**Exploratory Data Analysis on Global Terrorism**

**By - Harsh Jain**

**Data science trainee,**

**AlmaBetter, Bangalore**

**Abstract:**

The Purpose of this project is to analyze the terrorist activities over the past 47 years. And to gather some insights on different terrorist groups and their active attacking areas.

This Exploratory Data Analysis can help understand what is the trend of terrorist activities from the year 1970 to 2017. This Analysis is done with Python libraries like Pandas, Numpy, Matplotlib, and Seaborn. The Dataset used in this project is from the Global terrorism database (GTD) which is an open-source database that keeps records of terrorist incidents that have occurred during a particular time period throughout the globe.

***Keywords: Exploratory Data Analysis, Global Terrorism, Global Terrorism Database (GTD), Python Data Analysis***

**1. Problem Statement**

To explore and analyze the data on Global terrorism to discover key findings pertaining to terrorist activities.

The Dataset contains information on -

* Different terrorist groups
* Countries where these groups have executed attacks
* Date, Month, And year in which attack was executed.
* The number of people killed
* Number of People wounded due to these attacks.
* Motive behind the attacks
* Region-wise distribution of terrorist attacks
* City in which attacks were carried out.
* Latitude and longitude of the places where attacks have been carried out.
* Type of attacks
* Different Target sites for these attacks.
* Different Weapons used for executing attacks.

The main goal of this project is to gather some conclusions on terrorist activities over the previous 47 years and to evaluate the data to determine which areas are hot spots for terrorism, which weapons are most commonly utilized, and which nations, regions, and target categories are most frequently attacked.

These insights from exploratory data analysis on Global Terrorism Dataset can help to track terrorism throughout the world and learn more about different terrorist organizations and where they are most likely to be discovered, and which terrorist groups are most active.

**2. Introduction**

### Terrorism is one of the world's most serious challenges today. Politics, or the death of innocent people for political reasons, is one of the primary motivations for terrorism. So, in my project, I attempted to add some information such as attack patterns, most active organizations, and most affected cities and countries that can be used to develop strategic plans to combat terrorism.

**3. Data Cleaning**

### The Dataset that we are going to explore contains the Terrorist Activities from the year 1970 - 2017.

### The first step is to load the dataset and do various operations on it, such as checking the shape and number of rows and columns. On loading the dataset and checking its shape we get to know that there is a total of 135 columns in our dataset, which is a very large number. Then we checked the null values in the dataset just to know the count of null values in each column.

### By analyzing the count of null values of each column and checking out the entries of each column, many of the columns were rejected and a new dataframe was made with only those columns which are relevant for the analysis.

### Our goal here is to get insights from these relevant columns by using several exploratory data analysis techniques like visualizations and plotting the relations between 2 columns.

## **4. Libraries Used**

* Numpy
* Pandas
* Matplotlib
* Seaborn

Numpy for rounding of some values, pandas for loading the dataset and checking its shape, columns, and number of rows, and also for applying certain functionalities in order to extract valuable insights. Matplotlib and Seaborn are used for making attractive visualizations in order to gain clear relationships and plots from the data.

## **5. Different Regions -:**

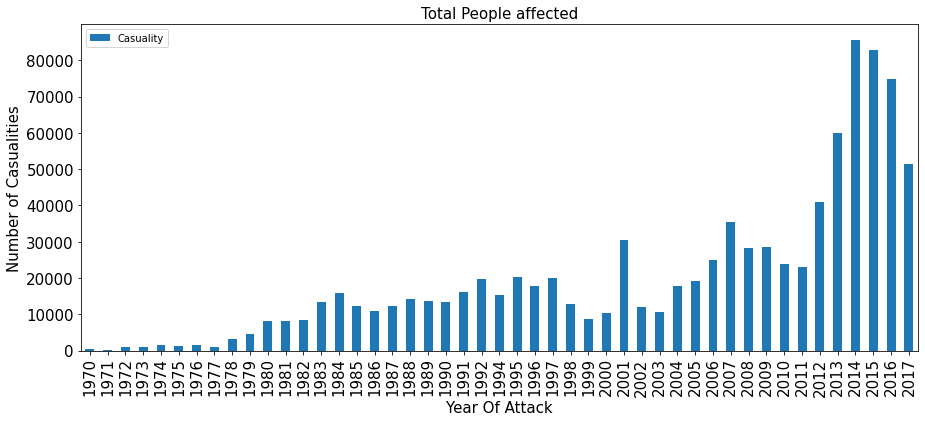
Different regions with terrorist activities that are included in this dataset are:

