**“Predicting Major Earthquakes with Historical Data Analysis in Python”**

**Team C**

DS522\_01\_ON Data Acquisition & Analytics   
USADL – Fall 2024-2025

Data Science Program

School of Technology and Computing

City University of Seattle

Irene Gwyneth Smith

SmithIreneGwyneth@cityuniversity.edu

Jared Graham

grahamjared@cityuniversity.edu

Amber Giza  
gizaamber@cityuniversity.edu

1. **ABSTRACT**

This project focuses on creating a predictive model using Python to forecast potential occurrences of major earthquakes, specifically those with magnitudes of 7.0 or higher. By analyzing historical earthquake data, the study emphasizes trends and patterns, including temporal, magnitude, and geospatial distributions, that may assist in predicting future seismic events. Leveraging machine learning techniques and data visualization tools, the goal is to provide actionable insights to improve earthquake preparedness and risk assessments in vulnerable areas.

**Keywords:** Data visualization, Python, OpenCV, Image Processing, Natural Disaster, Earthquake Detection, Pandas, Matplotlib, and Data Analysis.

**2. INTRODUCTION**

**Background.** Earthquakes are one of the most devastating natural disasters, capable of causing widespread destruction and loss of life. Understanding the underlying causes of earthquakes and their patterns is crucial for mitigating risks and improving preparedness. This project delves into historical earthquake data, leveraging data science tools to uncover meaningful trends. By focusing on seismic events of magnitude 9.0 and higher, we aim to identify predictive factors that could signal the likelihood of such events in the future. The study’s primary goal is to empower emergency responders and city planners with data-driven insights to reduce the impact of potential disasters.

**3.** **LITERATURE REVIEW**

The occurrence of earthquakes is primarily linked to tectonic plate movements along fault lines and subduction zones (Michigan Technological University, n.d.). Magnitudes are commonly measured using the Richter scale and Moment Magnitude Scale (US Geological Survey, n.d.). Historical events, such as the 1960 Valdivia earthquake in Chile and the 2011 Tohoku earthquake in Japan, illustrate the catastrophic consequences of seismic activity, including tsunamis and infrastructure collapse (Hammer, 2011). Advanced early warning systems, such as Japan’s Earthquake Early Warning system and the US ShakeAlert, have demonstrated the value of leveraging seismic data to save lives and minimize damage (Mittal et al., 2022). This project builds on these insights by using historical data to predict future risks.

**4.** **DATA**

The dataset used in this project was sourced from Kaggle and contains 1,137 records of earthquakes, with data dating back to 2150 B.C. Key attributes include magnitudes, depths, dates, locations (latitude and longitude), and associated damage estimates. Data preprocessing involved cleaning missing values, standardizing formats, and augmenting the dataset with additional columns for related disasters and economic impacts. This curated dataset forms the foundation for temporal, spatial, and predictive analysis.

Figure 1.0 Earthquake Magnitude Data

A graph with a bar graph

Description automatically generated with medium confidence

**5. METHODOLOGY**

The project methodology involved the following steps:

1. **Data Collection and Cleaning:** Raw data was collected from Kaggle and cleaned to address inconsistencies, missing values, and irrelevant entries. Augmentation included creating derived features, such as time intervals between events and cumulative damage estimates.
2. **Exploratory Data Analysis (EDA):** Visualization techniques, such as time-series plots and heatmaps, were used to identify patterns in seismic activity across regions and time periods.
3. **Predictive Modeling:** Machine learning algorithms from scikit-learn were used to develop a model predicting the likelihood of a high-magnitude earthquake. Various models were tested to select the most accurate one.
4. **Validation and Testing:** The model was evaluated using cross-validation and metrics like accuracy, precision, and recall to ensure reliability.
5. **Visualization and Interpretation:** Visualizations included temporal trends, magnitude distributions, depth relationships, and geospatial mappings.

**6. RESULTS**

**1. Visualization**

Earthquake occurrences over the years.

