Exploratory Data Analysis

Global Terrorism

BY :- WORKING WITH INSIGHTS

MEMBERS : NIKITA VERMA, FIRAZ THAKUR, ASHWINI KALAMKAR , TEJASWINI SAJU

DATA SCIENCE TRAINEES, ALMABETTER

1.Abstract

Terrorism is defined as using indiscriminate destruction to provoke fear or terror in people's hearts to attain financial, religious, ideological or political goals. In recent years, terrorism has expanded its appendages globally and is the favourite fanatics method to achieve what they aspire.

An analysis of a project is performed, which consists of a variety of studies and visualisations for interpreting patterns and trends. The visualisation tool that gives the user the ability to explore datasets. Diverse viewpoints result from a lack of comprehension and awareness of global terrorism and widespread misunderstandings among civilians. In this age of globalisation, there are adequate knowledge on this topic which can help boost our anti-terrorism measures, security issues, enforce improved economic policies, and expand the body of information about civilians.

2. Problem Statement

Data provided by the Global Terrorism Database.csv file is in unformatted manner, uneven data, and duplicate data and also some data columns it is irrelevant, because it's a piled-up data coming from various different countries. For doing the analysis on the data the data needs to be in correct format and well organized formed The main objective of the analysis is to obtain the meaning full information and facts from the given huge datasets as shown above, by cleaning the datasets, doing a proper analysis and visualization and plotting the useful information into different graph and charts so that the trend and relationship between the various indicators on which the analysis is done can be understand easily.

3. Dataset

The dataset contains data of more than 180000 terrorist events (no. of rows) happened between (1970- 2017) and has over 135 variables (no. of columns) describing each attack. Some of the key attributes consisting those variables which are taken under consideration for this project are listed below

DATA DICTIONARY

- Year - Year of the attack
- Month - Month of the attack
- Day - Day of the attack
- Region - Name of the region where the attack happened consists of values like East Asia, Western Europe, etc
- Country – Name of the countries where the attack happened consists of values like Mexico, India, Iraq, Portugal

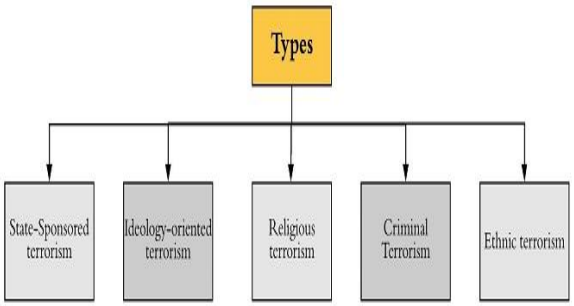
- State – Name of the provinces/states where the attack happened which consists of values like Lattakia , Manipur
- City – Name of the cities were attack happened consists of values like Baghdad, Mexico city, Imphal,
- Latitude - Latitude of the location
- Longitude- Longitude of the location
- Location -Gives information of where the incident took place
- Summary- Gives information about the attacks
- Attack_type - Consists of categories like explosion, armed assault, assassination, kidnapping.
- Target_type - Consists of categorical values like private citizens, military, police, government officials, transportation, education, religious institution, airports, etc
- Gang_name - Organization that claimed responsibility of an attack
- Weapon_type - Type of weapon used in the attack. Weaptype1 contains values like firearms, explosives, melee, vehicles etc.
- Killed - Number of people killed in any event.
- Wounded - Number of people wounded in any event.

4. Introduction

- World peace was one of the core reasons for forming the United Nations organization. Terrorism is the biggest hurdle to world peace. In this project, we focus on terrorism by analyzing the dataset provided by to explore meaningful patterns and statistics. Terrorism is an unsettled term.
- Terrorism is sporadic, widespread and inconsistent with time and nature. Because of these characteristics, International terrorism is difficult to summarize all aspects as a single conclusive solution and make this information available to be easily understood by most people. Exploring this dataset

can provide an insight into how different parameters are correlated with each other, which can help identify unknown hidden patterns. This exploration will also assert enough facts to provide justifications for some common misconceptions regarding terrorism. One of misconceptions is that more military can suppress and control terrorism. However, using the instrumental +variable approach, studies show that counter-terrorism solutions like more military spending is not enough to control terrorism and is also dependent on other factors like economy and national politics. Another popular misconception is that terrorism only affects the individuals directly involved in any terrorist event. Terrorism adversely impacts not only the economy of the victim country but also the countries financially associated with international terrorism

5. TYPES OF TERRORISM



➢

State-Sponsored terrorism

It is pursued in order to achieve such clearly stated foreign policy objectives.

