**Global Terrorism Analysis**

**Abstract:**

The Global Terrorism Database (GTD) documents more than 200,000 international and domestic terrorist attacks that occurred worldwide since 1970. With details on various dimensions of each attack, the GTD familiarizes analysts, policymakers, scholars, and journalists with patterns of terrorism. The GTD defines terrorist attacks as: The threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation*.* Data collection is ongoing and updates are published annually at [www.start.umd.edu/gtd](http://www.start.umd.edu/gtd).

The Global Terrorism Data analytics is qualitative and quantitative process and techniques use to find data about terrorism and identify areas or region that affected.

We’ve performed project right from proper data cleaning to analysis a problem statement. For better understanding we’ve accomplished project in general. The task of the analysis has helped a lot to understanding the motive of terrorism,

How terrorist attacks have been increased in recent years. Which type of equipment were used by terrorists? Working on different view

of analysis to make it more reliable which provide in this technical documentation.

EDA can benefit the government, investigating agency, UN, UN peace keeping missions, companies for their plant set up or

business or private investment. EDA also benefit local governing bodies to help curb on such radical group.

Keywords: The Global Terrorism Database, Exploratory Data Analysis

**EDA, Data Visualization**

1. **Problem Statement**

Terrorism is complex political and social phenomenon. Terrorist attacks have a significant threat to the safety and security of the international community and have become one of the greatest obstacles to the sustainable development of global social security. Antiterrorism is an important part of global security governance, which is sustainability issue that guarantees global security development.

Explore and analyze the data to discover key understanding:

* Which regions/area more suffering from terrorism?
* How many countries are victim of terrorism?
* How many terrorist organizations work in world?
* Which countries witness more terrorist attacks?
* Motive of terrorist attack?

1. **Introduction**

The Global Terrorism Database (GTD) is an open-source database including information on terrorist event around the world from 1970 through 2017. Unlike many event databases, the GTD includes systematic data on domestic as well as transnational and international terrorist incidents that have occurred during the time period and now include more than 180000 cases. For each GTD incident, information is available on the date and location of the incident, weapon use and nature of target, the no of casualties and group or individual responsible for this act.

## 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.

Terrorist attacks typically involve high lethality and destructive power and directly cause massive casualties and property loss. In addition they bring tremendous psychological pressure on people. In summary, terrorist attacks result in social unrest to a certain extent.

We will explore and visualize the dataset from GTD on world and also India using basic exploratory data analysis techniques.

We will find out Middle East and North Africa region witness more terrorism and terror activities are increase in this areas.

The goal here is to explore the data and find useful insights from data and find out different relations between the column **The Global Terrorism Database (GTD)**

This dataset has around 180000 Observation in it with 135 columns and it is mix of categorical and numerical values like date, year, country, location, motive. Exploring them will definitely help us have a very good understanding.

**Column Information**

* Year = Describe about year of incident
* Month = In which month attack happen
* Day = On which day attack executed
* Region = Describe region in the word
* City = City name where attack happen
* Latitude = place of location
* Longitude = place of location
* Attack type = which type of attack
* Killed = how many people kill
* Wounded = how many people injured
* Target = who is the prime target of attacker
* Summary = What happen in attack
* Group = which group is responsible for attack
* Target\_type = type of target
* Weapon\_type = types of weapon use
* Motive = motive of terror activity
* Success = Successful attack

# **Steps involved**

* **Setting up notebook**

The notebook is setup in Google Collaboratory platform. The google drive containing the notebook and it is loaded as a pandas data frame. The necessary library such as Numpy, pandas (for working on the data frame ), seaborn and matplotlib (for visualization) are imported.

* **Cleaning the Dataset**

**Null Values:** The dataset contains a large amount of null values. The column name, month, motive, claimed, region contain large no of null values. So, we tried to preserve as many rows as possible by replacing null values with suitable values.

**Data with error:** These data were considered as erroneous data and removed.

* **Exploratory Data Analysis**

During the preparation of the dataset for EDA, a basic statistical summarization was done. The information regarding the data type of each column was explored.

This was followed by a detailed exploratory data analysis. The relation of region and attacks, terrorist organization categories features and visualization. Relationship between variables was observed.

Throughout the analysis, we tried to answer question that help us understand terrorist activity trends.

