## Section 1: Introduction to the Case (4 marks)

*Clear Problem Definition*: The central problem we aim to address with our static data visualisation system is the lack of accessible and comprehensible earthquake data for the general public and researchers. Earthquakes are natural disasters with far-reaching impacts, and timely, informative data is crucial for monitoring and understanding seismic activity.

Our primary goal is to develop a real-time earthquake visualisation system with a user-friendly interface for accessing and analysing earthquake data. We want to empower users to gain meaningful insights from this data, contribute to public awareness, and support scientific research on seismic activity.

*Narrative*: The narrative of our project is centred on the idea of exploring the dynamic nature of the Earth's seismic activity. We'll take users on a journey through real-time earthquake events, historical trends, and the quest to identify patterns in seismic activity. By doing so, we aim to engage our audience in the fascinating world of earthquakes.

*Target Audience*: Our target audience includes the general public, students, educators, researchers, and anyone interested in earthquake monitoring and research. We want to make earthquake data accessible and comprehensible to a wide range of users.

## Introduction to the Dataset

The United States Geological Survey (USGS) Earthquake Catalogue API is a useful repository of seismic events that have been recorded across the globe. This dataset contains a rich history of earthquakes, tracing back to the early 20th century and extending to the most recent tremors felt today.

At its core, the USGS Earthquake Catalogue is more than just a collection of data points; it represents our planet's dynamic and ever-evolving nature. Each entry in the Catalogue is a testament to the Earth's tectonic movements, capturing details such as the time, location, depth, magnitude, and many more of each seismic event.

One of the standout features of this dataset is its real-time data acquisition, ensuring that researchers and enthusiasts alike have access to the most up-to-date information on earthquakes as they occur. This real-time capability, combined with its historical depth, offers a unique perspective on the patterns and behaviours of earthquakes over time. Furthermore, the global coverage of the dataset ensures a comprehensive view of seismic activities, irrespective of where they occur. From minor tremors in remote regions to major quakes in populous cities, the USGS Earthquake Catalog leaves no stone unturned.

The integrity and reliability of the data are paramount. With rigorous data collection and validation processes, the USGS ensures that the information presented is accurate and trustworthy. This commitment to data integrity makes the USGS Earthquake Catalog API an indispensable tool for researchers, policymakers, and anyone keen on understanding the seismic movements of our planet.

In choosing this dataset for our project, we aim to harness its depth, breadth, and reliability to derive meaningful insights and contribute to the broader understanding of earthquake phenomena.

# API Documentation

The Earthquake Catalog API is based on the FDSN Event Web Service Specification. It offers tailored searches for earthquake data using various parameters. For automated applications, it's recommended to utilize Real-time GeoJSON Feeds for optimal performance and availability.

## Base URL

[https://earthquake.usgs.gov/fdsnws/event/1/[METHOD[?PARAMETERS](https://earthquake.usgs.gov/fdsnws/event/1/%5bMETHOD%5b?PARAMETERS)]]

## Methods

* **application.json:** Retrieves known parameter values for the interface.
* **application.wadl:** Fetches the WADL for the interface.
* **catalogs**: Lists available catalogs.
* **contributors**: Lists available contributors.
* **count**: Counts data requests using the same parameters as the query method. Available in plain text, geojson, and xml formats.
* **query**: Submits a data request. Parameters for the URL are detailed below.
* **version**: Retrieves the full-service version number.

