**EXPLORATORY DATA ANALYSIS ON GLOBAL TERRORISM**

Nisha Ahire,

Data science trainee,

Alma Better, Bangalore

ABSTRACT

In this work, I present analytical results obtained by data mining on the START (Study of Terrorism and Response to Terrorism) dataset. The main objective is to visualize terrorism data and make it available to users in an easy-to-understand format.

Lack of understanding and awareness about global terrorism leads to diverse opinions and common misconceptions among civilians. In this age of globalization, sufficient information about this topic can help strengthen our counter-terrorism strategies, improvise security concerns, regulate better economic policies and enhance the knowledge base of civilians.

The primary dataset for this project is provided by START Consortium which contains data of terrorist events since 1970. Performing various data mining and data visualization techniques to interpret the nature of terrorism to better understand its trends and patterns in over 45 years of its recorded history.

DATASET CHALLENGES

Given dataset from the Global Terrorism Database have 135 columns,181691 rows and it is a mix of categorical and numeric values.

And given database is incomplete, inconsistent, contains many errors, missing attributes values, contains outliers, incorrect tags, and duplicate entries, name of the columns are not informatory enough to develop proper understanding. so, we will rename the columns, i.e., Data pre-processing and this can help us to resolve these discrepancies. and formation of informatory data so, that we can extract the required columns for exploring and analysing the database, to discover the key findings pertaining to terrorist activities.

PROBLEM STATEMENT

The problem statement here is to build a tool that can present processed information in the form of intuitive visual representation of analysed data. Implementation of this project involves system design, backend design, visual design, and user interface. System design includes the overall design plan of the whole project system which explains how each individual module is correlated with others. Backend design contains a series of data pre-processing steps to transform the raw dataset into a more meaningful and focused collection required for this project. This design module also includes scripts for analyses and other factual information derived from the dataset. Visual design mostly consists of analyses and visualization techniques to construct different graphics representing the end results in an easy-to-interpret format. User Interface is in the form of a user-oriented website to present and explain different visual analyses to provide a better understanding of the dataset. The user can interact with the dataset using the exploration tool.

INTRODUCTION

Terrorism in its broadest sense, is the use of criminal violence to provoke a state of terror, terrorism is the biggest hurdle to world peace. Terrorism is commonly ignored by the civilians who are not affected directly by the dangers. For the most part, terrorism is considered unpredictable and unfortunate calamity that strikes some parts of the world more than others. Based on the location of the events, people at large have very limited information about any such event happening in other parts of the world and hence react differently.

Steps involved in data cleaning

After analysing this dataset:

* There are **181691** rows and **135** columns in **GLOBAL TERRORISM DATABASE**.
* The database is mixture of float, integers, object.
* Many of the columns are not of much use for our analysis so they were not used and only the columns with informative, useful data are used for further analysis.
* There was a need of renaming the columns to make it readable for all the user of data.
* We have used all the rows of the database but only 24 columns were selected for this exploratory data analysis project.

Treatment of null value

Removal of null values is necessary for each row to get effective analysis output in order to terminate negative data and work on useful data.

Treatment of Duplicate Rows

After removal of null values another important task is the removal of duplicate row values if they exist. With the removal of duplicates, we can now move forward with the problem statement without having unnecessary multiplied data.

Replacing the missing values

This method is also important in terms of getting output visualization.

