EARTHQUAKE IMPACT ASSESSMENT USING GEOSPATIAL AND SEISMIC PARAMETERS

Jalla Srinidhi[1], Saniya Almas[2], Bandari Sanjana[3]

UG Scholar Department of AI&DS, Methodist College of Engineering and Technology, Hyderabad, India

Dr Diana Moses[4]

Professor, Department of CSE, Methodist College of Engineering and Technology, Hyderabad, India

ABSTRACT

This dataset documents significant global earthquake events, capturing key seismic attributes and geospatial data. The records span multiple high-magnitude earthquakes from mid-2023, detailing parameters such as magnitude, date and time, intensity (CDI/MMI), alert levels, tsunami warnings, and geographic coordinates. Each event is identified by its epicenter location, along with data on depth, seismic significance, and source network. The dataset includes both automated and reviewed information, providing insights into the seismic activity in regions such as Vanuatu, El Salvador, Argentina, and Alaska. This compilation supports earthquake risk assessment, geophysical studies, and emergency preparedness analysis by offering structured, timely data on impactful seismic events.

Introduction

Literature Survey:

S. Mostafa Mousavi et al., proposed STanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI. A large-scale earthquake and non-earthquake dataset enhances machine learning model training, improving earthquake detection and seismic monitoring efficiency. Its scale and accuracy pave the way for improved earthquake detection and deeper insights into Earth's dynamic processes [1].

Kristine L. Verdin et al., proposed Development of a Global Slope Dataset for Estimation of Landslide Occurrence Resulting from Earthquakes.The USGS PAGER system uses SRTM global slope data to assess landslide risk post-earthquake, with statistical methods filling data gaps for rapid impact assessment.A global topographic dataset from SRTM data helps assess landslide potential after earthquakes, with methods to fill gaps and extrapolate beyond 60°N. It supports global landslide modeling [2].

Yoshihisa Maruyama et al., proposed Fragility curves for expressway embankments based on damage datasets after recent earthquakes in Japan. After the 1995 Kobe earthquake, Japan improved its expressway seismometer network. Fragility curves, based on earthquake damage data, show major damage occurs when PGV exceeds 35.0 cm/s, aiding quicker disaster response.Fragility curves for expressway embankments were developed using earthquake damage data, showing major damage occurs when PGV exceeds 45.0 cm/s [3].

Subash Ghimire et al., proposed Testing machine learning models for seismic damage prediction at a regional scale using building-damage dataset compiled after the 2015 Gorkha Nepal earthquake.This study uses random forest regression to assess post-seismic damage from the 2015 Nepal earthquake, achieving 0.68 accuracy by combining building data with seismic intensity values. This study uses machine learning with basic building features for scalable seismic damage prediction, proposing crowd-sourced databases for data acquisition [4].

Roberto Paolucci et al., proposed BB-SPEEDset: a validated dataset of broadband near-source earthquake ground motions from 3D physics-based numerical simulations.The BB-SPEEDset is a broadband earthquake ground motion dataset from 3D simulations using the SPEED code. It includes time histories and ground motion measures, validated against real data. Its features align with the NESS v2.0 dataset.The BB-SPEEDset is a dataset of simulated near-source earthquake ground motions, validated against real data. It supports seismic hazard analysis and structural design [5].

R. H. Jones et al., proposed A method for determining significant structures in a cloud of earthquakes. This study proposes adjusting earthquake hypocenter locations within uncertainty ellipsoids to simplify and interpret 3D data. Applied to synthetic and real data from Rabaul Caldera, it reveals potential deviations from assumed fault structures. Sensitivity analysis shows the method’s robustness to assumptions.The method simplifies earthquake locations using uncertainty bounds to reveal underlying fault structures, as shown in Rabaul Caldera [6].

Weiqiang Zhu et al., proposed California Earthquake Dataset for Machine Learning and Cloud Computing. The CEED dataset compiles detailed California earthquake data from 2000–2024 to support deep learning and seismic research. It integrates records from NCEDC and SCEDC, offering rich event-based data for modeling and analysis.The CEED dataset unifies high-quality seismic data from California to support AI-driven earthquake detection and analysis. Its open-access format enables scalable, global research on seismic activity and fault processes [7].

Mufti Mahmud et al., proposed Application of Artificial Intelligence in Predicting Earthquakes: State-of-the-Art and Future Challenges. This study reviews 84 papers on AI-based earthquake prediction, analyzing rule-based, machine learning, and deep learning techniques. It compares their performance by datasets and metrics, guiding future research and technique selection. AI methods improve earthquake prediction accuracy, reducing damage.This review compares 84 studies (2005–2019) to guide future research [8].

Mark Stirling et al., proposed Comparison of Earthquake Scaling Relations Derived from Data of the Instrumental and Preinstrumental Era. The research updates regression relations for surface rupture displacement and magnitude, highlighting differences between preinstrumental and modern data due to natural censoring of smaller earthquakes. Preinstrumental regressions show higher estimates than W and C, with differences due to censoring. After adjusting, both era regressions align [9].

