**Global terrorism Analysis**

Rishi Gupta

Data Science Trainee,

Almabetter.

**Abstract**

The Global Terrorism Database (GTD) is an open-source database including information on terrorist attacks around the world from 1970 through 2017. The GTD includes systematic data on domestic as well as international terrorist incidents that have occurred during this time period and now includes more than 180,000 attacks. The database is maintained by researchers at the National Consortium for the Study of Terrorism and Responses to Terrorism (START), headquartered at the University of Maryland.

### Characteristics of the Database

* Contains information on over 200,000 terrorist attacks
* Currently the most comprehensive unclassified database on terrorist attacks in the world
* Efficiently covers attacks between 1970 to 17 September, 2019
* More than 4,000,000 news articles and 25,000 news sources were reviewed to collect incident data from 1998 to 2019 alone
* The attacks in the GTD are attributed to more than 2,000 named perpetrator organizations

**Problem Statement**

The Global Terrorism Database (GTD) documents more than 200,000 international and domestic terrorist attacks that occurred worldwide since 1970. An action must also be carried out for political, economic, religious, or social purposes to count as terrorism. The GTD defined as: "The unlawful use of force or violence against persons or property to intimidate or coerce a Government, the civilian population, or any segment thereof, in furtherance of political or social objectives." The GTD includes more than 83,000 bombings . It also includes more than 18000 assassinations and more than 11000 kidnappings.

We have to perform Exploratory Data Analysis to find out the following:

1. year Vs Number of terrorist attack?
2. which top 5 countries has highest terrorist attacks?
3. which 5 cities has hightest terrorist attacks?
4. which attack types mostly used?
5. which regions has hightest terrorist attacks?
6. which weapon is mostly used by terrorist?
7. find motive behind the terrorist attacks?
8. who are main targets of attacks?
9. which gropes are mostly active terrorist attacks?

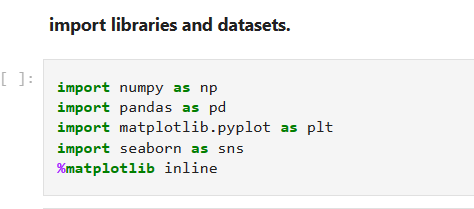
**Introduction**

Terrorism indeed overshadows every aspect of economic, social, cultural and political life. While it brings instability and disrupts peace and coexistence environments, it directly endangers the lives of people and brings every type of violence in the society. To appeal for wider attention on such heinous action and contempt for life, we are going to propose a visualization system, which will address different aspects of the danger of terrorism worldwide.

Our visualization helps the users to gain a more comprehensive view of the terrorism attacks happened during 1970-2017, as well as provides the functionalities to view accurate number and make comparisons across different countries.

