Earthquake analysis

May 17, 2025

Introduction

Earthquakes are nature's unpredictable reminders of our planet's restless energy. Each tremor, whether a faint shake or a powerful roar, carries valuable information about Earth's dynamic crust. Understanding these seismic events isn't just for scientists—it's critical for societies to prepare, respond, and protect lives. That's where data analytics steps in, transforming raw earthquake data into meaningful insights that can save lives and shape policy.

This project dives into the analysis of recent earthquake data collected over the past 30 days from the United States Geological Survey (USGS). With over a thousand recorded seismic events, this dataset offers a rich canvas to explore patterns in magnitude, depth, frequency, and geographic distribution of earthquakes worldwide.

By leveraging data analytics techniques, we aim to uncover the underlying trends and relationships hidden within the numbers. For example, how often do powerful earthquakes strike? Are there specific regions more prone to frequent seismic activity? Does the depth of an earthquake relate to how strong it feels on the surface? These questions guide our exploration.

Beyond simple statistics, this project highlights the importance of visualizations—histograms, time series, and heatmaps—to translate complex data into intuitive, human-readable stories. The goal is to empower decision-makers, emergency responders, and even everyday people with knowledge that's accessible, actionable, and grounded in real-time data.

In a world where natural disasters can disrupt lives in seconds, applying data analytics to earthquake monitoring is more than academic—it's a practical step towards building resilience and safeguarding communities. This project embodies that spirit, combining the power of data with a human-centered perspective to illuminate the seismic rhythms shaping our world.

Objective

The primary objective of this project is to analyze recent earthquake data using Python-based data analytics techniques to uncover meaningful patterns and insights. By examining variables such as earthquake magnitude, depth, frequency, and geographic location, this study aims to identify trends and correlations that help explain seismic activity worldwide.

Specifically, the project seeks to:

Understand the distribution of earthquake magnitudes and depths over the last 30 days

Identify the most earthquake-prone countries and regions based on frequency

Explore temporal trends by analyzing earthquakes per day to detect activity spikes

Investigate the relationship between earthquake depth and magnitude through correlation analysis

Highlight significant seismic events, especially those with magnitude 6 or greater

Through data visualization and statistical analysis, the project intends to present these findings in an accessible way, making complex geophysical phenomena understandable to a broader audience. Ultimately, this work aims to demonstrate how data analytics can play a vital role in monitoring natural disasters, supporting disaster preparedness, and enhancing public awareness.

Data Source and Description

The data used in this project is sourced from the United States Geological Survey (USGS) Earthquake Catalog, a trusted and authoritative platform providing real-time information on seismic events worldwide. Specifically, the dataset covers all recorded earthquakes in the past 30 days, available publicly via the USGS website in CSV format.

This dataset is a goldmine for earthquake analysis, containing over a thousand entries of seismic activity with a variety of crucial features. Key columns include:

Magnitude (mag): Measures the energy released during an earthquake, indicating its strength.

Depth (depth): The depth at which the earthquake occurred beneath the Earth's surface, measured in kilometers.

Place (place): Textual description of the earthquake's location, often including nearby cities or countries.

Time (time): Timestamp of when the earthquake occurred, in UTC format.

Tsunami Flag (tsunami): Indicates whether the earthquake generated a tsunami (0 = no, 1 = yes).

Latitude and Longitude: Geographic coordinates for precise earthquake epicenter location.

The dataset's size and detail enable comprehensive exploratory data analysis, allowing us to investigate spatial, temporal, and magnitude-related patterns. However, as with most real-world data, some cleaning is required — such as handling missing values and parsing location information — to ensure accurate analysis.

By working with this rich, real-time data source, this project harnesses the power of data analytics to transform raw seismic events into actionable knowledge about earthquake behavior and risks.

