Global Terrorism Exploratory Data Analysis

Yash Patil
Mohd Zahid Ansari
Pritam Rajput
Data science trainees,
AlmaBetter, Bangalore

Abstract:

Our analysis covers major aspects with respect to features in datasets and they are, Regions / Countries, Weapon Type, Terrorist Groups and Targeted Organizations.

Our Exploratory Data Analysis can used by curious civilians, security related policymakers, international organizations hosting worldwide events, foreign investors, and academic researchers for the purpose of understanding terrorism and its nature.

Keywords: Features Exploratory Data Analysis, Terrorist.

1.Problem Statement

The Global Terrorism Database (GTD) is an open-source database including information on terrorist attacks around the world from 1970 through 2017. The GTD includes systematic data on domestic as well as international terrorist incidents that have occurred during this time and now includes more than 180,000 attacks.

Features in EDA:

- Eventid: Unique Id assigned to a terrorist attack.
- Year: Year of the attack
- Month: Month of the attack.
- Day: Day of the attack.
- Country: Country in which attack took place.
- Region: Region in which attack took place.
- Latitude: Latitude co-ordinate w.r.t to world map.
- Longitude: Longitude co-ordinate w.r.t to world map.
- Attack: Type of attack.
- Target: Targeted facility of the attack.
- Killed: Number of people killed in this attack.
- Wounded: Number of people wounded in this attack.
- Summary: Attack description in short.
- Groupname: Terrorist Group name.
- Target_type: Name of the Specific entity suffered by the attack.
- Weapon_type: Weapon type used by the terrorists.
- Motive: Reason behind the attack.
- Damages: Damages incurred in dollars ($).
- Damage_txt: Scale of damage done (Minor, Major, Catastrophic)
- Suicide: Suicide number of terrorists.
- City: City in which attack took place.

2. Introduction

The Global Terrorism Data includes systematic data on Indian as well as international terrorist incidents that have occurred during this time and now includes more than 180,000 attacks.
As the first step for handling such huge dataset we started with data cleaning and data preprocessing. Performed data wrangling on dataset to get insights and understanding from the features for visualizing the data comprehensively.
I divided this analysis on 4 major aspects with respect to features in datasets and they are, Year of attack, City / Countries, Weapon Type and Terrorist Groups. In City and Countries features I took Indian city on which terror attacked happened and done visualization on these features to find out on which cities of India most terror attack happened and who are the top most terrorist group attacked on these cities. I also visualized which weapons were mostly used to attack by terrorist groups etc.
The goal of this project was to understand and interpret the nature of terrorism efficiently and comprehensively with the use of data visualizations.
Users can understand various patterns, trends, and correlation in terrorism through visual interpretation and its provided explanation.

3. Major Visualization Aspects

We divided this analysis on 4 major aspects with respect to features in datasets and they are:

- Regions / Countries
- Weapon Type
- Terrorist Groups
- Targeted Organizations.

4. Steps involved:

- **<u>Exploratory Data Analysis:</u>**

  Imported the Global Terrorism csv file. Identified the Data Format of all the features under the dataset. Compared all datatypes and understood it for further handling the data to get desired data for better visualization.

  target and the independent variables. It gave us a better idea of which feature behaves in which manner compared to the target variable.

- **<u>Data Wrangling</u>**

  - **Feature extraction**:
    - Performed Feature extraction on the dataset to filter out useful data only.

  - **Renaming Features:**
    - Renamed the columns for better understanding.

- **<u>Null values Treatment</u>**

  We replaced the NaN values with the desired values and dropped some features that was not interpolatable.

Our dataset contains a large number of null values which might tend to disturb our data pre-processing hence we dropped them at the beginning of our project in order to get a better result.

- **Feature Selection**
  Features from such a large dataset is selected on the basis of their requirement for data visualization and data-preprocessing.

5. Data-preprocessing:

After some data wrangling and null values handling in the Global Terrorism dataset. Before each visualization performed by the help of Matplotlib, Seaborn and Plotly we did the required data-preprocessing, isolated the features in some case to pass it on to the visualization libraries and edited row data to get specific values required for better data visualization.

6. Challenges:

Huge amount of data needed to be analyzed and understood to extract useful features from it.
Handling sparse dataset like Global Terrorism was challenging because of many features and many null values included in the dataset.
Pre-processed values in the features for better visualization.

7. Conclusion:

Strategic intelligence gives an insight into terrorist intent, capability useful in prioritizing risks and developing preventative measures also helpful in focusing on key vulnerabilities.
A visualization which can be used to calculate the total number of attacks per year by which terrorist group, total attack counts on Indian cities based on year provides interactive interface to explore this dataset. Users can understand various patterns, trends, and correlation in terrorism through visual interpretation and its provided explanation.
This work can be used by curious civilians, security related policymakers, international organizations hosting worldwide events, foreign investors, and academic researchers for the purpose of understanding terrorism and its nature.

References-

1. Almabetter Dataset
2. GeeksforGeeks
3. Kaggle
4. Plotly
5. W3schools
6. Pandas.pydata.org