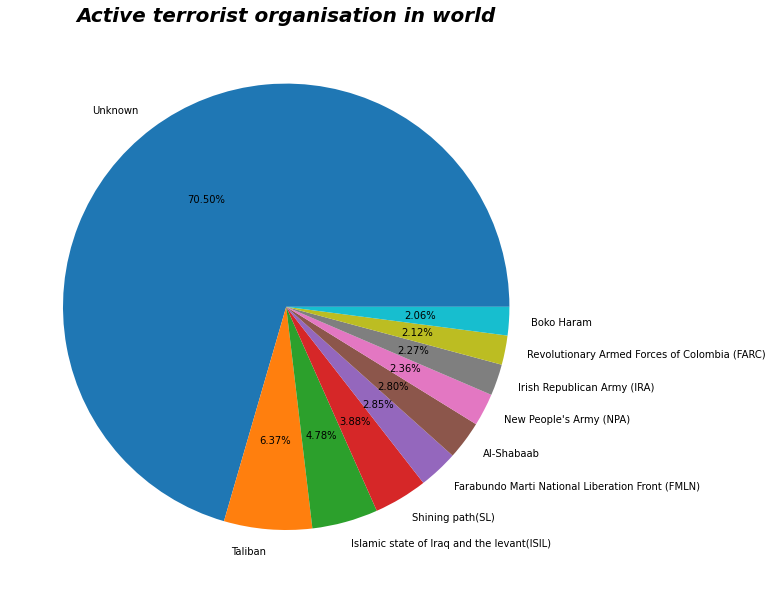
**OUTPUT FROM THE EDA**

* NUMBER OF ACTIVE TERRORIST GROUPS WHICH ARE SPREADING THE THREAT OF TERRORISM ACROSS THE WORLD.

There are total 3537 active terrorist groups in the world who are trying to threaten people, and spread terrorism across the peaceful global world.

* MOST ACTIVE ORGANISATION WHO HAS TAKEN RESPONSIBILITY OF ATTACKS.

Taliban has taken the majority of responsibility of attacks. They are Claiming that in the world majority of attacks has been done by their organisations .In the given dataset 70.50% of the attacks are claimed by unknown groups.



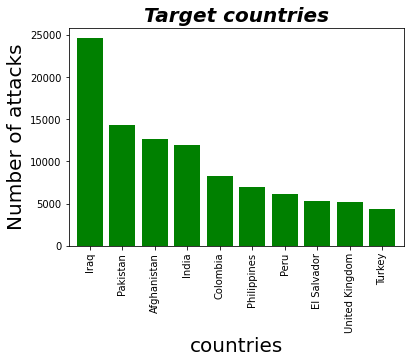
* SUCCESS RATE OF TERRORIST GROUPS

The finding suggest that successful attacks are huge in number i.e., 161632 success rate of attacks by terrorist groups is very high it is 88.95, and the unsuccessful attack is very less in number compared to successful attacks i.e. 20059, and the rate is 11.04

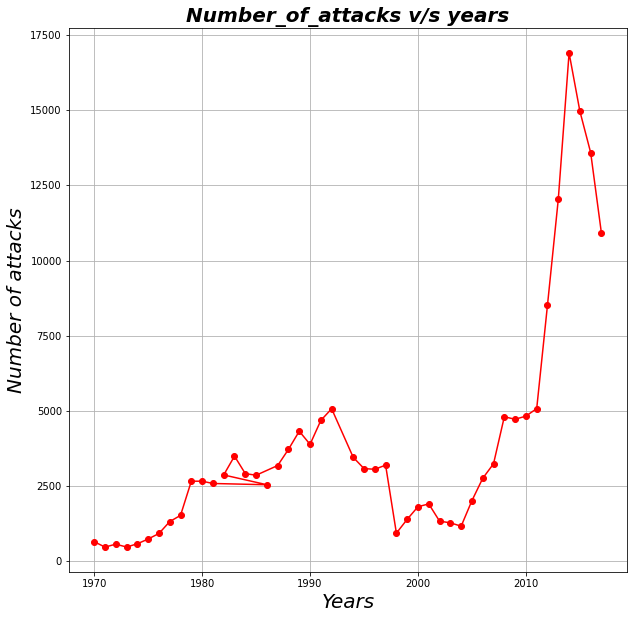
The success rate is an alarm that the terrorist attacks are very dangerous to humanism and for the peace across the globe .so we should immediately find the ways to tackle this groups and should be very cautious for the plannings planned by them in future activities.

* CITY AND COUNTRY AFFECTED BY THE ATTACKS MAJORLY

Iraq is the most affected country by terrorist attack. And Bagdad is the most affected city of Iraq, so we should be extra vigilant for suspicious situations, in this country. And we should also take care about the countries like Pakistan, India, Columbia as this are the top 5 targeted countries by the terrorist groups.

