Exploratory Data Analysis on Global Terrorism

ABSTRACT

There are over 200 different definitions of terrorism. Violence is the hallmark of terrorism. A qualifying terrorist incident is commonly associated with a political motive. Understanding the attributes involved in terror attacks is a profoundly complex task. Psychologists would agree that terrorists are not highly likely to volunteer as experiment test subjects to be psychoanalyzed. Examining their activities from afar can be quite misleading as assumptions have to be made. The psychology behind terrorism due to all these complexities proves to be challenging to say the least. Understanding the patterns and behaviors behind the attacks are important to unlocking the prediction power of attacks in the future and their mitigation.

The goal of this project is to use the wide range of historic data on Global Terrorism to draw comparisons and to summarize their main characteristics in order to identify patterns or meaning. From basic data visualizations to complex multivariate analysis, we have performed data analysis at every level of the dataset to draw out key insights about the terrorist activities.

INTRODUCTION

The threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation is terrorism.

Terrorism is defined in the Oxford Dictionary as “the unlawful use of violence and intimidation, especially against civilians, in the pursuit of political aims.” The key problem is that terrorism is difficult to distinguish from other forms of political violence and violent crime, such as state-based armed conflict, non-state conflict, one-sided violence, hate crime, and homicide.

Terrorism is the largest human-oriented criminal organization that has become a common problem globally. Nowadays, it is known that the underlying reason of most of today’s terrorist activities is strategic actions and international purposes. Terrorist attacks in a country can take place because of interior reasons (economic prosperity, socio-cultural, educational system) or other factors. The most important of factor is the geopolitical position of the country.

In our coverage of terrorism, we rely strongly on data from the Global Terrorism Database (GTD), which defines terrorism as “acts of violence by non-state actors, perpetrated against civilian populations, intended to cause fear, in order to achieve a political objective.”

To be considered an act of terrorism, an action must be **violent, or threaten violence**. As such, political dissent, activism, and nonviolent resistance do not constitute terrorism.

The Global Terrorism Database (GTD) includes information on terrorist events around the World (205 countries, more than 33000 cities, from 1970 through 2017 on at least 45 records for each case, with more recent incidents including information on more than 120 records.

PROPOSED SYSTEM

1. In the proposed system, we are going to use the Global Terrorism Database to analyze and derive insights on the various terrorism acts that have taken place. We will preprocess the database to get reliable and accurate patterns and insights and use descriptive statistics to draw conclusions such as — most affected regions of terrorism, major groups behind terrorism, categories of the targets, locations of terrorist attacks.

2. Using EDA, we will show the trend of the growth of terrorism in the world and how it spread to various parts of the world. Using graphs, we study how one factor behind terrorism is interrelated to the other.

3. We also do predictive analytics as to what the magnitude of the terrorism acts can be in the future years.

OBJECTIVE OF ANALYSIS AND STUDY

The GLOBAL TERRORISM INDEX defines **terrorism** as “the threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation”.

The objectives of this project are:

- to draw a clear understanding of the data to see if we can locate useful insights about where and when terror attacks occur worldwide

- to create an opinion on the relatively risky and secure regions based on past experience, by in-depth analysis of the terrorist attacks worldwide.

- to study and analyze the different terrorism acts that have taken place in various parts of the world, at different times.

- different types of terrorism acts

- categories of terrorism acts

- distribution of terrorism acts

- statistical conclusions

DATA

- Global Terrorism Data (GTD)

Description

The Global Terrorism Database (GTD) is an open-source database including information on domestic and international terrorist attacks around the world from 1970 through 2017, and now includes more than 181,691 cases. For each event, information is available on the date and location of the incident, the weapons used and nature of the target, the number of casualties, and–when identifiable–the group or individual responsible. There are 18,1691 records and 135 columns including date, time, location, number of hostages, killed, wounded, if there was a ransom, the outcome, if there was a suicide attack, claims, weapons used. Unit of analysis: Attack.

Characteristics of the GTD

- Contains information on over 181,691 terrorist attacks
- Currently the most comprehensive unclassified database on terrorist attacks in the world
- Includes information on more than 91,000 bombings, 20,000 assassinations, and 13,000 kidnappings and hostage events since 1970
- Includes information on at least 45 variables for each case, with more recent incidents including information on more than 120 variables

- More than 4,000,000 news articles and 25,000 news sources were reviewed to collect incident data from 1998 to 2017 alone

Content and Data fields

Geography: Worldwide

Time period: 1970-2017

Unit of analysis: Attack

Variables: >100 variables on location, tactics, perpetrators, targets, and outcomes

- eventid: Incidents from the GTD follow a 12-digit Event ID system.
- iyear: This field contains the year in which the incident occurred.
- imonth: This field contains the number of the month in which the incident occurred.
- iday: This field contains the numeric day of the month on which the incident occurred.
- country: This field identifies the country code country or location where the incident occurred.
- region: This field identifies the region code in which the incident occurred.
- city: This field contains the name of the city, village, or town in which the incident occurred.
- latitude: This field records the latitude (based on WGS1984 standards) of the city in which the event occurred.

- longitude: This field records the longitude (based on WGS1984 standards) of the city in which the event occurred.
- success: Success of a terrorist strike is defined according to the tangible effects of the attack. Success is not judged in terms of the larger goals of the perpetrators.
- suicide: This field records the number of attacks which were in the form of suicide bombing.
- nkill: This field records the number of people killed due to terrorism.
- nwound: This field records the number of people injured due to terrorism.
- gname: This field records the name of the terrorist group.
- targtype1_txt: This field contains the name of the target community of the terrorist which became the victim.
- weaptype1_txt: This field records the name of the weapon used for spreading terrorism.
- attacktype1_txt: This field specifies the attack type used by terrorists.

LIBRARIES (Packages Used in the Project)

1. Pandas: Pandas is a high-level data manipulation tool. It is built on the Numpy package and its key data structure is called the Data Frame. Data Frames allow you to store and manipulate tabular data in rows of observations and columns of variables.

2. Numpy: NumPy is a python library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices. NumPy stands for Numerical Python.

3. Matplot: Matplotlib is an amazing visualization library in Python for 2D plots of arrays. Matplotlib is a multi-platform data visualization library built on NumPy arrays and designed to work with the broader SciPy stack. One of the greatest benefits of visualization is that it allows us visual access to huge amounts of data in easily digestible visuals. Matplotlib consists of several plots like line, bar, scatter, histogram etc.

4. Seaborn: Seaborn is a library for making statistical graphics in Python. It is built on top of matplotlib and closely integrated with pandas data structures. A dataset-oriented API for examining relationships between multiple variables. Specialized support for using categorical variables to show observations or aggregate statistics. Options for visualizing univariate or bivariate distributions and for comparing them between subsets of data. Automatic estimation and plotting of linear regression models for different kinds dependent variables. Convenient views onto the overall structure of complex datasets.

5. Plotly: The plotly Python library (plotly.py) is an interactive, open-source plotting library that supports over 40 unique chart types covering a wide range of statistical, financial, geographic, scientific, and 3-dimensional use-cases.