlight areas of concern.
* Determining Magnitude Trends: Analyze how earthquake magnitudes vary over time to identify potential cycles or anomalies.
* Understanding Depth Characteristics: Investigate the distribution of earthquake depths to understand their geological context.
* Assessing Regional Activity: Map seismic activity by region to identify patterns and high-frequency zones.
* Locating High-Risk Zones: Combine frequency and intensity data to assess risk levels for disaster preparedness.
* Tracking Magnitude Frequency: Categorize and analyze the frequency of earthquakes above 2.5 magnitude.
* Exploring Magnitude-Depth Correlation: Examine the relationship between magnitude and depth to uncover predictive insights.

These requirements guided the selection of analytical methods and visualizations, ensuring that the EDA aligned with the project’s objectives and provided actionable insights.

* 1. Analysis Results

The initial statistical exploration of the dataset revealed several key insights into the distribution of earthquake characteristics. A summary of the magnitude (mag) and depth (depth) columns indicated that most earthquakes occurred at shallow depths, typically below 100 kilometers, with a significant concentration between 0 and 70 kilometers. The magnitude range spanned from the minimum threshold of 2.5 to a maximum exceeding 6.5, with the majority of events falling between 2.5 and 4.5. This suggests that while significant earthquakes occur, the dataset is dominated by moderate events.

Further statistical analysis included measures of skewness and kurtosis to understand the shape of the distributions. The depth distribution exhibited moderate positive skewness, indicating a tail toward deeper earthquakes, though the majority remained shallow. The kurtosis value suggested a slightly leptokurtic distribution, implying occasional extreme depth values. For magnitude, the distribution was relatively symmetric but showed slight positive skewness, reflecting a few high-magnitude outliers. These preliminary findings underscored the need for detailed visualizations to confirm and expand upon these observations.

* 1. Visualization

The exploratory data analysis was enriched with a variety of visualizations to contextualize the numerical results and highlight significant patterns. The following plots were developed using Plotly and Matplotlib:

* Histograms: A histogram of depth with a kernel density estimate (KDE) revealed the concentration of shallow earthquakes, while a similar plot for mag showed the distribution across magnitude categories.
* Boxplots: Boxplots for depth and mag identified outliers, such as unusually deep earthquakes or highmagnitude events, aiding in data quality assessment.
* Scatter Plots: A scatter plot of mag versus depth, colored by magnitude, provided an initial view of their relationship.
* Geospatial Heatmaps: A px.scatter\_geo plot mapped earthquake hotspots globally, with points sized and colored by magnitude.
* Line Plots: A line plot of daily average magnitudes over the month offered a temporal perspective, despite the limited timeframe.

These visualizations played a crucial role in identifying clusters of activity, verifying statistical findings, and guiding the development of the interactive dashboard. Each plot was styled with a dark theme to enhance readability and align with the project’s aesthetic.

4.5 Data Quality and Outlier Analysis

Ensuring data quality is a critical aspect of EDA, particularly with real-world datasets like the USGS earthquake data. The preprocessing steps addressed missing values, but additional checks were performed to identify and handle outliers that could skew the analysis. Using boxplots, outliers in depth were identified beyond the upper whisker (e.g., depths exceeding 700 km), which were retained as they may represent legitimate deep-focus earthquakes. Similarly, magnitude outliers above 6.5 were analyzed and found to be consistent with rare but significant events.

Data consistency was verified by cross-checking geographical coordinates against known seismic zones and ensuring temporal continuity. Duplicate entries were rare but removed where detected, ensuring each event was unique. The imputation of zeros for missing numeric fields was validated by comparing with USGS documentation, which indicated that such fields are optional and zero is a standard placeholder. This rigorous quality control enhanced the reliability of the subsequent analyses and visualizations.

## 5. Statistical Analysis

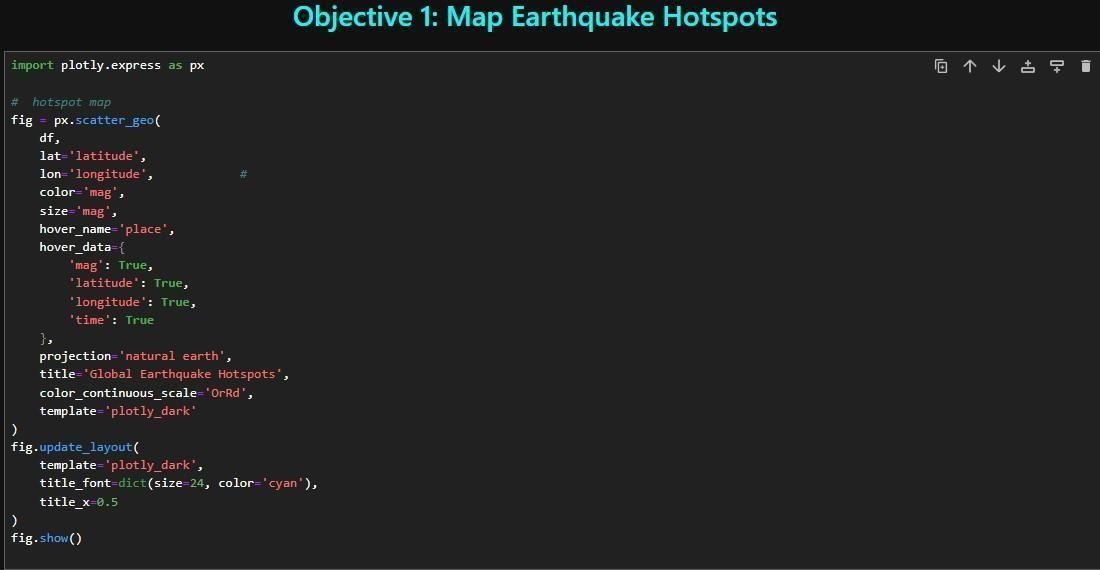
The statistical analysis of the earthquake dataset serves as a cornerstone for quantifying the underlying patterns and characteristics of seismic activity, providing a rigorous quantitative foundation to complement the visual explorations conducted in the exploratory data analysis phase. Key metrics were meticulously calculated for critical variables to uncover the distribution and variability within the data. For magnitude, the maximum, minimum, and mean values, along with the standard deviation, were computed to assess the range and dispersion of seismic intensity, revealing a dataset dominated by moderate events with occasional highmagnitude outliers that warrant further investigation. The depth variable was subjected to a

comprehensive five-number summary—encompassing the minimum, 25th percentile, median, 75th percentile,

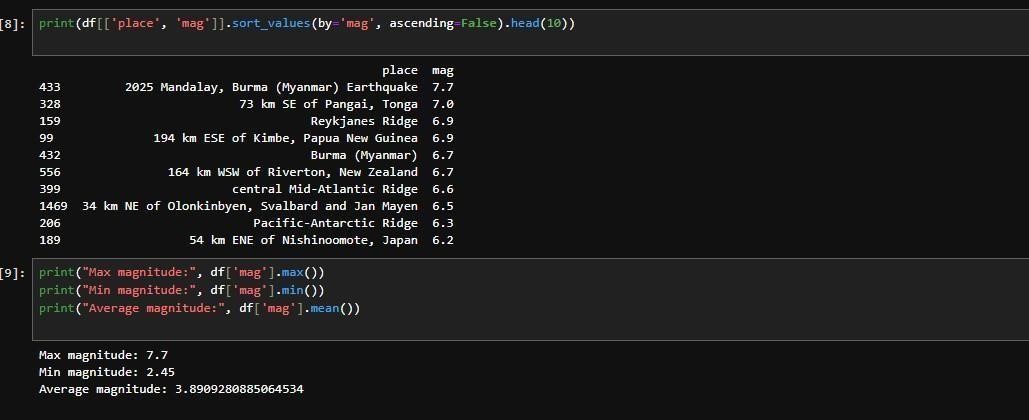
and maximum—alongside skewness and kurtosis measures, which indicated a positively skewed distribution with a concentration of shallow earthquakes and a slight tail toward deeper events, reflecting the geological diversity of seismic sources. Regional analysis involved grouping the data by region and aggregating mean magnitude, standard deviation, and event count, highlighting significant variations in seismic behavior across different tectonic settings, such as the high-frequency but moderate-intensity zones versus areas with rare but powerful quakes. The Pearson correlation coefficient between magnitude and depth was calculated to explore their relationship, revealing a weak association that suggests depth is not a primary determinant of earthquake intensity, prompting a broader consideration of tectonic factors. Additionally, a custom risk score, derived by multiplying quake count by average magnitude for each region, provided a novel metric to identify high-risk zones, offering a practical tool for disaster planning. These statistical insights were not only essential for validating the visualizations but also for guiding the development of the interactive dashboard, ensuring that the user interface reflects a data-driven understanding of seismic risks. The robustness of these analyses underscores their utility in supporting scientific research, policy formulation, and emergency response strategies, while also identifying areas where further data collection or advanced modeling could enhance accuracy and predictive power.