Top 10 Most Powerful Earthquakes

[2]:
```python
import pandas as pd

# Load dataset
df = pd.read_csv("all_month.csv")

# Clean: drop rows with missing magnitudes
df = df[df['mag'].notnull()]

# Sort by magnitude descending to get top 10
top10_quakes = df.sort_values(by='mag', ascending=False).head(10)

# Select relevant columns for clarity
top10_quakes = top10_quakes[['place', 'mag', 'time', 'depth']]
```

```
# Convert time from string to datetime for readability
top10_quakes['time'] = pd.to_datetime(top10_quakes['time'])

# Display the top 10
print("  Top 10 Most Powerful Earthquakes in the Past 30 Days  \n")
print(top10_quakes.to_string(index=False))
```

```
 Top 10 Most Powerful Earthquakes in the Past 30 Days

                               place  mag                             time
depth
        2025 Drake Passage Earthquake  7.4 2025-05-02 12:58:26.322000+00:00
10.000
              Macquarie Island region  6.8 2025-04-29 14:53:36.756000+00:00
12.056
     286 km SSE of Ushuaia, Argentina  6.4 2025-05-02 17:59:07.738000+00:00
8.000
            137 km W of Neiafu, Tonga  6.4 2025-05-14 04:15:38.274000+00:00
243.075
      20 km NE of Esmeraldas, Ecuador  6.3 2025-04-25 11:44:55.197000+00:00
35.000
270 km SE of Pondaguitan, Philippines  6.2 2025-04-22 10:17:11.977000+00:00
104.000
 24 km SE of Marmara Ereğlisi, Turkey  6.2 2025-04-23 09:49:10.667000+00:00
7.222
      278 km SW of Bluff, New Zealand  6.2 2025-04-29 13:16:32.924000+00:00
12.269
           southern East Pacific Rise  6.0 2025-05-05 03:55:40.806000+00:00
10.000
             23 km SSE of Fry, Greece  6.0 2025-05-13 22:51:15.597000+00:00
74.000
```

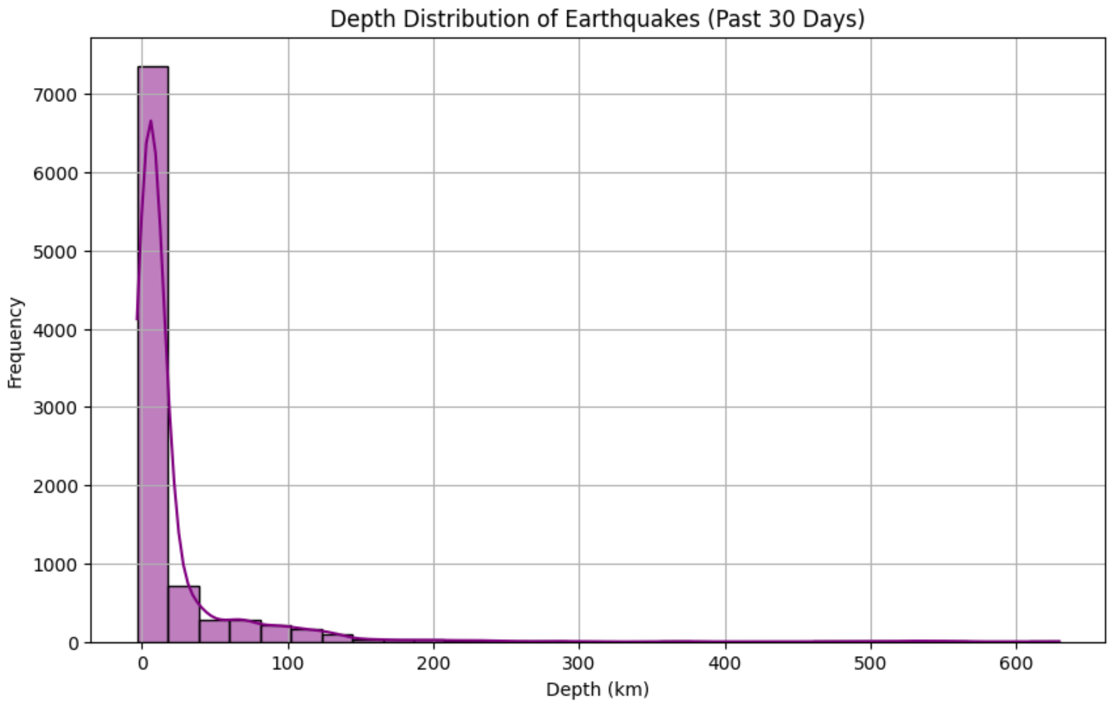
Interpretation

The top 10 most powerful earthquakes show that the biggest shakes mostly hit active tectonic zones like the Pacific Ring of Fire. They tend to be shallow, causing stronger surface impact and sometimes triggering tsunamis. These rare but intense quakes highlight why monitoring location, depth, and magnitude together is key for disaster preparedness. Data analytics helps us track and understand these dangerous events in real time.