## **Analysis on Terrorist Attacks Worldwide**

Pandas df.nunique() function returns object containing unique value count in the dataset. The result object will be describe the no of country present in the dataset. In this dataset

250 countries present that means in the world 250 country affected bt terrorism.

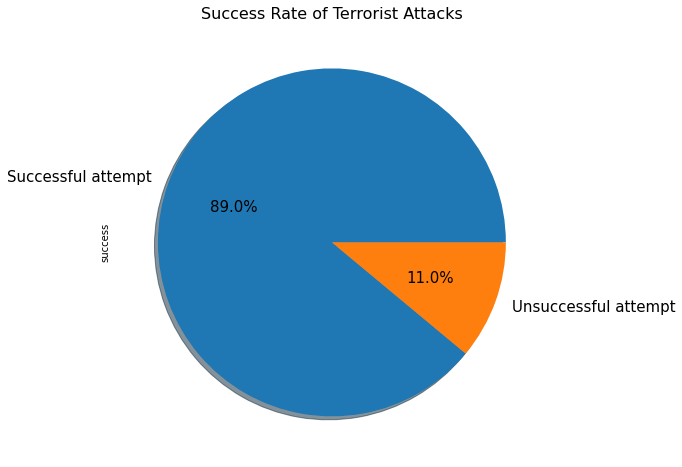
1. **How many successful attacks happened in the world?**

Pandas df.map() function use to map series in pandas. The result of this function describe the successful and unsuccessful number of activity.

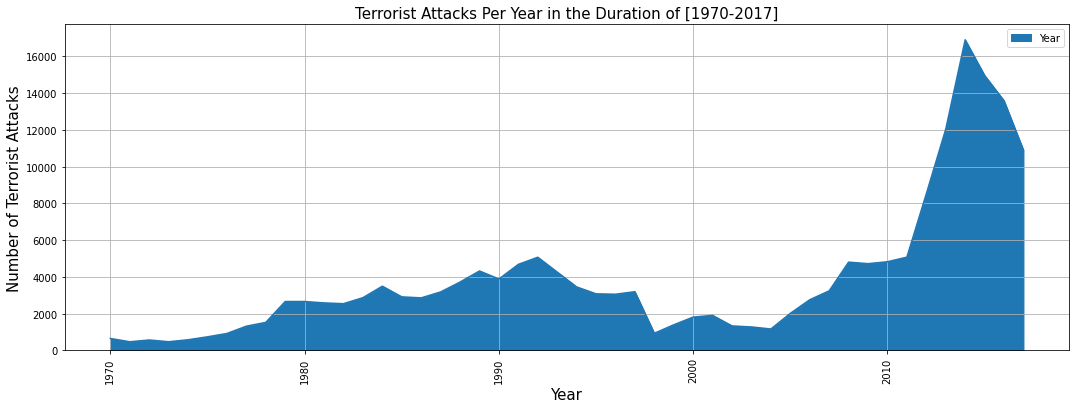
Basic Pie Chart Matplotlib

Pie charts don’t have the best reputation in the visualization community, but there are times when you want one anyway.