Ideology-oriented terrorism

Ideology-oriented terrorism is typically categorized into two: left-wing and right-wing terrorism.

Left-wing terrorism

It is violence against the ruling class, mostly by the lower classes, motivated by leftist ideology.

Right-wing Terrorism

Right-wing groups tend to seek to protect the status quo or to return to some past situation that they feel should have been preserved. Examples of this include: fascism in Italy, nazism in Germany,

Religious terrorism

Terrorist groups are notably motivated by religion. Religious terrorism is more destructive in nature. These groups are motivated either in whole or in part by a religious imperative that considers violence as a sacred duty. The theology of ISIS is in accordance with the Wahabi theological ideology.

Criminal Terrorism

Terrorist activities are used to aid in crime and criminal profit. For instance, in narco-terrorism, narcotics traffickers attempt to influence the policies of the Government by systematic threat or use by violence.

Ethnic terrorism

It is deliberate violence by a subnational ethnic group to advance its cause. Such violence usually focuses either on the creation of a separate state or on the elevation of the status of one ethnic group over others.

6.Reasons for Terrorism

There are many reasons which make people or a group terrorist. Those reasons are political, religious, poverty, and lack of education.

Political

The main cause of terrorism is perceived socio-political or historical injustice and a belief that violence will lead to change. People who choose this path when they have been stripped off their land or rights are denied the same.

Religion

Terrorist groups use a specific religious ideology to inspire people to join terrorist groups. For example, ISIS and Al-Qaeda use Islamic ideology making people follow them.

Intolerance

Because of the increasing population and decreasing resources, intolerance is growing in society. Increasing globalization of the society come to transcend national boundaries spreading terrorism.

Governance

The ineffective anti-terrorism legislation and misplaced judicial activism are somehow also responsible for growing terrorism.

7. Dataset challenges

This dataset presents a significant problem due to the fact that individual investigations lead to contradictory outcomes. Current deficiencies and limits in data gathering methodologies, arguments over definitions, and irregularities in coding and processing give rise to disagreements among researchers, so invalidating their conclusions . A heuristic casual model showing relationships between globalisation and terrorism must be supported by an acceptable level of theoretical and empirical study . Critical disagreement over the definitional arguments surrounding major terrorist acts has a negative impact on the growth of this area . This issue necessitates exercising the requirement for common grounds that can be acknowledged by the majority of specialists and relevant authorities in order to agree on what may be the standard norms and method to be deemed a valid piece of information on terrorism upon which proper study can be conducted .

8. Steps involved:

DATA PREPROCESSING

After data collection, the first stage is data preprocessing. It is a series of operations done on dataset to change unclear data that can impede analytical conclusions. Raw data is essentially a collection of interconnected information. Oftentimes, raw data is unstructured and contains a great deal of information that is unnecessary to the project's requirements. Methodologies for data preparation facilitate the transformation of these raw data into a more meaningful, focused, interpretable, and readable manner.

The dataset from the Global Terrorism Database is insufficient, inconsistent, and contains several errors, missing attributes,

values, outliers, improper tags, and duplicate entries. These disparities can be resolved through data preparation.

DATA CLEANING

Filling in missing numbers, removing outliers, and handling irregularities in data constitute the data cleaning process. Numerous categories, such as 'motives' and 'responsible organisations,' are absent from the terrorist dataset due to a lack of information or because the field was irrelevant to the incident in question

9. Technology Used

Python is an advanced programming language that supports multiple platforms, including Windows, Linux, Mac, and Raspberry Pi, among others. Python can be used to build online applications, database systems, manage large amounts of data, and conduct complicated mathematical calculations. Python is object-oriented, functional, and procedural. These are some of the Python packages utilised by this project.

Mathplotlib is a 2D plotting tool that includes the necessary modules and functions. A developer can adjust font properties, styles, axis properties, etc.

Pandas is used for data manipulation and analysis. Pandas can transform data structures and dataset formats to data frames on which operations like as loading data, renaming attributes, mapping, crosstab, sub-data frames, etc. can be carried out.

NumPy offers structures for multidimensional array objects as well as tools for associated operations. Typically, NumPy is employed for high-performance scientific computations.

Seaborn: It is a library that uses Matplotlib underneath to plot graphs. It will be used to visualize random distributions.

Google Colab: Colaboratory, or “Colab” for short, is a product from Google Research. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education

10. ANALYSIS

1. CORRELATION BETWEEN THE DATAFRAMES

Correlation refers to a process for establishing the relationships between two variables. we are using the Pearson method with heat map (to check the intensity of the correlation0

Pearson Correlation coefficient used for linear dependency between the data sets. The value of the coefficient lies between -1 to +1. When the coefficient comes down to zero, then the data is considered as not related. While, if we get the value of +1, then the data are positively correlated, and -1 has a negative correlation.