Papiya Debnath et al., proposed Analysis of Earthquake Forecasting in India Using Supervised Machine Learning Classifiers. Machine learning classifiers were used to forecast earthquake magnitudes in three categories: fatal, moderate, and mild. Simple Logistic and LMT performed well across datasets from India. This paper used machine learning algorithms to forecast earthquake types in India, with Simple Logistic and LMT classifiers achieving the highest accuracy rates [10].

Materials And Methods

Dataset:

The dataset appears to contain detailed records of earthquakes, likely significant events from around the world. It contains **499 rows** and **18 columns**, each representing an individual earthquake event.The dataset is organized into rows and columns, where each **row** represents a single earthquake event.

Attribute overview:

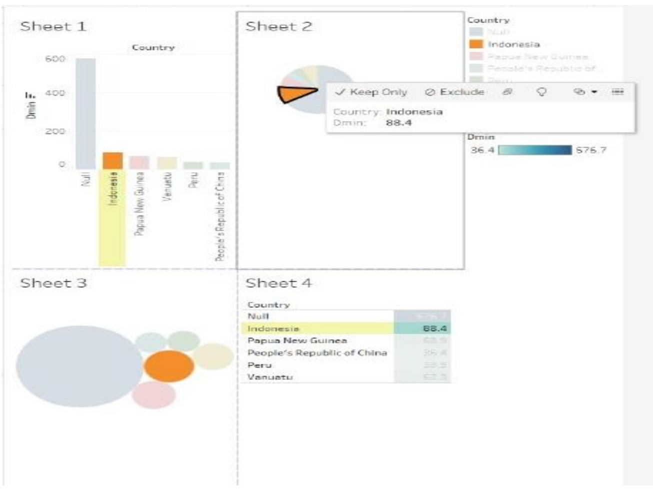
Sl. No.	Attribute Name	Description	Example	Datatype
1	title	Summary of the event including magnitude and location	"M 6.5 - 42 km W of Sola, Vanuatu"	String
2	magnitude	Magnitude of the earthquake	6.5	Float
3	date_time	Date and time of the event	"2023-08-16 12:47:00"	Datetime
4	cdi	Community Internet Intensity	7	Integer/Float
5	mmi	Modified Mercalli Intensity	4	Integer/Float
6	alert	Alert level issued based on impact severity	"green", "yellow"	String
7	tsunami	Indicates if a tsunami was generated (1 = yes, 0 = no)	0	Integer
8	sig	Significance score of the event	657	Integer
9	net	Network ID that reported the event	"us", "at"	String
10	nst	Number of seismic stations that detected the quake	114	Integer
11	dmin	Distance to nearest station (in degrees)	7.177	Float
12	gap	Azimuthal gap between stations (in degrees)	25.0	Float
13	magType	Type of magnitude calculation used	"mww", "Mi"	String
14	depth	Depth of the earthquake (in kilometers)	192.955	Float
15	latitude	Latitude of the epicenter	-13.8814	Float
16	longitude	Longitude of the epicenter	167.1580	Float
17	location	Description of the earthquake's location	"Sola, Vanuatu"	String

Sl. No.	Attribute Name	Description	Example	Datatype
18	continent	Continent where the event occurred (may contain missing values)	"South America"	String
19	country	Country of the epicenter (may contain missing values)	"Argentina"	String

Software:

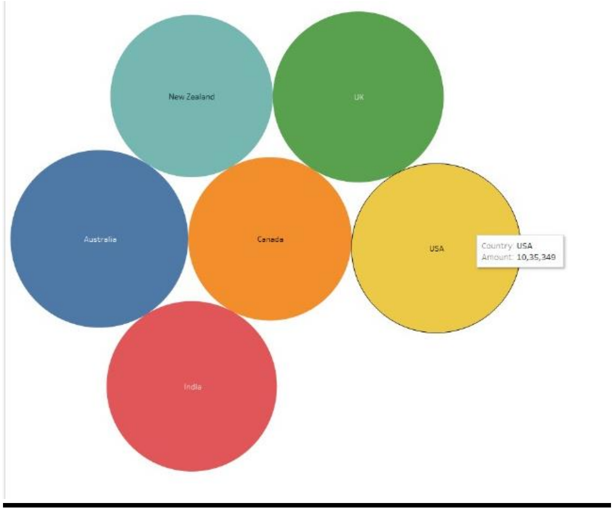
Tableau is a powerful data visualization and business intelligence tool that helps users create interactive, easy-to-understand charts, graphs, dashboards, and reports from complex datasets. It allows for real-time data analysis and supports connectivity to a wide variety of data sources like spreadsheets, databases, and cloud services. Tableau is widely used in industries for making data-driven decisions, as it emphasizes visual analytics, enabling users to explore and share insights quickly without needing advanced programming skills. Its features include drag-and-drop functionality, data blending, and advanced calculations, making it a preferred choice for both beginners and professionals in data analytics.