## Query Parameters

* **Formats**: If no format is specified, quakeml is the default. Available formats include CSV, GeoJSON, KML, plain text, and XML.
* **Time:** Uses the ISO8601 Date/Time format. If no timezone is mentioned, UTC is assumed.
* **Location:** Can specify a rectangle or circle for location-based searches.
* **Other:** Includes parameters like catalog, contributor, eventid, and more.
  + catalog:
    - ***Type:*** String
    - ***Description:*** Filters events from a specific catalog. To see available catalogs, use the Catalogs Method. If both catalog and contributor are omitted, the most preferred information from any catalog or contributor for the event is returned.
  + contributor:
    - ***Type:*** String
    - ***Description:*** Filters events contributed by a specific entity. To see available contributors, use the Contributors Method.
  + eventid:
    - ***Type:*** String
    - ***Description:*** Selects a specific event by its ID. Event identifiers are specific to the data center.
  + includeallmagnitudes:
    - ***Type:*** Boolean
    - ***Description:*** If set to true, all magnitudes for the event are included. By default, only the preferred magnitude is suggested.
  + includeallorigins:
    - ***Type:*** Boolean
    - ***Description:*** If set to true, all origins for the event are included. By default, only the preferred origin is suggested.
  + includearrivals:
    - ***Type:*** Boolean
    - ***Description:*** If set to true, phase arrivals are included. However, this is not currently implemented.
  + includedeleted:
    - ***Type:*** Boolean or "only"
    - ***Description:*** If set to true, deleted products and events are included. If set to "only", only deleted events are returned.
  + includesuperseded:
    - ***Type:*** Boolean
    - ***Description:*** If set to true, superseded products are included. This also includes all deleted products.
  + limit:
    - ***Type:*** Integer (range: 1 to 20,000)
    - ***Description:*** Limits the results to a specified number of events. The maximum limit is 20,000.
  + maxdepth and mindepth:
    - ***Type:*** Decimal (range for maxdepth: -100 to 1,000 km, for mindepth: -100 to 1,000 km)
    - ***Description:*** Filters events based on depth. Maxdepth limits to events with depth less than the specified maximum, while mindepth limits to events with depth more than the specified minimum.
  + maxmagnitude and minmagnitude:
    - ***Type:*** Decimal
    - ***Description:*** Filters events based on magnitude. Maxmagnitude limits to events with a magnitude smaller than the specified maximum, while minmagnitude limits to events with a magnitude larger than the specified minimum.
  + offset:
    - ***Type:*** Integer (range: 1 to ∞)
    - ***Description:*** Returns results starting at the specified event count, beginning at 1.
  + orderby:
    - ***Type:*** String
    - ***Description:*** Orders the results based on specified criteria such as time or magnitude.

## Extensions

Parameters like alertlevel, callback, eventtype, and more can be used to further refine the search.

You can refer to the original documentation for a deeper dive into each method and parameter. Link: [API Documentation - Earthquake Catalog (usgs.gov)](https://earthquake.usgs.gov/fdsnws/event/1/)

## Data Analysis

## Overview of dataset

The dataset contains retrieved from USGS API for the last 10 years. For the magnitude above 5, it resulted in information about 16,939 earthquakes; for each earthquake, there are 27 different attributes or data points.

Figure 1 gives a quick overview of the dataset's content and structure as follows:

* **Identify Columns**: Preview of the dataset's columns and sample data.
* **Check Data Quality**: Initial assessment for any missing values or inconsistencies in the top rows.
* **Recognize Geospatial Data**: Spot the type of spatial data, typically in the 'geometry' column.
* **Observe Patterns**: Quick glance at any evident trends or patterns.

A close-up of a white page

Description automatically generated

Figure 1. Overview of the Earthquake Data

* **Determine Dataset Scope**: Initial insight into the temporal or spatial range of the earthquake data. We notice that most of the features are not relevant to the objective of the visualisation system. After selecting the features desired ('geometry', 'mag', 'place', 'time', 'type', 'title'), we get the following dataset overview in figure 2.

A screenshot of a computer

Description automatically generated

Figure 2. Overview of the filtered Earthquake Data

## Statistical Summary

Figure 3 provides a concise statistical summary that offers insights into the distribution, spread, and quality of the dataset as follows:

* **Central Tendency**: Understand average values, like mean earthquake magnitude (5.34).
* **Data Spread**: View the range of data from minimum to maximum values and interquartile ranges. The min is magnitude 5, the same as the query we requested from the server; the max is 8.3 in the last 10 years.
* **Data Distribution**: Assessing the distribution through standard deviation and percentiles. We can tell that the magnitudes are right-skewed distributions.
* **Data Quality**: The feature ‘time’ should not be in the statistical summary. It requires more attention.

A black screen with numbers and numbers

Description automatically generated

Figure 3. Statistical Summary

## Datatypes and Missing Values

Figure 4 and Figure 5 guide how we analyse the earthquake data, and missing values can influence the insights we derive from the USGS dataset.

