BACSE101 Problem Solving using Python

PROJECT REPORT

On

GLOBAL EARTHQUAKES ANALYZER AND SAFETY GUIDE

Prepared by

HARSH AGARWAL – 25BCE0995

Under the supervision of

Professor Thirumoorthy Krishnan

School of Computer Science and Engineering
Vellore Institute of Technology, Vellore.

November 06, 2025

Table of Contents

ABSTRACT

1. INTRODUCTION
   - 1.1 Domain Information
   - 1.2 Software libraries used
   - 1.3 Contributions
   - 1.4 GitHub Repository
   - 1.5 Challenges faced

2. PROBLEM STATEMENTS AND OBJECTIVES
   - Problem Statement
   - Objectives

3. IMPLEMENTATION / FEATURES
   - Data Loading and Pre-processing
   - Feature 1: Plot Global Hotspots (Ring of Fire)
   - Feature 2: Show Top 10 Strongest Earthquakes
   - Feature 3: Country-Specific Analysis
   - Feature 4: Show Advanced Analysis
   - Feature 5: Show Earthquake Safety Guide

4. DEMO SCREENSHOTS

5. CONCLUSION

Abstract

*The **"Global Earthquake Analyzer & Safety Guide"** is a command-line application built in Python to analyse global seismic activity and promote public safety. Leveraging a large dataset of significant earthquakes (M5.0+ from 2015-2025), the program uses the pandas, NumPy and matplotlib libraries to perform a comprehensive, text-based analysis. Key features include the ability to identify the top 10 most seismically active "grid squares" on Earth, list the strongest earthquakes on record, compare quake statistics between different countries, and analyse time-based patterns. In addition to this data-driven analysis, the tool provides a vital social welfare component: a clear, static guide on precautionary measures to take before, during, and after an earthquake, making it a comprehensive resource for both data exploration and risk awareness.*

1. Introduction

This project, the "Global Earthquake Analyzer and Safety Guide," is a command-line application built using Python. It is designed to address the real-world challenge of understanding global seismic activity and promoting public safety. The program transforms a large, raw CSV dataset of earthquake events (M5.0+ from 2015-2025) into an interactive and understandable tool. It provides users with powerful data analysis features, such as identifying the strongest quakes and busiest hotspots, as well as a vital social welfare component by offering a clear, accessible guide to earthquake safety measures.

1.1 Domain Information

Earthquakes are a major natural hazard. Understanding their patterns (frequency, location, magnitude) is the critical work of seismology. This field relies on analyzing vast datasets to identify high-risk zones, map tectonic boundaries like the "Ring of Fire," and inform emergency preparedness. This project serves as an accessible tool to explore these same global seismic trends.

1.2 Software Libraries Used

- **csv:** Built-in module used for initial file reading (open() and csv.reader()).
- **pandas:** The core analysis library for data loading, cleaning (type conversion, feature engineering), filtering, grouping, and aggregation.
- **numpy:** Foundational library for numerical operations, powering pandas and used for categorizing data (pd.cut()).
- **matplotlib:** The primary plotting library, used to create the scatter plot map of earthquake locations.
- **geopandas:** Geospatial library used to load and draw the world map background from a GeoJSON file.

1.3 Contributions by Team Members

This project was developed by **Harsh Agarwal**, under the valuable guidance and supervision of **Professor Thirumoorthy Krishnan**. As the sole contributor, all aspects of the project—from conceptualization and data acquisition to coding the analysis functions and writing this report—were handled by Harsh Agarwal.

1.4 GitHub Repository

The complete source code, documentation, and query.csv dataset for this project are available on GitHub at the following repository:

https://github.com/HarshAgarwal07/GLOBAL-EARTHQUAKE-ANALYZER-AND-SAFETY-GUIDE.git

1.5 Challenges Faced

1. **Data Acquisition:** USGS download limits required using a 10-year, M5.0+ dataset (17,000+ records) as a manageable compromise.
2. **Data Cleaning:** The text-based `place` column was inconsistent and required a custom function to parse and extract a usable `country` for analysis.
3. **Data Type Conversion:** Manually converted text-based numbers (mag, depth) to numeric types using `pd.to_numeric` after loading via the `csv` module.
4. **Map Visualization:** The `geopandas` built-in map was deprecated, requiring a fix to load the world GeoJSON from a reliable online URL to serve as the map background.

2. Problem Statement and Objectives

- **Problem Statement**

Raw seismic data from sources like the USGS is simply a massive, multi-column spreadsheet (CSV) containing thousands or millions of rows of numbers and text. For an average user, student, or even a professional, this raw file is unreadable and provides no immediate insight. It is impossible to quickly see which areas are most prone, identify the strongest quakes, or understand trends over time.

