

# Global Terrorism Data Analysis

Mr. Chanakya. Adari

## 1. Abstract:

Terrorism poses major threat to the safety and security of the international community and have become one of the greatest obstacles to the sustainable development and its analysis and prediction are imperative. According to the given data 181,691 identified terrorist incidents in the GTD dataset from 1970 to 2017. Terrorist attacks typically involve high lethality and destructive power and directly cause massive casualties and property losses. The analysis and prediction of terrorist attacks provide insights on most affected nations, states, cities and can identify the dangerous terrorist organizations across the world, weapons used by them etc.

## 2. Executive Summary:

### Characteristics of the GTD

Duration of Attack	1970 to 2017
Number of Data Points	135
Number of Records	181,691
Geography	Worldwide
Variables (Data Point)	Data is collected for variables like location, tactics, success of attack, target group name, number of victims killed or wounded, number of perpetrators killed or wounded, targets and outcomes.
Size	155 MB

### 1) GTD ID and Date:

- Event Id: Incidents from the GTD follow a 12-digit Event ID system.
- Year, Month, Day, Approximate Date
- Extended Incident: whether the duration of an incident extended more than 24 hours or not.

### 2) Incident Information:

- Incident Summary: A brief narrative summary of the incident, noting the "when, where, who, what, how, and why."
- Inclusion Criteria

### 3) Incident Location:

- Country, region, state/province, city, vicinity, Location Description.
- Latitude and longitude
- Geocoding specificity

### 4) Attack Information:

- Attack Type: 8 categories + unknown.
  - Assassination, Hijacking, Kidnapping, Barricade Incident, Bombing/Explosion, Armed Assault, Unarmed Assault, Facility/Infrastructure Attack, and Unknown.
- Suicide Attack

### 5) Weapon Information:

- Weapon Type: 12 categories + unknown.
- Several sub weapons types.

### 6) Target Information:

- Target Type: 22 categories
- Several specific target/victim information, including names, nationalities, etc.

### 7) Perpetrator Information:

- Perpetrator Group Name: the name of the group that carried out the attack
- Several sub-group information, including number, claim, motive, etc.

## 8) Casualties and Consequences:

- Total Number of Fatalities and Number of Perpetrator Fatalities
- Total Number of Injured and Number of Perpetrators Injured
- Property Damage, including damage extend, values and comments
- Total Number of Hostages/Kidnapping Victims, or Kidnapping Victims, kidnapping hours, countries, total ransom amount demanded, and number released/escaped/rescued.

Out of the 135 columns, 62 columns have more than 90% of data is missing. Around 40 columns have data which have missing values of less than 20%. Out of which we have selected below columns for our data analysis.

During this prediction task we used the following features:

- Year of the attack (Numeric)
- Month of the attack (Numeric)
- Day of the attack (Numeric)
- Region (Text)
- Country (Text)
- State (Text)
- City (Text)
- Latitude (Numeric)
- Longitude (Numeric)
- Attack type (Text)
- Target Type (Text)
- Target Subtype (Text)
- Gang name (Text)

- Weapon Type (Text)
- Number of people killed (Numeric)

## 3. Data Cleaning and Preparation

Preprocessing is important into transitioning raw data into a more desirable format. Undergoing the preprocessing process can help with completeness and compellability. For instance, you'll see if certain values were recorded or not. Also, you'll see how trustable the info is. It could also help with finding how consistent the values are. We need preprocessing because most real-world data are dirty. Data can be noisy i.e. the data can contain outliers or simply errors generally. Data can also be incomplete i.e. there can be some missing values.

The available data is raw and unusable for Exploratory data analysis, so before we do anything with the data we will have to explore and clean it to prepare it for data analysis.

We have done our data cleaning as below

Step 1: finding out missing values in each selected column.