* **Datatype**
* Determines the kind of data (e.g., earthquake magnitude as a number, location as coordinates).
* The feature ‘time’ is in an easy-to-read format. It requires conversion.
* **Missing Values**
* Only the feature ‘place’ shows 207 missing values.
* The missing values will not affect the visualisations. Since “Geometry” is the most important feature we will use. As a result, we don’t see the need to apply any imputation or removal of the missing values.

A black screen with white text

Description automatically generated A screenshot of a computer code

Description automatically generated

Figure 4. Data Types Figure 5. Missing Values

## Earthquake Type

The 'type' column differentiates seismic events into three categories:

* **Earthquake (**16882**)**: Natural seismic events caused by tectonic processes.
* **Volcanic Eruption (**55**)**: Seismic activities associated with volcanic actions.
* **Nuclear Explosion (**3**)**: Seismic events triggered by human-made nuclear detonations.

This differentiation helps distinguish between natural and man-made seismic activities and their potential impacts.

## Temporal Analysis

**Hypothesis:**

*"Has the frequency of earthquakes changed over the specified time, and if so, is there an observable trend in the increase or decrease of occurrences?"*

Figure 6 visually represents the frequency of monthly earthquakes over the last decade. The x-axis represents time (in months), while the y-axis represents the number of earthquakes that occurred in each month.

**Trend Over Time**

* Upward line: Increasing earthquake frequency. (Highest in 2021)
* Downward line: Decreasing frequency. (Lowest towards the end of 2013)

A graph with blue lines

Description automatically generated

Figure 6. Monthly Earthquake Count

Over the past decade, the frequency of earthquakes has varied. However, overall, earthquake frequency has no consistent visual trend over the entire period.

**Hypothesis:**

*Are there specific months or seasons during which earthquakes occur more frequently?"*

Figure 7, heatmap, provides a quick visual summary of when (which months) and how strong (magnitude) earthquakes occurred over the 10 years. A concise breakdown of the visualisation is as follows:

* X-axis: Months (Jan to Dec).
* Y-axis: Denotes the earthquake magnitude, grouped into specific ranges such as 5-6, 6-7, etc.
* Darker shades: Indicate months with a higher frequency of earthquakes. For example, March recorded 1462 earthquakes, September had 1401, and August reported 1394.
* Lighter shades: Represent months with a lower number of earthquakes. For instance, June had 1146 earthquakes, July saw 1172, and October experienced 1187.
* The majority of earthquakes are clustered in the lower magnitude range, specifically 5-6, where the occurrences are in the thousands. In contrast, the 6-7 magnitude range witnesses a drop to the hundreds.
* Earthquakes with an extremely high magnitude, such as 9-10, are absent from the record.

A chart with numbers and a few months

Description automatically generated with medium confidence

Figure 7. Earthquake Frequency by Month and Magnitude

Figure 8 presents a bar chart depicting the monthly count of earthquakes, segmented by their respective magnitudes. Superimposed on this is a line graph showing the monthly average magnitudes of these earthquakes. A concise breakdown of the visualisation is as follows:

* The magnitude categories within each month are distinguished by varying colours in the stacked bar chart.
* A prominent red line traversing the bar chart illustrates each month's average magnitude of earthquakes.
* Peaks or upward surges in the line graph denote months where earthquakes had, on average, a higher magnitude.
* The chart utilises dual Y-axes: the left Y-axis displays the number of earthquake occurrences, while the right Y-axis showcases the average magnitude.
* Months characterised by taller bars, such as March and August, witnessed a more frequent occurrence of earthquakes.
* In instances where the red line climbs to greater heights, it suggests a month with a notably higher average earthquake magnitude. For instance, August recorded an average magnitude of 5.36, indicating a more potent seismic activity month.

A graph with a red line

Description automatically generated

Figure 8. Monthly Frequency and Average Magnitude

There is a potential seasonal variation in the frequency of earthquakes, with certain months like March and August having a higher occurrence. While most earthquakes fall within the lower magnitude range, August experienced a higher frequency and notably increased average earthquake magnitude, suggesting more potent seismic activity during this month.