- **Objectives**

The primary objective of this project is to solve this problem by creating a Python program that acts as a powerful and user-friendly interface to this complex data.

The specific objectives are:

- To load and pre-process a large raw CSV file using the csv and pandas libraries.
- To clean and enrich the data by converting data types, handling missing values, and engineering new features like country, year, and depth_category.
- To create a simple, menu-driven command-line interface for the user.
- To provide a key geospatial visualization (plot_ring_of_fire) that plots earthquake locations on a real world map using matplotlib and geopandas.
- To implement text-based analysis features, including finding the top 10 strongest quakes, performing country-specific statistical analysis, and comparing different nations.
- To fulfill a social welfare goal by including a clear, static, and accessible guide on earthquake safety measures.

3. Implementation

The program is implemented as a series of functions, all driven by a main loop that displays the user menu.

3.1 Data Loading and Pre-processing (load_data)

This is the first and most critical function.

1. It uses Python's open() and csv.reader() to read the query.csv file row by row.
2. The header and data are loaded into a pandas DataFrame.
3. It then performs crucial pre-processing:
   - **Type Conversion:** pd.to_numeric is used to convert mag, depth, latitude, and longitude from text to numbers.
   - **Time Conversion:** pd.to_datetime is used to convert the time string into a datetime object.
   - **Feature Engineering (Country):** The extract_country function is applied to the place column to create a new, clean country column.
   - **Feature Engineering (Time):** The datetime object is used to extract year, month, and hour into new columns for time-based analysis.
4. Finally, it drops any rows with missing essential data (.dropna()) and returns the clean DataFrame.

3.2 Feature 1: Plot Global Hotspots (Ring of Fire)

(plot_ring_of_fire) provides the project's main visualization.

1. It loads a world map GeoJSON from a URL using gpd.read_file().
2. It creates a matplotlib figure and axes (fig, ax).
3. It plots the world map onto the ax as a light gray base layer.
4. It then uses ax.scatter() to plot the longitude (x-axis) and latitude (y-axis) of every earthquake on top of the map.

5. Key visualization parameters are used:
   - c='red': Makes the dots red.
   - alpha=0.3: Makes the dots semi-transparent, so areas with many dots (hotspots) appear as a darker, solid red.
   - s=df['mag'] * 1.5: The size of the dot (s) is made proportional to the earthquake's magnitude.
6. plt.show() opens this map in a new window.

3.3 Feature 2: Show Top 10 Strongest Earthquakes:

This is a straightforward pandas analysis.

1. It takes the DataFrame and sorts it using .sort_values(by='mag', ascending=False).
2. It selects the top 10 rows using .head(10).
3. It then loops through these 10 rows and prints a formatted report of the mag, time, place, and depth for each.

3.4 Feature 3: Country-Specific Analysis (analyze_countries)

This feature is a sub-menu that demonstrates powerful grouping and filtering.

1. **Top 10 Active Countries:** This filters out any "Unknown" countries, then uses the .value_counts() method on the country column to get a count for each, and finally shows the .head(10).
2. **Report for Specific Country:** This prompts the user for a country name. It filters the DataFrame by that country (e.g., df[df['country'].str.lower() == 'japan']) and then runs the .describe() method on the mag and depth columns to show a full statistical summary (count, mean, min, max, etc.).
3. **Compare Two Countries:** This performs the same filtering as the single report but for two countries. It then prints a custom, side-by-side comparison of their total quake counts, average magnitudes, and max magnitudes.

3.5 Feature 4: Show Advanced Analysis (show_advanced_analysis)

This feature demonstrates how to categorize data using pd.cut.

1. **Depth Analysis:** It categorizes the depth column into 'Shallow', 'Intermediate', and 'Deep' bins and prints the .value_counts() for each, showing that most quakes are shallow.
2. **Time Analysis (Year):** It prints the .value_counts() for the year column (sorted by year) to show the number of quakes per year in the dataset.
3. **Time Analysis (Day/Night):** It uses pd.cut() on the hour column to create 'Day' and 'Night' categories and then prints the total count for each, answering whether more quakes occurred during the day or at night.