	Description	Missing Values
0	Year	0
1	Month	0
2	Day	0
3	Country	0
4	Region	0
5	State	421

6	City	434
7	latitude	4556
8	longitude	4557
9	Attack_type	0
10	Target_type	0
11	Gang_name	0
12	Weapon_type	0
13	number of people killed	10313

Step 2:

There are 421 values in the State column is missing and 4292 values are 'Unknown'. Hence, we market the 421 values as 'Unknown'. Both putting together 4793 which is approximately 3% of total number of rows. Hence, it is in the permissible limit.

Step 3:

There are 434 values in the city column is missing and 9822 values are 'Unknown'. Hence, we market the 434 values as 'Unknown'. Both putting together 10256 which is approximately 6% of total number of rows. Hence, it is in the permissible limit.

Step 4:

We have 4556 values are missing, we need exact location of latitude to plot our data. Hence, we dropped such locations.

Step 5:

We have 4557 values are missing, we need exact location of longitude to plot our data. Hence, we dropped such locations.

Step 6:

There are 10313 values are missing, hence, we considered number of people killed in each instance as 0.

## 4. EXPLORATORY DATA ANALYSIS

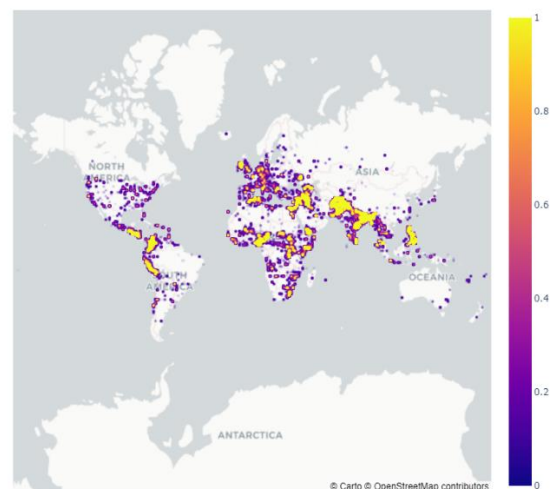
Exploratory Data Analysis, or EDA, is an important step in any Data Analysis or Data Science project. EDA is the process of investigating the dataset to discover patterns, and anomalies (outliers), and form hypotheses based on our understanding of the dataset.

EDA involves generating summary statistics for numerical data in the dataset and creating various graphical representations to understand the data better for this we have extensively used matplotlib, scipy, numpy, pandas, seaborn, plotly for computation and visualization purposes.

## 5. Frequency of Attacks all over the world

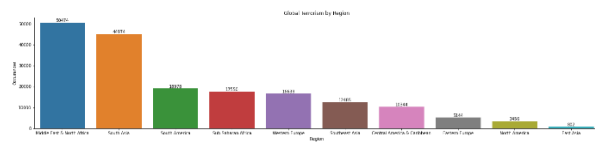
We have calculated cumulative number of attacks on a particular latitude and longitude and prepared a graph to visualize the occurrences of terror attacks across the globe.

Frequency of Attacks all over the world



**Fig1: Frequency of Attacks all over the world**

## 6. Terrorist Attacks across the regions:

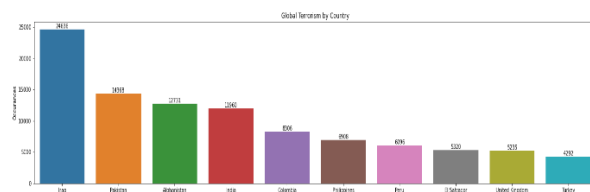


**Fig 2: Region vs Number of terror attacks**

We have grouped each region and calculated the number of terror attacks for each region.

From the above figure it is clear that than 50% of the total number of global terror attacks are happening in the Middle east & North Africa and South Asia regions.

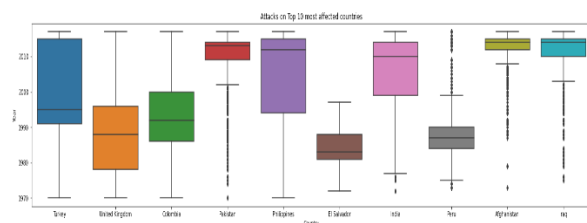
## 7. Top 10 most affected Countries:



**Fig 3: Country vs Number of terror attacks**

We have grouped each country and calculated the number of terror attacks for each country and taken top 10 most affected countries.