Data Visualization



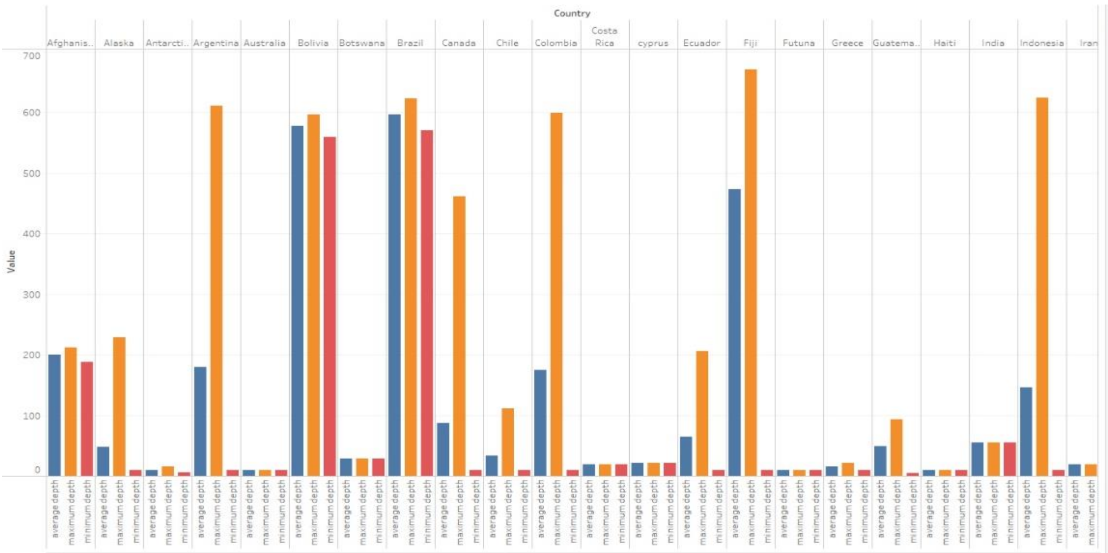
This image shows a Tableau dashboard combining multiple visualizations (bar chart, pie chart, bubble chart, and table) to display country-wise data. Indonesia stands out with a value of 88.4, highlighted in both the pie chart and data table.

The dashboard visualizes and compares the "Dmin" metric across countries, with Indonesia highlighted showing a Dmin value of 88.4.



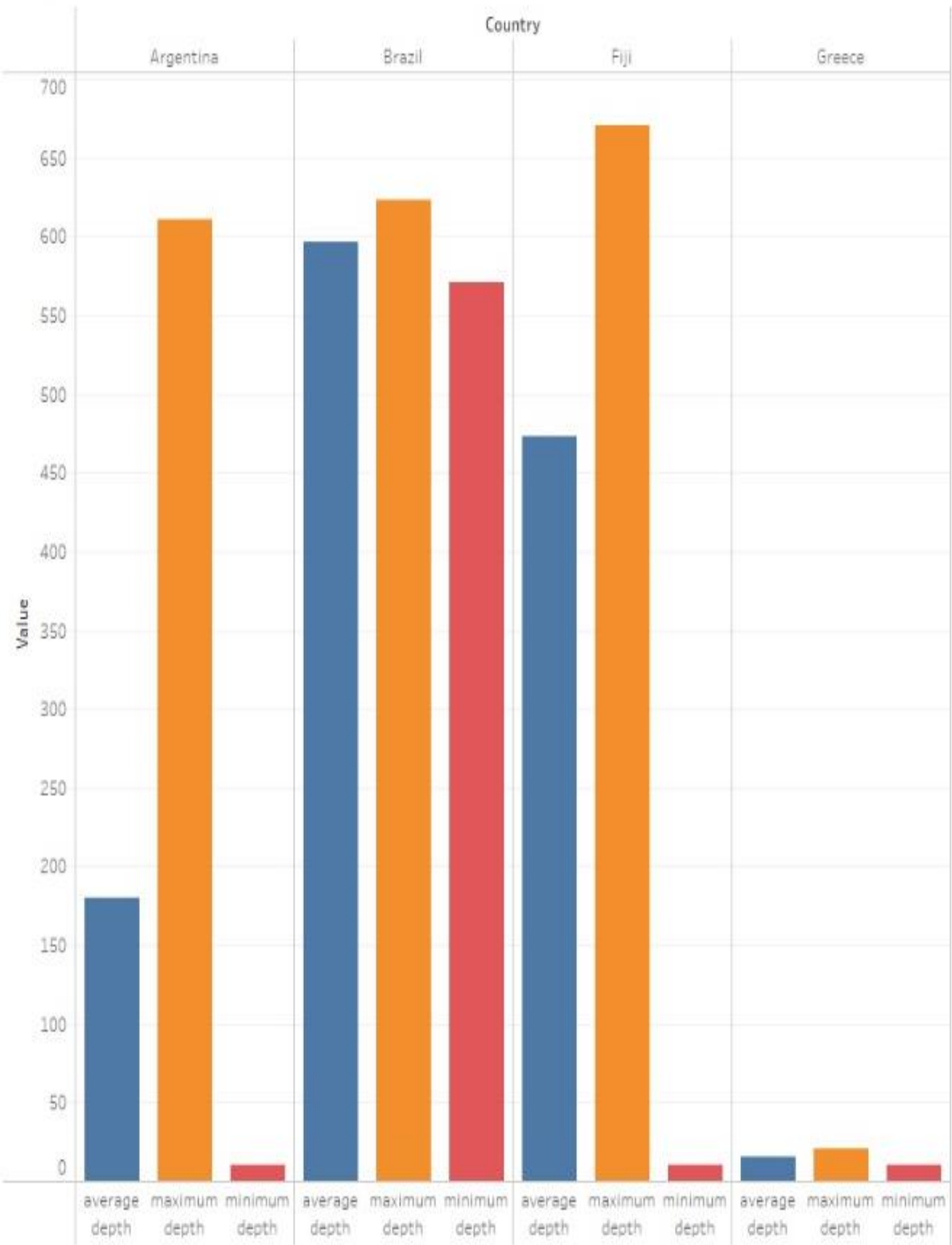
This image shows a bubble chart created in Tableau, representing the sum of amounts by country. Each bubble's size corresponds to the total amount, with the USA having the highest value displayed as 10,35,349.