## Spatial Analysis

**Hypothesis:**

*"Are there particular geographic regions or 'hotspots' where earthquakes occur more frequently compared to other areas?"*

Figure 9 showcases global earthquake occurrences over the past decade, identifying both high-risk seismic zones and less active areas. Key takeaways include:

* Seismic Zones: Regions with dense earthquake occurrences signify active seismic zones.
* Magnitude:
  + Blue (5-6): Common and widespread; usually causes minimal damage.
  + Green (6-7): Less frequent; can cause significant localised damage.
  + Yellow (7-8): Major quakes with wider damage radii.
  + Orange (8-9): Rare, extremely powerful, with vast damage zones.
  + Red (9-10): Extremely rare but devastating.
* High-Risk Areas: Regions with orange and red indicators witnessed the most powerful quakes in the last decade.
* Diverse Seismic Activity: Areas with a colour mix experienced varied magnitudes over the decade.
* Safe Zones: Sparse or absent markers signify minimal seismic activity over the decade.

A map of the world

Description automatically generated

Figure 9. Seismic Zones

Figure 9 maps out seismic zones in relation to tectonic plate boundaries, underscoring the significant interplay between plate movements and earthquake occurrences.

A map of the world

Description automatically generated

Figure 10. Seismic zones in the Ring of Fire.

Figure 10 delves deeper into zones with the highest magnitudes, revealing a marked concentration around the "Ring of Fire." This is an important observation, as the "Ring of Fire" is renowned for its frequent earthquakes and numerous active volcanoes, making it a critical region for seismic research and monitoring.

A map of the world with red dots

Description automatically generated

Figure 11. Top 10 highest magnitude earthquakes

Figure 11 further solidifies this point by showcasing the top 10 highest magnitudes, all pinpointed within the "Ring of Fire." Such concentrated seismic activity in this region underscores its vulnerability and the potential risks to the communities residing along its perimeter. To complement these visual insights, a table is provided detailing specific information on these top 10 seismic events, which can serve as a valuable reference for researchers and policymakers alike.

A screenshot of a black and white screen

Description automatically generated

Table 1. Top 10 earthquakes in the last 10 years.

The analysis strongly supports the hypothesis. It's evident from Figure 9 that seismic zones are closely aligned with tectonic plate boundaries, suggesting a direct relationship between plate movements and earthquake occurrences. Most notably, the "Ring of Fire", which is a major area in the Pacific Plate where several tectonic plates converge, exhibits a significantly higher frequency of earthquakes, especially those of higher magnitudes, as highlighted in Figures 10 and 11. This confirms that specific geographic regions, influenced by their tectonic settings, experience more seismic activity than others.

## Magnitude Analysis

**Hypothesis:***“Is there a relationship between depth and magnitude of the earthquake?”*

Figure 12 presents a histogram illustrating the distribution of earthquake magnitudes. The distribution is right-skewed, indicating that lower magnitude earthquakes are more frequent. A significant majority of earthquakes, particularly in the magnitude range of 5 to 5.5, dominate the dataset.

A graph of earthquake magnitude

Description automatically generated

Figure 12. Distribution of earthquake magnitudes

Figure 13 categorises the earthquakes based on their magnitude:

* Moderate (5.0 - 5.9): These earthquakes make up the bulk of the data with 15,556 occurrences, accounting for 91.84% of the total.
* Strong (6.0 - 6.9): They are less frequent, with 1,241 instances contributing to 7.33% of the dataset.
* Major (7.0 - 7.9): Such events are rarer, recorded 134 times, making up just 0.79%.
* Great (8.0 and above): The rarest of the categories, these extremely powerful earthquakes have occurred only 8 times, representing a mere 0.05% of the total.

A graph with blue squares

Description automatically generated

Figure 13. Categorisation of earthquake magnitudes

In summary, while lower magnitude earthquakes (especially those between 5 and 5.5) are predominant, higher magnitude earthquakes are exceedingly rare, emphasising the exceptional nature of extremely powerful seismic events.

## Depth Analysis

Figure 14 offers a histogram detailing the distribution of earthquake depths. This right-skewed distribution showcases that shallower earthquakes are substantially more common than their deeper counterparts. Most of these earthquakes have depths of less than 100 Km. Interestingly, there's a notable uptick in the frequency around the 600 Km depth, suggesting another layer of seismic activity at this specific depth.