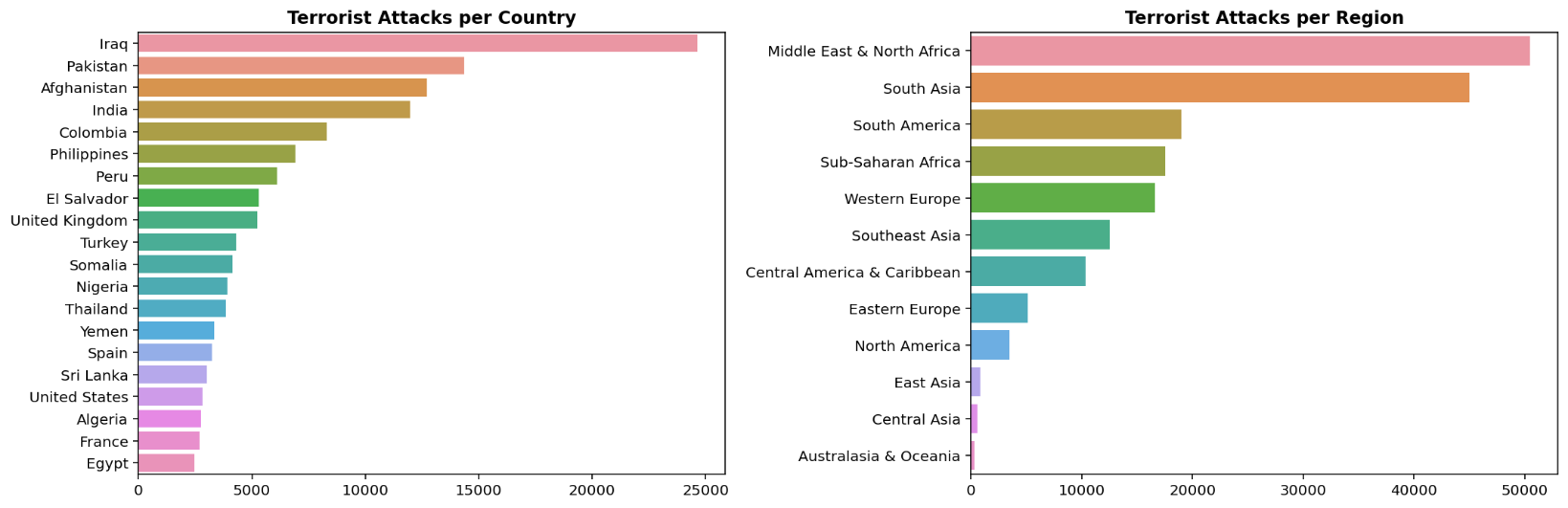
Result of map function use in the pie chart, So it very useful to directly analysis in the percentage of successful and unsuccessful attacks done by terrorist.



**Insights:** Thesuccess rateof theseterrorist attacks is huge that is89%**,** this is a very big issue of concern for the security agencies around the world.



### **Year wise terrorist attacks around the World**

df.isnull() function write all null values present in the dataset. For proper analysis need to remove null value with help of this function. df.unique.year give uniqueall unique yearfrom 1970 to 2017.We can observe that the data for the year **1993** is missing from the df.groupby function use to make group of similar values. With the help of above function we calculate terrorist attack per year.With help of matplotlib library plot 

area graph to visualize trends of terrorist attack in the world.

#### **Insights:** The number of attacks increased exponentially through the years from 1970 to 2017 that is around more than 15 times**.** In 1990 world witness rise in terrorism then gradually in decrease in 2000 but after 2010 terrorist activity exponentially increase.

### **Number of terrorist attacks in each country and region**

Plt.subplots() this matplotlib library function is use to plot a graph by dividing space for

different-different graph. This function has a capability show all kind of graph on single page. In subplot we use bar plot to show the trends of terrorist activity per year

#### The most number of attacks took place till 2017 are in **Iraq**approximately 25000which is huge in last 50 years.

#### Number of terrorist attacks are more in**Iraq, Pakistan** and **Afghanistan**and region wise it is more in **Middle East and North Africa** **regions**.

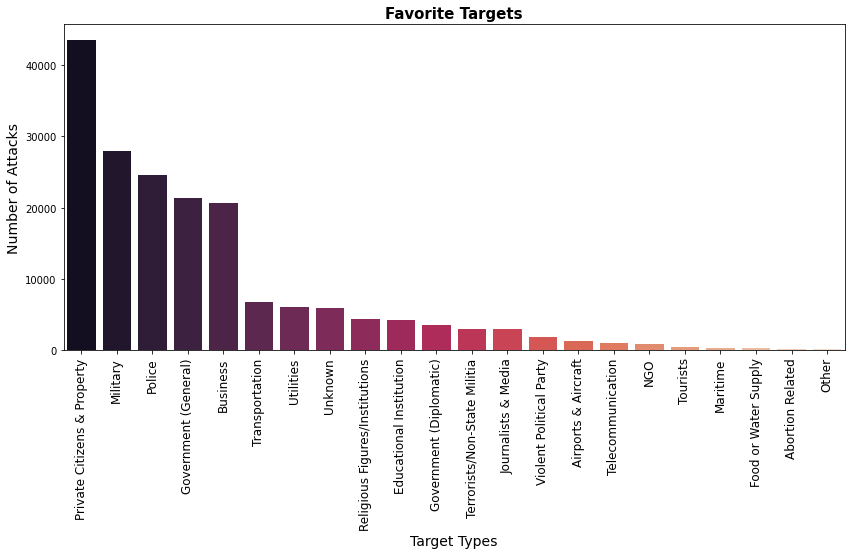
### **Finding out the terrorist organizations that carried out the operations in different countries and methods used in different attacks.**

df.value\_counts() function use to count the values in the respective column. plt.subplot() use to show to two pie chart in the graph.autopct='%1.1f%%' method is used to write percentage in pie chart.