**Findings**:

* Earthquake occurrences show variability across years.
* No evident seasonal trends, suggesting that seismic activity does not follow a monthly or seasonal pattern.

The following sections highlight the main findings and corresponding visuals.

Figure 2.0 : Earthquake occurrences over the years.

A graph of earthquake occurrence

Description automatically generated

**2. Magnitude Distribution**

Exploratory data analysis (EDA) focused on uncovering trends in magnitude and location.

Figure 3.0 : Distribution of earthquake magnitudes.

A graph with orange lines and black text

Description automatically generated

**3. Depth vs. Magnitude**

Scatter plot of earthquake depth vs. magnitude.

**Findings**:

* Earthquakes occur at various depths, with higher magnitudes often associated with deeper seismic events.
* Depth appears to influence the magnitude of an earthquake.

Figure 4.0 : Depth vs. Magnitude of Earthquakes.

A graph with orange dots

Description automatically generated

**4. Geospatial Insights**

Scatter plot of latitude and longitude, colored by magnitude.

**Findings**:

* Earthquakes are concentrated along tectonic plate boundaries.
* High-magnitude events align with fault lines and subduction zones, consistent with tectonic theory.

Figure 5.0 : Geospatial distribution of earthquakes.

A diagram of a distribution of earthquakes

Description automatically generated

**7. DISCUSSIONS**

The analysis highlights temporal variability in earthquake occurrences, magnitude distributions, and the influence of depth on seismic activity. The geospatial analysis confirms the clustering of earthquakes along tectonic plate boundaries.

**7. CONCLUSIONS**

This study demonstrates the power of data visualization in uncovering key insights about earthquakes. While precise predictions remain challenging, the trends identified can inform risk assessment and disaster preparedness.

**8. ACKNOWLEDGEMENTS**

We extend our gratitude to the professors and mentors at City University of Seattle for their guidance and support. Special thanks to the Kaggle community for providing access to the dataset and to our teammates for their dedication and collaboration throughout this project.

**9. FOOTNOTES**

This project draws on credible sources that combine historical and scientific insights into earthquakes. Michigan Technological University (n.d.) explains earthquake magnitude measurement, while the US Geological Survey (n.d.) explores "megaquakes" and their impacts. Historical events, such as the Great Japan Earthquake of 1923, are detailed by Hammer (2011) and Denawa (2005), showcasing seismic disasters' consequences. Mittal et al. (2022) highlight effective early warning systems, and Kaggle (n.d.) provides the historical data used for analysis. McKinney (2022) offers essential Python techniques for data wrangling, forming the project's methodological foundation. These sources collectively support the project's goal of enhancing earthquake preparedness through data-driven insights.

**10.** **REFERENCES**

Denawa, M. (2005). The Great Kanto Earthquake of 1923. Retrieved from [library.brown.edu](<https://library.brown.edu/cds/kanto/denewa.html>)

Hammer, J. (2011). The Great Japan Earthquake of 1923. Smithsonian. Retrieved from [smithsonianmag.com](<https://www.smithsonianmag.com/history/the-great-japan-earthquake-of-1923-1764539>)

Kaggle Dataset: Historical Earthquake Data.  
   
McKinney, W. (2022). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O’Reilly. (ISBN: 978-1-098-10403-0) 

Michigan Technological University. (n.d.). How Do We Measure Earthquake Magnitude? Retrieved from [mtu.edu](<https://www.mtu.edu/geo/community/seismology/learn/earthquake-magnitude>)

Mittal, H., Yang, B. M., & Wu, Y.-M. (2022). Earthquake Early Warning in Taiwan. Geoscience Letters.Retrieved from [springeropen.com](<https://geoscienceletters.springeropen.com/articles/10.1186/s40562-022-00251-w>)

US Geological Survey. (n.d.). Can "MegaQuakes" Really Happen? Retrieved from [usgs.gov](https://www.usgs.gov/faqs/can-megaquakes-really-happen)

## 