A graph of a earthquake

Description automatically generated

Figure 14. Distribution of earthquake depths

Figure 15 classifies earthquakes according to their depth:

* Shallow Earthquakes (Less than 70Km): This category accounts for the largest portion of the dataset with 14,055 occurrences, making up 82.98% of the total.
* Intermediate Earthquakes (between 70Km and 300Km): They are less frequent with 2,233 instances, comprising 13.18% of the earthquakes.
* Deep Earthquakes (More than 300Km): These are the rarest depth category, with only 650 instances, translating to 3.84% of the dataset.

A graph with blue rectangular bars

Description automatically generated with medium confidence

Figure 15. Categorisation of earthquake depths

While shallow seismic events are predominant, deeper ones are considerably less frequent. The notable increase around 600 Km depth might interest seismologists, potentially pointing to specific geophysical phenomena or characteristics at that depth.

Figure 16 visualises the relationship between earthquake depth and magnitude. Scatter plot data points are enhanced by histograms at the top (depth distribution) and the right (magnitude distribution). These histograms confirm that earthquakes are more frequent at shallower depths and lower magnitudes.

A graph of a number of dots

Description automatically generated with medium confidence

Figure 16. Magnitude by Depth with Distribution of Depth and Magnitude

Figure 17 uses a color-coded scatter plot to represent this relationship further. Most data points are in the lower-left corner, indicating frequent, shallow, and lower-magnitude earthquakes. A smaller cluster in the top-right reveals that high-magnitude earthquakes also occur at significant depths.

A graph of a graph

Description automatically generated with medium confidence

Figure 17. Magnitude by Depth

From these visualisations, we infer that dangerous earthquakes (magnitude > 7.5) can occur at very shallow and deep depths. Despite the Pearson correlation coefficient being a mere 0.10, suggesting a weak linear relationship, the plots emphasise the importance of specific depth zones where high-magnitude earthquakes are more prevalent.

## Section 4: Data Storytelling (20 marks)

### Narrative:

The story takes the audience on a journey from the natural and inherent dynamism of the Earth's geological processes, specifically earthquakes, to humanity's relationship with these powerful forces. It progresses from understanding the phenomena to appreciating human efforts in predicting, preparing for, and mitigating the impacts of earthquakes.

The story concludes with a call to action, urging the audience to actively participate in the collective quest for understanding, preparedness, and resilience against earthquakes. While not explicitly stated, implicit actions may include supporting research, advocating for effective early warning systems, promoting earthquake education, and adopting individual preparedness measures.

Earthquakes are a fundamental part of our planet's geological processes and can occur anywhere.

The importance of smaller, frequent earthquakes in understanding Earth's seismic behaviour.

The role of tectonic plates in triggering earthquakes.

The correlation between an earthquake's depth and its magnitude.

The advancements and importance of early warning systems in earthquake-prone regions.

* A sense of awe and respect for the power and complexity of the Earth.
* A personal connection or empathy to those affected by earthquakes.
* An urgency and responsibility to participate in understanding and preparing for earthquakes.
* Hope and optimism stemming from human achievements in early warning systems and the potential to mitigate earthquake impacts.

To ensure their own safety and the safety of their communities.

To participate in a broader, collective effort to mitigate the impacts of natural disasters.

Inspired by humanity's progress and ingenuity in the face of challenges.

The story showcases the power and unpredictability of nature juxtaposed with human tenacity, innovation, and adaptability. Despite the might of earthquakes, humanity's unrelenting quest for knowledge and preparedness stands as a beacon of hope. The narrative's tone reinforces the idea that we can turn potential tragedies into stories of resilience and preparedness.

The language used in the narrative is evocative and poetic. It weaves scientific facts with a tone of reverence and admiration for both the Earth's natural processes and human determination. Phrases like "Earth's seismic dance," "relentless pursuit of knowledge," and "our collective strength, determination, and ingenuity" stir emotions and emphasize the interconnectedness of Earth's processes and human endeavours. The language serves to educate, evoke emotion, and inspire action.

### Storyline:

The storyline of your data visualization presentation can be understood as a structured progression, moving from introducing a broad topic (earthquakes) to delving deeper into the intricacies of its occurrence, human interactions with it, and ultimately a call to action. Here's a summarized version of your storyline:

Introduction: Earth's Dynamic Nature

* Introduction to earthquakes as timeless storytellers.
* Emphasis on their global significance through a World Heatmap.