Plt.subplot() function make one row and two column and figsize() method give the dimension to the graph. In the group all terrorist organization present with help of value\_count() method count particular groups involvement in terrorist activity and give limit ten so we only find top ten terrorist organization according to their involvement in activity. Kind = pie is suggested here pie chart autopct() method return in pie chart percentage value base on value\_count. We use subplot so need to identify space where this graph to be drawn So we use index method {ax=axes[0]}.

df.set\_title()method use to give title to the graph. axes[0] return most active terrorism and axes[1] return attacking method by terrorist.

#### **Insights:** The most active terrorist organization is **Taliban** and the favorite attacking technique used by these organizations is Bombing/Explosion.



### **Favorite targets of terrorist groups**

Here seaborn library is used to show the graph. On x-axis target\_type and on the Y-axis there is number of attacks.

#### **Insights:** The most targeted assets of these organizations arePrivate Citizens & Propertyfollowed byMilitaryandPolice**.**

## **Analysis on Terrorist Attacks in India**

df.sort\_values() function is use to sorting values in ascending order. In case we provide in parenthesis descending= True then it work in reverse manner.

Plt.plot(kind=line) function is use to draw line graph and it is very useful command in mantplotlib.

Here we sorted value on year so get from 1970 to 2017 and that value plot into line graph.

Plt.xticks() are useful function. It take very useful argument like rotation, fontsize like that also similarly plt.ysticks()

plt.title() use to give title to the graph also its take others argument.

plt.xlabel() give title to the x-axis

plt.ylabel() give title to the y-axis

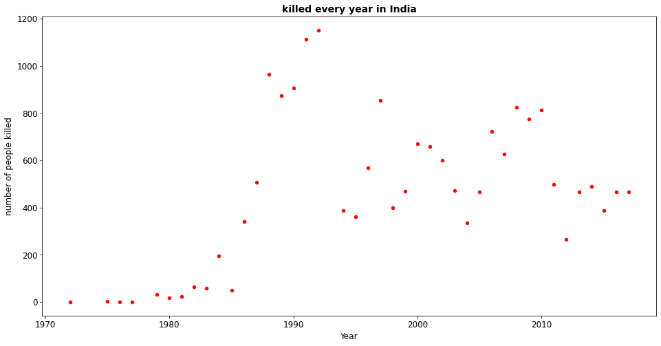
**Insights:** We can seeanexponential growthin the number of attacksand it's the maximum in theyear 2016with1025 attacks**.**

### **Killings happened in India every year**

plt.scatter() chart is shown below is very useful to show the information. Here we have used groupby function on year and sum() apply on killed column so we get per year killed people by terrorist.

#### **Insights:** In the **year 1992** there were

highest number of deaths in India.

**6. Conclusions** 

Finally able to answer some really important question about Terrorism analysis using this dataset. By studying the trends of the past terror activity, we can take steps to prevent terrorist attack and understand where to focus the efforts the most to take to make the most positive result.

* Thesuccess rateof theseterrorist attacks is huge thatis89%**,** this is a very big issue of concern for the security agencies around the world
* 205 countriesin the world areaffected by terrorism according to the Dataset.
* The most number of attacks took place till 2017 are inIraqapproximately 25000which is huge in last 50 years.
* Number of terrorist attacks are more inIraq, Pakistan and Afghanistanand region wise it is more inMiddle East and North Africaregions.
* The most targeted assets of these organizations arePrivate Citizens & Propertyfollowed by MilitaryandPolice**.**
* Talibanis the terrorist organization thattargeted Afghanistanthe most.
* Srinaga**r** was the mostfrequently attacked city in India which caused540 deaths**.**
* The most active terrorist groups are SikhExtremists followed by CPI-Maoist causing the maximum number of deaths in India.