The values indicating the correlation in this heatmap

- 0.9 - 0.7 INDICATES STRONG CORRELATION
- 0.7 -0.5 INDICATES MEDIUM CORRELATIO
- 0.5-0.3 INDICATES WEAK CORRELATION

From the heatmap it can be concluded that the columns of dataframes mostly have positive correlation as the negative values are bare minimum .

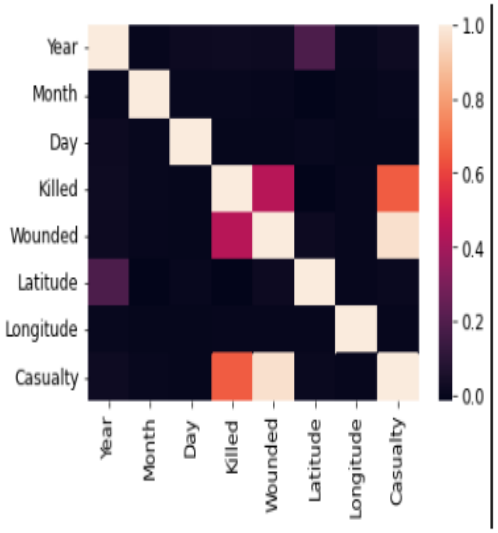


Fig 1- Correlation between the dataframes

2. NUMBER OF TERRORIST ATTACKS IN DIFFERENT REGIONS

In the region wise attacks we focused on year and the region features to find how the distribution of attacks in the region across the years.

This distribution will helps us to better understand that which are the region have to face most of the attacks through out the years

We can see that recently middle east and north Africa are the worst affected regions followed by South asia.

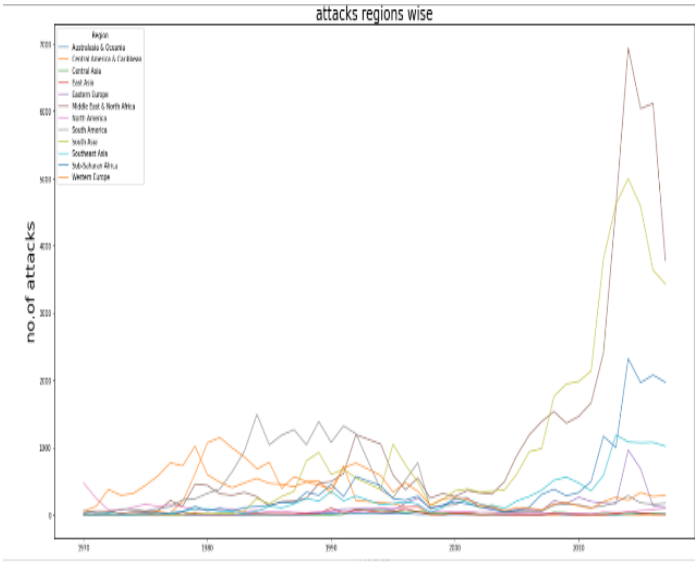
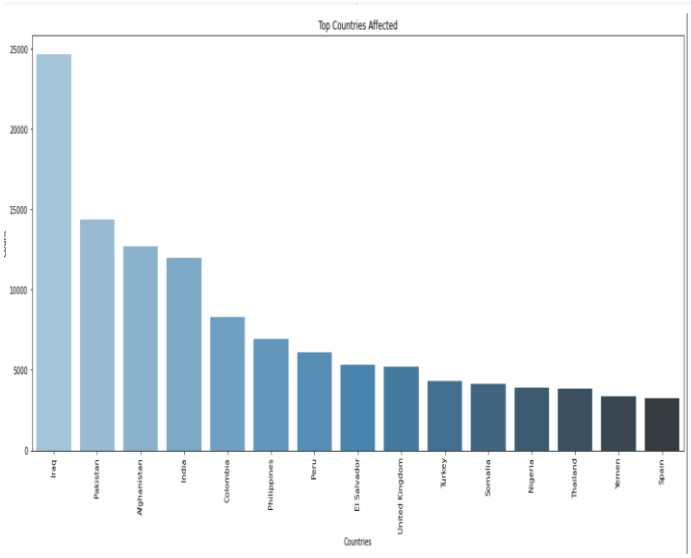


fig2 : Attack over the regions

3. NUMBER OF TERRORIST ATTACKS IN DIFFERENT COUNTRIES

This analysis was carried out to see how each country in our dataset is affected by these attacks.We have calculated the number of times the country has been there in the different



terrorist attacks and through that we could find the total attack on each country.