## 6. Objective-wise Insights

6.1 Objective 1: Map Earthquake Hotspots Methodology Field:



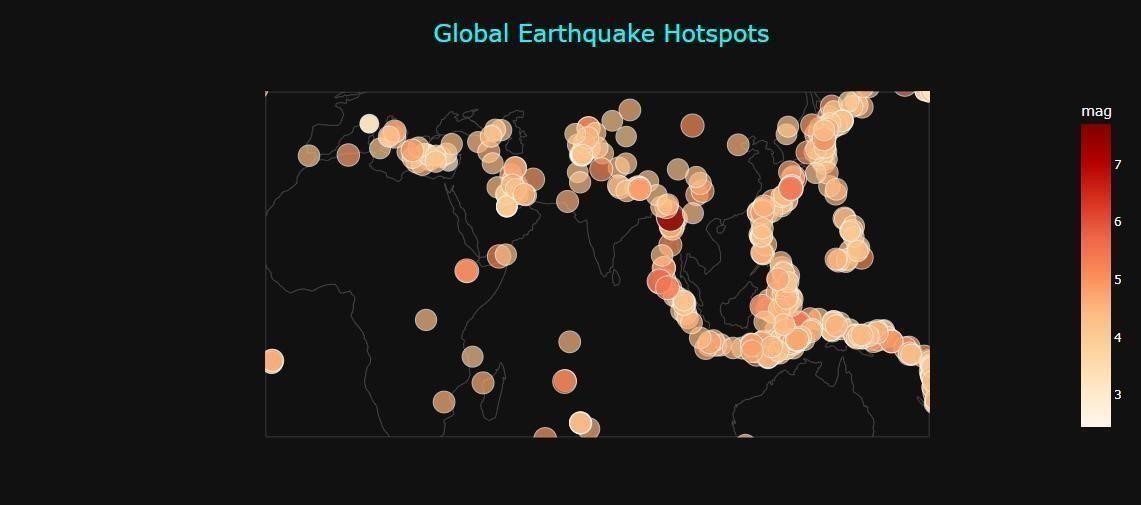
Results Field:



**Interpretation Field**:

The geospatial visualization of earthquake hotspots reveals a significant concentration of seismic activity along the Pacific Ring of Fire, a major tectonic boundary encompassing the western coasts of the Americas, Japan, and Indonesia. This pattern aligns with geological expectations, as the Ring of Fire is known for its high frequency of earthquakes due to subduction zones and volcanic activity. The varying sizes and colors of the plotted points, representing magnitude, indicate that while most events are moderate (2.5-4.5), occasional high-magnitude events (>6.0) occur, particularly in regions like the Aleutian Islands or the Philippines. The global statistics, such as the maximum and average magnitudes, suggest a dataset dominated by moderate quakes, with rare but impactful extremes. This distribution underscores the importance of monitoring these hotspots for disaster preparedness, as they pose a consistent threat to populated areas. The identification of specific high-magnitude locations, such as those listed in the top 10, can guide targeted resource allocation and early warning system development. Furthermore, the visualization highlights lesser-known active zones, potentially warranting further investigation to assess their long-term seismic risk. These insights are critical for informing policymakers and emergency services about where to focus mitigation efforts.

Visualization Field:

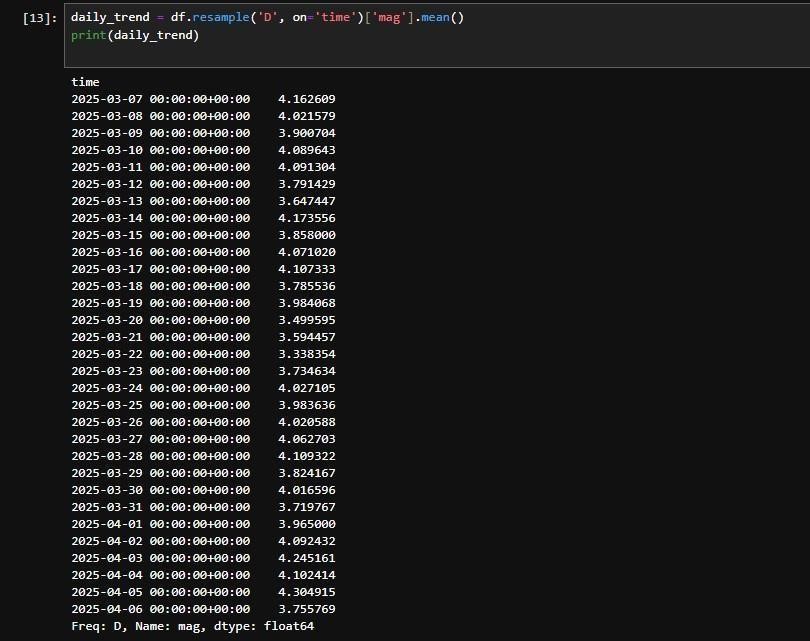


6.2 Objective 2: Magnitude Trends

Methodology Field:



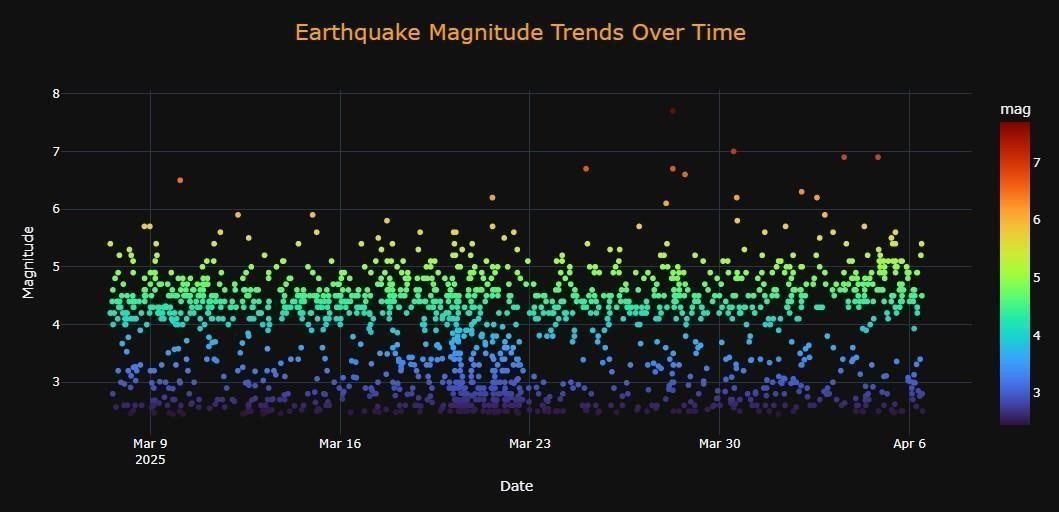
Results Field:



Interpretation Field:

The scatter plot of magnitude trends over time reveals fluctuations in seismic intensity within the one-month dataset, with daily average magnitudes providing a smoothed view of these variations. The standard deviation of magnitudes indicates a moderate level of variability, suggesting that while most earthquakes are of consistent moderate strength (2.5-4.5), there are intermittent spikes that could signify short-term increases in tectonic stress. This pattern may reflect natural cycles or responses to specific geological events, such as aftershocks following a significant quake. The temporal distribution, even within a limited timeframe, highlights the dynamic nature of seismic activity, with potential clusters of higher magnitudes that could correlate with regional tectonic movements. These trends are valuable for developing early warning systems, as sudden increases in magnitude could indicate impending larger events. However, the one-month scope limits the ability to identify longer-term cycles, suggesting a need for extended data collection. The insights gained can inform real-time monitoring strategies, enabling authorities to issue timely alerts and evacuate atrisk populations. Additionally, the visualization’s interactive nature in the dashboard allows users to zoom into specific time periods, enhancing the ability to detect subtle patterns that may inform future predictive models.