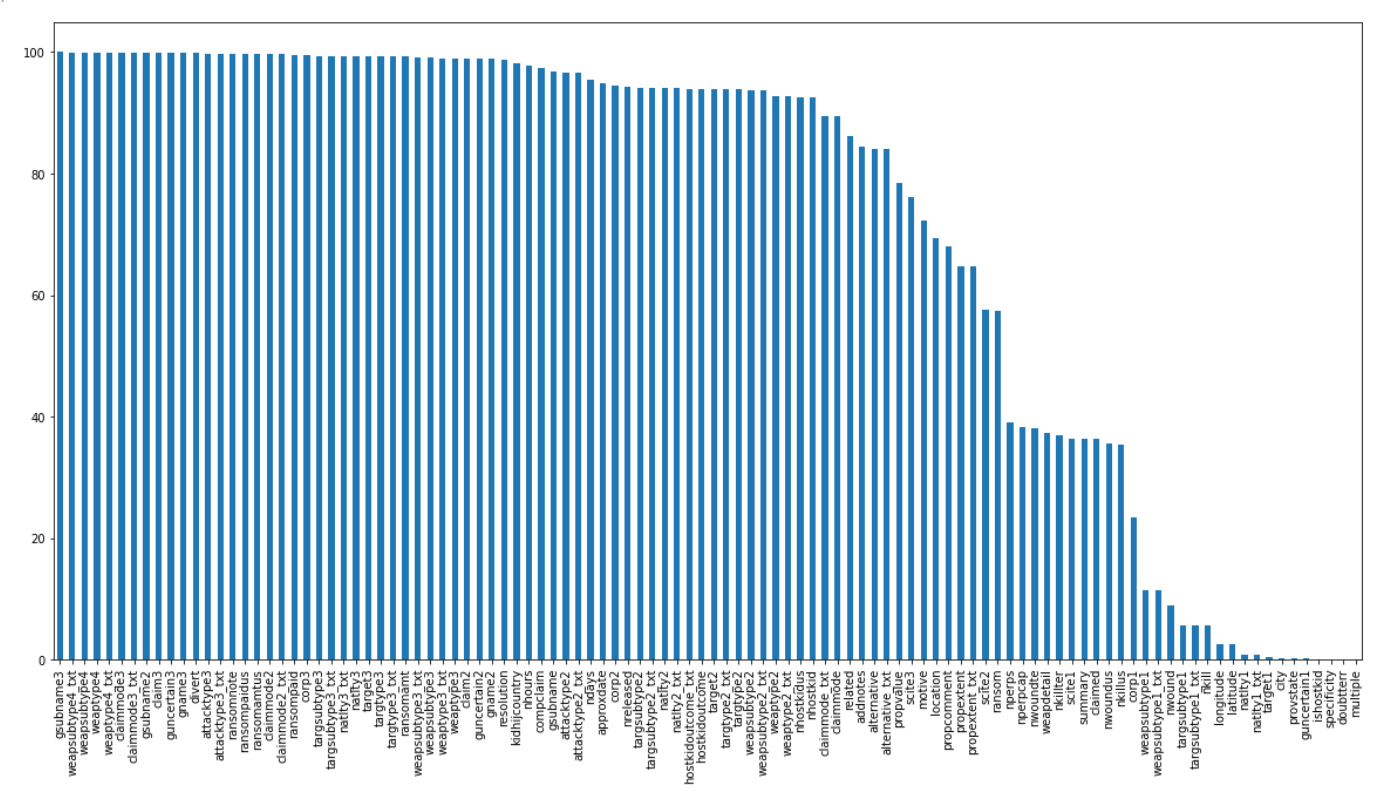
**Data Preparation**

Step 1- Importing the required libraries. The entire research is performed on Google collab-Python programming language. Python has earned the title of one of the most popular language for machine learning task due to its vast collection of libraries. The machine learning libraries used in this research are Numpy, Pandas, Seaborn and Matplotlib.



Step 2- Importing the Dataset. The dataset is present in csv format consisting of tabular data stored in plain text. The read\_csv() method of pandas library is used to create a data frame of given dataset.