We have plotted the top 10 countries with the highest attack attempts.

Iraq is the most attacked country followed by Pakistan, Afghanistan, India, Columbia, Philippines, Peru, El-Salvador, United Kingdom, and Turkey

Fig 3 : Attack over different countries

4. NUMBER OF TERRORIST ATTACKS IN DIFFERENT CITIES

This analysis was carried out to see how each city in our dataset is affected by these attacks.We have calculated the number of times the city has been there in the different terrorist attacks and through that we could find the total attack on each country.

We have plotted the top 6 city with the highest attack attempts.

Baghdad is the most attacked city followed by Karachi, Lima, Mosul, Belfrost, Santigo

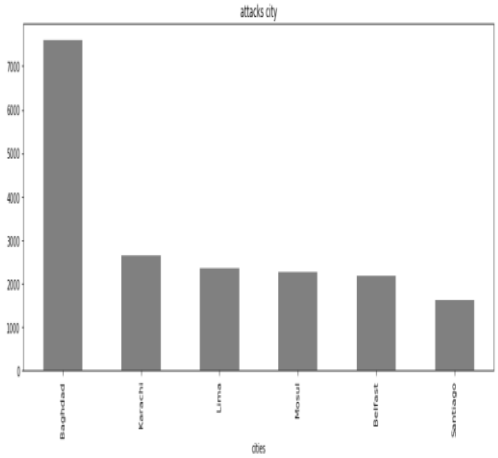


Fig 4: Attacks on the cities

5. NUMBER OF TERRORIST ATTACK EACH YEAR

From this analysis, all activities that occurred in a given year in each available geographical location in the world are represented. This study allows us to determine whether terrorist activities are increasing or decreasing per year. We can see that in the year 2014 we had the most amount attacks and least amount attack in the year 1971. We can see that the

number of attacks occurring increasing as the year progresses from 1970-2017 ,

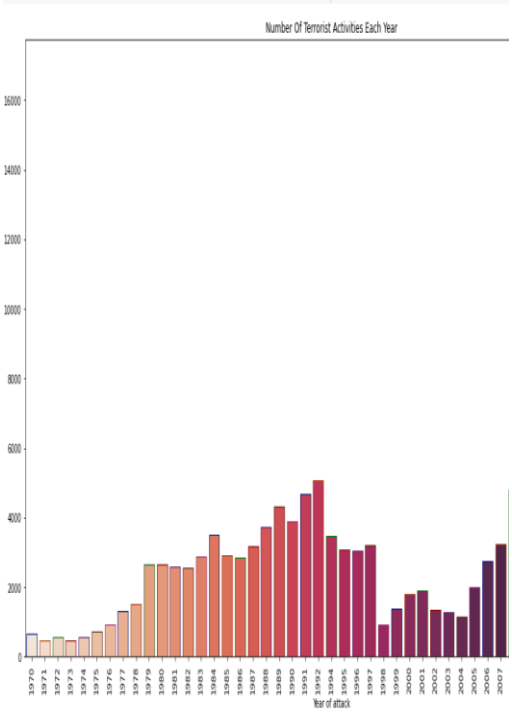


Fig 5 : Attacks over the years

6. METHOD OF ATTACK

This analysis was carried out to see mode of attack preferedd by the terrotist groups . We have calculated and categorized the no of attacks and presented it according to the type of attack

We have plotted 9 types of attacked in which bombing and explosion is the most used means of attack by the terrorist groups followed by armed assualt and assassination

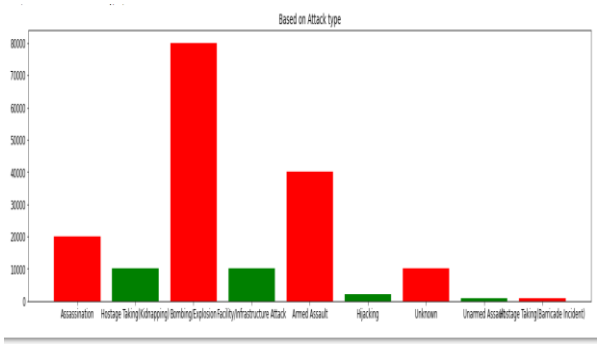


Fig 6 : Based on Attack type

7. TYPE OF WEAPONS USED BY THE TERRORIST GROUPS

This analysis was carried out to see the types of weapons in our dataset was used for the attacks We have calculated different types of weapons used for the attacks over the years.