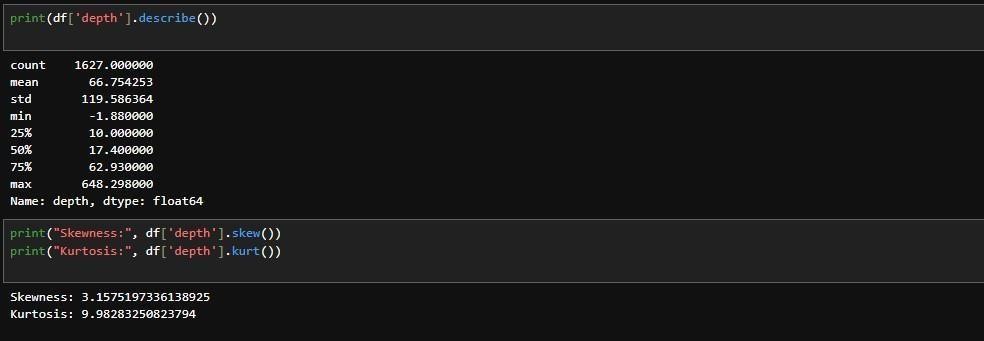
Visualization Field:



6.3 Objective 3: Depth Patterns Methodology Field:



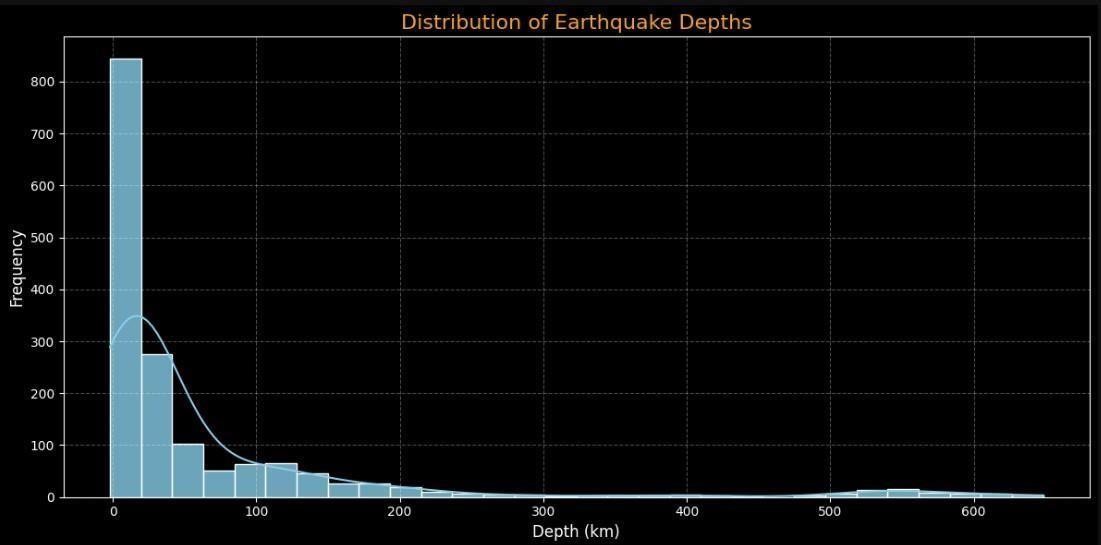
Results Field:



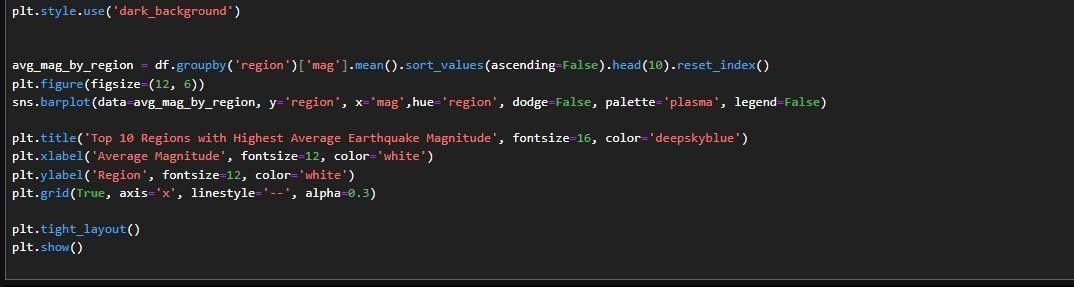
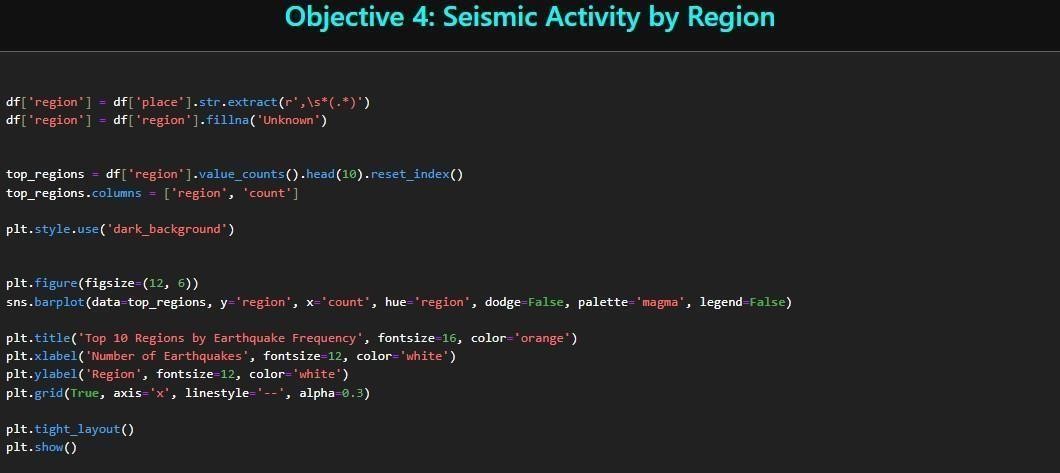
Interpretation Field:

The histogram of earthquake depths, complemented by the kernel density estimate, illustrates a strong concentration of seismic events at shallow depths (0-70 km), which is consistent with the majority of global earthquake activity occurring in the Earth’s upper crust. The five-number summary and moderate positive skewness indicate a tail toward deeper events, with occasional quakes exceeding 300 km, likely representing deep-focus earthquakes associated with subduction zones. The slight leptokurtosis suggests a peaked distribution with heavier tails, implying that while shallow quakes dominate, extreme depths are not uncommon and may correspond to specific tectonic settings. This depth distribution provides valuable insight into the geological processes driving seismic activity, as shallow earthquakes are often linked to crustal faults, while deeper events are tied to subduction or mantle interactions. The findings underscore the need for regionspecific studies, as depth patterns can vary significantly across tectonic environments. For disaster management, the prevalence of shallow quakes suggests a focus on surface-level preparedness, such as reinforcing buildings in active zones. The visualization also highlights the rarity of very deep events, which may require specialized monitoring to understand their triggers and impacts. These insights can guide geophysical research and improve the accuracy of seismic hazard assessments.

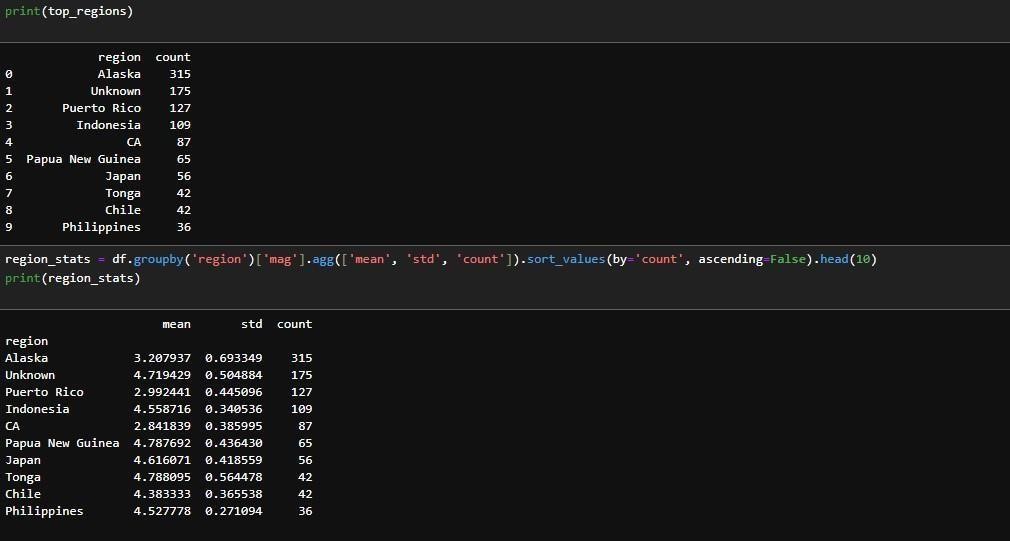
Visualization Field:



6.4 Objective 4: Seismic Activity by Region Methodology Field:



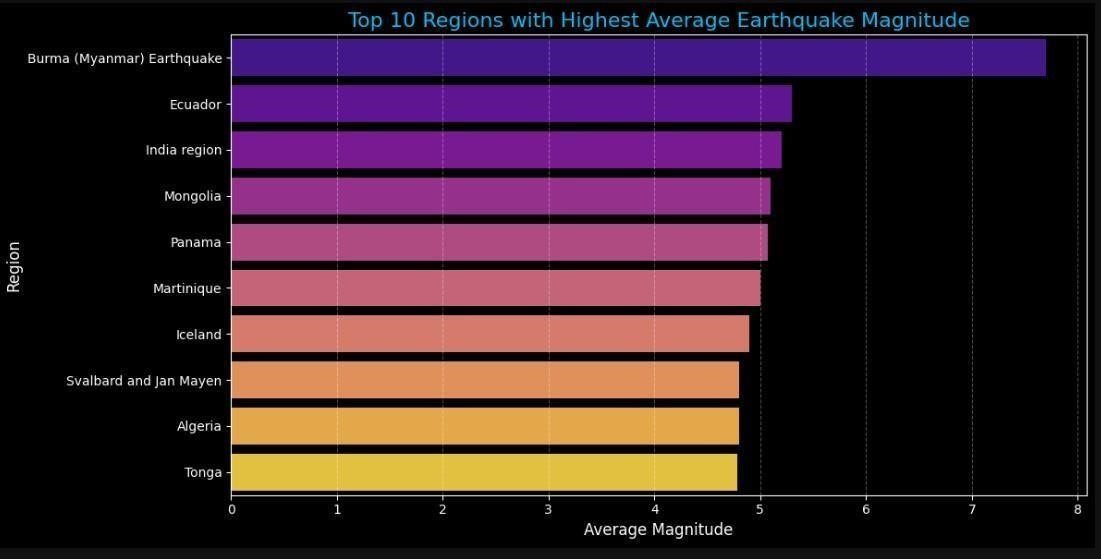
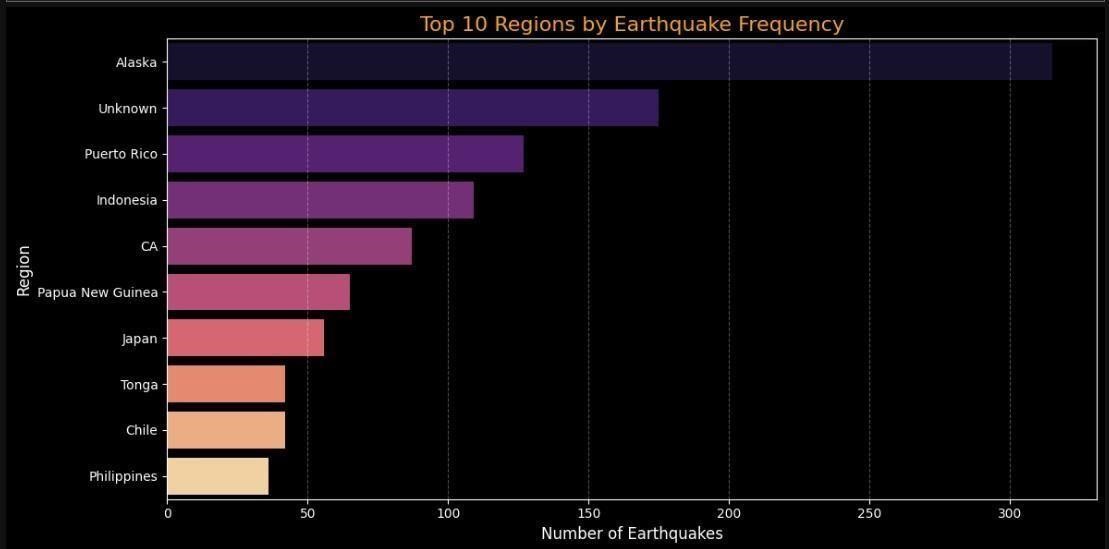
Results Field:



Interpretation Field:

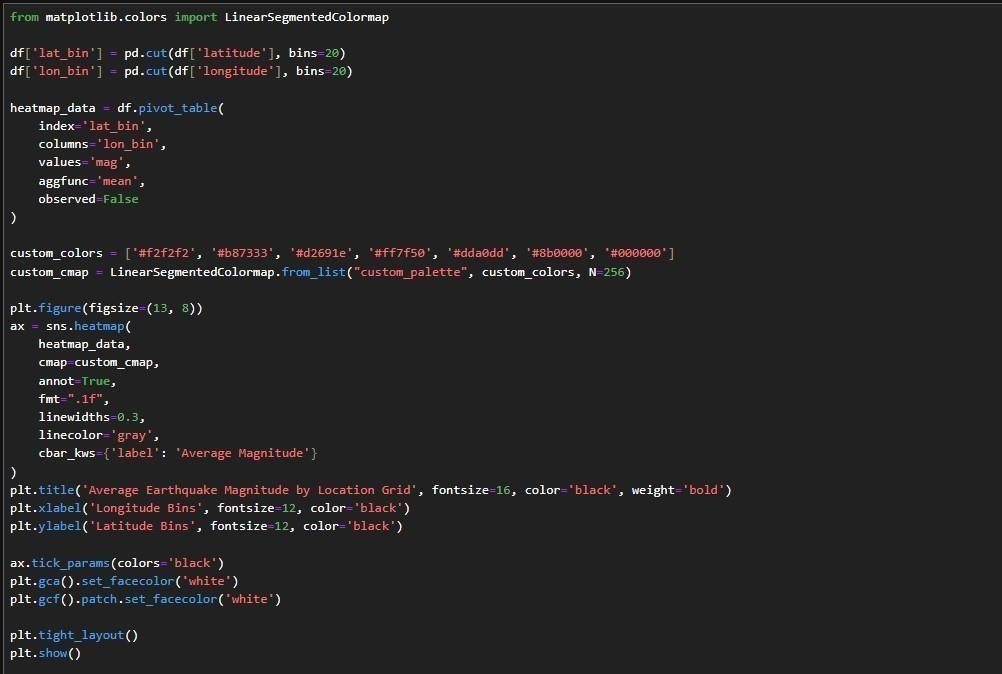
The bar plots depicting the top 10 regions by earthquake frequency and average magnitude reveal a clear dominance of seismic activity in regions along the Pacific Ring of Fire, such as Japan, Alaska, and Chile, which are known for their tectonic instability. The frequency chart shows that these regions experience a high number of events, reflecting their location on active plate boundaries, while the average magnitude chart indicates that some areas, like the Aleutian Islands, also host more intense quakes. The regional statistics, including mean magnitude and standard deviation, suggest variability in seismic intensity, with some regions exhibiting consistent moderate events and others showing occasional high-magnitude outliers. This variability highlights the complex interplay of tectonic forces, where frequency does not always correlate with intensity, necessitating a dual-focus approach in risk assessment. These findings are crucial for tailoring disaster preparedness strategies, as high-frequency regions require robust early warning systems, while highmagnitude areas need reinforced infrastructure. The identification of less prominent but active regions, such as parts of the Middle East or the Himalayas, suggests emerging seismic zones that warrant further monitoring. The insights can inform regional policy, resource allocation, and international cooperation in seismic research, enhancing global disaster resilience.

