# Global Terrorism Database

**Junaid A R**
**Data science trainee,**
**AlmaBetter, Bangalore**

## Abstract:

Terrorist attacks have become a serious source of risk affecting the security of the international community. Using the Global Terrorism Database (GTD), in order to quantitatively study past terrorist attacks and their temporal and spatial evolution the analytic hierarchy process (AHP) was used to classify the degree of damage from terrorist attacks. The various factors influencing terrorist attacks were extracted and represented in three dimensions. Subsequently, using MATLAB for analysis and processing, the grading standards for terrorist attacks were classified into five levels according to the degree of hazard. Civilians are the targets most at risk for terrorist attacks, and the corresponding risk index is considerably higher than it is for other targets.

***Keywords: terrorist attacks; global terrorism database; quantitative analysis;***

## 1.Problem Statement

- **Misconception about Terrorism:** 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. 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
- **Factors Affecting Terrorism:** Identifying dependent factors of terrorism is one of the goals of this project. There are parameters like religion or nationalism which are not defined in the dataset but have a major influence on contemporary terrorism
- **Dataset Challenges:** Global Terrorism Database (GTD) has a marginally low occurrence of events occurred at the same geolocation. Most of the events are not consistent or do not occur frequently [15]. Hence difficulty arise in making quantitative projections with varying degrees of similar events. Another major challenge while working on this dataset is that individual studies lead to different conclusions. Current shortcomings and limitations in data collection techniques, definition debates, irregularity in coding and analysis give rise to disagreements among researchers and in turn ruling out their conclusion

## 2. Introduction

World peace was one of the core reasons for forming the United Nations organization. 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. In this project, we focus on terrorism by analyzing the dataset provided by START (Study of Terrorism and Response to Terrorism) Consortium to explore meaningful patterns and statistics.

## 3. Data Summary

Original Data : Originally we have the data in the size of 181691 rows and 135 columns. There were following three types of data
float(64) = 55
Int64 =22
object =58
*Data Cleaning:* After dropping the columns which have null values greater than 30% we got our data in the size of 181691 rows and 48 columns. Again, we removed some unwanted columns to finally get our data matrix in the size of 181691 rows and 21 columns. We renamed some of the columns to the best suitable names. Following were the data types of the of different columns
float64 =5
int64=8
Object =8

*Columns.*

**Year**: This column contains the year in which the incident occurred. In the case of incident(s) occurring over an extended period, the column will record the year when the incident was initiated.

**Month**: This column contains the months in which the incident occurred. In the case of incident(s) occurring over an extended period , the column will record the year when the incident was initiated.

**Day**: This column contains the numeric day of the month on which the incident occurred.
Extended (Categorical variable):
1=" yes" The duration of an incident extended more than 24 hours.
0=" No" The duration of an incident extended less than 24 hours.

**Country**: This column identifies the country where the incident occurred. In the case where the country in which an incident occurred cannot be identified, it is coded as "Unknown".
**Region** : This column identifies in which the incident occurred. The regions are divided into the 12 categories, and dependent on the country code.
**City** : This column contains the name of the city, village or town in which the incident occurred.
**Latitude** :This column records the latitude ( based on WGS1984 standards) of the city in which the event occurred.
**Longitude**: This column records the longitude ( based on WGS1984 standards) of the city in which the event occurred.
**Attack_type** : This column captures the general method of attack and often reflects the broad class of the tactics used. It consists of nine categories which are as follows.
1. Assassination
2. Hijacking
3. Kidnapping
4. Barricade Incident
5. Bombing/Explosion
6. Armed assault
7. Unarmed Assault
8. Facility/Infrastructure Attack
9. Unknown

**Weapon type** : Up to four weapon types are recorded for each incident. This column records the general type of weapon used in the incident.
**Target type**: The target/victim column captures the type of target/victim.

**Nationality**: This column represents the nationality of the target that was attacked, and is not necessarily the same as the country in which the incident occurred, although in most cases it is.
**Group**: This field contains the name of the group that carried out the attack.

# 4. Steps Involved

***Reading the GTD data set:***
We have mounted the given data set and could see that it holds 181691 rows and 135 columns. It holds vast parameters that can be further used and help in analyzing the data set and come to the conclusion of required data set.

***Data Exploration:***
In this we first found the head and tail of the DataFrame. Then we went for the shape, info of the DataFrame and finally we found the basic description (likecount,mean,std,min,1st quartile,2nd quartile,etc) of each column.

***Data Cleaning***, In this step we removed the columns which had null values more than 30% and created a new Data Frame to get the shape of (181691,48). After this also we removed some unnecessary columns which are not required for our visualization to finally get the Data Frame of shape (181691,21). After that we found out the head and tail of our final Data Frame and printed the columns to show the percentage of null values in each column. Finally in the data cleaning step we print out the data types of each column.

***Analyzing and visualizing the data:***

***We have***

No of Attacks per year in descending order

Countries that were mostly attacked

Plotting the top 20 countries which were attacked most.

Different types of attack

Visualising the attack types that were most common with respect to the period.

Plotting the year and number of attacks that year

We can observe that there is a sudden spike in the graph after 2004.So We are deep diving into the data after 2004 to find out the reason behind the spike.

**6. Conclusion:**

1) The main growth of terrorism was from 2005-2014 and then it started declining till the data is available.
2) Iraq, Afghanistan, Pakistan & India were among top affected countries by terrorism.
3) In Iraq, Afghanistan, Pakistan approx. 85% attacks took place after 2005 but in case of India only 67% attacks took place after 2004, this means that terrorism was there in India even before 2004 and hence, it was a major problem from past many decades .
4) Iraq had reported most casualties.
5) It was the time (2004) when major terrorist groups ISIL in Iraq and Taliban in Afghanistan was emerging.
6) In India there are two major groups active, Communist Party of India – Maoist (CPI-M) and Maoists.
7) If we want to reduce terrorism from the world then we have to try to suppress the major groups of these top four most affected countries.

**References- Online Resources** >
https://stackoverflow.com/
> https://www.geeksforgeeks.org/
>https://www.start.umd.edu/gtd
>https://towardsdatascience.com/mapping-with-matplotlib-pandas-geopandas-and-basemap-in-python-d11b57ab5dac
>https://jakevdp.github.io/PythonDataScienceHandbook