The bubble chart displays different countries sized by an "Amount" metric, with the USA having an amount of 30,75,340.



This bar chart compares the number of cases across various countries using three different colored bars, likely representing different categories or time periods.

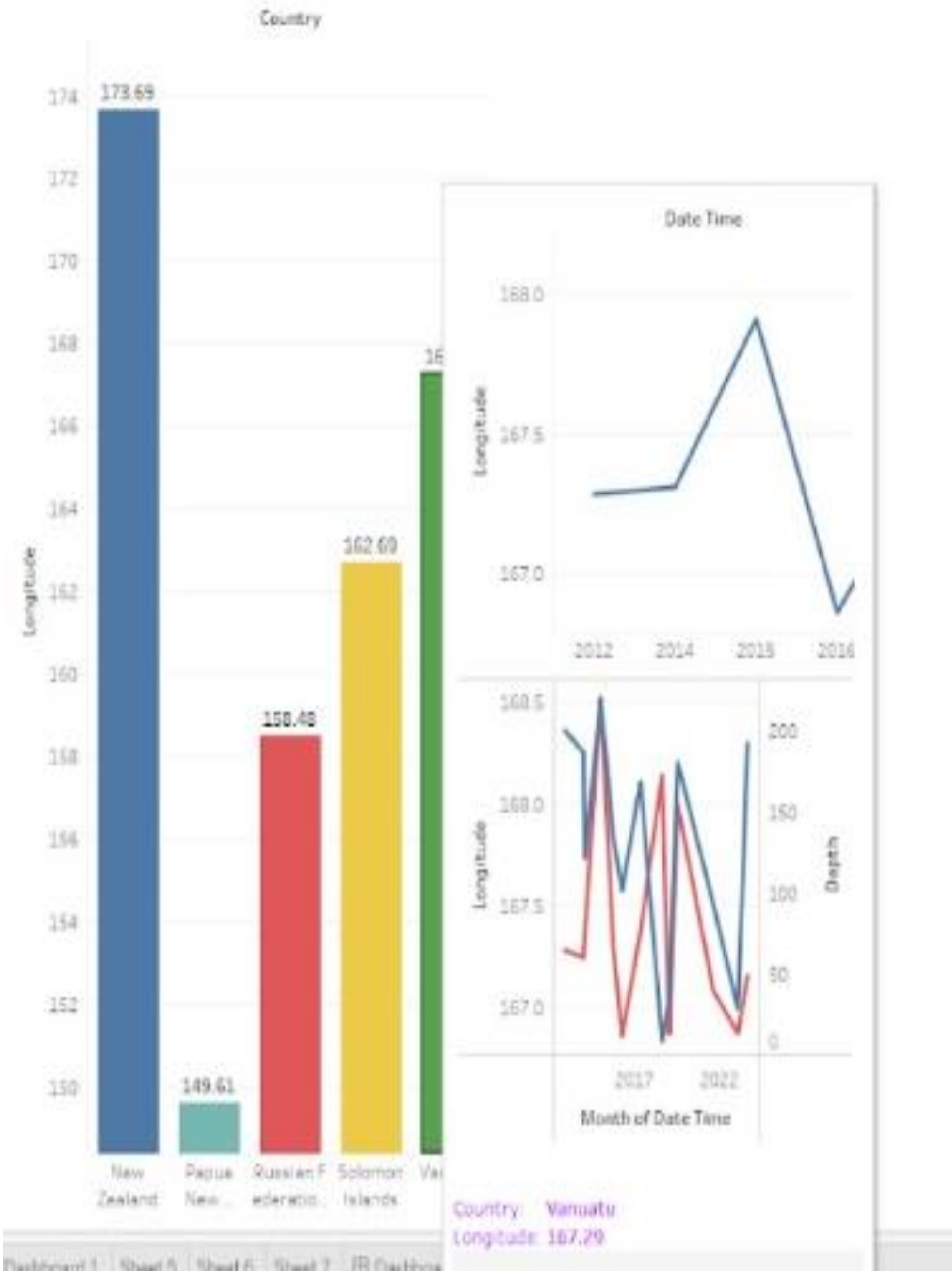
The bar chart compares values across multiple countries and categories, with noticeable variations in heights indicating differing data distributions.



ISSN: 2583-4584

This bar chart compares average, maximum, and minimum depth values for countries like Argentina, Brazil, and Egypt. Each group of three bars per country likely represents those depth metrics with different colors.

This bar chart compares the average, maximum, and minimum depths across four countries, showing Fiji with the greatest maximum depth and Greece with the shallowest overall values.