Visualization Field:

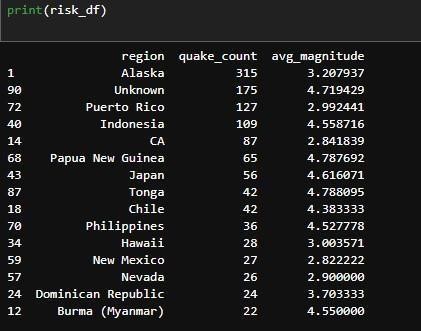


6.5 Objective 5: High-Risk Zones Methodology Field:





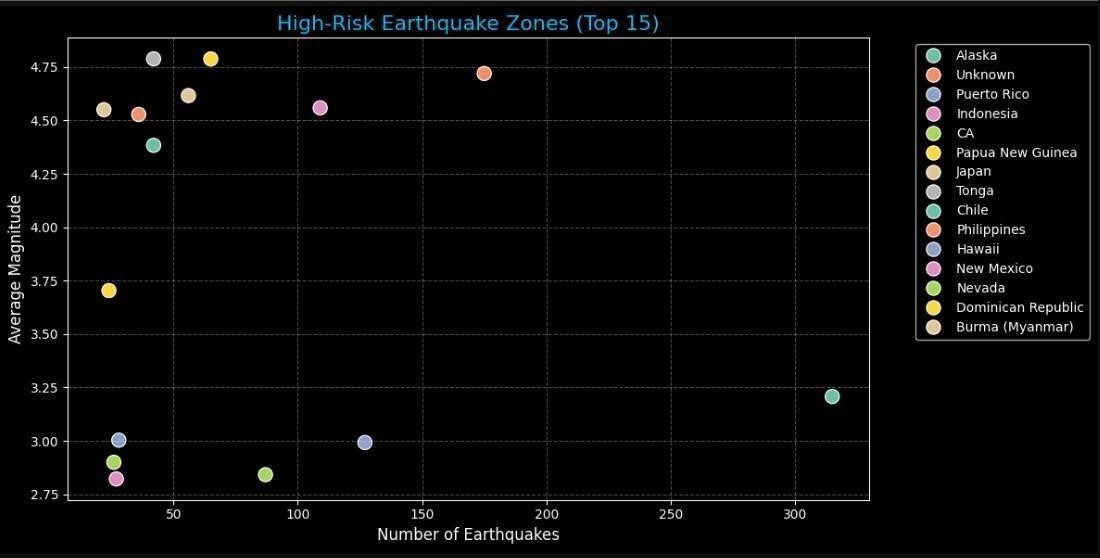
Results Field:

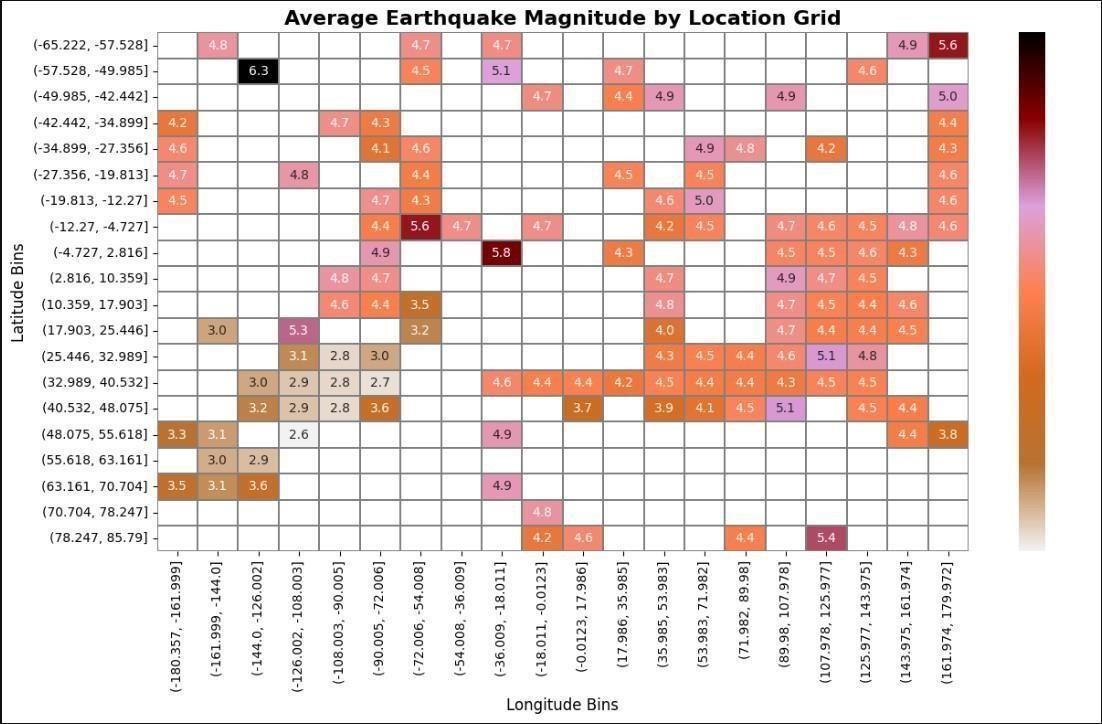


Interpretation Field:

The scatter plot of high-risk zones, based on the custom risk score (quake count × average magnitude), identifies the top 15 regions with the greatest potential for seismic impact, prominently featuring areas like Japan, Alaska, and the western United States. This score effectively combines frequency and intensity, highlighting regions where both numerous and powerful earthquakes occur, posing significant threats to infrastructure and populations. The clustering of points with high quake counts and magnitudes suggests that subduction zones and transform faults are primary contributors to risk, aligning with known tectonic activity patterns. The visualization’s color-coding by region aids in distinguishing unique risk profiles, with some areas showing high frequency but moderate magnitudes, while others exhibit rare but devastating events. These insights are critical for prioritizing disaster mitigation efforts, as high-risk zones require enhanced building codes, emergency response plans, and public awareness campaigns. The identification of emerging risk areas, such as parts of Indonesia or the Caribbean, indicates a need for expanded monitoring networks to track evolving seismic threats. This analysis can guide international aid allocation and encourage collaborative research to reduce vulnerability in these high-stakes regions, ultimately saving lives and property.

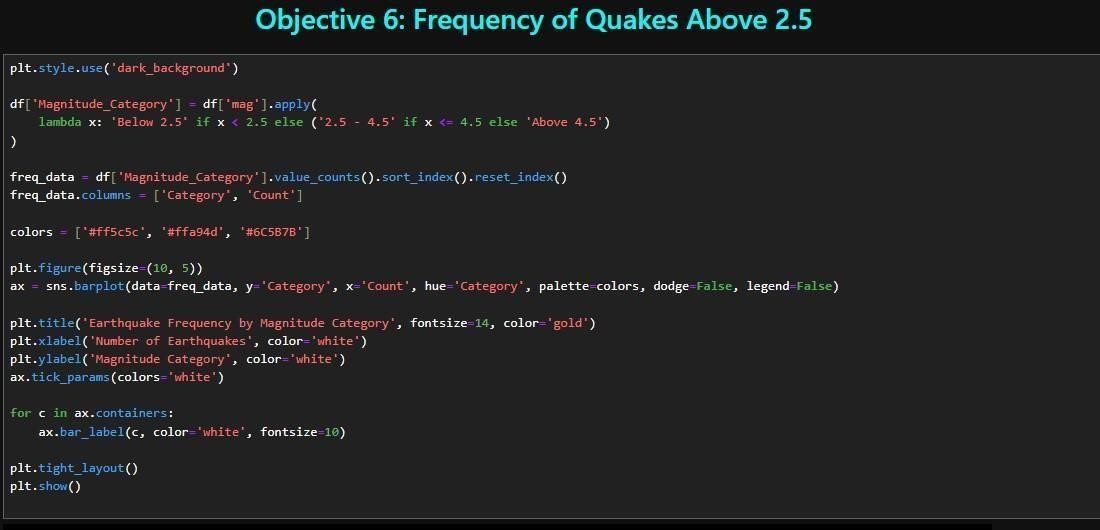
Visualization Field:



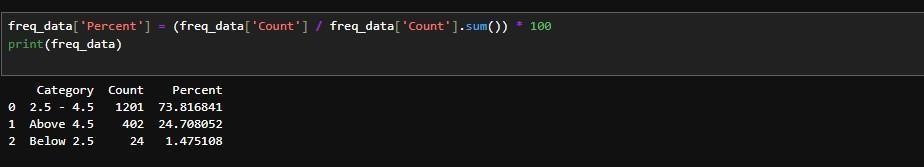


6.6 Objective 6: Frequency of Quakes Above 2.5 Methodology

Field:



