Exploratory Data Analysis on

Global Terrorism Analysis

Chaitanya Inamdar, Jagdish Tawar, Pritam Gangwar

Data Science Trainees, AlmaBetter, Bangalore

**Abstract** :

Terrorist attacks affect the confidence and security of citizens; it is a violent form of a political struggle that ends in the destruction of order. In the current decade, along with the growth of social networks, terrorist attacks around the world are still ongoing and have had potential growth in recent years. Consequently, it is necessary to identify where the attacks were committed and where is the possible area for an attack.

The objective is to provide Exploratory Data Analysis of this data. The input would be a database that has a systematic record of worldwide terrorist attacks from 1970 to 2017. As a final result, it is necessary to know the number of terrorist attacks in the world, the most frequent types of attacks and the number of seizures caused by region; furthermore, to be able to

predict what kind of terrorist attack will occur and in which areas of the world. Finally, this research aims to help the scientific community use Exploratory Data Analysis to find the issues related to Global Terrorism .

***Keywords- Exploratory Data Analysis, Global Terrorism, Terrorists, Casualties etc.***

**1 ] Problem Statement-**

Analysing the given Dataset and to extract important data from it to provide required information. Different data frames related to this analysis are –

* global\_terrorism\_df – It contains all the information regarding Total attacks, Countries, Regions of different world,Casualties because of attack, Terrorist Groups from all over world from the year 1970 to 2017.
* country\_attacks – It contains the information regarding the Countries in which attack happened.
* Region\_df – It contains different world wide Regions in which Terrorist activities happened. E.g. Eastern Europe, Middle East etc.
* Attack\_type\_df – It gives information regarding the type of attack done by Terrorists.
* Group\_df - It gives information regarding different Groups of Terrorists which do attacks globally. E.g.- Taliban, ISIS etc.
* city\_df – It gives information regarding the different cities from countries where the most attacks and Casualties happened.

**2 ] Introduction –**

The Technological progress has benefited millions of people worldwide; being informed about any event is much faster and easier in comparison to previous decades. Thanks to the Internet, communication and social interaction is much more fluid, yet not all are profitable. Since technological growth in the world began to grow a higher degree of dissatisfaction about current events, as explained by the research of, which indicates an increase in recent years associated with terrorist attacks and various assaults in the world due to dissatisfaction with the political system. Terrorist attacks, as reported by , are considered as such if they are occasioned by political, religious, economic, or social reasons. To be able to visualize and predict these types of events, Exploratory Data Analysis must be used since this is one of the most modern technology can be used to analyse the data.

Since the study of terrorist attacks can be extended to many areas of knowledge and can contribute to providing strategies to combat them, this paper aims to use Exploratory Data Analyse to visualize and predict terrorist attacks from 1970 to 2017, to contribute to the scientific community related to global events.

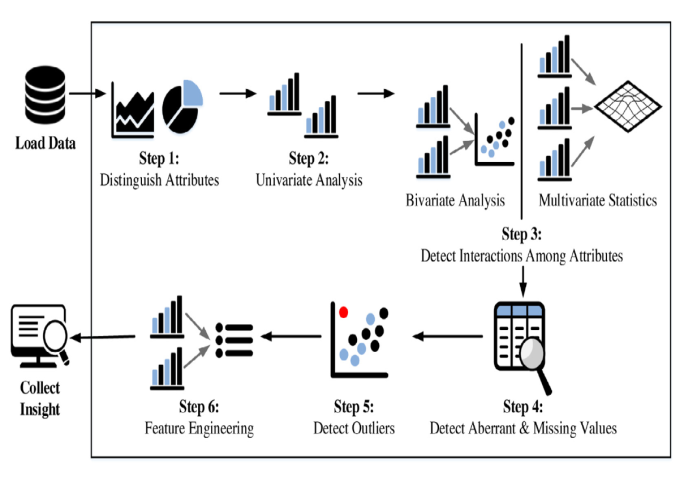
3 ] Exploratory Data Analysis-

Exploratory Data Analysis (EDA) consists of techniques that are typically applied to gain insight into a dataset before doing any formal modelling.

EDA helps us to uncover the underlying structure of the dataset, identify important variables, detect outliers and anomalies, and test underlying assumptions. With EDA, we identify relevant variables, their transformations, and interaction among variables with respect to the model we want to build.

 EDA can also point out missing data as may be relevant to building desired models.

EDA uses techniques of statistical graphics but has a broader scope. It's an approach rather than just a set of techniques.

