Kyrgyzstan Earthquake analysis

Project Requirements (Summary)

1. Data Collection (20%)

- Gather a dataset related to Kyrgyzstan (min. 3 variables).
  Use sources like the National Statistical Committee, World Bank, UN Data, Kaggle, etc.

2. Data Preprocessing (20%)

- Clean data: handle missing values, outliers, duplicates.
- Normalize/standardize variables if needed.
- Create new variables if applicable.

3. Data Analysis (30%)

- Conduct exploratory data analysis (EDA).
- Use statistical analysis (correlation, regression, hypothesis testing).
- Summarize findings with statistics, tables, and charts.

4. Data Visualization (20%)

- Create at least **5 different** visualizations (bar charts, histograms, scatter plots, heatmaps, etc.).
- Use **Python (Matplotlib, Seaborn, Plotly), Excel, or Tableau**.
- Ensure clarity, accuracy, and proper labeling.

5. Report Writing (15%)

- **Sections:**
  1. **Introduction** – Background and dataset selection.
  2. **Data Description** – Source, variables, key details.
  3. **Data Analysis** – EDA findings, statistics, insights.
  4. **Visualizations** – Charts with explanations.
  5. **Conclusion** – Key findings and implications.
  6. **References** – Data sources and resources.
- Proper formatting with headings and a table of contents.

6. Presentation & Defense (15%)

- **10-minute presentation** covering:
  - Dataset overview and key insights.
  - Data analysis and visualizations.
  - Conclusion and takeaways.

- Be prepared to **defend** your methods and answer questions.

Kyrgyzstan earthquake analysis

Introduction:

The Kyrgyz Republic is located in a highly seismic region, prone to devastating earthquakes that have led to deaths, destroyed homes and destroyed households in the past and most recently. Although earthquakes are less common than other natural hazards, such as floods and landslides, they cause the largest share of natural disaster losses across the country (World Bank, 2008). Due to the rapid urbanization and evolving nature of the economy in the Kyrgyz Republic, there is a strong incentive to invest in a national strategy to reduce seismic risk as the most effective way to mitigate the potential impact of natural disasters and reduce expected losses.

Data Description:

Source - [USGS Earthquake Catalog]

Variables:

1. **time:** The exact date and time (UTC) when the earthquake occurred
2. **latitude:** The geographic latitude of the earthquake's epicenter in degrees
3. **longitude:** The geographic longitude of the earthquake's epicenter in degrees
4. **depth:** The depth of the earthquake's hypocenter below the Earth's surface in kilometers (km). Shallow: <70 km, Intermediate: 70–300 km, Deep: >300 km.
5. **mag:** The magnitude of the earthquake, measuring energy released
6. **magType:** The method used to calculate magnitude (e.g., "mb" for body-wave magnitude, "mw" for moment magnitude).
7. **rms:** Root mean square residual (seconds), measuring the fit of the seismic model to observed data. Lower values indicate better data quality
8. **place:** A text description of the earthquake's location, often including distance and direction from a landmark
9. **depthError:** The uncertainty in the depth measurement (km), indicating location reliability.
10. **magNst:** The number of seismic stations used to calculate the magnitude.

Key values: **Analysis (EDA)** contains four sections

- Loading dataset
- General info
- Data Preprocessing

- Data analysis and Visualization

Data Analysis:

In the EDA phase, we explored the dataset with the following focus areas:

Time of earthquakes ranges from 2020 to 2025(most recent)

1. Descriptive Statistics:

	Mean	**Min**	**Max**
Magnitude (mag)	**4.3**	**3**	**7.2**
Depth	**54**	**0**	**285**

Most quakes were in the range of 3.5 to 4.5, indicating moderate seismic activity.
Most earthquakes were shallow (<70 km), which tend to cause more surface damage.

RMS (Root Mean Square):

Low RMS values (0.14 to 1.92) suggest good model fit, ensuring reliable data quality.

Temporal Analysis:

Earthquake frequency was analyzed by year and month.

Peaks in earthquake frequency were observed in certain seasons, suggesting possible seasonal/geological patterns.