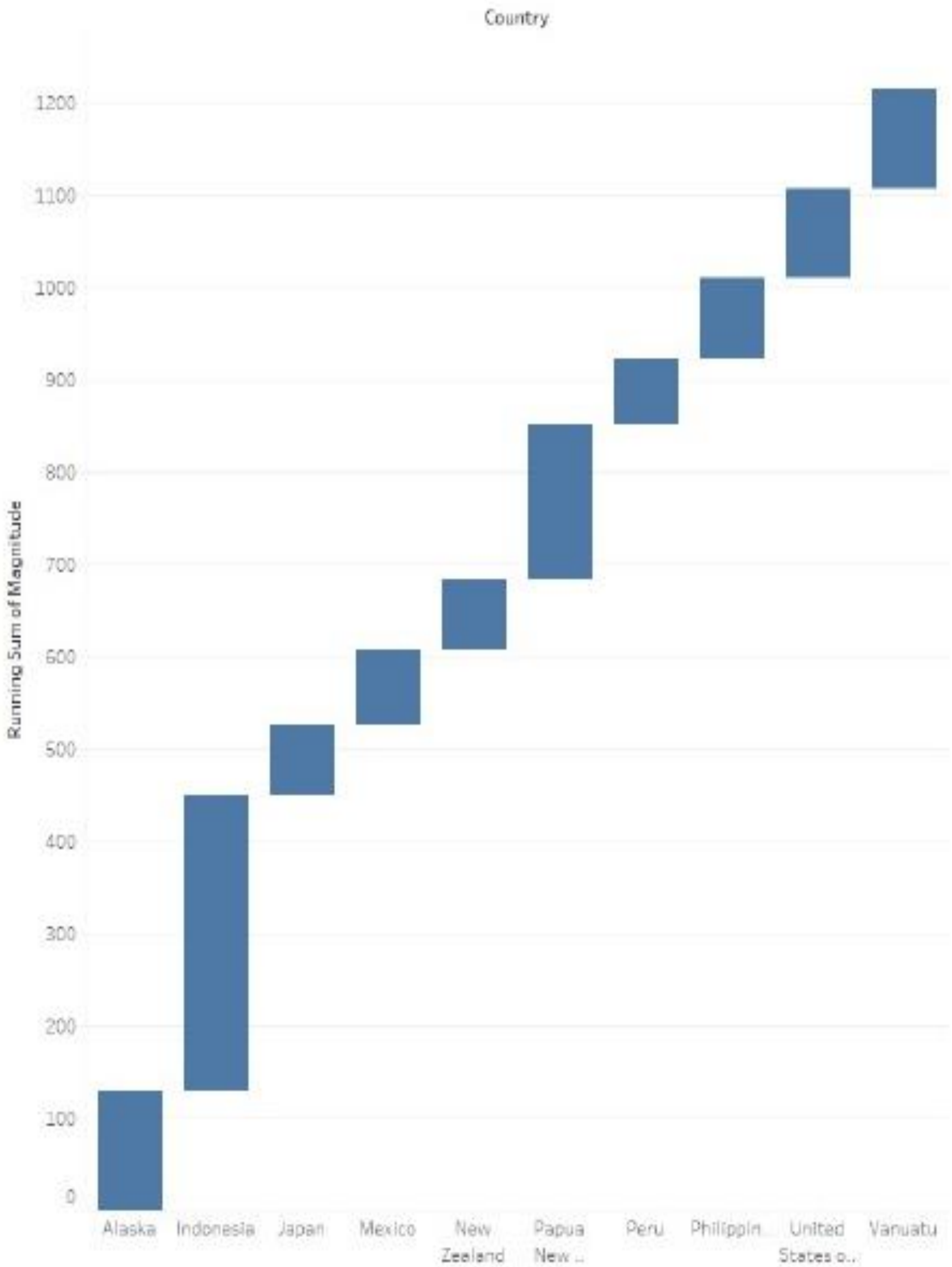


This Tableau dashboard visualizes longitudinal data by country using bar and line charts. Vanuatu is highlighted with a longitude of 167.89, and time-series charts show variations across years and months.

This visualization shows longitudinal data by country and over time, highlighting that New Zealand has the highest longitude value and Vanuatu exhibits notable fluctuations in depth and longitude over time.