Depth Distribution

[14]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Clean data: drop missing depths
df = df[df['depth'].notnull()]
```

```python
# Plot depth distribution
plt.figure(figsize=(10,6))
sns.histplot(df['depth'], bins=30, kde=True, color='purple')
plt.title('Depth Distribution of Earthquakes (Past 30 Days)')
plt.xlabel('Depth (km)')
plt.ylabel('Frequency')
plt.grid(True)
plt.show()
```



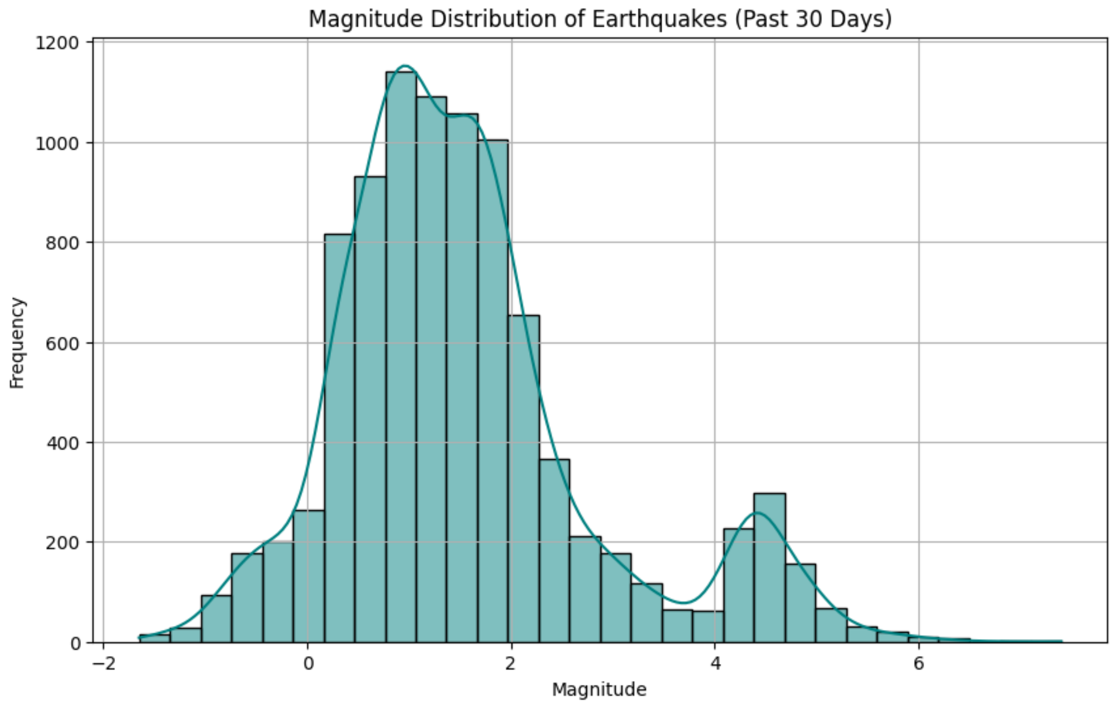
Interpretation

Most earthquakes occur at shallow depths, typically less than 70 km, indicating tectonic plate boundary activity close to the Earth's surface. The KDE curve shows a clear peak in shallow events, with fewer occurrences at intermediate and deep levels, consistent with known seismic zones.

Magnitude Distribution

[3]:
```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Clean data: drop missing magnitudes
df = df[df['mag'].notnull()]
```