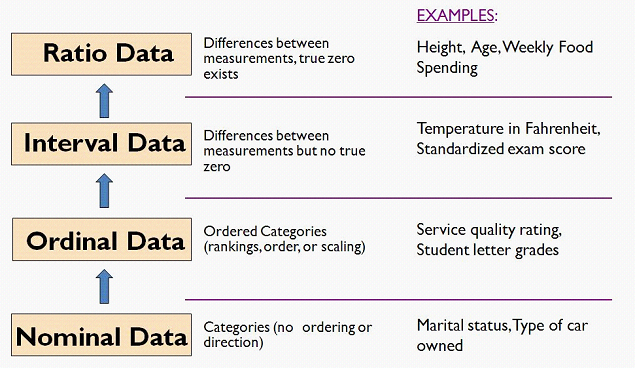
****

**3.1 ] Process for EDA –**

Look at the structure of the data: number of data points, number of features, feature names, data types, etc.

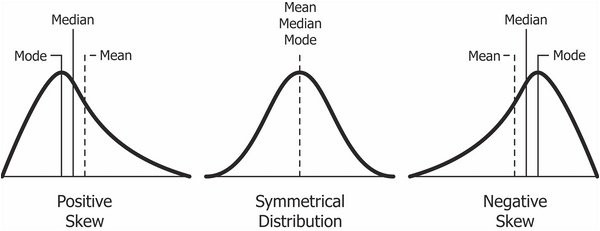
* When dealing with multiple data sources, check for consistency across datasets.
* Identify what data signifies (called measures) for each of data points and be mindful while obtaining metrics.
* Calculate key metrics for each data point (summary analysis): a. Measures of central tendency (Mean, Median, Mode); b. Measures of dispersion (Range, Quartile Deviation, Mean Deviation, Standard Deviation); c. Measures of skewness and kurtosis.
* Investigate visuals: a. Histogram for each variable; b. Scatterplot to correlate variables.
* Calculate metrics and visuals per category for categorical variables (nominal, ordinal).
* Identify outliers and mark them. Based on context, either discard outliers or analyze them separately.
* Estimate missing points using *data imputation techniques*.

3.2 ] Data Types in EDA –



* **Nominal**: This is qualitative, not quantitative; eg. Type of Vehicle Preference: 1 = Bus, 2 = Truck, 3 = Car, 4 = Two Wheeler, 5 = Other.
* **Ordinal**: An ordinal scale that indicates ordering or direction in addition to providing nominal information; eg. Low/Medium/High or Faster/Slower are examples of ordinal levels of measurement. Ranking an experience as a "nine" on 1-10 scale tells us that it was higher than an experience ranked as a "six".
* **Interval**: Interval scales provide information about order, and also ability to compare ranges; eg. temperature measured either on a Fahrenheit or Celsius scale: measured in Fahrenheit units, the difference between a temperature of 46 and 42 is the same as the difference between 72 and 68.
* **Ratio**: In addition to possessing the qualities of nominal, ordinal, and interval scales, a ratio scale has an absolute zero, a point where none of the quality being measured exists; eg. income, years of work experience, number of children.

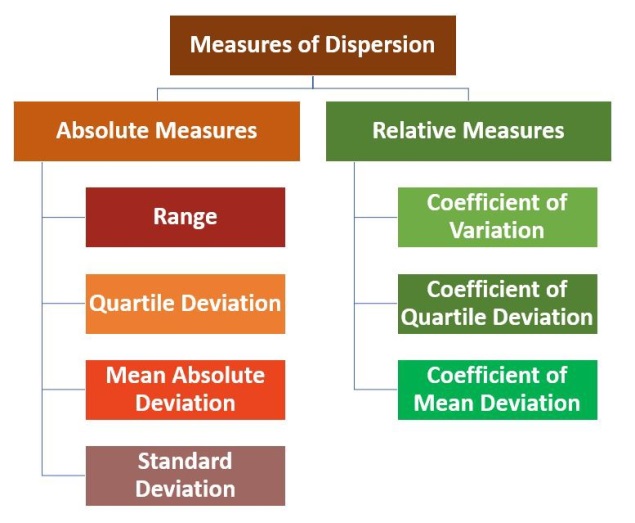
**3**.3 ] Central Tendency –

****

A measure of central tendency is a single value that attempts to describe a set of data by identifying the central position within that set of data. These include the following:

* **Mean**: Mean is equal to the sum of all the values in the data set divided by the number of values in the data set. This is also called *arithmetic mean*. Other means such as *geometric mean* and *harmonic mean* are also sometimes useful.
* **Median**: Median is the middle score for a set of data that has been arranged in order of magnitude. For example, given an ordered list of student marks, [14 35 45 55 55 56 58 65 87 89 92], median is 56 because it is the middle mark since there are 5 items before it, 5 items after it.
* **Mode**: Mode is the most frequent score in our data set. For the above data set of student marks, mode is 55 because 55 is repeated for the maximum number of times.