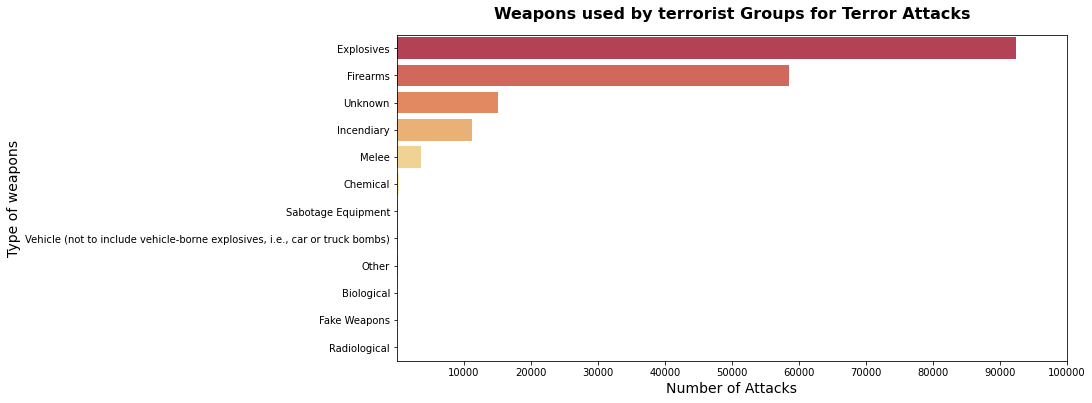


* PATTERN OF GROWTH IN NUMBER OF ATTACKS



Summarizing all the terrorist attacks over the years can provide us an idea about how terrorism has evolved and what rate has it impacted the world each year. Graph shows data from 1970 to 2017 for the total number of attacks happened each year. Terrorist attacks were quite low in numbers in the year 1970. Terrorism then had a fairly rise in the 1980s and early 1990s and was considerably low in the next decade but then terrorism has substantially rose from early the 2005 and highest attacks were in 2014.

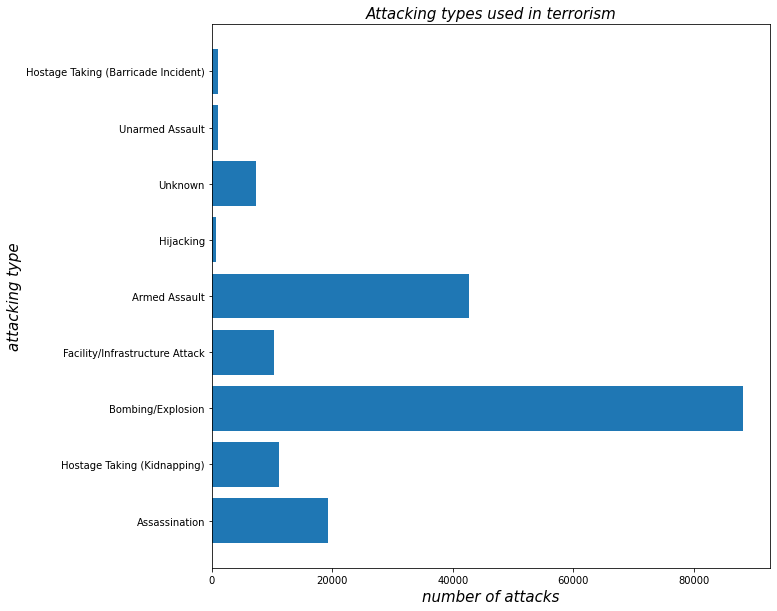
* WEAPONS USED



In the following graph we can see that Explosives are used at maximum times for the attacks, it shows that the terrorist groups are very much aware with the use of explosives, and may have the reservoirs of explosives in abundance. Terrorist groups are also aware about firearms, incendiary, melee.

Many other types of weapons are also used by them, but they are in very less number as compared to above mentioned weapons, so in this graph they are shown as blank.