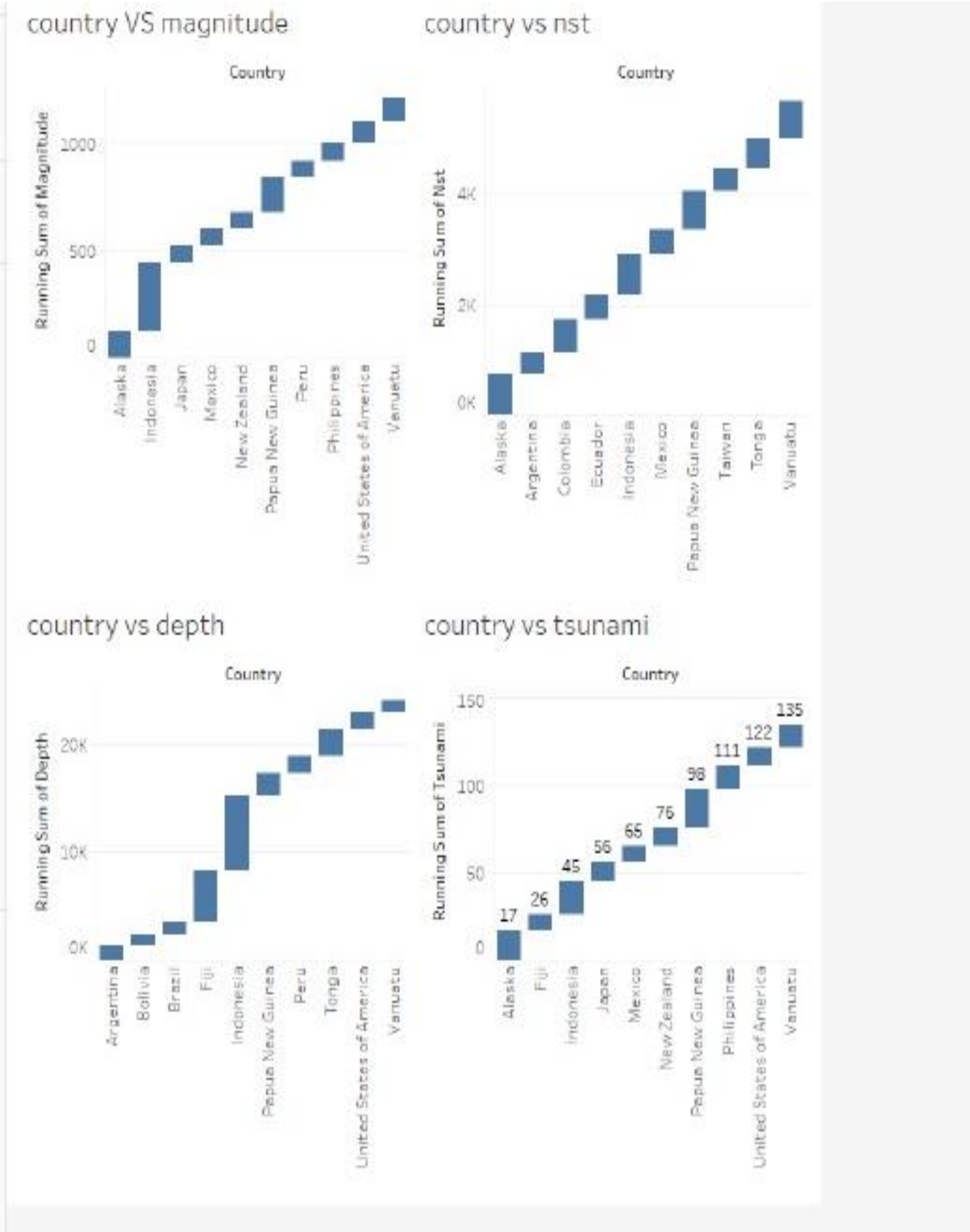
This heatmap from a Tableau dashboard displays longitudinal data by country from 2016 to 2023, with color intensity indicating value magnitude. Notably, Indonesia in 2019 and Vanuatu in 2016 show specific highlighted values (61.80 and 6.50 respectively).

This heatmap shows the intensity of values (possibly earthquake depths or frequencies) across countries and years, with Indonesia in 2019 having the highest recorded value of 61.80.