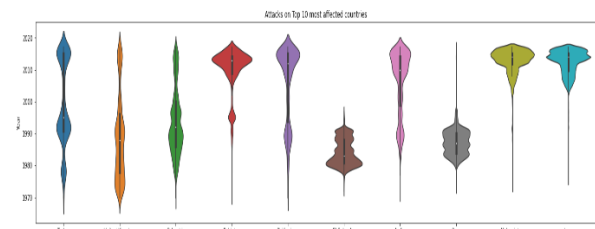
Iraq, Pakistan, Afghanistan and India are the most affected countries in the world for the period 1970-2017.



**Fig 4: Box plot (Country vs Number of terror attacks)**

The left and right sides of the box are the lower and upper quartiles. The box covers the interquartile interval, where 50% of the data is found. The vertical line that split the box in two is the median.

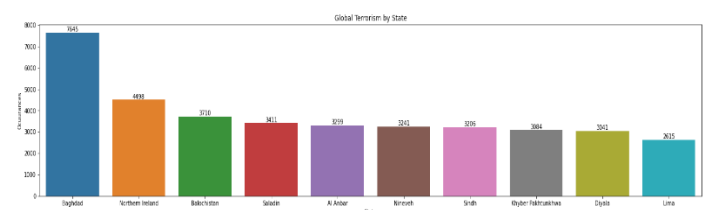
Iraq, Pakistan, Afghanistan, India and the Philippines data is skewed towards 2017 thereby indicates these countries are currently experiencing Terrorist activities whereas countries like the UK, Colombia, Turkey, El Salvador and Peru experienced attacks in the past.



**Fig 5: Violin plot (Country vs Number of terror attacks)**

Similar to the box plot we can also deduce the same information from here, we can identify the string of attacks happening the country. The more the width of violin shape indicates the significant growth in the number of attacks in a short duration. We can see identify the number of terrorist attacks are reducing in most of the countries.

## 8. Top 10 most affected States:

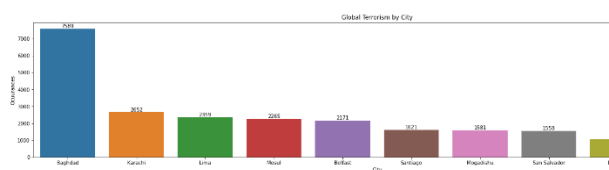


**Fig 6: State vs Number of terror attacks**

We have grouped each State and calculated the number of terror attacks for each State and taken top 10 most affected States.

Baghdad has witnessed most number of attacks followed by Northern Ireland and Balochistan

### 9. Top 10 most affected Cities:



**Fig 7: Cities vs Number of terror attacks**

We have grouped each City and calculated the number of terror attacks for each City and taken top 10 most affected Cities.

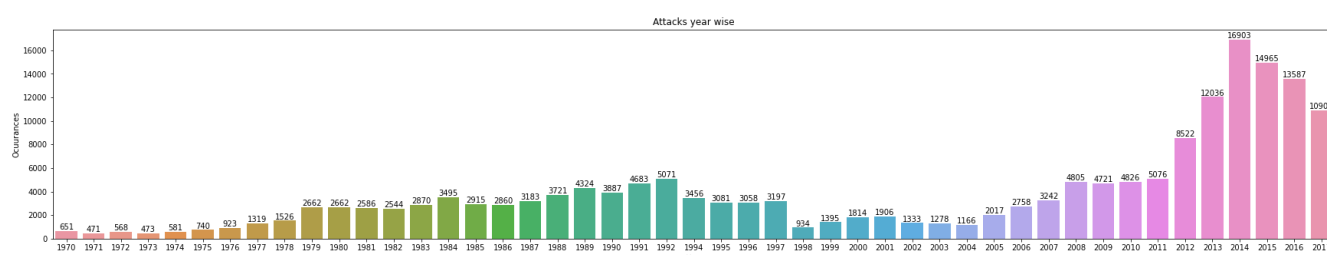
Baghdad has witnessed the most number of attacks followed by Karachi and Lima

We have found out 75<sup>th</sup> percentile of attacks and considered attacks occurring more than 75percentile needs more attention to reduce the number of attacks happening in those months.