Unearthing the Details: Seismic Patterns

* Highlighting both major and minor seismic events.
* Showcasing the consistent seismic activity via a Magnitude Distribution Histogram.
* Underlining the importance of even the smallest tremors in the larger context.

The Culprits: Tectonic Plates

* Explanation of tectonic plates and their movements.
* A visualization of tectonic plate boundaries to provide insight into earthquake-prone regions.
* The Science: Depth and Magnitude Correlation
* Delving into the relationship between the depth of an earthquake's origin and its magnitude.
* Using a scatter plot to illustrate the depth-magnitude relationship and its implications.

Human Adaptation: Prediction and Preparedness

* Tracing the evolution of earthquake early warning systems.
* Celebrating human achievement and resilience in the face of seismic threats.

The Practical Impact: Real-World Implications

* Explanation of Earthquake Early Warning Systems (EEWS) and their importance.
* Demonstrating the tangible, life-saving benefits of these systems.

Conclusion and Call to Action

* Recap of the narrative's major points.
* Encouraging the audience to contribute to the shared goal of understanding and preparing for earthquakes.

This storyline effectively transitions from a macro understanding of earthquakes as global phenomena to a micro-analysis of their causes and implications. By the end, it melds both perspectives to emphasize the human element and responsibility, compelling the audience to action.

### Story structure

**1. Beginning:**

**Setting the Stage: The Power of Earthquakes.** This introduces the grand theme of earthquakes as essential to the Earth's narrative and showcases their global prevalence through a World Heatmap.

**2. Middle:**

**Delving Deeper: Intensity and Frequency.** This segment dives into the importance of both minor and major seismic activities.

**Tectonic Foundations.** Here, the story addresses the root causes of earthquakes, the movement of tectonic plates, and their interactions.

**The Connection of Depth and Magnitude.** This portion explores the scientific relationship between the depth at which earthquakes occur and their resulting magnitudes.

**3. End:**

**The Human Quest: Predicting the Unpredictable.** The story shifts to human interaction with earthquakes, detailing our advancements in predicting and preparing for them.

**The Real Impact: Testimonies and Stories.** This highlights the real-world implications and the tangible benefits of Earthquake Early Warning Systems (EEWS).

**Conclusion and Call to Action.** This serves as a wrap-up, tying all the discussed elements together and urging the audience toward awareness and action.

Status Quo, Challenge, Proposal, Conflict, and Real Solution:

**Status Quo**: Earth is a dynamic planet with seismic events occurring regularly. These events have been integral to Earth's history and topography, as visualized by the World Heatmap.

**Challenge:** While major earthquakes capture global attention, it's the smaller, frequent ones that are often overlooked, despite their significance. Also, understanding the root causes (like tectonic interactions) is crucial.

**Proposal:** By studying and visualizing the entire spectrum of earthquake magnitudes, as well as the underlying tectonic movements, we can gain a more comprehensive understanding of seismic activities.

**Conflict:** Despite the advancements in understanding earthquakes, predicting them remains a considerable challenge. The unpredictable nature of earthquakes poses risks to lives and properties.

**Real Solution:** Through the combined efforts of technology, research, and community resilience, systems like the Earthquake Early Warning Systems (EEWS) have been developed. These systems, while not predicting earthquakes per se, offer a critical window of preparation, potentially saving lives and mitigating damages.

The storyline and structure effectively guide the audience from understanding the intrinsic nature of earthquakes to realizing the potential of human innovation and collaboration in addressing the challenges they pose.

### Creating an effective layouts

## Justification of Visuals/Graphics

It is essential to justify the choice of visual representations to ensure that the selected visual representations effectively convey the intended message to the audience. The right visuals can make data more understandable, insightful, and actionable. For our visualisations:

1. Style/Format:

a. Histograms (Figures 12 & 14):

Histograms are powerful tools to show the distribution of data. They allow for easy identification of central tendencies, spread, and skewness.

b. Heatmaps (Figure 7):

Heatmaps are ideal for representing data where colour intensity can signify magnitude or frequency. It makes it easier to spot trends or high/low points over a grid, like months in this case.

c. Scatter Plots (Figures 16 & 17):

Scatter plots are ideal for comparing two quantitative variables and identifying potential relationships or clusters within the data.

d. Bar Charts & Line Graphs (Figure 8):

Rationale: Bar charts are great for comparing discrete data points across categories, while line graphs are excellent for showcasing trends over a continuous domain, like time.