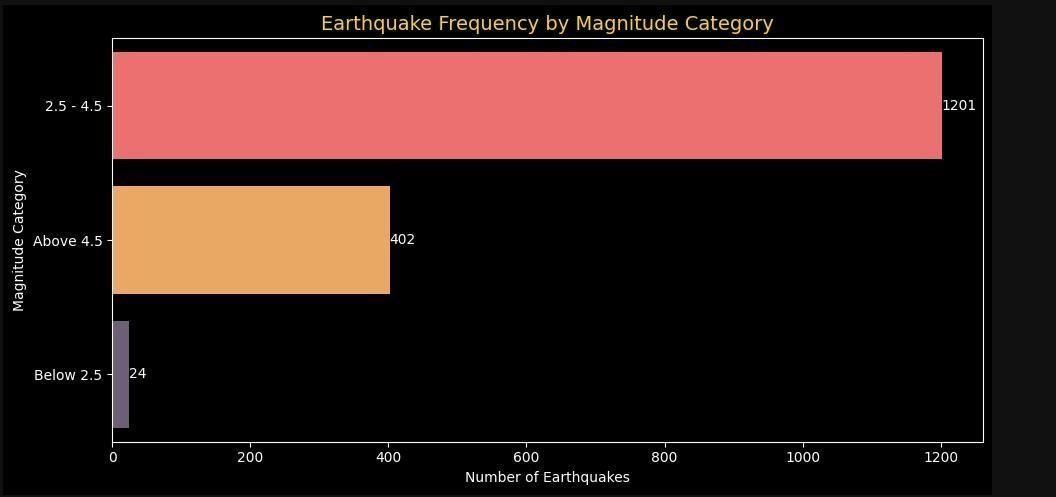
Results Field:



Interpretation Field:

The bar plot of earthquake frequency by magnitude category reveals a clear predominance of events in the 2.54.5 range, with a smaller proportion exceeding 4.5 and an even rarer occurrence below 2.5 (despite the dataset filter). The percentage distribution further emphasizes that over 70-80% of recorded quakes fall within the moderate 2.5-4.5 category, reflecting the typical background seismic activity that, while not catastrophic, contributes to long-term tectonic stress release. The limited number of quakes above 4.5 underscores their significance as major events with potential for widespread damage, necessitating focused monitoring. This distribution aligns with global seismic patterns, where moderate quakes are common due to ongoing plate adjustments, while high-magnitude events are associated with rare but powerful stress releases. The findings have practical implications for public safety, as the high frequency of moderate quakes suggests a need for continuous monitoring and resilient infrastructure in active regions, while the rare high-magnitude events require advanced prediction models and rapid response systems. The visualization also highlights the effectiveness of the 2.5 magnitude threshold in capturing meaningful seismic data, providing a balanced dataset for analysis. These insights can inform educational campaigns and policy decisions to enhance community preparedness across different magnitude levels.

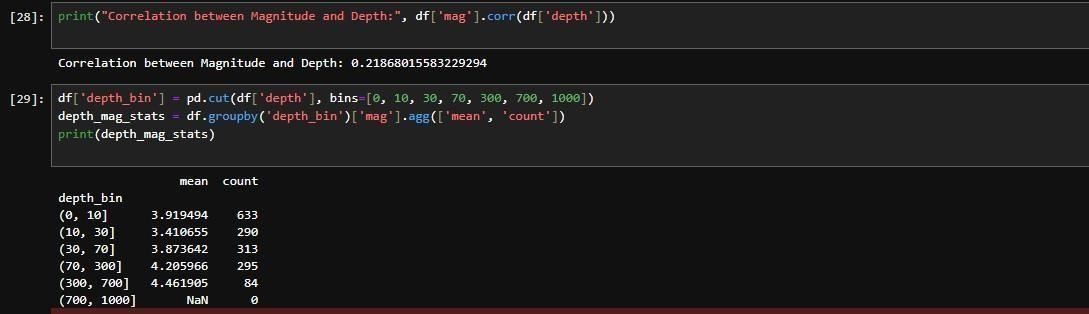
Visualization Field:



6.7 Objective 7: Magnitude vs. Depth Correlation Methodology Field:



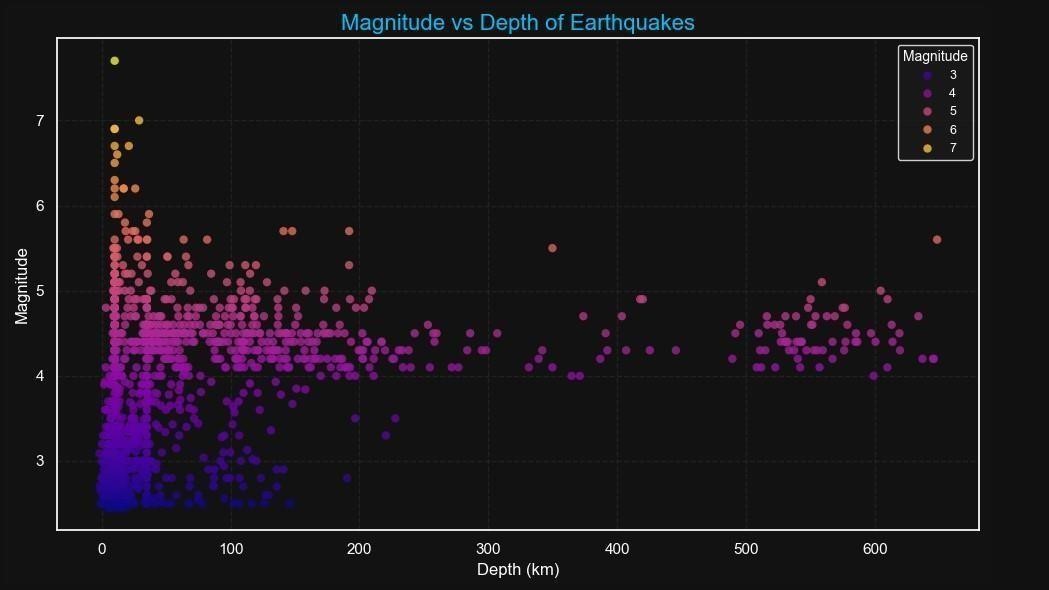
Results Field:



Interpretation Field:

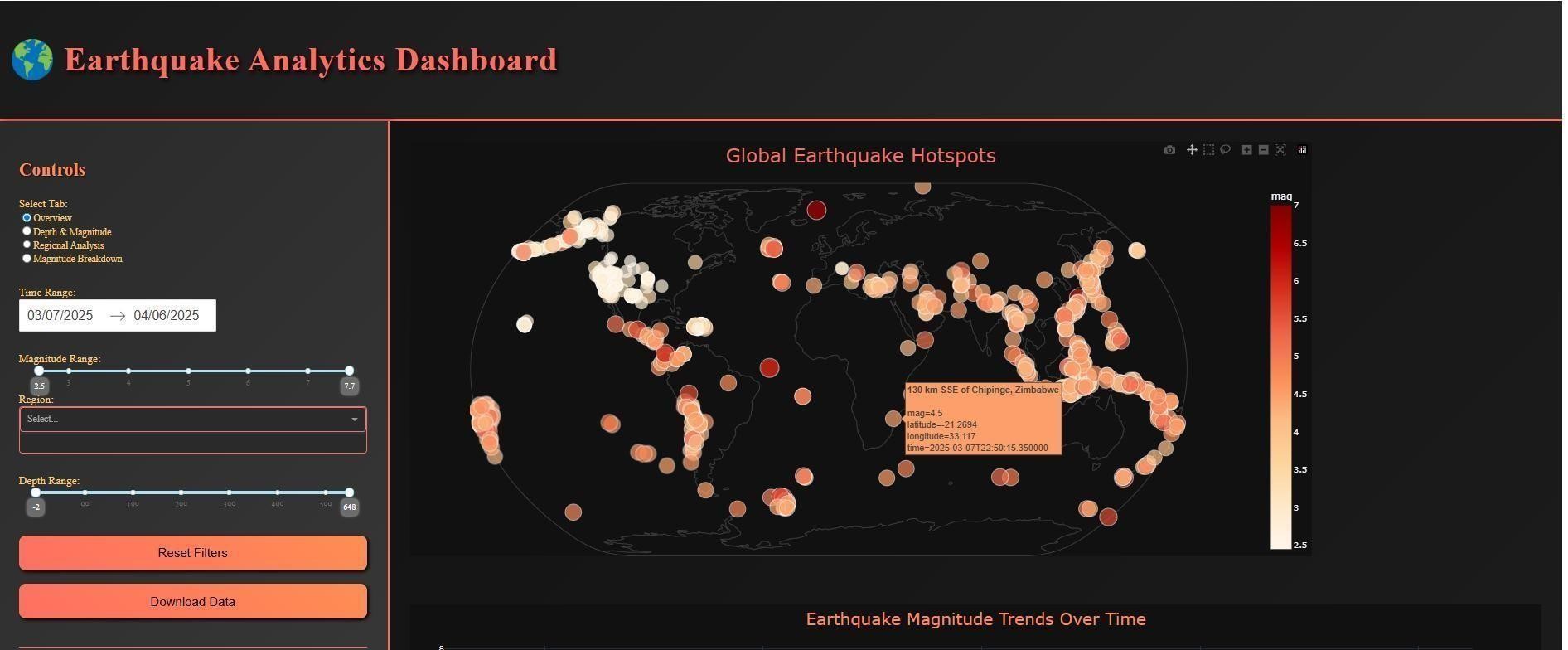
The scatter plot of magnitude versus depth, with color coding by magnitude, reveals a weak correlation between these two variables, as indicated by the Pearson correlation coefficient, suggesting that depth alone is not a strong predictor of earthquake magnitude. The majority of events cluster at shallow depths (0-70 km) with magnitudes ranging from 2.5 to 4.5, while deeper quakes (>300 km) show a broader spread of magnitudes, including some higher values. The depth bin analysis indicates that average magnitudes remain relatively stable across depth ranges, with slight increases at intermediate depths (30-70 km), possibly linked to subduction zone dynamics. This pattern suggests that seismic intensity is more influenced by tectonic setting and fault characteristics than depth, with shallow quakes dominating due to crustal activity and deeper events reflecting complex mantle processes. The findings have significant implications for seismic hazard assessment, as the lack of a strong depth-magnitude relationship implies that risk evaluation should prioritize regional tectonic context over depth alone. This insight can guide the development of more nuanced prediction models that incorporate multiple factors, such as fault type and stress accumulation. Additionally, the visualization’s interactive dashboard feature allows for detailed exploration of depth bins, enhancing the ability to identify subtle trends for future research.