As we can see 75th percentile value is 8184.50, and the attacks in the months of April, May, June, July, August, October, and November are almost equal to or above the 75th percentile. Extra security measures are to be taken during these months with emphasis on May, July August.

In a similar manner we have found 4,8,11,13,14,15,21,27 more attacks are happening.

## 10. Attacks year wise from 1970 to 2017

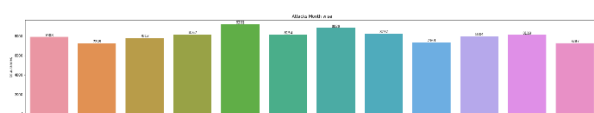


**Fig 7: Years vs Number of terror attacks**

We have grouped each year and calculated attacks happening in each year.

A significant increase in the number of attacks can be seen from 2011 to 2014 and there is a reduction in the number of attacks thereafter.

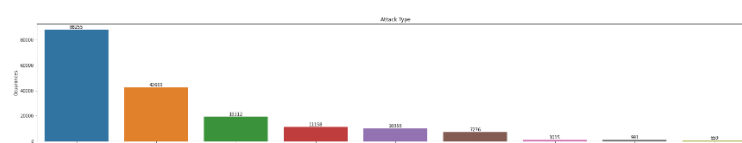
### 11. Attacks month wise from 2008 to 2017



**Fig 8: Months vs Number of terror attacks**

We have grouped each attacks happening month wise from 2008 to 2017 and calculated its sum.

## 12. Type of attacks employed by Terrorists (1970-2017)

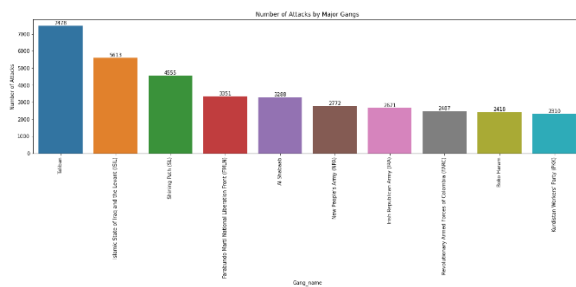


**Fig 9: attacks type vs Number of attacks occurrences**

We have grouped attack type employed by terrorists.

Bombing/Explosion is significant attack type followed by Armed Assault and Assassination

### 13. Number of attacks carried out by top 10 terrorist organizations (1970-2017)

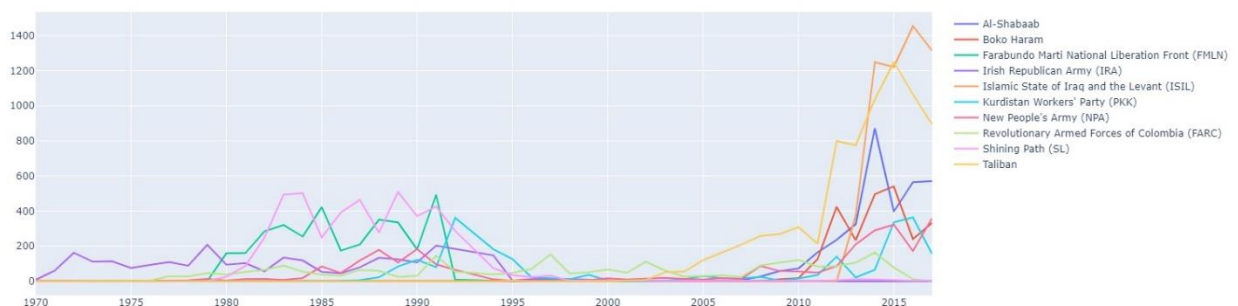


**Fig 10: Terrorist organizations vs Number of attacks**

We have grouped number of attacks done by each terror group and calculated individual sum.

