Global Terrorism Analysis - EDA

Vimal Kumar Hoon
Data Science Trainee
Almabetter, Bangalore

Abstract:

This project aims to perform an exploratory data analysis on the Global Terrorism Database, which contains information about terrorist attacks that occurred worldwide from 1970 to 2017. The main objective is to gain insights into patterns and trends of terrorism incidents, as well as to provide a comprehensive understanding of the factors that contribute to terrorism. To achieve this, the project involves several steps such as data cleaning, preprocessing, and exploratory data analysis using Python and popular libraries such as pandas, numpy, seaborn, and matplotlib. The analysis includes creating visualizations to represent the frequency of terrorist attacks over the years and gaining a deeper understanding of the characteristics and factors associated with terrorism. By conducting this project, we hope to contribute to the broader efforts aimed at understanding and addressing the problem of terrorism.

The project involves analyzing the Global Terrorism Database to gain insights into the distribution of terrorist attacks across different regions and countries, the types of weapons used, and the most targeted sectors. The analysis reveals that the frequency of terrorist attacks has significantly increased over the past few decades, with the Middle East and North Africa being the most affected regions. Furthermore, the most common types of terrorist attacks are bombings

and explosions, with civilians being the most targeted group. These findings provide a better understanding of the nature of terrorism and can help in developing effective strategies to prevent and combat terrorism. By conducting this project, we hope to contribute to the broader efforts aimed at addressing this pressing global issue.

INTRODUCTION

Terrorism is a global issue that has affected many countries in recent years, and it's crucial to understand its patterns and trends to develop effective strategies to combat it. The Global Terrorism Database (GTD) is a comprehensive dataset that provides information on terrorist attacks around the world from 1970 to 2017. In this project, we aim to analyze the GTD dataset using Python programming language and various libraries like NumPy, Pandas, and Matplotlib to perform data analysis and visualization. Our objective is to explore the dataset, clean the data, and conduct exploratory data analysis (EDA) to gain insights into the patterns and trends of terrorism worldwide. The project will be divided into different stages, from importing and cleaning the data to performing EDA and drawing conclusions from the insights gained. Through this project, we hope to contribute to the efforts aimed at addressing this significant global issue.

PROBLEM STATEMENT

The focus of this project is to analyze the Global Terrorism Database and gain insights into the patterns and trends of terrorist attacks around the world. With terrorism being a critical issue globally, it is essential to understand the underlying causes, motivations, and locations of such activities. The analysis of this dataset can help identify areas with high terrorist activity, attack types, and weapons used, and the

impact of these attacks on different regions and countries. The main objective of this project is to provide a comprehensive analysis of the global terrorism dataset to identify such trends and patterns.

DATA SUMMARY

The Global Terrorism Database is a widely used dataset that provides comprehensive information on terrorist attacks across the globe from 1970 to 2017. With a total of 181,691 observations and 135 variables, the dataset covers crucial details such as the location, date, target type, weapon used, and casualties related to the attacks. The information was collected from various sources, including government reports, news articles, and public information. However, due to the diverse sources and large size of the dataset, it requires thorough data cleaning and preprocessing to address issues such as missing values, inconsistent values, and redundant information before performing any analysis.

STEPS INVOLVED

Here are the general steps involved in completing this project:

1. Importing the necessary libraries: Before starting the data analysis, the required libraries such as Pandas, NumPy, Matplotlib, and Seaborn must be Imported.

2. Loading the dataset: The global terrorism dataset must be loaded Into the Jupyter Notebook using the Pandas library.

3. Data Cleaning: The dataset may contain missing values, duplicates, and Irrelevant columns that must be removed. The data must be cleaned to make it ready for the analysis.

4. Exploratory Data Analysis (EDA): The EDA process Involves visualizing and analyzing the

data to find patterns, relationships, and Insights. Different types of plots such as bar plots, line plots, scatter plots, and heat maps must be used to explore the data.

5. Feature Engineering: Feature engineering Involves creating new features or modifying existing features to Improve the performance of the machine learning model. This step is optional and can be skipped depending on the project's requirements.

6. Data Preparation: After completing the EDA and feature engineering, the data must be prepared for the machine learning model. This step may Involve scaling, encoding, and splitting the data Into training and testing sets.

7. Model Selection: Different machine learning models such as Logistic Regression, Decision Trees, Random Forest, and SVM can be used for classification or regression tasks. The best model must be selected based on the project's requirements and the model's performance on the testing set.

8. Model Evaluation: The performance of the model must be evaluated using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC score.

9. Conclusion: Based on the Insights gained from the EDA and the model's performance, conclusions can be drawn, and future work can be suggested.

There are several types of exploratory data analysis (EDA) Involved in this project, including:

Univariate Analysis: This involves analyzing each variable in the dataset separately. In this project, univariate analysis was used to examine the distribution of various variables such as the number of terrorist incidents, types of attacks, countries affected, and so on.

Bivariate Analysis: This involves analyzing the relationship between two variables. In this project, bivariate analysis was

used to explore the relationship between variables such as the number of terrorist incidents and the countries most affected.

Multivariate Analysis: This involves analyzing the relationship between multiple variables. In this project, multivariate analysis was used to examine the relationship between variables such as the type of attack, number of fatalities, and the countries most affected.

Temporal Analysis: This involves analyzing data over time. In this project, temporal analysis was used to examine trends in the number of terrorist incidents over the years.

Geospatial Analysis: This involves analyzing data in relation to geography. In this project, geospatial analysis was used to examine the geographic distribution of terrorist incidents, as well as to create maps to visualize the data.

CONCLUSIONS

Based on the analysis of the Global Terrorism Database, some conclusions that can be drawn are:

1. Terrorism has been a global problem, with Incidents reported in nearly all countries.

2. The number of terrorist attacks Increased sharply after the 9/11 attacks in 2001, peaking In 2014 and declining slightly in recent years.

3. Iraq and Afghanistan have been the most affected countries by terrorism, with the highest number of attacks and fatalities.

4. Terrorist attacks are not

evenly distributed across the world, with certain regions like the Middle East, South Asia, and North Africa being more affected.

5. Bombing/explosions are the most common type of terrorist attack, followed by armed assaults and assassinations.
6. Terrorist groups like the Tallban, Islamic State, and Boko Haram are responsible for the majority of attacks and fatalities.

7. The analysis also indicates that there is a significant correlation between the number of terrorist Incidents and the number of fatalities.