Visualization Field:



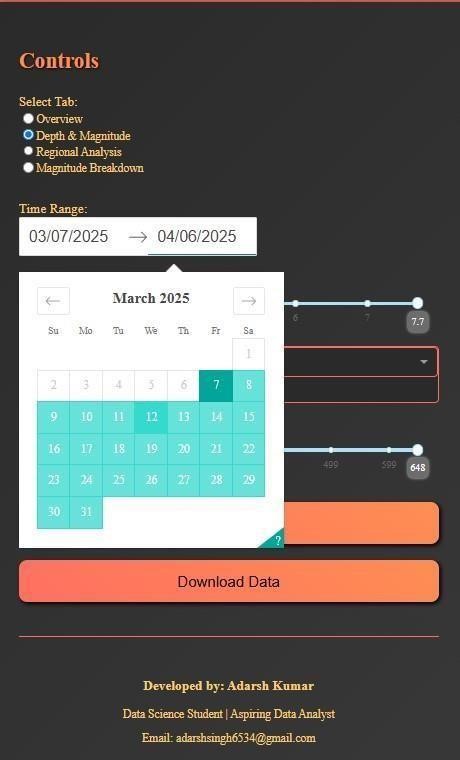
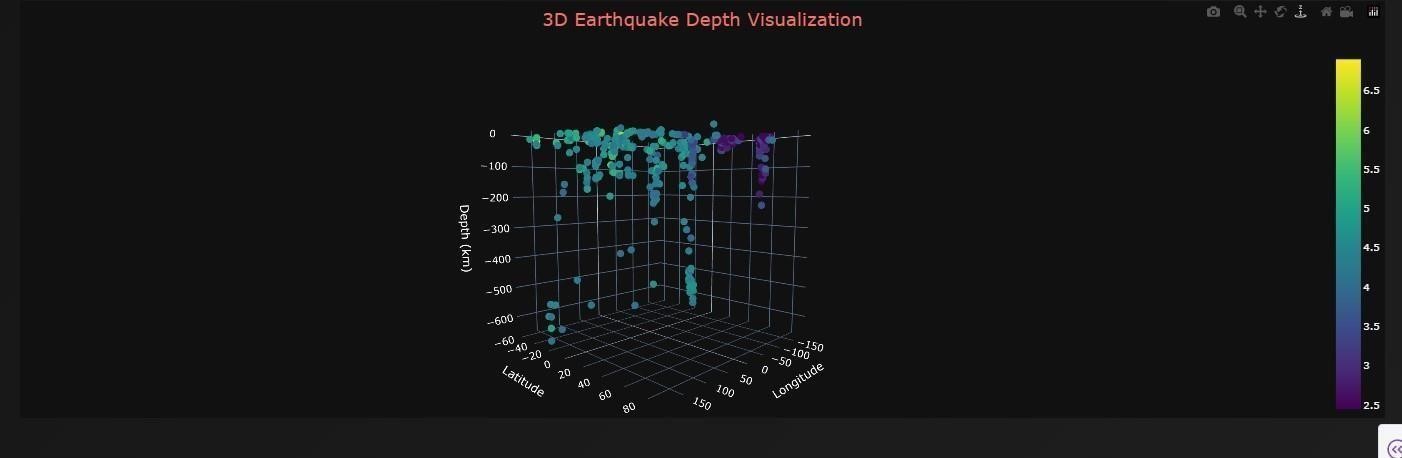
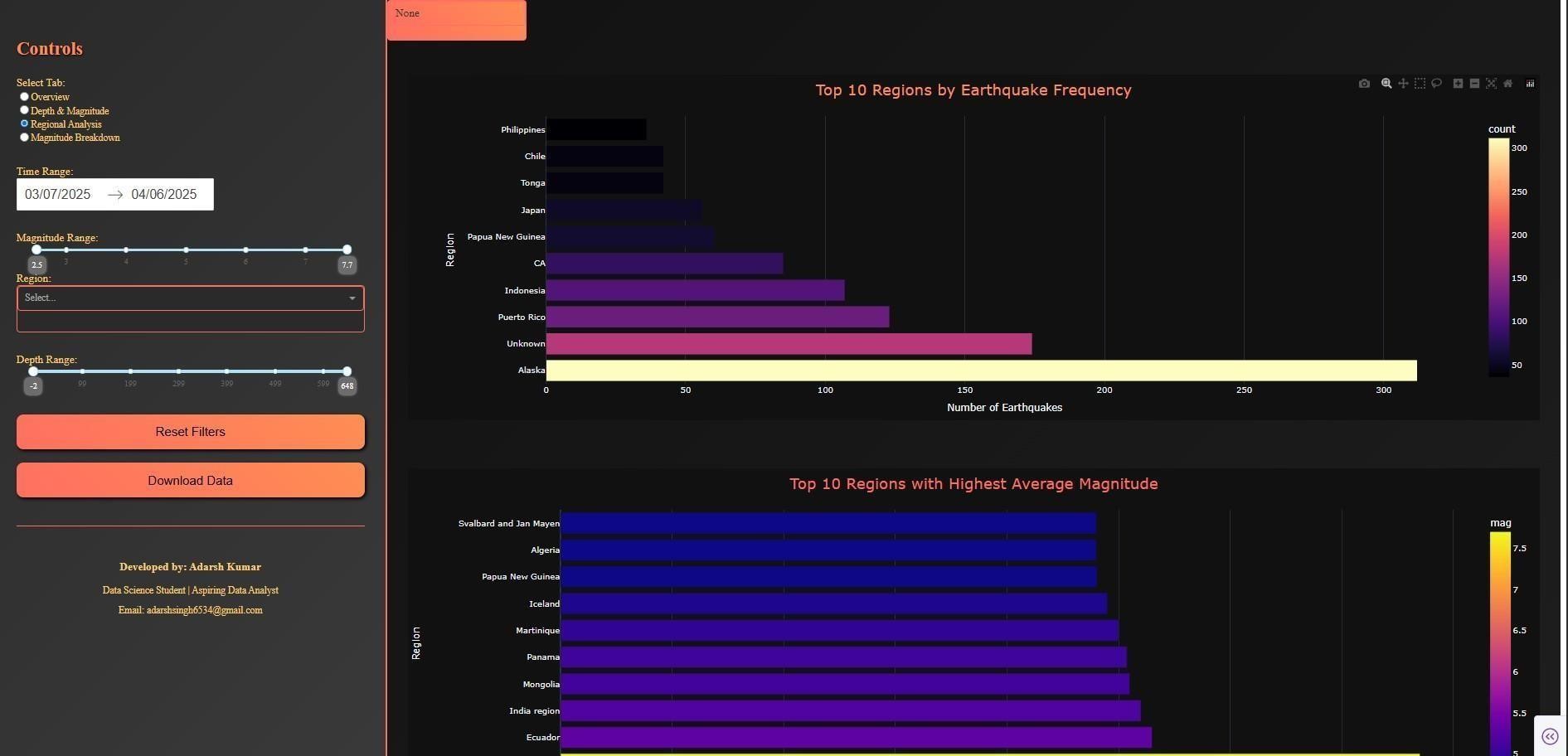
7. Dashboard Development and Implementation