**3.5 ] Measures of Dispersion-**

****

**3.6 ] Visual Aids for Exploratory Analysis-**



Data can be represented visually in many ways with programming languages and visualization packages. Programming languages such as R, Python, Matlab, SAS, etc. provide libraries for creating data visuals. In JavaScript, we have D3.js, NVD3, FusionCharts and Chart.js. In Python, we have Matplotlib, Seaborn, Bokeh and Plotly.

There are dedicated visualization platforms such as Tableau, Qlikview, and PowerBI in the market that even non-programmers and traditional data analysts can use to make visuals.

Histograms and scatterplots are widely used for exploratory analysis to quickly understand the structure of data and inter-relations of variables. However, numerous other charts can be used to create visuals that have repeat purpose and long shelf life.

4 ] Null values Treatment-

Our dataset contains a large number of null values which might tend to disturb our

accuracy hence we dropped them at the beginning of our project in order to get a better result.

|  |  |  |
| --- | --- | --- |
| Sr.No. | Attack Type | Description |
| 1. | Assassination | An act intended primarily to murder one or more specific and prominent individuals. It is usually carried out on individuals of some significance, such as high-ranking military officers, government officials, celebrities, etc. It does not involve attacks on non-specific target group members. The killing of a police officer would be an armed robbery, unless there is reason to believe that the perpetrators have targeted a particularly prominent officer for assassination. |
| 2. | Armed assault | An attack primarily aimed at causing physical harm or death to humans directly through the use of a firearm, incendiary, or sharp instrument (knife, etc.). It does not cover attacks that involve the use of fists, stones, sticks or other (less lethal) hand weapons. It also includes attacks involving certain kinds of explosive devices in addition to firearms, incendiary or sharp instruments. The subcategories of explosive devices included in this classification include grenades, projectiles, and unknown or other explosive devices that are thrown. |
| 3. | Bombing/explosion | An attack in which the primary effects are produced by an energetically non-stable material that rapidly decomposes and delivers a pressure wave resulting in physical damage to the surrounding environment. It may include high or low explosives (including a dirty bomb) but does not extend to a nuclear explosive device that releases fission and/or fusion energy, or an incendiary device in which decomposition occurs at a much slower rate. If an attack involves certain classes of explosive devices in conjunction with firearms, incendiary or sharp objects, then the attack is coded as an armed assault only. The subcategories of explosive devices covered by this classification are grenades, projectiles and unknown or other explosives. |
| 4. | Hijacking | An act designed to take control of a vehicle such as an airplane, ship, bus, etc. in order to redirect it to an unscheduled destination, force the release of prisoners, or some other political objective. Getting a ransom payment should not be the sole purpose of a kidnapping, but may be an aspect of the incident as long as other objectives have been declared as well. Kidnappings are distinguished from hostage-taking because the objective is a vehicle, regardless of whether there are people/passengers in the vehicle. |
| 5. | Hostage taking (kidnapping) | An act committed for the purpose of taking possession of hostages so as to attain a political goal by means of concessions or the interruption of normal activities. Kidnappings are different from barricade incidents (the type of attack detailed above) as they involve the transfer and retention of hostages in another location. |
| 6. | Infrastructure attack | Infrastructure attacks may involve acts that are intended to damage a facility, but also harm the surrounding people in an incidental manner (e.g., an arson attack whose primary objective is to damage a building, but which causes injury or death in the process). |

5 ] Conclusion –

Terrorist attacks are among the causes of national instability. A clear understanding of how this event is occurring will help us to conduct more in-depth investigations. The focus of future research will be on performing a quantitative analysis of the countries in each region to conduct further research. Other future work to be done is the use of Big Data techniques for sentiment analysis, which will extract information from social networks to determine possible threats of cyber terrorism. Thus the investigation would use large volumes of data .

Through this research, it is possible to conclude that the use of Exploratory Data Analysis techniques was able to visualize and predict terrorist attacks. The results section shows that there has been a considerable growth in terrorist attacks since 2010 and that due to the classification models, it was possible to determine the probability of which region and type of attack may occur. Concerning the number of attacks by region, it was obtained that there is a probability that they will happen in the Middle East & North Africa and followed by South Asia.