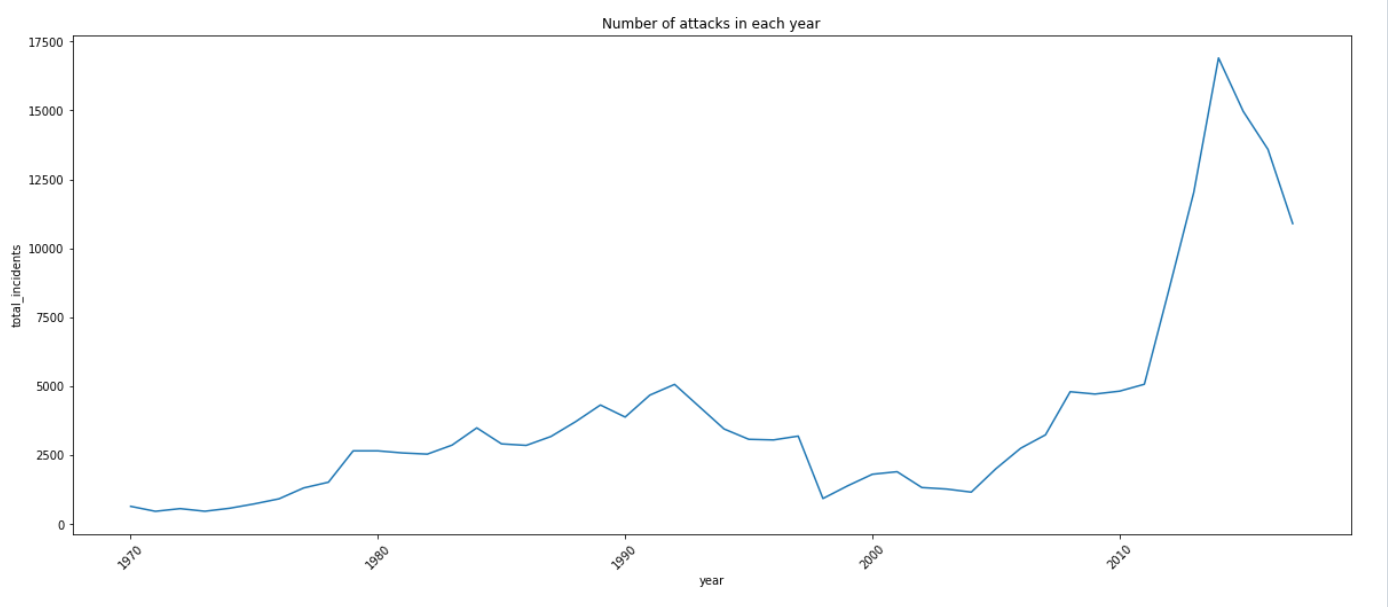
Step 3- Handling the missing values. In the dataset we can see lots of variables have more than 99 percent data has missing, so : Theoretically, 25 to 30% maximum missing values are allowed, beyond which we might want to drop the variable from analysis. Practically totally depends upon the importance of the variable.



In the dataset 87 those variable have more than 30 percent missing values. We have replaced the missing values with “unknown’ value. Now our dataset is ready and we can perform Exploratory Data analysis on it.