* Central America & the Caribbean
* North America
* Southeast Asia
* Western Europe
* East Asia
* South America
* Eastern Europe
* Sub-Saharan Africa
* Middle-East & North Africa
* Australasia and Indiana
* South Asia
* Central Asia

# **6. Different Visualizations Used**

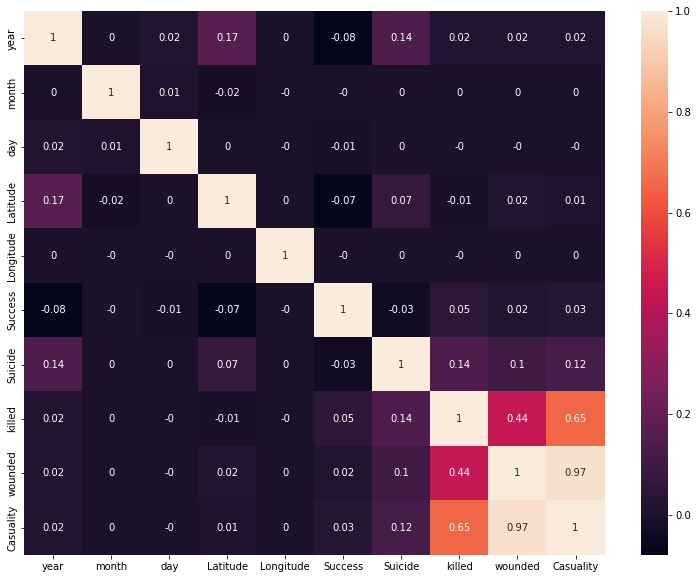
## **Bar-Chart (Matplotlib)**

It is the most commonly used visualization approach in the project, in which we draw the graph between columns and counts to acquire a better understanding of the top entries and top features and observations from the dataset.



## **Correlation Analysis**

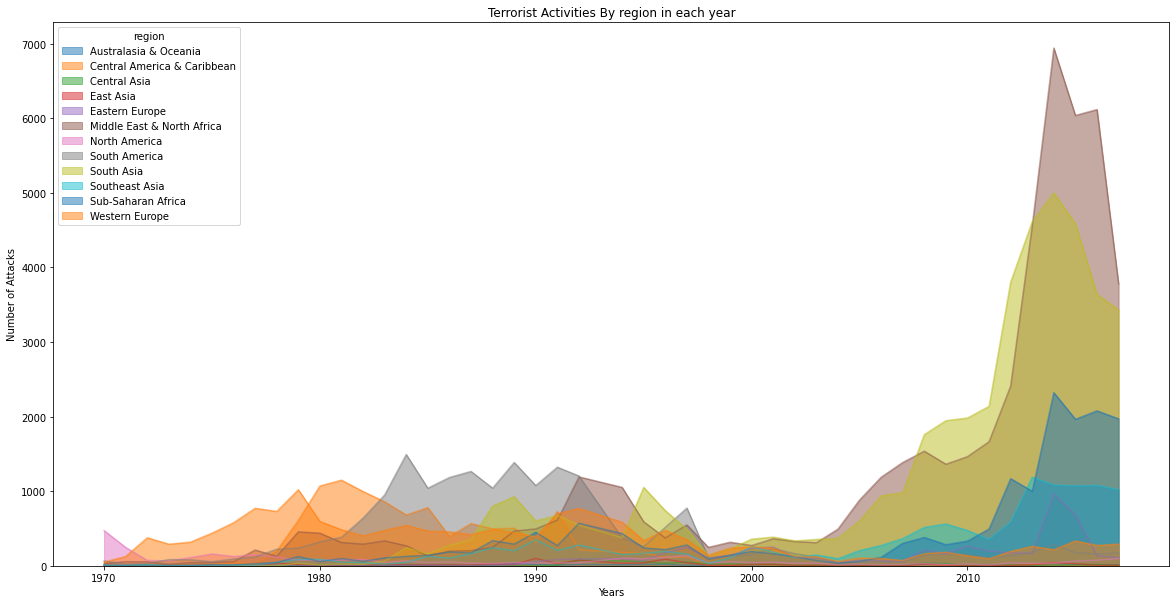
It is the visualization technique used to check the correlation between the features of the dataset. The most correlated features in this dataset were killed (Number of people killed) and wounded (Number of People wounded due to attacks).