We have plotted a pie diagram which shows the percentage of weapon type used which concludes :

- Explosives : 51.4%
- Firearms : 34.3%
- Incendiary : 5.7%
- Melee : 2.9%

Explosives were the most prefered type of weapon of the terrorist organizations for the attacks

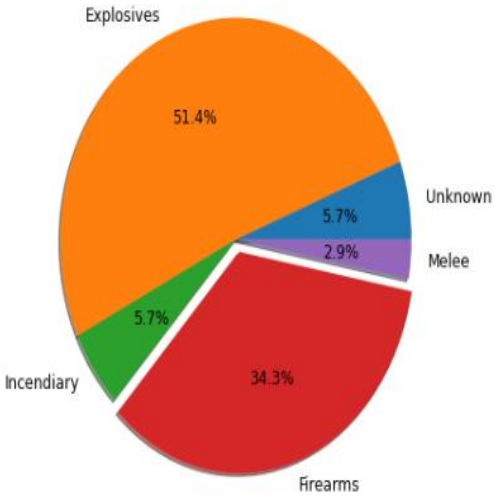


Fig 7 : Pie Diagram

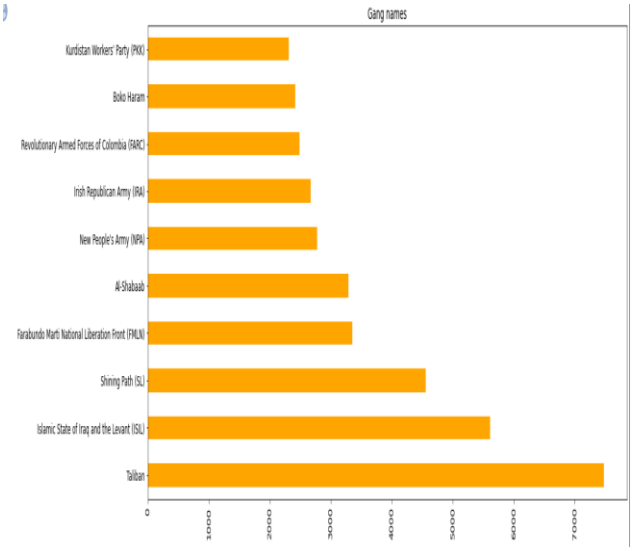


Fig 8 : Terrorist Group Names

8.TOP TERRORIST GROUPS

This analysis will tell us which are the most dangerous terrorist organisations in world. Here we are using features like terrorist group,success, number of attacks as our features to find the ranking of the terror organisation with respect to successfully executed missions

Taliban is the group which ranks at the top followed by

9. NO OF CASUALITIES OVER THE YEARS

From this analysis, all casualities that occurred in a given year in each available geographical location in the world are represented. This study allows us to know the number of people killed and wounded throughout the year We can see that in the year 2014 we had the most number of people killed and least number was in the year 1971 and also the number of wounded people which is most in the year 2015 and least in the year 1971

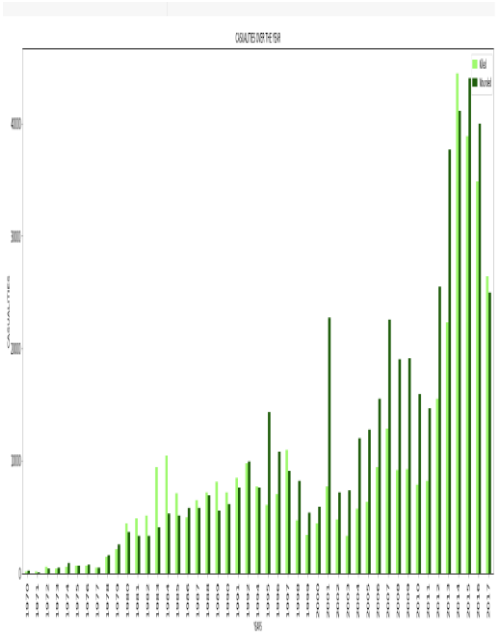


Fig 9 Casualities over the years

11. Conclusion:

That's it! We reached the end of our exercise.

Starting with loading the data so far we have done EDA , null values treatment.

In all of these models our accuracy revolves in the range of 70 to 74%.

So the accuracy of our best model is 73% which can be said to be good for this large dataset. This performance could be due to various reasons like: no proper pattern of data, too much data, not enough relevant features.

12. References:

Github

https://github.com

Kaggle

https://www.kaggle.com

Geeks for Geeks

https://www.geeksforgeeks.org