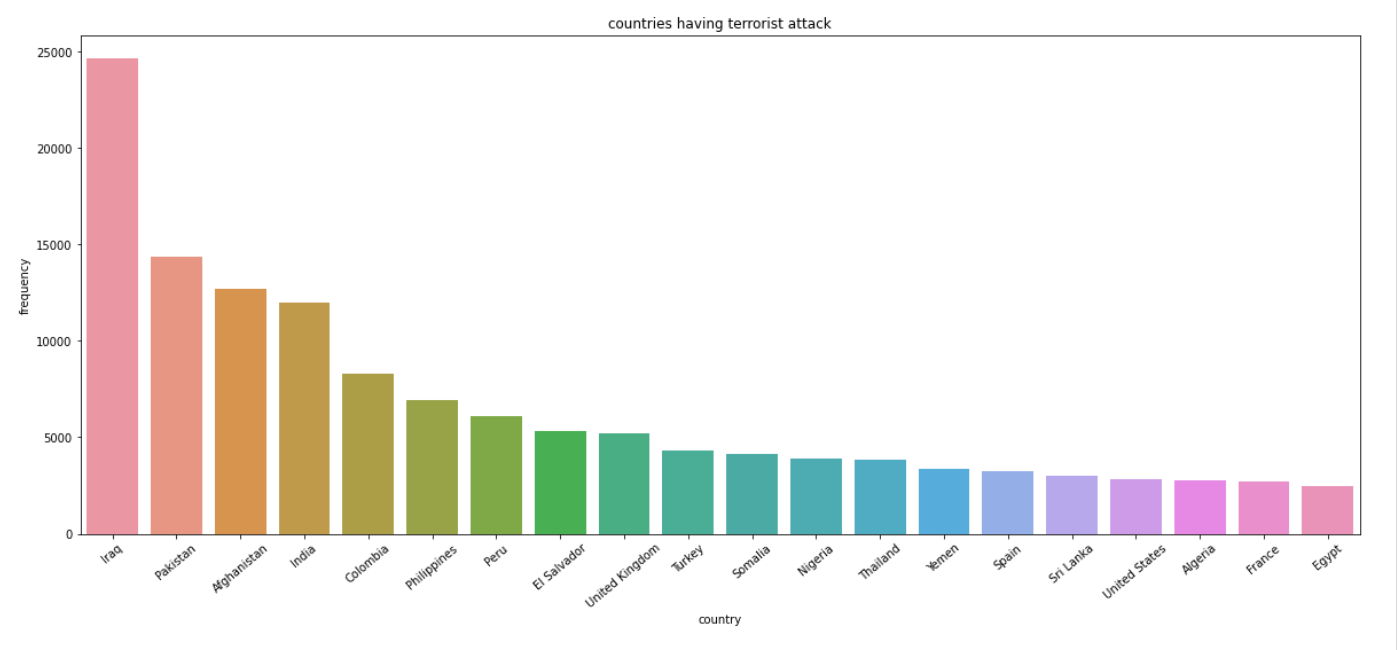
**Exploratory Data Analysis**

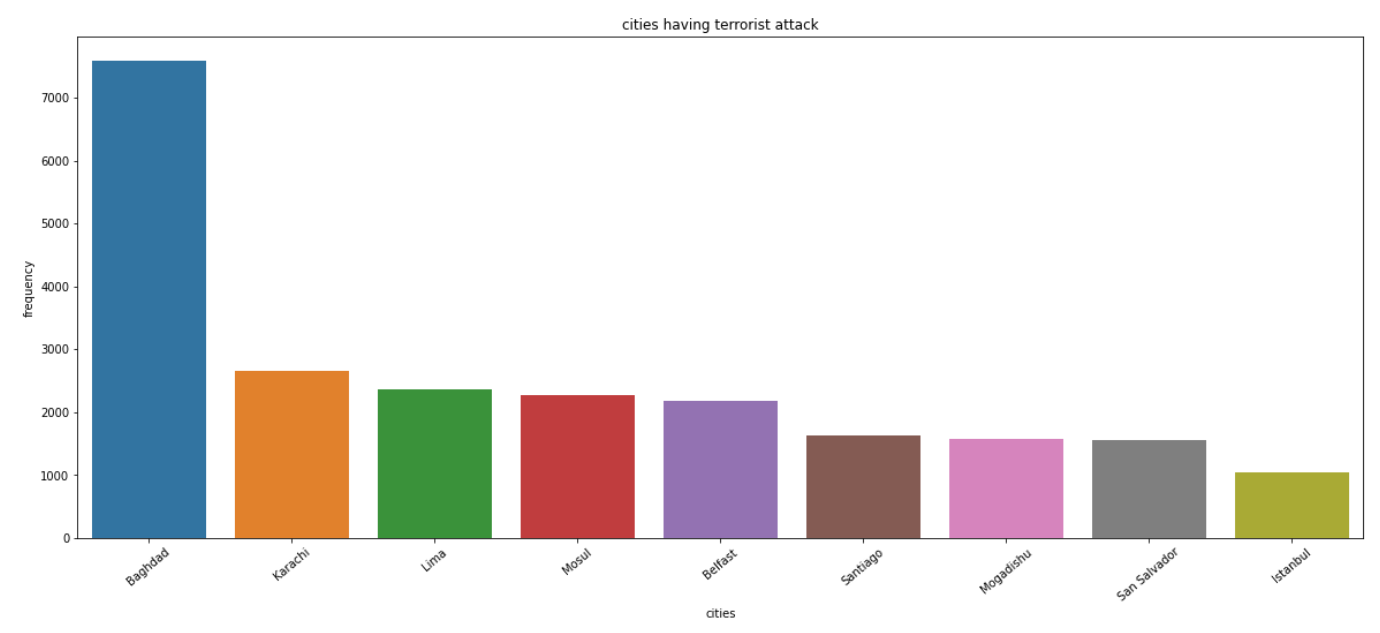
Exploratory Data Analysis (EDA) is an analysis approach that identifies general patterns in the data. These patterns include outliers and features of the data that might be unexpected.EDA is an important first step in any data analysis. Understanding where outliers occur and how variables are related can help one design statistical analyses that yield meaningful results.