## **Area-Plot (Matplotlib)**

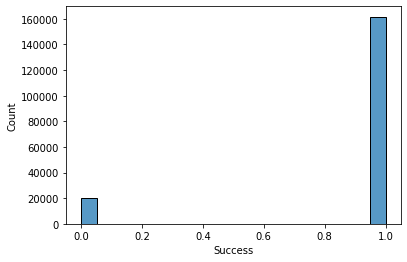
This is the plotting approach used in the project to describe the number of attacks in relation to different regions in our dataset. The area plot displays quantitative data visually.

This plot helped in extracting the insight that there is a peak increase in attacks after 2010 in almost every region, with the Middle East and North African nations topping the list due to the highest number of attacks, and then comes South asian countries at the second position.



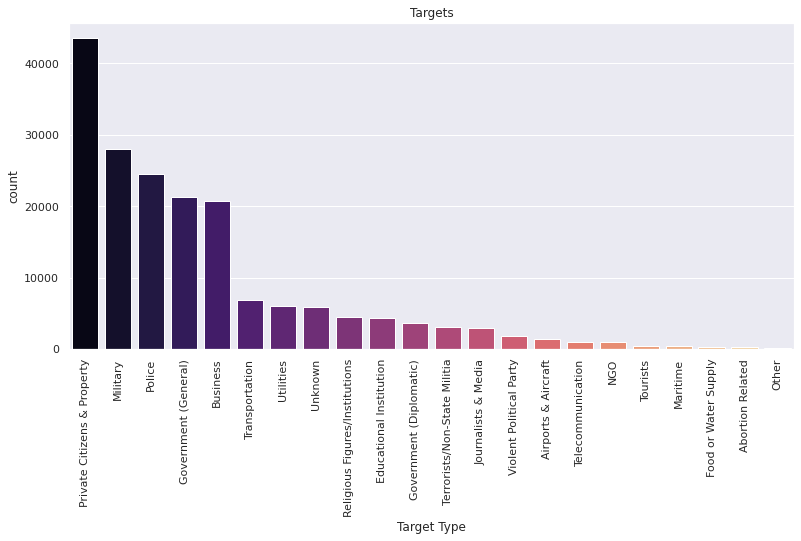
## **HistPlot**

Here histplot (from seaborn library) is used for plotting the histogram plot between the successes and failures of all the terrorist attacks from 1970 to 2017, terrorists were successful in executing 161632 attacks but failed in 20,059 attacks.



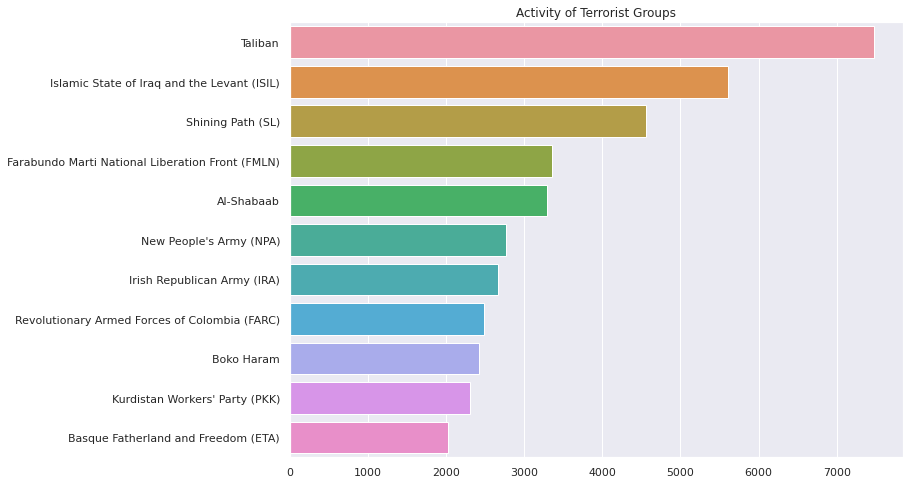
* **Countplot (Seaborn)**

Sns.countplot counts the number of observations in each category and displays the results as a bar chart. The key observation retrieved in the project using Countplot is the main attack target places. Private properties, citizens, military, police, government, and businesses are the primary targets of terrorists to plan attacks.



