Analysis Report:

Tsunami Study (1750 -2023)

By: Raul Chavez

December 2023

Introduction

This study investigates tsunami events from 1750 to 2023, a period that begins with the onset of the Industrial Revolution—a time of significant human impact on the environment. Tsunamis are large, potentially devastating ocean waves usually triggered by seismic activities such as underwater earthquakes, volcanic eruptions, or landslides. The analysis aims to chart the frequency, magnitude, and impacts of tsunamis, exploring how these factors have changed over the epochs marked by increasing human activity. We examine regions most frequently affected by tsunamis and delve into the geophysical processes responsible for their inception. Additionally, we consider whether industrialization and subsequent human interventions have influenced the incidence or severity of these natural phenomena. Our goal is to discern any patterns and trends that can enhance our understanding of tsunamis for improved prediction and mitigation strategies in the future.

Methodology

This study was conducted using data from the National Centers for Environmental Information (NCEI) and the National Oceanic and Atmospheric Administration (NOAA), providing a robust catalog of tsunamis from 1750 to 2023. The data was retrieved as a TSV file, chosen for its comprehensive coverage and authoritative sourcing. Initial data preparation and cleaning were performed in Jupyter Notebooks, an environment well-suited for such tasks in data science. This phase involved meticulous inspection for data completeness, consistency, and formatting. Redundant columns were removed, missing values were addressed through median imputation or exclusion, and data types were converted to formats conducive to analysis. These steps were critical in ensuring the dataset's quality before the analysis.

The analysis was carried out using Python's rich ecosystem of libraries, including pandas for data manipulation, numpy for numerical computations, and matplotlib and seaborn for visualization. Univariate analysis on numerical variables provided insights into their distributions, while multivariate analysis explored relationships between variables, with a particular focus on identifying correlations that could indicate patterns or trends. The final stage of visualization was conducted in Tableau, where interactive maps and charts were created to depict the geographical spread and temporal changes of tsunami events. This two-pronged approach, combining data analysis with visualization techniques, allowed for an in-depth exploration of the tsunami data, facilitating the extraction of meaningful insights.

Data Cleaning and Preparation

In the initial stage of data exploration, cleaning, and wrangling, the analysis began with a DataFrame comprising 2393 entries and 51 columns. This inspection revealed significant missing data across several columns and a variety of data types. To refine the dataset, a series of cleaning and preprocessing steps were undertaken. Columns with high percentages of missing values or those deemed irrelevant for the analysis, such as 'Unnamed: 0', 'Id', and 'Hr', were dropped. Adjustments were also made to the data types of specific columns to facilitate easier analysis. For instance, columns like 'Year' and 'Deaths' were converted to integers, and others like 'Tsunami Event Validity' were categorized. Additionally, columns with predominantly missing values, like 'Tsunami Magnitude (Abe)', were identified and considered for exclusion. Addressing the issue of missing values was a multifaceted approach: missing values in natural phenomena measurements, such as 'Earthquake Magnitude', were replaced with median values, missing 'Area' values were filled in with 'NA', and the mean values of similar entries were used to fill in gaps in 'Latitude' and 'Longitude'. These steps were crucial in ensuring the reliability and accuracy of the subsequent analysis.

Data Analysis

The data, sourced from the National Geophysical Data Center (NGDC) and the National Oceanic and Atmospheric Administration (NOAA).

In the data analysis phase of the study, a thorough examination of the tsunami dataset was conducted using various statistical techniques. Univariate analysis was performed on numerical variables using histograms and box plots, which provided detailed insights into the distribution, range, and skewness of key variables such as 'Earthquake Magnitude' and 'Maximum Water Height'. This analysis helped identify characteristics like central tendencies and outliers within individual data points. For the categorical variables, count plots were utilized to analyze the frequency distribution of categories, notably in fields like 'Tsunami Event Validity' and 'Tsunami Cause Code'. These plots offered a clear visual representation of the prevalence of different categories, highlighting the most common characteristics of tsunami events. Additionally, a correlation matrix was generated to delve into the relationships between various numerical variables. This matrix was particularly insightful, revealing significant correlations, such as the strong relationship between 'Deaths' and 'Maximum Water Height (m)', which suggested a potential link between the height of the tsunami wave and the resulting fatalities. This part of the analysis was critical in understanding the underlying patterns and relationships within the data, providing a foundation for more complex analyses and interpretations.

Findings and Insights

Increase in Tsunami Occurrences
The temporal analysis underscored a discernible increase in tsunami frequency over time, a trend that became pronounced in the 20th century. This upswing could potentially be linked to heightened seismic activity or advancements in detection technology.

Geographic Hotspots for Tsunamis
The study identified the Pacific Ring of Fire as a significant hotspot for tsunami activity. Countries along this rim, particularly Japan, Indonesia, and the United States, are frequently impacted. This finding is crucial for directing future resources and preparedness efforts.

Causation Analysis
Consistent with scientific consensus, earthquakes emerged as the primary trigger for tsunamis. The data showed that seismic activity, especially those with significant magnitudes and shallow focal depths, was intimately linked to the generation of these devastating waves.

Correlation of Tsunami Characteristics with Impact
A pivotal discovery was the strong correlation between tsunami deposits and the severity of their impact, both in terms of human fatalities and economic damages. This finding underscores the potential of geological surveys to inform risk assessments and early warning systems.

The Human and Economic Toll
The study laid bare the economic and human toll of tsunamis, particularly in the most affected regions. The deadliest tsunamis recorded in the dataset were those with the highest number of fatalities and the most extensive property damage. Notably, the 2004 tsunami in Indonesia marked one of the deadliest in recent history, with the project's analysis providing a detailed breakdown of the damages incurred during such catastrophic events.

Limitations and Recommendations

The tsunami project's limitations include incomplete historical data, particularly from the 18th and 19th centuries, which may skew trend analysis due to unrecorded or inaccurately reported events. Recent improvements in data collection methods could give a false impression of an increase in tsunami frequency. While immediate effects like fatalities and economic damage are recorded, long-term impacts on communities and ecosystems are not thoroughly examined. The dataset may also lack critical variables that contribute to a tsunami's impact, such as local population density or infrastructure resilience. Additionally, the influence of statistical outliers and the distinction between correlation and causation are areas that require careful consideration. Advanced statistical or predictive modeling could further strengthen the analysis by illuminating factors that contribute to tsunamis and enhancing forecasting and mitigation strategies.

Conclusion

This study has provided valuable insights into the historical patterns and impacts of tsunamis, using data from NGDC and NOAA. The methodologies employed have facilitated a detailed understanding of tsunami events, highlighting areas for future improvement. As the study concludes, the findings serve as a basis for future research, which should focus on developing predictive models to further our understanding and preparedness for tsunami events.

This report has employed a systematic approach to analyze and interpret tsunami data, providing a comprehensive understanding of the trends and patterns of these natural events. By building on the work presented here, future studies can enhance our ability to predict and mitigate the risks associated with tsunamis, ultimately contributing to the safety and resilience of coastal communities worldwide.