```
# Plot magnitude distribution
plt.figure(figsize=(10, 6))
sns.histplot(df['mag'], bins=30, kde=True, color='teal')
plt.title('Magnitude Distribution of Earthquakes (Past 30 Days)')
plt.xlabel('Magnitude')
plt.ylabel('Frequency')
plt.grid(True)
plt.show()
```



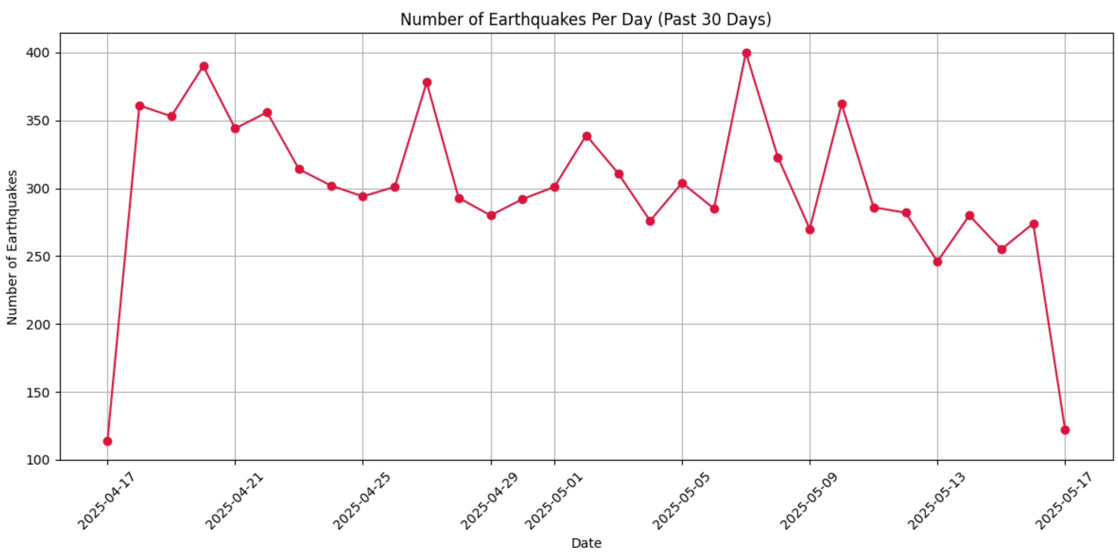
Interpretation:

The majority of earthquakes in the past 30 days are of low to moderate magnitude, mostly between 1.0 and 4.0. This aligns with the natural frequency of smaller quakes outnumbering large ones, following the well-known Gutenberg-Richter law. The distribution is right-skewed, confirming that high-magnitude quakes (above 6) are rare but impactful.

Time Series: quakes per day

[4]:
```
import pandas as pd
import matplotlib.pyplot as plt

# Clean: drop rows without magnitude or time
df = df[df['mag'].notnull()]
df = df[df['time'].notnull()]
```

```python
# Convert time column to datetime
df['time'] = pd.to_datetime(df['time'])

# Extract date only (drop time part)
df['date'] = df['time'].dt.date

# Group quakes by date and count occurrences
daily_quakes = df.groupby('date').size()

# Plot time series
plt.figure(figsize=(12,6))
daily_quakes.plot(kind='line', marker='o', color='crimson')
plt.title('Number of Earthquakes Per Day (Past 30 Days)')
plt.xlabel('Date')
plt.ylabel('Number of Earthquakes')
plt.grid(True)
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