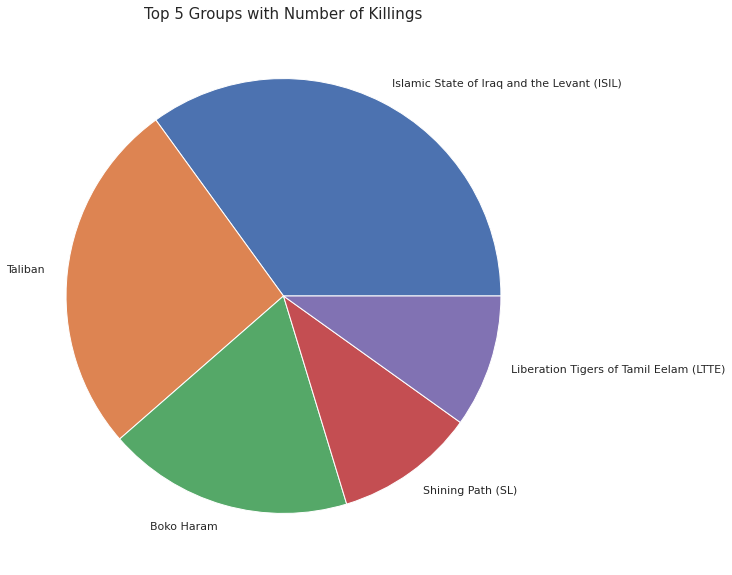
* **Bar Plot (Seaborn)**

Barplot function from the Seaborn library creates new bars between specified x and y values. Key insights that have been derived using bar plot were the number of attacks that every terrorist group has executed in which Taliban tops the list followed by the Islamic State of Iraq and the Levant (ISIL).



* **Pie Chart (Matplotlib)**

The pie function in the Matplotlib Library makes a pie diagram from data stored as an array. It is used to show the distribution of distinct observations in terms of their counts, by plotting them in different colors. The major discovery drawn from the piechart in this project is that (ISIL) is the group with the highest number of killings, followed by the Taliban.



# **7. Value Counts and GroupBy**

These are the two most often utilized functionalities in this EDA on Global Terrorism Dataset. The value counts() method provides objects containing counts of unique values, which can be used to create visually appealing bar charts for comparing different unique values.

While groupby on the other hand groupby() was helpful in splitting the data into separate groups to perform computations for better analysis.

# **8. Steps Involved-:**

* **Exploring Data and Data Filtering-:**

First I have explored data and decided on how to proceed with problem statement. And then according to the dataset I decided on the features that I will be needing in data analysis which also includes dropping of columns with large number of null values.

* **Creation Of Column - Casualties-:**

New column named Casuality was added in order to get the information on total number of people who suffered due to the terrorist attacks. Casuality is equal to the killed + wounded.

* **Yearwise Visualisation-:**

In the next step we did applied value\_counts on the column year in order to get the yearwise count of attacks. Then a matplotlib bar chart was created to show the year vs. the number of attacks in each year.

* **Correlation Analysis-:**

In the next step Correlation analysis was done in order to see what features are correlated with what percentage. The insights from this part could be further useful in model building part.

* **Area-Plot Of attacks by Region-:**

Further we used pd.crosstab functionality from pandas library in order to get the number of attacks region-wise with respect to the year of attack, and then area plot was build using matplotlib. Middle East and North African countries suffered most terrorism.

* **Percentage Increase in attacks-:**

Further we used value\_counts() and to\_dict() functions in order to create dictionary that contains the year as key and number of attacks in that particular year as value. And then some basic maths applied to get the percentage, and as a final result we observe this - 651 attacks happened in 1970 & 10900 attacks happened in 2017. So the percentage at which the number of attacks have increased from 1970 to 2017 is: 94.03 %

* **Most Used Weapons-:**

With value\_counts() and matplotlib, we draw bar chart for the weapons used with number of attacks, we observed that Bombing Weapons and Explosives were mostly used weapons in attacks followed by Armed Assault

* **Successes and Failures-:**

With value\_counts() and histplot from seaborn, it was observed that Terrorists were successful in executing 161632 attacks but failed in 20,059 attacks.