2. Visual Components:

a. Dual Y-axes (Figure 8):

Dual Y-axes allow for the comparison of two metrics with different scales in a single chart. It can aid in identifying correlations or disparities between the two datasets.

b. Color-Coding (Figure 17):

Different colours can represent different categories or magnitudes, making data segmentation easier to comprehend visually.

c. Enhancements with Histograms (Figure 16):

Adding histograms on the scatter plot's axes can provide additional information about the distribution of both variables.

d. Geospatial Data Visualization (Figure 9):

Spatial visualisations allow for understanding geographic patterns, clusters, or anomalies. They are essential when the data has a geographical context.

Choosing the right visual representation for the data is paramount to effective communication. The visual choices in the dataset overview are supported by literature and are grounded in best practices for data visualisation (Gilbert, 2014; Kirk, 2012; Knaflic, 2015).

## Section 6: Justification of Toolsets (3 marks)

*Toolset Explanation*: We selected Python as our primary programming language due to its robust data manipulation and visualization libraries. Libraries such as Matplotlib, Seaborn, Plotly, and Folium offer extensive capabilities for data visualization, making Python a suitable choice for our project.

Python is renowned for its simplicity and readability, making it an excellent choice for both beginners and experienced developers. It is easy-to-understand syntax allows for rapid development and efficient collaboration among team members. Python offers powerful libraries like Pandas, which excel in data manipulation tasks. We used Pandas which provides data structures like DataFrames, making it seamless to clean, transform, and manipulate data. The preprocessing tasks such as handling missing values, and structuring the data become easy with Pandas. Further, Python integrates seamlessly with web frameworks like Flask and Django. This integration enables to deploy interactive visualizations as web applications, making them accessible to a wider audience. Python's compatibility with Jupyter Notebooks also facilitates the creation of dynamic, interactive reports.

In addition Python provides a rich ecosystem of visualization Libraries such as pandas, matplotlib, seaborne etc which are discussed in detail below:

## Standard Libraries for Data Manipulation:

The toolsets and libraries used to analyse the earthquake dataset are among the most popular and efficient in the Python ecosystem (Jupyter Notebooks), especially for data analysis, visualisation, and geospatial analysis. Here's a justification for each:

**pandas (pd):** Data manipulation and analysis.

Pandas is one of the most widely used libraries in the data science field. Its primary data structures, Series and DataFrame, enable data manipulation tasks ranging from subsetting to aggregation.

**numpy (np):** Mathematical computations and operations on arrays and matrices.

Numpy is the core library for numerical computation in Python. It provides support for arrays (including multidimensional arrays) and an assortment of mathematical functions to operate on these arrays. Its tight integration with pandas makes it invaluable for data analysis.

## Libraries for Visualization:

**matplotlib (plt):** Visualization and plotting.

Matplotlib: Matplotlib is one of the oldest and most popular Python libraries for data visualization. It provides a solid foundation for creating static plots, including line charts, bar charts, and histograms. It is a comprehensive library for creating static, interactive, and animated visualisations in Python. It is foundational and provides extensive flexibility in creating various plots and charts.

**seaborn (sns):** Statistical data visualisation.

Seaborn is a data visualisation library based on Matplotlib. It provides a high-level interface for drawing attractive statistical graphics. Its integration with pandas DataFrames makes it a natural choice for visual exploratory data analysis. It simplifies the process of creating aesthetically pleasing and informative statistical graphics, enhancing the visual appeal of your plots.

## Geospatial Libraries and Tools:

**mpl\_toolkits.basemap (Basemap):** Geospatial visualization and mapping.

Basemap is an extension to Matplotlib that allows for the creation of geospatial maps. It offers various map projections, and its integration with Matplotlib makes it easy to overlay data on maps.

**geopandas (gpd):**

Handling geospatial data in Python.

