Global Terrorism Analysis 1970-2017

Vikas Chaudhary
Data science Trainee,
AlmaBetter

Introduction:

The given Exploratory Data Analysis (EDA) project was performed on “The Global Terrorism Database (GTD) 1970-2017”. It is an open-source database prepared and managed by the National Consortium for the Study of Terrorism and Responses to Terrorism (START), at the University of Maryland, USA. The given database consists of records of terrorist attacks globally from 1970 to 2017.

Problem statement:

Our task is to perform the Exploratory Data Analysis (EDA) technique on “The Global Terrorism Database (GTD) 1970-2017” using Python language on a Google Colab book and get some insights to understand the fundamentals and evolution of Terrorism across the globe over the years.

Methodology(EDA):

The Global Terrorism Database (GTD) contains nearly 1.8 lakh rows and 135 columns. In order to understand the database and get some intuition we need to perform EDA on it in the given order:

- Import all the required libraries
    - E.g. *pandas, numpy, matplotlib, plotly*
- Get the overview of the database
- Retain the necessary data
- Explore terrorism at the Global level (overview)
- Explore the terrorism in South Asian Region
- Explore the terrorism in India

Data Prepping:

Since the given database has some flaws that could generate inappropriate outcomes. So before doing any visualization and operations we need to remove or substitute the unwanted data. The given below are some of the key operations required to do before diving deep into the dataset:

- Dropping the columns that are not required
- Renaming the remaining columns with accurate names

- Filling the NaN or null values with
    - “Unknown” in columns with Object datatype that represents ‘text’ data.
    - “0” in columns with Integer or Float datatype that represents ‘numeric’ data.
- The EDA of the Global arena can be done directly using the cleaned dataset
- To perform EDA on South Asian Region, India and the Taliban, separate datasets should be put in place with precise names to avoid any mingling of datasets.
- The rectification in the names also required
- The following 14 columns are selected:
    - iyear: Renamed as ***‘Year’***, it contains the year of the incident.
    - country_txt: Renamed as ***‘Country’***, it contains the country or location of the incident.
    - provstate: Renamed as ***‘State’***, it contains the 1st order subnational administrative region or State where the incident
    - city: Renamed as ***‘City’***, it contains the name of cities or towns where terrorist occurred.
    - region_txt: Renamed as ***‘Region’***, it contains the geographical regions with a group of countries.
    - attacktype1_txt: Renamed as ***‘Attack’***, it contains the method or tactics of attacks that these organizations follow.
    - targtype1_txt: Renamed as ***‘Target’***, it contains the targets (victims) of attacks for these organizations.
    - gname: Renamed as ***‘Group_Name’***, it contains the name of the terrorist organizations that had carried out attacks.
    - weaptype1_txt: Renamed as ***‘Weapon’***, it contains the weapon type used by terrorists to carry out attacks.
    - nkill: Renamed as ***‘Death’***, it contains the number of people (victims + perpetrators) who died in the attacks.
    - nkillter: Renamed as ***‘Perpetrator_Death’***, it contains the number of perpetrators who died in the attacks.
    - nwound: Renamed as ***‘Wounded’***, it contains the number of people (victims + perpetrators) who got injured in the attacks.
    - nwoundte: Renamed as ***‘Perpetrator_Wounded’***, it contains the number of perpetrators who got injured in the attacks.
    - ransomamt: Renamed as ***‘Ransom_Amt’***, it contains the ransom that was demanded during any incident.

Challenges Faced:

- The dataset has a large number of columns with a sufficient number of “null” values, retaining those columns would hardly add any value.
- This dataset is mostly dedicated to the framework of The US Security Policy and also looking at world security through the lens of The US.
- There are several columns entirely recoding the information for The US and its citizens, the information that these columns contain is ineffectual for other countries or regions.
- The names of several columns are intricate and it is hard to deduce any piece of information after looking at them.

Approach Used:

- The information that the given dataset contains could be unfathomable so it is better to carve out the small subset that contains a few columns.
- We have following the given workflow for our project:
    - Analyzing terrorism at the global scale e.g. most affected regions, countries, large terrorist organizations, most volatile year.
    - Analyzing terrorism at the regional scale e.g. South Asian Region.
    - Analyzing terrorism at the national scale e.g. India
- The project also emphasizes putting some light on the recent geopolitical shift in Afghanistan after the pullout of the American Security Forces by exploring ‘Afghanistan’, ‘Pakistan’, ‘Taliban’ and also a few points about ‘ISIS-K’.

Tools Used:

Since the given project is limited to the EDA, we only need a few tools or libraries to achieve our objectives. The following Python libraries are required to explore and visualize the dataset:

- *numpy:* Numpy is the core library for scientific computing in Python. It provides a high-performance multidimensional array object and tools for working with these arrays.
- *pandas:* Pandas is an open-source library that is built on top of the NumPy library. It is a Python package that offers various data structures and operations for manipulating numerical data and time series.
- *matplotlib:* Matplotlib is probably the most used Python package for 2D graphics. It provides both a quick way to visualize data from Python and publication-quality figures in many formats.
- *plotly:* Plotly is the Python Library for interactive data visualizations. Plotly is insanely powerful at explaining and exploring data.

Analysis:

The following approach for the EDA has been followed:

1. EDA on the Global Scenario: The project tries to find out the answers to the following points:
    a. The total number of attacks, deaths and casualties in those attacks.

    b. Most violent regions
    c. Method of attacks
    d. Most violent organizations
    e. Involvement of ransom
    f. Targets of attacks

2. EDA on the South Asian Region: In South Asia, the project gives answers to the following points:
    a. Number of attacks, death and casualties involved in those attacks
    b. Terrorist activities in India, Pakistan, Afghanistan separately and later combined
    c. Comparison of attacks in India with other remaining South Asian countries combined
    d. Most violent organizations
3. EDA on Taliban: Since the Taliban are at front of current geopolitical discourse so it becomes mandatory to do some analysis on their modus operandi. This project gives the answers to the following points:
    a. Total number of attacks
    b. Trajectory of attacks
    c. Geography of attacks i.e. Target countries
    d. Weapons used
    e. Who are their targets
4. EDA on India: Since India is one of the major victims of terrorism and also encircled by hostile nuclear-armed neighbors, it is mandatory for our establishment to efficiently deal with this menace. This project gives answers to the following points:
    a. Number of attacks in every State/UT
    b. Most affect State/UT
    c. Number of death and casualties in every State/UT
    d. Most violent organizations
    e. Weapons used for attacks
    f. Targets of attacks

Conclusion:

1. Terrorism is present in every nook and corner of the world, there is hardly any place that had not felt the repercussions of terrorism.
2. Although, the presence of terrorism is not uniform in every geography.
3. There are certain places like South America where terrorism had gone down and while places like the Middle East & North Africa and South Asia had seen a rise in terrorism.
4. Terrorist organizations are using all available weapons and methods to carry out attacks. These include the use of chemical and biological weapons.
5. Thorough scrutiny must be done to find out the reasons why some regions are relatively peaceful so governments and other administrative bodies can emulate those governing methodologies that are keeping terrorism at bay.
6. Global efforts must be done to uproot the sinister designs of these terror organizations.

Reference:

1. GTD Codebook
2. GeeksforGeeks
3. Analytics Vidhya
4. Python Graph Gallery (python-graph-gallery.com)
5. Stack Overflow
6. https://towardsdatascience.com/
7. Youtube, etc