Geospatial Trends:

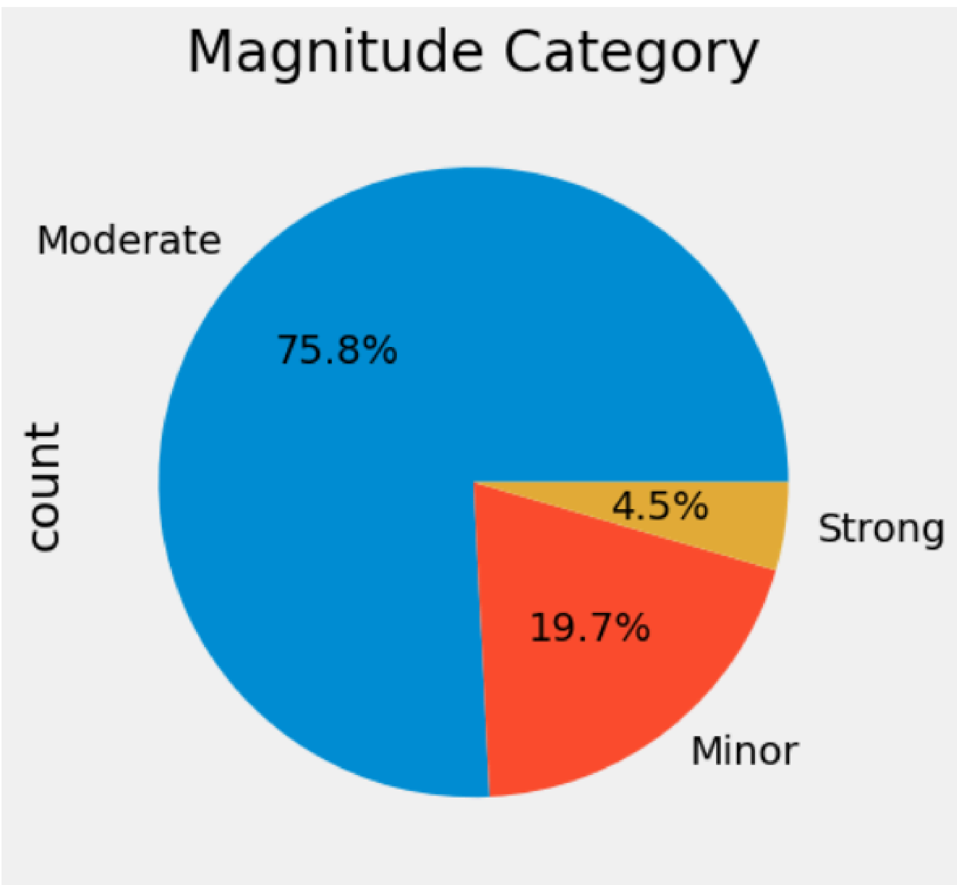
Mapping latitude vs. longitude showed clusters of activity, mainly in the Tien Shan mountain region, a known fault zone.

Correlation Analysis:
Weak correlation found between magnitude and depth ($r \approx -0.21$), implying that stronger earthquakes don’t necessarily occur deeper.

Visualizations

Here are key visuals included in your analysis:

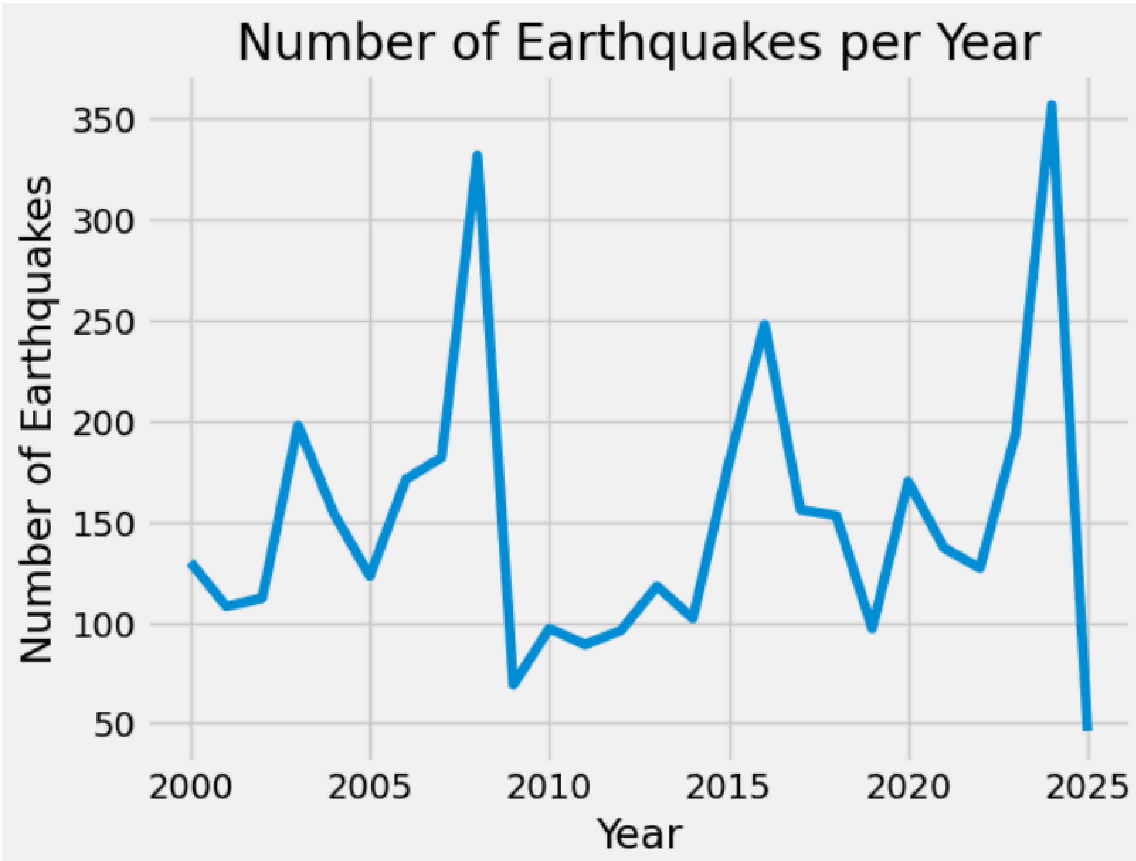
1. **Magnitude category pie chart**



minor: <4, moderate: 4-5, strong: >5

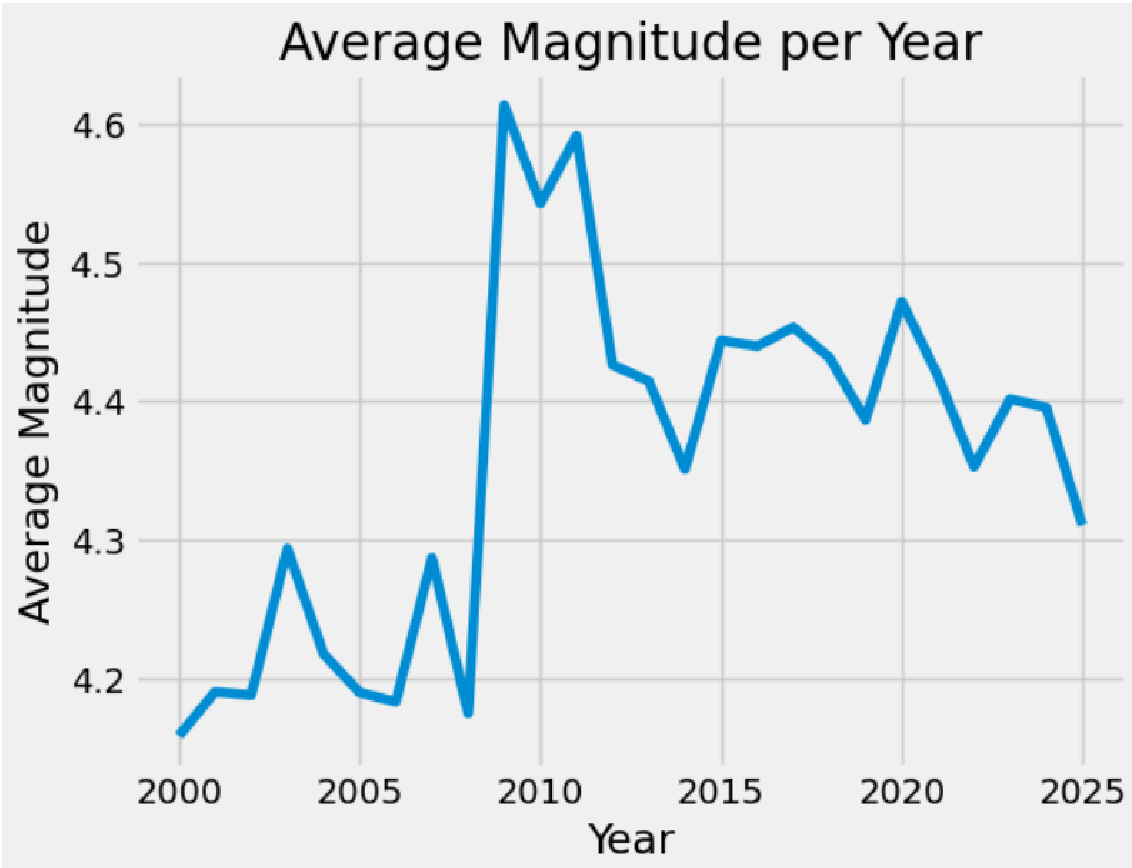
Plot shows that 75.8% of earthquakes had moderate magnitude, 19.7% minor, 4.5% strong