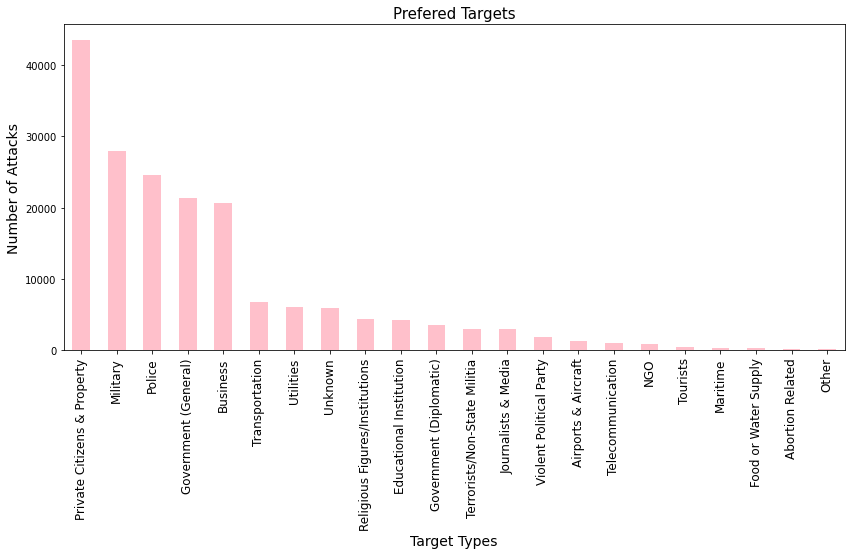
* FAVOURABLE ATTACKING TYPE

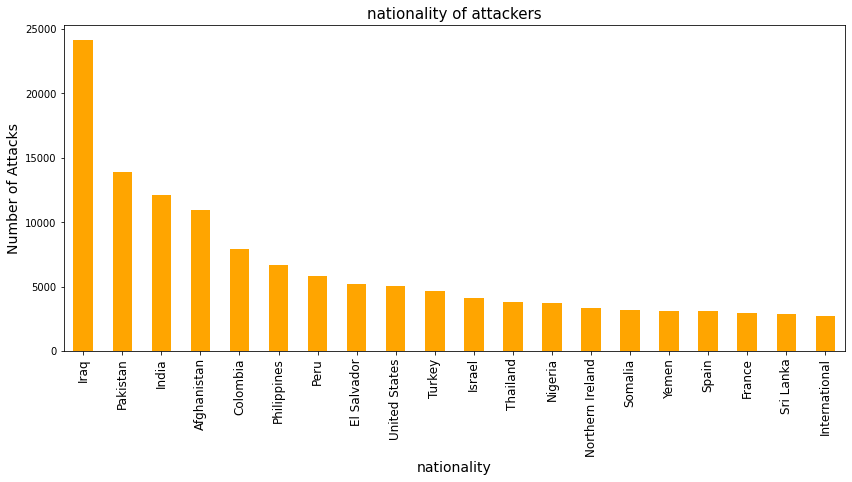


Bombing is the favourite attacking style of terrorist, as they have used this specific style in highest number of attacks.

* FAVOURABLE TARGETS

Analyzing the type of target will help us understand their objective and most likely their motives.Terrorism is driven by an ideology that tries to make a change or impose anideology therefore terrorist are attacking mostly on private citizens and property as the citizens are the largest part of the country.



As the graph is showing that most of the attackers belong to Iraq. and Iraq is indubitably one of the countries most affected by terrorism. So, Iraq should aim to develop and implement comprehensive national responses to prevent and counter-terrorism.

* TARGETS IN PARTICULAR COUNTRIES

The targets in every country is different, the graph is a stacked bar graph showing the attacks in that particular country.

* CONCLUSION

The goal of this project was to understand and interpret the nature of terrorism. After doing the analysis we founded the following insights from the database-

1.TALIBAN is the most active organisation who is taking responsibility of many of the attacks that had happened in the world.

2. IRAQ is the most affected by the attacks, so security groups should be extra cautious in this country.

3. IN 2014, maximum number of attacks has taken place in the world, so it was the dark year of terrorism.

4. IT is seen from the database that the explosives were majorly used by the terrorist groups for the attacks.so security groups should try to locate their sources from where they acquire this explosive.

5. After doing the analysis we founded that that citizens are the major targets of the terrorist group, and the noticeable part is that IRAQ is most affected by the terrorist groups, and maximum number of members of terrorist group belong to the nationality of IRAQ.