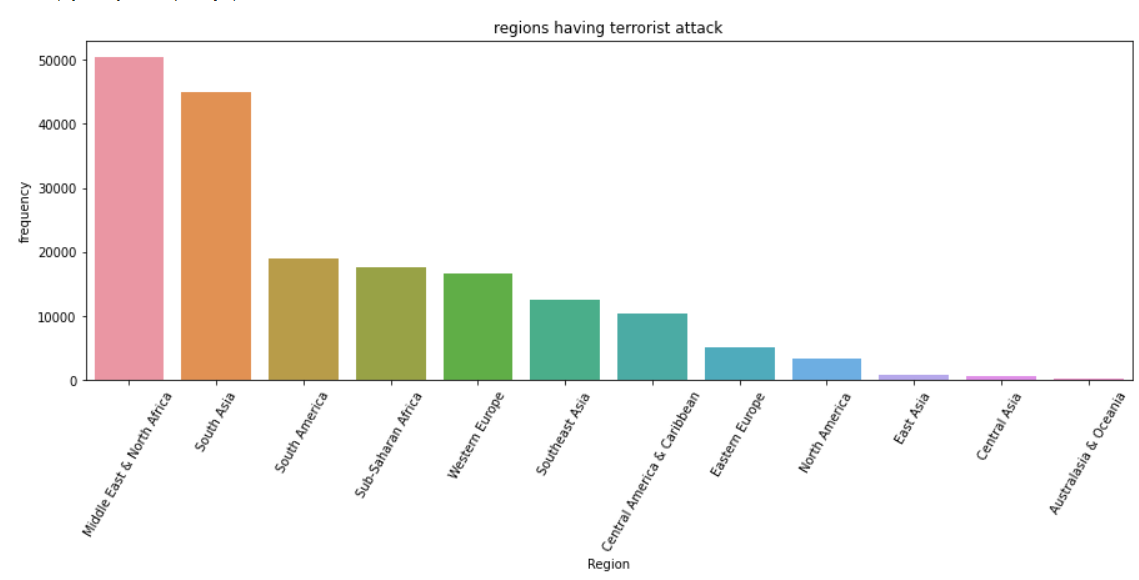
1)Plotting number of attacks each year with the help of line graph.

The terrorist attacks gradual increase in 2004.The highest peak of attacks in 2014. After 2014 attacks decreases.

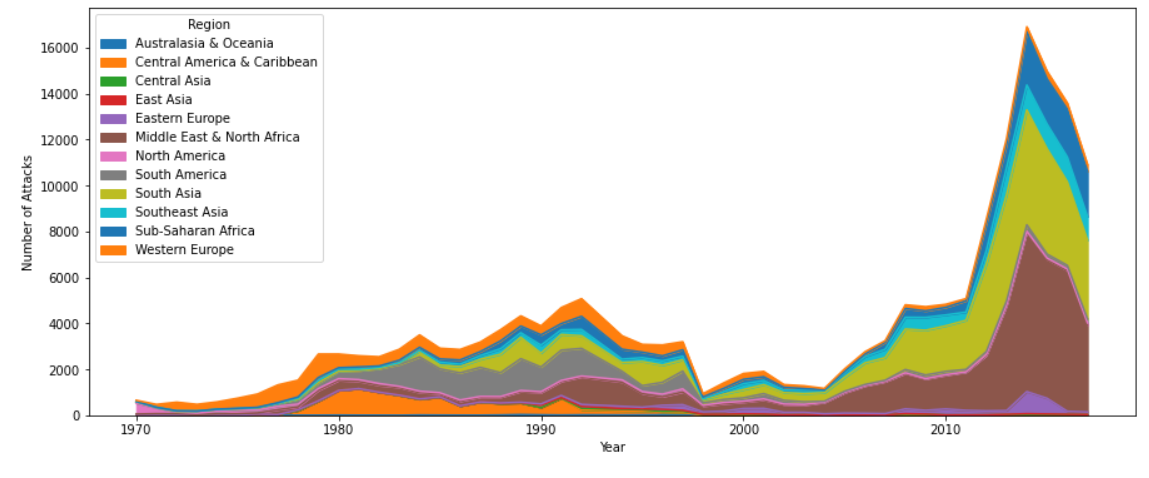
2) Finding out countries, cities and regions with highest terror attacks.