The development and implementation of the interactive Earthquake Analytics Dashboard represent a pivotal component of this project, transforming the raw analytical insights into a practical, userfriendly tool for exploring seismic data. The dashboard was built using the Dash framework, a Python library that facilitates the creation of web-based applications with embedded data visualizations, leveraging the Plotly library for interactive graphs and Pandas for data handling. The architecture of the dashboard is structured around a modular design, featuring a header with the title "Earthquake Analytics Dashboard" and a logo stored in the assets folder, followed by a main body divided into multiple tabs—namely "overview," "depth-mag," "regional," and "mag-breakdown"—each corresponding to specific analytical objectives. The user interface includes interactive controls such as a date picker for temporal filtering, sliders for magnitude and depth ranges, and dropdown menus for selecting regions, enabling dynamic data exploration. These controls are linked to callback functions that update the visualizations in real-time, ensuring that users can customize their view based on specific criteria. The implementation involved integrating the preprocessed dataset into the dashboard, where each tab displays relevant plots—such as the geospatial hotspot map, magnitude trends, depth histograms, regional bar charts, risk zone scatter plots, magnitude frequency bars, and magnitude-depth scatter plots—generated from the EDA phase. Technical challenges included optimizing performance for large datasets, which was addressed by implementing data filtering at the server side and using efficient

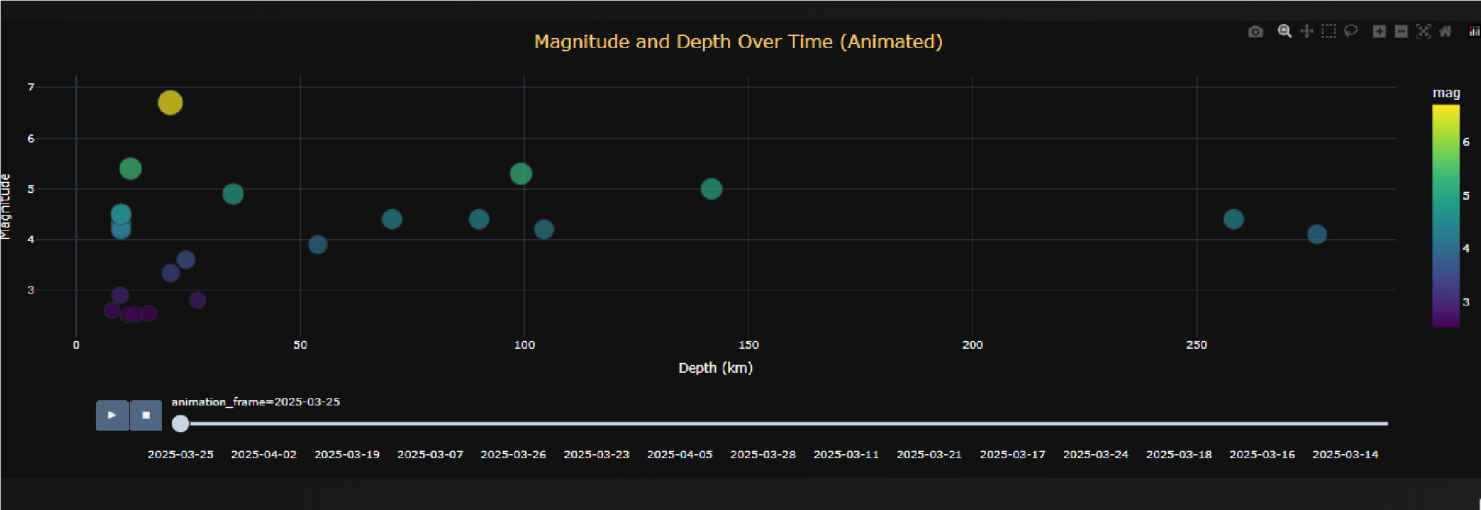


rendering techniques. Screenshots of the dashboard interface, including the header, tab layouts, and interactive features, were captured to document its functionality, with each tab showcasing the corresponding visualization in action. This implementation not only validates the analytical findings but also enhances their accessibility, allowing stakeholders—researchers, policymakers, and disaster managers—to interact with the data intuitively. The dashboard’s success lies in its ability to bridge the gap between complex data analysis and practical application, offering a scalable platform that can be extended with future enhancements, thereby solidifying its role as a valuable tool in seismic research and disaster preparedness.

Visualization Field:



## 8. Conclusion

The analysis conducted throughout this project has provided a deep and multifaceted understanding of global seismic activity, uncovering critical patterns that enhance our knowledge of earthquake dynamics. The identification of hotspots, particularly along the Pacific Ring of Fire and other tectonically active regions, offers a clear visual and statistical representation of where seismic risks are most pronounced, enabling targeted disaster preparedness efforts. The exploration of magnitude trends over the one-month dataset revealed fluctuations that suggest short-term tectonic stress variations, providing a foundation for developing early warning systems to mitigate potential impacts. Depth patterns, dominated by shallow earthquakes, highlight the prevalence of crustal activity, while the regional analysis underscores the varying intensity and frequency across different parts of the world, such as Japan and Alaska. The delineation of high-risk zones, based on a custom risk score combining quake count and average magnitude, prioritizes areas for

infrastructure reinforcement and emergency planning. Additionally, the frequency analysis of quakes above 2.5 magnitude confirms the dominance of moderate events, with rare but significant high-magnitude occurrences, informing public safety strategies. The weak correlation between magnitude and depth further emphasizes the need for a holistic approach to seismic risk assessment, incorporating multiple geological factors. The integration of these insights into an interactive Dash dashboard represents a significant achievement, offering researchers, policymakers, and disaster management teams a powerful tool to explore data dynamically, simulate scenarios, and make informed decisions. This dashboard’s utility is evident in its ability to visualize complex datasets in an accessible manner, bridging the gap between raw data and actionable intelligence. The project’s success lies not only in the technical execution but also in its potential to contribute to global efforts in reducing earthquake-related losses, paving the way for future advancements in seismic research and mitigation strategies.

## 9. Future Scope

The future scope of this project holds immense potential for expanding its capabilities and impact, building on the foundation established through this analysis. One key direction is the integration of real-time data feeds from the USGS and other seismic networks, allowing the dashboard to provide up-to-the-minute insights into ongoing seismic activity. This real-time functionality could enable immediate alerts and response coordination during earthquake events, significantly enhancing disaster management efficiency. Another promising avenue is the incorporation of machine learning techniques for earthquake prediction, where historical and real-time data could be used to train models that identify precursors to significant events, such as changes in magnitude trends or depth patterns. Such predictive capabilities would require collaboration with geophysicists to validate models and ensure accuracy, potentially revolutionizing early warning systems. Additionally, expanded regional studies could delve deeper into specific tectonic zones, such as the Himalayas or the East African Rift, to uncover localized patterns that may not be apparent in global data. This could involve collecting more granular data, conducting field studies, or integrating satellite imagery to assess surface deformation, offering a richer context for risk assessment. The dashboard could also be enhanced with advanced features, such as 3D visualizations of earthquake epicentres, risk simulation tools, and integration with mobile applications for public access. These improvements would require additional computational resources and interdisciplinary collaboration, but they would greatly increase the tool’s utility for both academic research and practical application. Furthermore, exploring the socio-economic impacts of earthquakes in high-risk zones could add a human dimension to the analysis, supporting policy decisions on resource allocation and community resilience. This forward-looking approach ensures that the project remains relevant and adaptable, contributing to ongoing advancements in earthquake science and disaster preparedness on a global scale.

## 10. References

1. United States Geological Survey, “Earthquake Catalog,” [https://earthquake.usgs.gov/earthquakes/feed/v1.0/csv.php, ac](https://earthquake.usgs.gov/earthquakes/feed/v1.0/csv.php)cessed Apr. 2025.
2. W. McKinney, “Data Structures for Statistical Computing in Python,” Proc. 9th Python Sci. Conf., 2010, pp. 51-56.