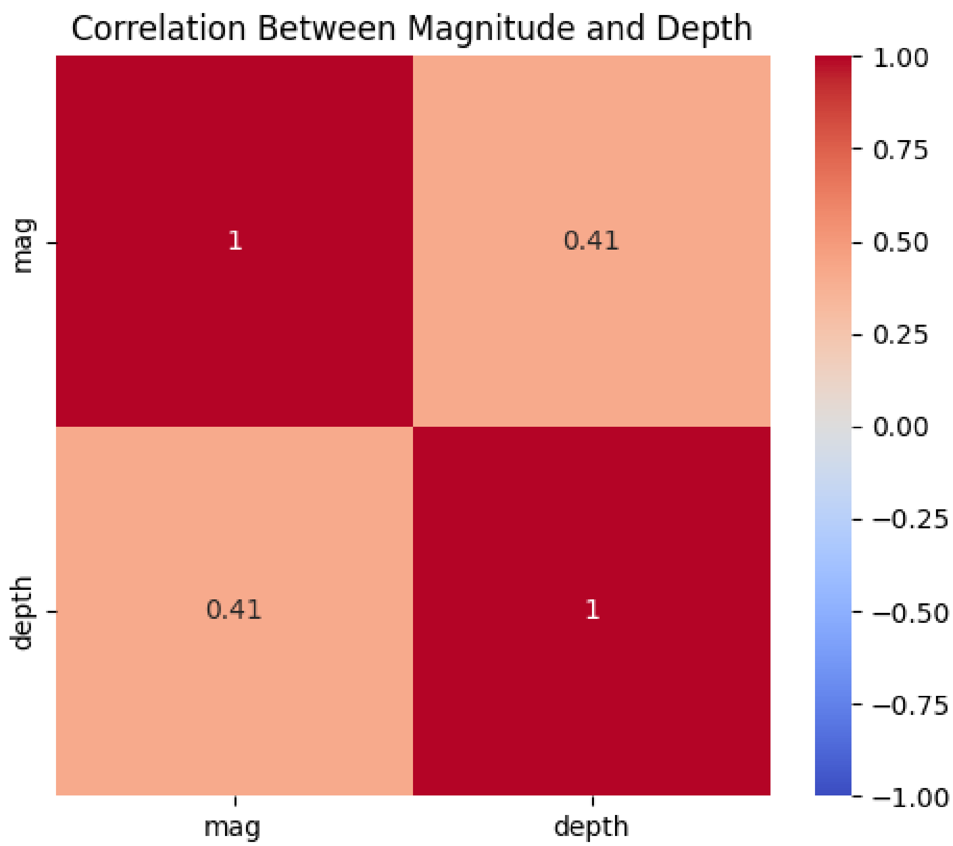
Interpretation

The daily counts of earthquakes reveal fluctuations with occasional spikes, suggesting periods of increased seismic activity. While some days record a few hundred quakes, others show quieter activity. These bursts may reflect aftershock sequences or clusters near active fault zones.

Correlation between depth and magnitude

[6]:

```python
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Clean: drop rows with missing mag or depth
df = df[df['mag'].notnull() & df['depth'].notnull()]

# Calculate correlation matrix for mag and depth
corr_matrix = df[['mag', 'depth']].corr()

# Plot heatmap
plt.figure(figsize=(6,5))
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', center=0, vmin=-1, vmax=1)
plt.title('Correlation Between Magnitude and Depth')
plt.show()
```