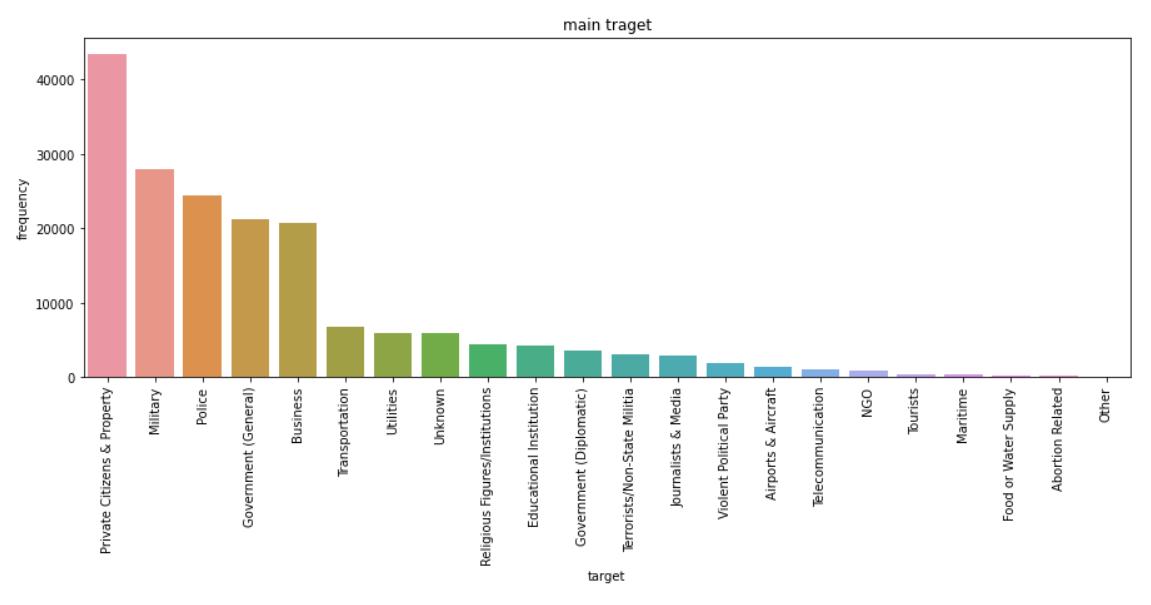


3)Finding out terror activity in each year w.r.t region



From barplot and cross tab we can see that Middle East & North Africa, South Africa are highest terrorist attack region. Australia and Oceania and Central America has maximum attack after 2004. There have been less terrorist attack in East Asia and north America.

4)Places mostly targeted by terror groups



Private Citizens & Property, Military, Police, Government (General), Business these are major targets for terrorist groups.

**Conclusion:**

* The time interval between 1979 to 1997, few terrorist attacks but after 2005 attacks are increased year by year.
* Iraq (24636), Pakistan (14368), Afghanistan (12731), India (11960) and Colombia (8306) these countries has highest terrorist attacks.
* Middle east&North Africa are most prone to terror attacks
* Baghdad, Karachi, Lima, Mosul, Belfast these cities has highest attacks in world.
* Australia and Oceania and Central America has maximum attack after 2004.
* Islamic State of Iraq and the Leavant (ISIL) this group has no activities 1070 to 2011 but after 2011 this group is active.
* Taliban is continuously active.
* Till 2001 Iraq and Afghanistan has approximately 0 attacks but after 2002 Iraq has more attacks as compare to Afghanistan.