2. **Number of earthquakes by year**



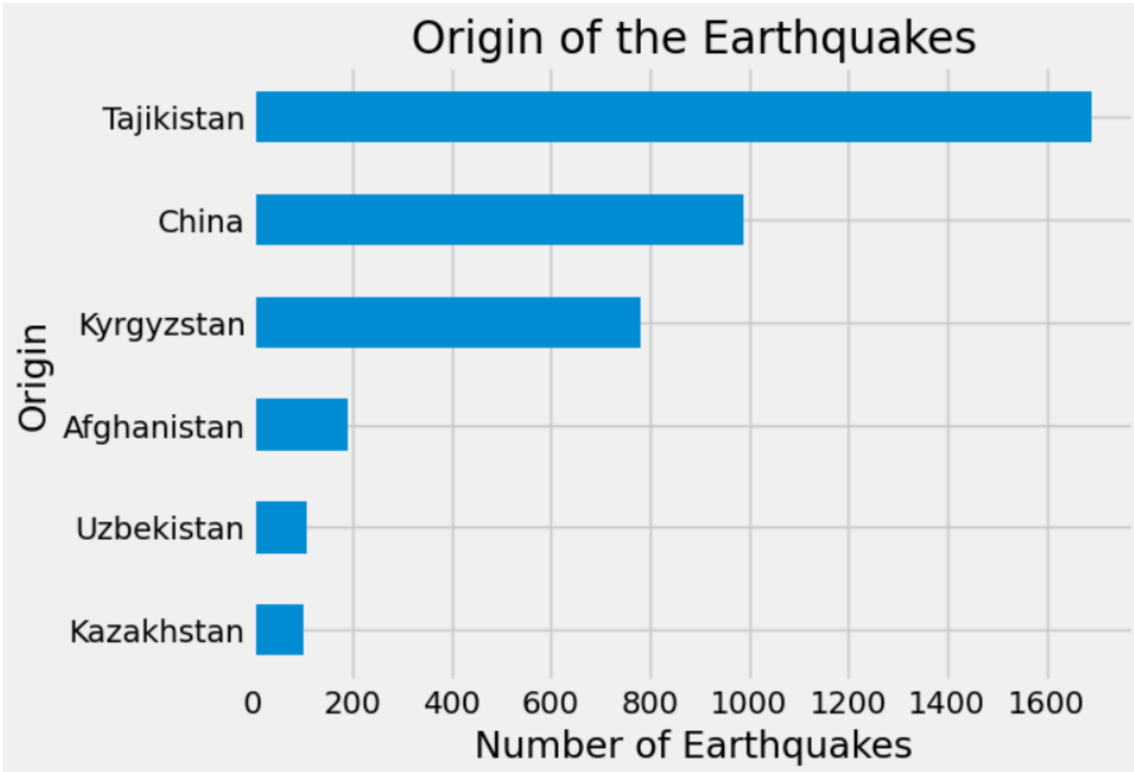
Peak years: 2024 and 2008, more than 300 earthquakes happened these years, we can see that number of earthquakes is not stable