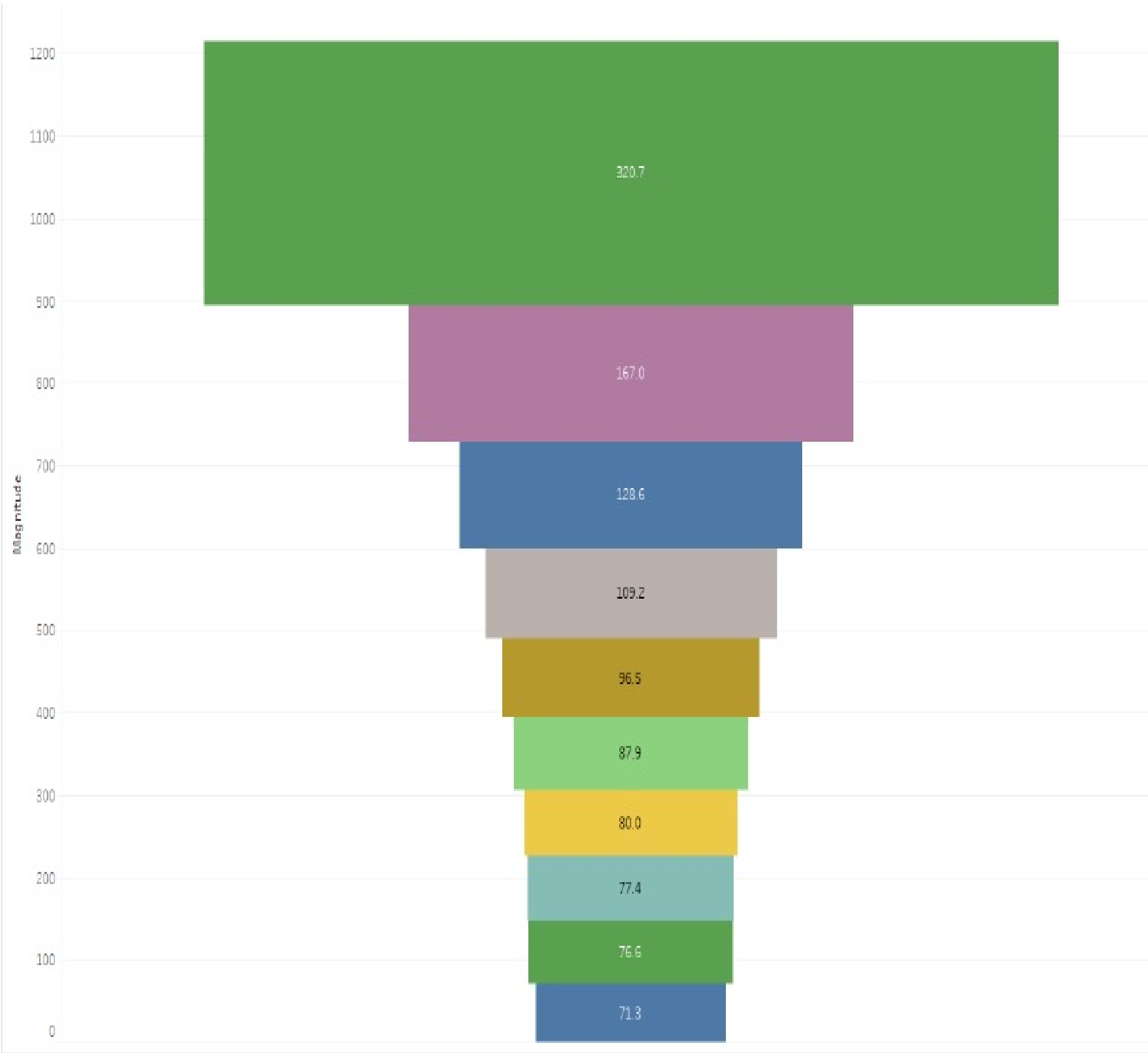
This bar chart shows the cumulative sum of a variable (possibly cases or events) across countries, arranged in ascending order. Vanuatu and the United States appear to have the highest totals, while Alaska has the lowest.

This waterfall chart displays the running sum of earthquake magnitudes by country, showing Indonesia contributing the largest individual increase to the cumulative total.



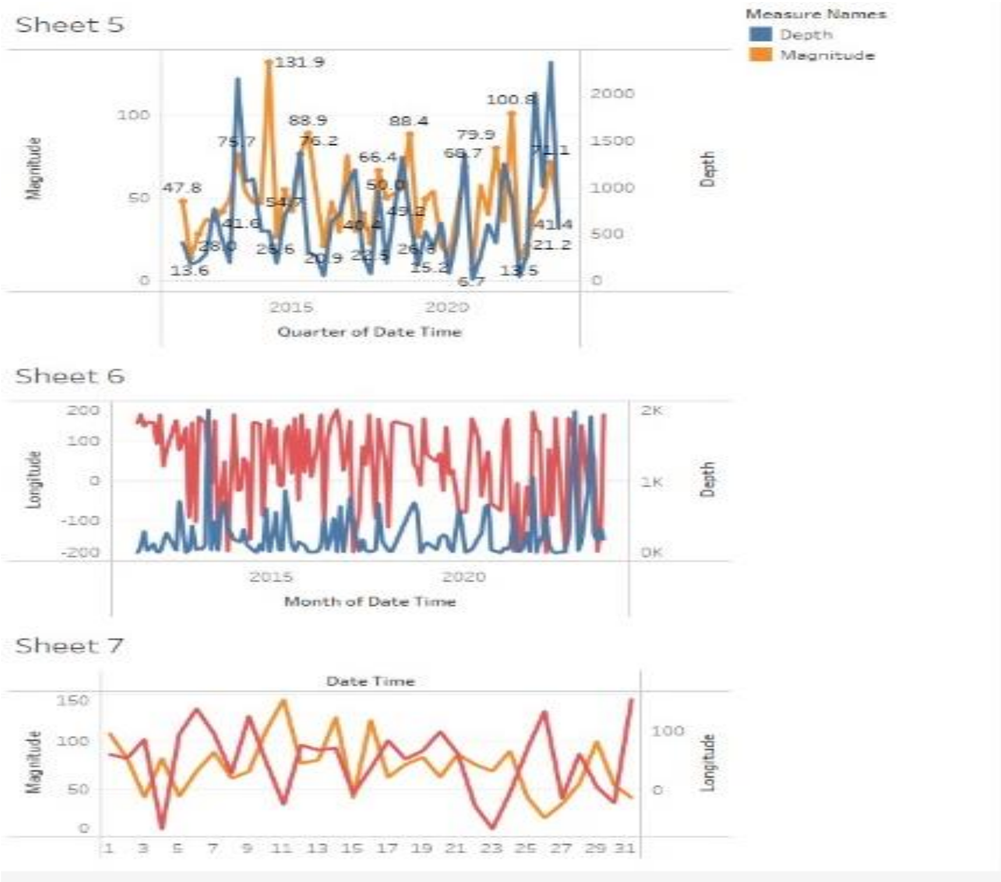
The charts compare countries by total magnitude, stations, depth, and tsunami events, with Vanuatu and the U.S. leading most metrics.