Geopandas extends the pandas library to allow spatial operations on geometric types. Geometric operations are performed by Shapely, and GeoPandas also depends on Fiona for file access and Matplotlib for plotting.

**shapely.geometry:**

Manipulation and analysis of geometric objects.

Shapely is used for geometric operations. It is particularly useful for shape manipulations, spatial queries, and geometric aggregations.

## Utilities:

**io.BytesIO:** Read or write binary data in memory.

BytesIO provides a convenient means of working with byte data in memory, often used when fetching data from web requests or APIs.

**requests:** Making HTTP requests.

When retrieving data from online sources, like APIs, the requests library is among the most popular and user-friendly tools to get this task done in Python.

**datetime (dt):** Handling of dates and times.

The datetime module supplies classes to work with date and time. Given that the earthquake dataset would contain temporal data (like timestamps), this module becomes essential for parsing, formatting, and manipulating these timestamps.

## Making it Interactive

In Group Major Assignment Part 2, we plan to make our static data visualisation system interactive by implementing features such as zoom and pan capabilities on maps, user-friendly filtering options for date range and magnitude range, and the ability to overlay geological features on the map.

By combining the functionalities of Python libraries such as Seaborne, Plotly, Folium etc. we can create a comprehensive and interactive visualisation of earthquake data. While Matplotlib and Seaborn are handy for static visualisations, and Plotly and Folium would be employed for interactive and dynamic data visualisations. A brief overview of the interactive elements and features are listed below:

1. **Selection of Python Library and Basic Map Creation:**

Choose Bokeh as the primary Python library for visualisation due to its extensive interactive features and capabilities for creating dynamic web-based reports and dashboards. Bokeh would be employed to create a basic interactive map as the foundation of the interactive visualisation.

1. **Implementation of Interactive Features:**

*Tooltips and Hovers:* When users hover over data points, we would add tooltips and hovers to display relevant information (such as earthquake magnitude, location, and date).

*Filters:* We would implement user-friendly filtering options for date range and magnitude range, allowing users to filter earthquakes based on specific time periods and magnitude levels.

*Location-Based Filters:* This feature would allow users to filter earthquakes based on specific g*eographical locations, regions or countries.*

1. **Zoom and Pan Capabilities:**

Then, we would implement zoom and pan capabilities to let users explore the map in greater detail. Users should be able to zoom in/out and pan across the map to focus on specific regions.

1. **Overlay Geological Features:**

We would also like to provide an option to overlay geological features (such as fault lines and tectonic plate boundaries) on the map. This would help enhance the context of earthquake occurrences.

1. **Real-time Streaming and Live Updates:**

We would also like to implement real-time streaming of earthquake data and live update visualisation to provide users with up-to-the-minute information.

1. **Deployment Options:**

We would deploy the interactive visualisation as a web application or within a Python environment (such as Jupyter Notebook).

1. **Interactive tool to be employed:**

**We would be primarily using Plotly and Folium for creating in**teractive data visualisations.

We would use interactive Heat Maps tools using Plotly, which is particularly powerful for creating interactive visualisations. This tool would help us create interactive heat maps where users can zoom in, hover over data points to see details and customise the view according to their preferences. Plotly's heatmap function and its interactive features provide a dynamic way to explore earthquake data geographically.

Plotly can also create interactive scatter plots, where users can hover over points to view specific earthquake details, enabling an interactive and engaging experience.

Folium is another powerful Python library designed specifically for visualising geospatial data. We would create interactive maps with earthquake data plotted on specific geographical locations. Folium supports various toolsets, allowing us to customise the map's appearance. Additionally, we can overlay markers, circles, and polygons on the map to effectively represent earthquake occurrences and their magnitudes.

## 

## References

Gilbert, M. (2014). *Data stories - how to combine the power storytelling with effective data visualization*. Www.slideshare.net. https://www.slideshare.net/miriamgilbert08/data-stories-workshop-34390209

Kirk, A. (2012). *Data visualization : a successful design process : a structured design approach to equip you with the knowledge of how to successfully accomplish any data visualization challenge efficiently and effectively.* Packt Publ.

Knaflic, C. N. (2015). *Storytelling with data : a data visualization guide for business professionals*. Wiley.