3.6 Feature 5: Show Earthquake Safety Guide (show_safety_info)

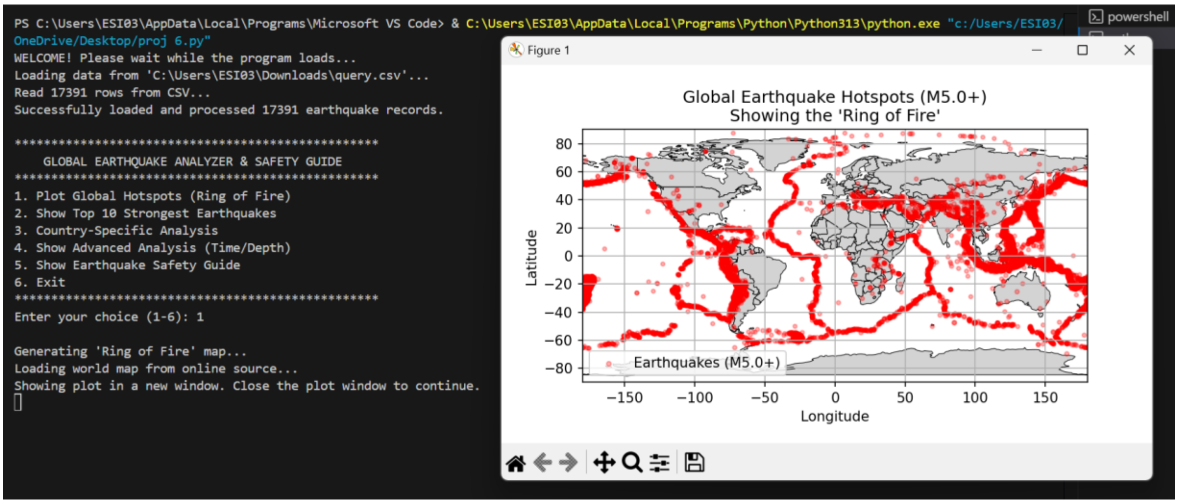
This function provides the project's social welfare component. It contains no data analysis. It simply defines a Python dictionary where the keys are "Before", "During", and "After", and the values are lists of safety tips. The function loops through this dictionary and prints the static information in a clean, easy-to-read format for the user.

4.Demo Screenshots

1. GENERATING GLOBAL HOTSPOTS AND PACIFIC RING OF FIRE

```
PS C:\Users\ESI03\AppData\Local\Programs\Microsoft VS Code> & C:\Users\ESI03\Ap
OneDrive/Desktop/proj 6.py"
WELCOME! Please wait while the program loads...
Loading data from 'C:\Users\ESI03\Downloads\query.csv'...
Read 17391 rows from CSV...
Successfully loaded and processed 17391 earthquake records.

**************************************************
    GLOBAL EARTHQUAKE ANALYZER & SAFETY GUIDE
**************************************************
1. Plot Global Hotspots (Ring of Fire)
2. Show Top 10 Strongest Earthquakes
3. Country-Specific Analysis
4. Show Advanced Analysis (Time/Depth)
5. Show Earthquake Safety Guide
6. Exit
**************************************************
Enter your choice (1-6):
```



2. GENERATING TOP 10 STRONGEST EARTHQUAKES

```
Enter your choice (1-6): 2

========================================
   TOP 10 STRONGEST EARTHQUAKES
========================================

Magnitude: 8.8
     Time: 2025-07-29
    Place: 2025 Kamchatka Peninsula, Russia Earthquake
    Depth: 35.0 km

Magnitude: 8.2
     Time: 2021-07-29
    Place: Alaska Peninsula
    Depth: 35.0 km

Magnitude: 8.2
     Time: 2017-09-08
    Place: 2017 Tehuantepec, Mexico Earthquake
    Depth: 47.4 km

Magnitude: 8.2
     Time: 2018-08-19
    Place: 267 km E of Levuka, Fiji
    Depth: 600.0 km

Magnitude: 8.1
     Time: 2021-08-12
    Place: South Sandwich Islands region
    Depth: 22.8 km

Magnitude: 8.1
     Time: 2021-03-04
    Place: Kermadec Islands, New Zealand
    Depth: 28.9 km
```

```
Magnitude: 8.1
     Time: 2021-03-04
    Place: Kermadec Islands, New Zealand
    Depth: 28.9 km

Magnitude: 8.0
     Time: 2019-05-26
    Place: 78 km NE of Navarro, Peru
    Depth: 122.6 km

Magnitude: 7.9
     Time: 2017-01-22
    Place: 35 km WNW of Panguna, Papua New Guinea
    Depth: 135.0 km

Magnitude: 7.9
     Time: 2018-01-23
    Place: 261 km SE of Chiniak, Alaska
    Depth: 14.1 km

Magnitude: 7.9
     Time: 2016-12-17
    Place: 140 km E of Kokopo, Papua New Guinea
    Depth: 94.5 km
========================================

Press Enter to return to the main menu...
```