* **Most Targeted Sites-:**

In the next step, with the use of sns.countplot() we observed that the main target sites are Private Properties, Citizens, Military, Police, Government, Businesses

* **Countrywise Distribution-:**

In the next step a bar chart was plotted to get the countrywise distribution of terrorist attacks with Iraq on top with 24,636 attacks and India at the 4th position with 11,960 attacks.

* **Activities of Different Groups-:**

Next we used sns.barplot in order to get the number of attacks that each terrorist organisation has executed. In this Taliban topped followed by Islamic State of Iraq and Levant (ISIL)

* **People Killed vs Terrorist groups-:**

Next we used groupby, sum() and sort\_values() on group\_name feature and killed, we plotted bar chart for the same and observed that ISIL has killed most number of people. And then we also plotted the pie chart of top 5 groups with number of killings.

* **Citywise Distribution of Attacks-:**

In next step we applied value\_counts on city column and sorted them in descending order and kept only top 10 observations for the plot. And we observed that baghdad is the city that suffered most terrorism with 21,151 people killed from 1970-2017.

* **Groups vs Countries-:**

This was done to get the insights on which terrorist organisation is most active in which country. And we observed that Taliban is most active in Afghanistan and ISIL is most active in Iraq while Maoists are active groups in India.

* **Attacks Carried Out in India-:**

By using bar chart visualisation we observed that In India Jammu Kashmir faced the most terrorism that is 2454 attacks followed by Assam and Manipur. Srinagar and Imphal are top 2 cities which suffered most terrorism between 1970-2017

Sikh Extremists, Maoists and Muslim Separatists were the most active terrorist groups between 1970-2017.

**Conclusions-:**

* In our EDA we used different libraries such as pandas, NumPy, matplotlib, and seaborn.
* Attractive Visualisations were made with the help of Matplotlib and Seaborn to display the findings in an attractive and engaging way.
* Different functionalities like value\_counts(), groupby(), sort\_values() etc. were mostly used in order to plot useful insights.
* The Problem Statement was divided into several parts in order to ease things and the process of data analysis.
* Thus from the findings of this analysis, National Security Agencies can plan on planting their agents in the most active sites for terrorism in order to get early insights on the plan of terrorist groups and to avoid any future attacks.
* These findings can further help security agencies on where to find the most active terrorist groups, and the weapons they mostly used for carrying out attacks.
* These insights can help monitor the activities that take place in the most active countries and cities for terrorist attacks.

***Some of the key findings from the overall analysis are-:***

* Iraq is the country where the most terrorist attacks took place in the past years between 1970 - 2017.
* The Middle East and North African Countries are the Prime targets for terrorist organizations.
* Maximum Number of Attacks happened in the year of 2014, i.e. 16903, it was also the year with maximum number of casualities. And the minimum Number of Attacks happened in the year of 1971, i.e. 471.
* From the Area Plot it can be seen that after 2010 there was a peak increase in attacks in almost every region.
* The percentage at which the number of attacks have increased from 1970 to 2017 is: 94.03 %
* Most used weapons were bombing and explosives and the prime target sites for terrorism are citizens and public properties, police, and military.
* Terrorists were successful in executing 161632 attacks but failed in 20,059 attacks.
* In India from 1970 to 2017, the number of People Killed is equal to - 19,341 because of terrorist activities.
* The main Groups that carried out several attacks in India were Maoist and Sikh Extremists and Muslim Separatists
* Jammu Kashmir was the most active site for terrorist activities in the year 1970-2017.
* With 7478 attacks, the Taliban is the most active terrorist organization, it is mostly active in Afghanistan.
* (ISIL) is the group with the most killings, followed by the Taliban.
* Baghdad is the city that suffered the most terrorism.
* Total Number of Casualities from the year 1970 to 2017 because of terrorism was: 9,35,737.0
* Total Number of People killed from the year 1970 to 2017 because of terrorism was: 4,11,868.0
* Total Number of Wounded People from the year 1970 to 2017 because of terrorism was: 5,23,869.0

**References-**

1. StackOverflow
2. GeeksforGeeks
3. Kaggle