Taliban has carried out the greatest number of terror attacks in the world followed by ISIL and SL.

### 14. Number of attacks carried out by top 10 terrorist organizations (1970-2017)



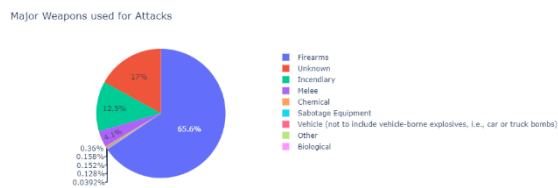
**Fig 11: Terrorist organizations vs Number of attacks vs year**

It is imperative to find current global terrorist organizations, to know about them more.

We have found top ten terrorist organization with respect to the number of attacks.

Taliban, ISIL and Al-shabaab are the cause of more than 50% total number of terror attacks all around the world from 2010 to 2017.

## 15. Weapons used by terrorist organizations (1970-2017)

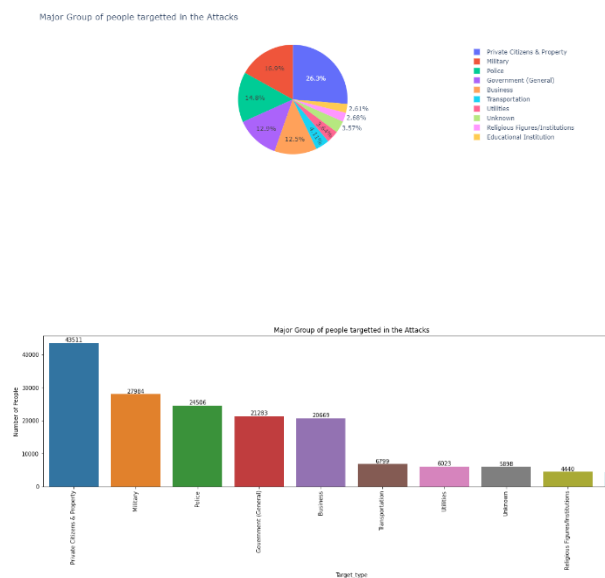


**Fig 12: Percentage of weapons used by terrorist organizations.**

It is very important to find out the weapons used by terrorist organizations and way of controlling the weapons getting into the wrong hands can reduce the number of attacks.

We found out around 65% of the total world attacks are caused due to fire arms.

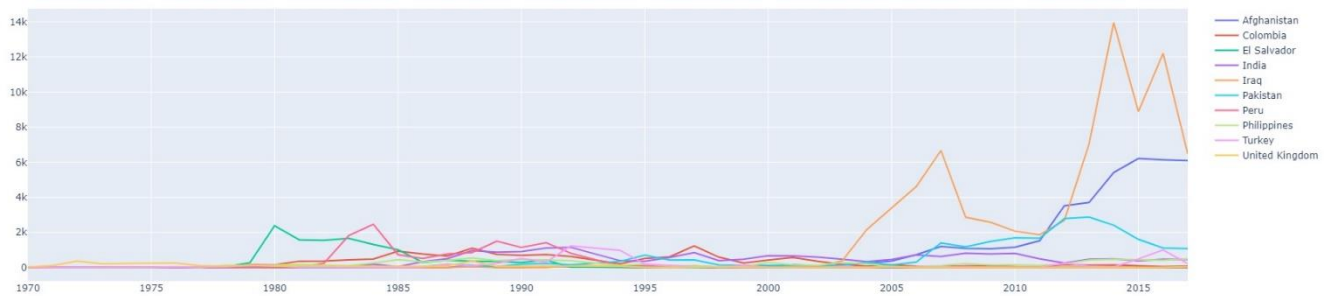
## 16. Major Group of people targeted by terrorist organizations (1970-2017)



**Fig 12: Percentage of Major Group of people targeted by terrorist organizations.**

Private citizens are the primary target for the terrorist followed by Military and Police personal.

