Earthquake Data Analysis - Project Summary

# Introduction

This project aimed to analyze global earthquake data to understand the patterns, impacts, and potential predictive factors associated with seismic events. By conducting a comprehensive analysis, the project sought to extract valuable insights that could inform future research and mitigation efforts.

# Project Brief

The Earthquake Data Analysis project involved several key steps:  
1. \*\*ETL Process\*\*: Extracting, transforming, and loading the raw earthquake data to prepare it for analysis.  
2. \*\*Data Statistics\*\*: Providing descriptive statistics to summarize the central tendencies and variability in the data.  
3. \*\*Exploratory Data Analysis (EDA)\*\*: Visualizing distributions, correlations, and other patterns within the dataset.  
4. \*\*Clustering and Predictive Modeling\*\*: Segmenting the data into meaningful clusters and developing predictive models to assess economic impacts.  
5. \*\*Hypothesis Testing\*\*: Performing statistical tests to validate findings and explore differences between earthquake clusters.  
The project utilized Python for data processing, analysis, and visualization, leveraging libraries such as Pandas, Seaborn, and Scikit-learn.

# ETL (Extract, Transform, Load) Process

The data was extracted from raw files, cleaned by handling missing values, converting data types, and normalizing key variables. This process ensured that the data was ready for analysis.

# Data Statistics

Descriptive statistics provided insights into the central tendencies, dispersion, and distribution of the data. Key metrics such as average magnitude, depth, and economic impact were calculated.

# Exploratory Data Analysis (EDA)

EDA involved visualizing the distribution of earthquake magnitudes, economic impacts, and the correlations between key variables. This helped in understanding the underlying patterns in the data.

# Clustering and Predictive Modeling

K-Means clustering was used to segment the earthquakes into different clusters based on magnitude, depth, and economic impact. A linear regression model was developed to predict the economic impact based on earthquake characteristics, though the model's R-squared value indicated a weak predictive power.

# Hypothesis Testing

ANOVA was performed to test if there were significant differences in economic impact across different clusters of earthquakes. The results showed statistically significant differences, indicating that cluster-based analysis could be useful.

# Conclusion

The analysis provided valuable insights into the patterns and impacts of earthquakes. While the predictive modeling showed limited accuracy, the clustering and hypothesis testing revealed important differences between earthquake groups. This project demonstrates the power of data analysis in understanding natural phenomena and guiding future research.