3. **Average number of earthquakes**



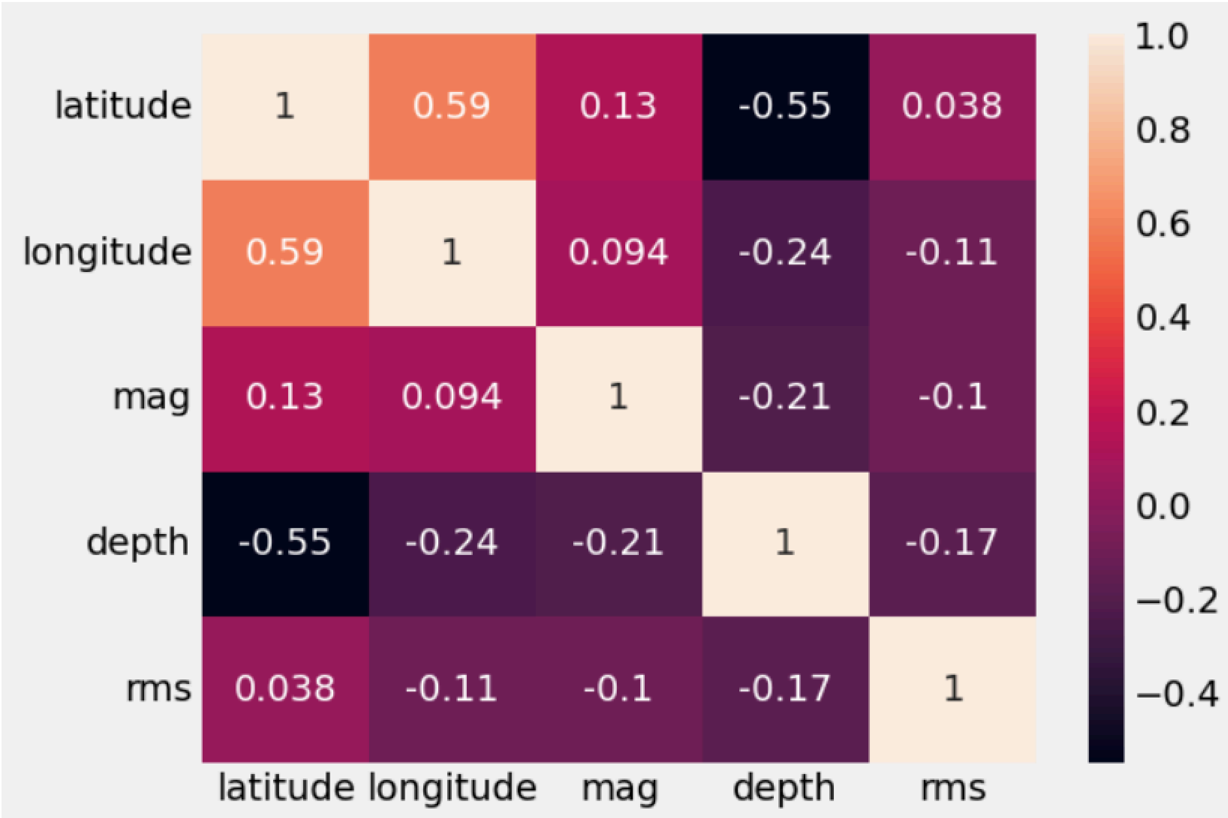
After 2008 year, the average magnitude of earthquakes increased, but since 2020 is decreasing

4. Origin of earthquakes



As we can see most of the earthquakes originate from our neighbors: Tajikistan and China, I would say more than 60% of earthquakes

5. Heatmap of numeric variables



There is not so much correlation between variables, only latitude and longitude, latitude and depth, if latitude and longitude correlation is obvious, but latitude - depth negative correlation is definitely insight

Conclusion

The EDA and visual analysis revealed several important insights:

Shallow earthquakes dominate the region, increasing potential surface impact.

The average magnitude is moderate (~4.3), but occasional high-magnitude events indicate a latent seismic risk.

Earthquake frequency patterns may be influenced by geological or seasonal factors.

Spatial clustering in the Tien Shan region aligns with known tectonic zones, confirming fault-line vulnerability.

Data quality appears high (based on low RMS), supporting reliable conclusions.

🔍 **Implications:**
The study supports the need for proactive urban planning in high-risk areas.

Emergency response and infrastructure resilience should focus on shallow earthquakes.

Continuous seismic monitoring and public education are essential for disaster risk reduction.

Reference

Central-Asian Institute for Applied Geosciences - link
USGS Earthquake Catalog - link