## 17. Number of People killed in the most affected countries (1970-2017)



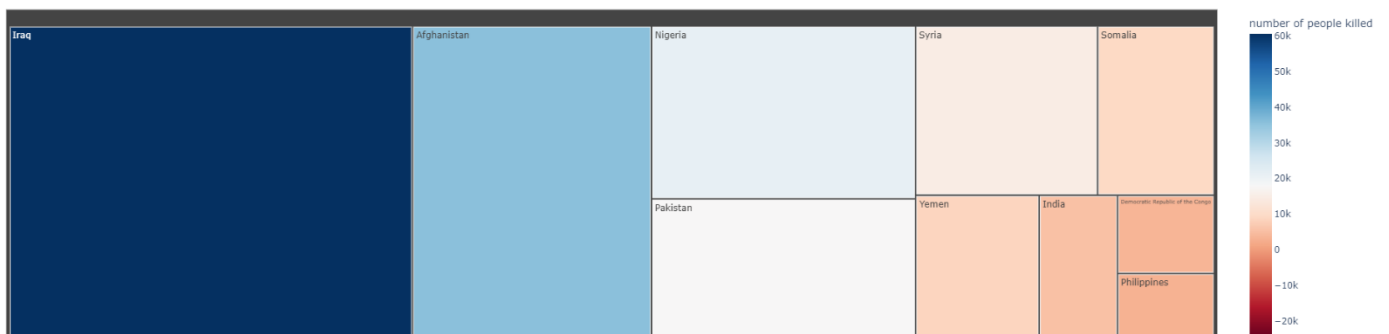
**Fig 12: Terrorist organizations vs Number of people killed vs year**

We have calculated top most affected countries in the world and calculated the number of people killed in each country.

There is significant reduction in the number of attacks when compared to 2014 but conditions in Iraq and Afghanistan remains alarming.

Combined deaths occurred in Iraq and Afghanistan due to terrorism is more than half of the total deaths resisted during the period 2008 to 2017.

## 18. Number of People killed in the most affected countries (2008-2017)





### Fig 13: Number of People killed (2008-2017)

It is very important to know terrorist find out the most affected countries in the recent times to focus more on those nations/counties.

Deaths in Iraq, Afghanistan, Nigeria, Pakistan and Syria combined represents more than 80% total deaths during 2008 to 2017.

Combined deaths occurred in Iraq and Afghanistan due to terrorism is more than half of the total deaths resisted during the period 2008 to 2017.

#### 19. Major findings in our Global Terrorism Analysis

1. The Middle East & North Africa region has witnessed the most number of attacks followed by South Asia and South America.
2. Iraq has witnessed the most number of attacks followed by Pakistan and Afghanistan.
3. Baghdad has witnessed the most number of attacks followed by Northern Ireland and Balochistan.
4. Baghdad has witnessed the most number of attacks followed by Karachi and Lima.
5. Bombing/Explosion is significant attack type followed by Armed Assault and Assassination.
6. A significant increase in the number of attacks can be seen from 2011 to 2014 and there is a reduction in the number of attacks thereafter.
7. Maximum number of attacks are happening in the months of April, May, June, July, August, October, and November.
8. Taliban has conducted the most number of terror attacks followed by ISIL and SL.
9. Around 65% of attacks in the world are caused by firearms.

10. Private citizens & property is being targeted by terrorists followed by Military people and Police personal.
11. Countries from the Middle East & North Africa and South Asia are currently witnessing major attacks in the world.
12. There is a significant increase in the number of attacks 2005 in the Middle East & North Africa and South Asian countries
13. Though the number of attacks all over the world are witnessing downfall the situation in Iraq, Afghanistan, Pakistan looks alarming.
14. More than 50% of deaths are occurring from Iraq and Afghanistan from 2008 to 2017.

#### 20. References

1. GeeksforGeeks
2. Stackoverflow
3. Towards data science
4. Library documentations