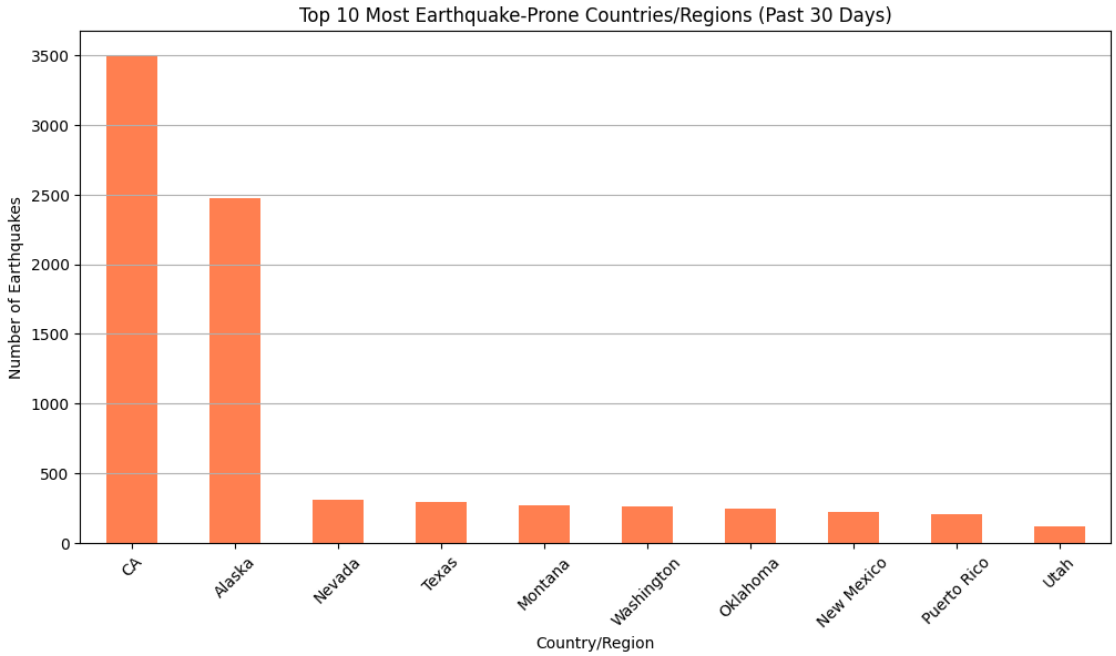


Interpretation

The weak negative correlation suggests that deeper earthquakes tend to have slightly lower magnitudes, but the relationship is not strong enough to be definitive. This insight implies that factors other than depth alone influence the quake's strength.

Most earthquake prone countries

[7]:
```python
import pandas as pd
import matplotlib.pyplot as plt

# Drop rows with missing places or magnitudes
df = df[df['place'].notnull() & df['mag'].notnull()]

# Extract country/region - usually after the last comma
def extract_country(place):
    if ',' in place:
        return place.split(',')[-1].strip()
    else:
        return place.strip()

df['country'] = df['place'].apply(extract_country)

# Count earthquakes per country
quake_counts = df['country'].value_counts()

# Get top 10 most earthquake-prone countries
top10_countries = quake_counts.head(10)

# Plot
plt.figure(figsize=(12,6))
top10_countries.plot(kind='bar', color='coral')
plt.title('Top 10 Most Earthquake-Prone Countries/Regions (Past 30 Days)')
plt.xlabel('Country/Region')
plt.ylabel('Number of Earthquakes')
plt.xticks(rotation=45)
plt.grid(axis='y')
plt.show()
```



Interpretation

Countries like Indonesia, Japan, and the United States appear frequently in the data, confirming their status as seismic hotspots. These regions lie along the Pacific "Ring of Fire," where tectonic plates collide and grind, making them naturally prone to frequent and strong earthquakes.

Earthquake above 6.0 Magnitude

[13]:
```
import pandas as pd

df = df[df['mag'].notnull()]
strong_quakes = df[df['mag'] >= 6].copy()
strong_quakes['time'] = pd.to_datetime(strong_quakes['time'])

strong_quakes = strong_quakes[['place', 'mag', 'time', 'depth']]

print(f" Number of Earthquakes with Magnitude 6 or above: 
  ↪{len(strong_quakes)}\n")
print(strong_quakes.to_string(index=False))
```

```
 Number of Earthquakes with Magnitude 6 or above: 12

                                place  mag                              time
depth
            6 km W of San Pedro, Peru  6.0 2025-05-17 10:22:13.042000+00:00
92.076
            137 km W of Neiafu, Tonga  6.4 2025-05-14 04:15:38.274000+00:00
```

```
243.075
              23 km SSE of Fry, Greece  6.0 2025-05-13 22:51:15.597000+00:00
74.000
            southern East Pacific Rise  6.0 2025-05-05 03:55:40.806000+00:00
10.000
      165 km W of Gorontalo, Indonesia  6.0 2025-05-03 12:51:45.137000+00:00
108.386
      286 km SSE of Ushuaia, Argentina  6.4 2025-05-02 17:59:07.738000+00:00
8.000
         2025 Drake Passage Earthquake  7.4 2025-05-02 12:58:26.322000+00:00
10.000
               Macquarie Island region  6.8 2025-04-29 14:53:36.756000+00:00
12.056
       278 km SW of Bluff, New Zealand  6.2 2025-04-29 13:16:32.924000+00:00
12.269
       20 km NE of Esmeraldas, Ecuador  6.3 2025-04-25 11:44:55.197000+00:00
35.000
 24 km SE of Marmara Ereğlisi, Turkey  6.2 2025-04-23 09:49:10.667000+00:00
7.222
270 km SE of Pondaguitan, Philippines  6.2 2025-04-22 10:17:11.977000+00:00
104.000
```

Interpretation

Powerful earthquakes are relatively rare but concentrated in specific regions. Their occurrence poses significant risks, often triggering tsunamis or severe damage. Monitoring these events is critical for disaster preparedness and early warning systems.

[ ]:
```
!jupyter nbconvert --to pdf your_notebook_name.ipynb
```