3. COUNTRY SPECIFIC ANALYSIS

```
Enter your choice (1-6): 3

========================================
      Country-Specific Analysis
========================================
1. Show Top 10 Most Active Countries
2. Show Report for a Specific Country
3. Compare Two Countries
4. Return to Main Menu
Enter your choice (1-4): 1

--- Top 10 Most Active Countries (by M5.0+ quake count) ---
country
Indonesia           1736
Papua New Guinea    1071
Russia               815
Philippines          812
Japan                750
Tonga                733
Vanuatu              536
Alaska               515
Chile                478
New Zealand          435
Name: count, dtype: int64

Press Enter to return to the Country Menu...
```

```
========================================
      Country-Specific Analysis
========================================
1. Show Top 10 Most Active Countries
2. Show Report for a Specific Country
3. Compare Two Countries
4. Return to Main Menu
Enter your choice (1-4): 2
Enter country name (e.g., Japan, Indonesia, USA): Australia

--- Analysis Report for Australia ---
Total Quakes (M5.0+): 15
             mag      depth
count  15.000000  15.000000
mean    5.433333   9.466667
std     0.463938   2.669047
min     5.000000   0.000000
25%     5.100000  10.000000
50%     5.200000  10.000000
75%     5.750000  10.000000
max     6.600000  12.000000

Press Enter to return to the Country Menu...
```

```
========================================
      Country-Specific Analysis
========================================
1. Show Top 10 Most Active Countries
2. Show Report for a Specific Country
3. Compare Two Countries
4. Return to Main Menu
Enter your choice (1-4): 3
Enter first country name: Japan
Enter second country name: India

========================================
--- Comparison: Japan vs. India ---
Total Quakes:    Japan = 750,    India = 124
Avg. Magnitude:          Japan = 5.33,   India = 5.19
Max. Magnitude:          Japan = 7.30,   India = 6.70
Avg. Depth (km):         Japan = 45.0,   India = 28.9
========================================

Press Enter to return to the Country Menu...
```

4. ADVANCED ANALYSIS

```
Enter your choice (1-6): 4

========================================
        Advanced Analysis
========================================

--- Analysis by Depth ---
Quake Count by Depth:
depth_category
Shallow (0-70km)           14439
Intermediate (70-300km)     2298
Deep (300km+)                654
Name: count, dtype: int64

(Note: Shallow quakes are often the most destructive.)

--- Analysis by Year ---
Total Quakes per Year in this Dataset:
year
2015     287
2016    1694
2017    1556
2018    1751
2019    1629
2020    1435
2021    2211
2022    1725
2023    1780
2024    1507
2025    1816
Name: count, dtype: int64

--- Analysis by Time of Day ---
Quake Count by Time of Day:
day_night
Day (07:00 - 18:59)      8468
Night (00:00 - 06:59)    5151
Night (19:00 - 23:59)    3772
Name: count, dtype: int64
```

5. EARTHQUAKE SAFETY GUIDE

```
Enter your choice (1-6): 5

========================================
      EARTHQUAKE SAFETY GUIDE
========================================

--- BEFORE AN EARTHQUAKE ---
  1. Create an emergency plan with your family.
  2. Prepare a 'Go Bag' with water, first-aid, food, and a flashlight.
  3. Secure heavy furniture (bookshelves, TVs) to the wall.

--- DURING AN EARTHQUAKE ---
  1. DROP, COVER, and HOLD ON.
  2. If indoors, stay inside. Move away from windows and falling objects.
  3. If outdoors, move to an open area away from buildings and power lines.

--- AFTER AN EARTHQUAKE ---
  1. Check yourself and others for injuries. Provide first aid.
  2. Be prepared for aftershocks. Drop, cover, and hold on.
  3. If in a damaged building, get out. Do not use elevators.
========================================

Press Enter to return to the main menu...
```

5. Conclusion

This project successfully achieved its goal of transforming a large, raw seismic dataset into a user-friendly and insightful "Global Earthquake Analyzer and Safety Guide." By leveraging the analytical power of pandas and numpy, the program effectively cleans, processes, and enriches the data, allowing users to perform complex analyses—such as country-specific comparisons and time-based studies—through a simple text menu. The integration of matplotlib and geopandas proved critical, providing a powerful visual confirmation of the "Ring of Fire" that text-based analysis alone could not.

The project demonstrates the power of Python's data science stack in making complex information accessible. More importantly, it fulfils its social welfare objective by directly embedding a practical safety guide, successfully bridging the gap between pure data analysis and real-world application.

Future enhancements could involve integrating a live USGS API to analyze real-time data, or developing a graphical user interface (GUI) with interactive, zoomable maps to further improve user accessibility.