These four charts show cumulative values by country, indicating that Indonesia and Vanuatu significantly contribute to earthquake magnitude, depth, tsunami counts, and sensor readings (nst).



This funnel chart visually represents a decreasing trend across categories, with the top segment having the highest value and each subsequent segment getting progressively smaller. It highlights the largest drop-off occurring between the first and second stages.

This funnel chart displays a decreasing sequence of values, with the highest value being 1027 and each subsequent segment representing progressively smaller quantities.



These three time-series charts display trends over time: Sheet 5 compares magnitude and depth quarterly, Sheet 6 shows monthly changes in depth and longitude, and Sheet 7 presents daily longitude fluctuations.

These charts show fluctuations in earthquake magnitude, depth, and longitude over time, highlighting no consistent pattern but evident variability across different time frames.

Result & Discussion

The dataset contains a total of 499 earthquake events with magnitudes ranging from 6.5 to 8.6 and an average depth of approximately 100.2 kilometers. Out of these, 325 earthquakes were associated with tsunami alerts, indicating a significant portion had the potential to impact coastal regions. The events occurred across 45 different countries, highlighting a broad global distribution of seismic activity. Most earthquakes were recorded by the US network and exhibited varying alert levels, with many categorized under "green" or "yellow" warnings. The number of stations used in detection varied, as did the distance from the epicenter and the calculated signal significance, suggesting a diverse set of seismic occurrences in both remote and populated areas.

The dataset highlights the global occurrence and potential impact of strong earthquakes, most of which fall within the 6.5 to 8.6 magnitude range. The average depth of around 100 kilometers suggests many of these quakes are intermediate-focus, which can lead to significant shaking over a wide area. The high number of tsunami alerts (65% of the events) emphasizes the relevance of this data for coastal hazard assessment. With 45 countries affected, the geographic distribution underscores seismic vulnerability around tectonic boundaries, particularly in regions like the Pacific Ring of Fire. The presence of missing or misencoded data in location names and countries points to a need for data cleaning to ensure accurate analysis. Overall, the dataset provides meaningful insights into the patterns, risks, and characteristics of recent major earthquake events worldwide.

Conclusion

This earthquake dataset covers events with magnitudes ranging from 6.5 to 8.6 and an average depth of 100 km. It captures global seismic activity, particularly along tectonic plate boundaries like the Pacific Ring of Fire. 65% of recorded events triggered tsunami alerts, highlighting coastal risks. The dataset spans 45 countries and includes key parameters like signal strength and alert levels. It has limitations such as missing values and character encoding issues. With data cleaning and visualization, it can aid research and improve earthquake preparedness.

The dataset tracks significant earthquakes globally, focusing on those with magnitudes between 6.5 and 8.6, including data on depth, tsunami alerts, and network involvement. It spans 45 countries, offering insights into seismic risks, especially along tectonic plate boundaries. Its benefits include improving early warning systems, aiding research, and supporting disaster response and planning. However, it has limitations like missing values, which can be addressed with data cleaning.

References

[1]. Mousavi, S. Mostafa, et al. "STanford EArthquake Dataset (STEAD): A global data set of seismic signals for AI." IEEE Access 7 (2019): 179464-179476.

[2]. Verdin, Kristine L., et al. Development of a global slope dataset for estimation of landslide occurrence resulting from earthquakes. No. 2007-1188. US Geological Survey, 2007.

[3]. Teamah, Abd-Elmonem AM, Ahmed A. Elbanna, and Ahmed M. Gemeay. "FRÉCHET-WEIBULL DISTRIBUTION WITH APPLICATIONS TO EARTHQUAKES DATA SETS." Pakistan Journal of Statistics 36.2 (2020).

[4]. Ghimire, Subash, et al. "Testing machine learning models for seismic damage prediction at a regional scale using building-damage dataset compiled after the 2015 Gorkha Nepal earthquake." Earthquake Spectra 38.4 (2022): 2970-2993.

[5]. Paolucci, Roberto, Chiara Smerzini, and Manuela Vanini. "BB-SPEEDset: A validated dataset of broadband near-source earthquake ground motions from 3D physics-based numerical simulations." Bulletin of the Seismological Society of America 111.5 (2021): 2527-2545.

[6]. Jones, R. H., and R. C. Stewart. "A method for determining significant structures in a cloud of earthquakes." Journal of Geophysical Research: Solid Earth 102.B4 (1997): 8245-8254.

[7]. Zhu, Weiqiang, et al. "California Earthquake Dataset for Machine Learning and Cloud Computing." arXiv preprint arXiv:2502.11500 (2025).

[8]. Al Banna, Md Hasan, et al. "Application of artificial intelligence in predicting earthquakes: state-of-the-art and future challenges." IEEE Access 8 (2020): 192880-192923.

[9]. Stirling, Mark, David Rhoades, and Kelvin Berryman. "Comparison of earthquake scaling relations derived from data of the instrumental and preinstrumental era." Bulletin of the Seismological Society of America 92.2 (2002): 812-830.

[10]. Debnath, Papiya, et al. "Analysis of earthquake forecasting in India using supervised machine learning classifiers." Sustainability 13.2 (2021): 971.