



# **EXPLORATORY DATA ANALYSIS (EDA)**

PROJECT ON

# **GLOBAL TERRORISM ANALYSIS**

PROJECT PERFORMED BY

SUNIL KUMAR



## Table of Contents

• Summary.....	1)
• Problem statement.....	2)
• Introduction.....	3)
• Technical Work.....	4)
i. Library used.	
ii. Function and Method used.	
• Step involved in exploratory data analysis.....	5)
• Observations.....	6)

## Summary:

Terrorism is commonly understood to refer to acts of violence that target civilians in the pursuit of political or ideological aims. The problem of terrorism is not only a problem to country but also It is problem to worldwide due to its human act violations and destructive nature. The dataset used for exploratory data analysis is drawn from Global Terrorism Database which essentially contains the data about terrorist attacks from 1907-2017. The attacks were almost 180000 and also massacre. It has distressed the political and economical stabilities on globally.

This exploratory data analysis part is to understand attacks and terrorist group and their trends what they follow for the attacks and as to how and when they are active and what sort of weapon and timing they prefer likewise for entire dataset.

## Problem Statement

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.

Look the data and try to understand below points:

- 1) Find most active terrorist group.
- 2) Weapon used commonly.
- 3) Most attacked country, region.
- 4) No. Of people wounded.

- 5) Check the suitable correlation with terrorist attacks.

Understanding above points requires a strong data analysis so that the factual representation is strong and based on which action can be recommended for a country to make it more safe and secure.

## Introduction

Terrorism clearly has a very real and direct impact on human rights, with devastating consequences for the enjoyment of the right to life, liberty and physical integrity of victims. In addition to these individual costs, terrorism can destabilize Governments, undermine civil society, jeopardize peace and security, and threaten social and economic development. All of these also have a real impact on the enjoyment of human rights. Terrorism is not only breaching the human rights but also it has been impactful for the economical growth as well as developmental activities globally. Hence gaining the data and understanding is very important as it reveal the pattern and facts to in order to secure the lives and assets by taking predefined measured towards the terrorism organization also to impose the amount of security for a country which has been targeted by terrorist too often.

## Technical Work

- *Library used*
- *Functions and methods used.*

### Library used

- **Pandas** Used in data analysis and manipulation and importing files.
- **Matplotlib** Used in for data visualization, graphs & plotting.
- **Seaborn** Used in making data visualisation in more colourful and meaningful way.
- **Folium** Used in drawing world map to locate any desired location..
- **Ipywidgets** Used in dynamic representation of our graphs and plots.
- **os** Used in provides functions for creating and removing a directory folder.

## Common functions and methods used

**pd.read\_csv():**used in to import data in CSV format. This function has a number of arguments, but the only essential argument is file, which specifies the location and filename.

**summary():**function return a summarized representation of the Index.

**pwd():** Finds present working directory.

**os.chdir():** Changes your directory to your data directory.

**head():** Used in checking first rows of the dataset.

**tail():** Used in fetching last rows of the dataset.

**shape():** Used in getting the total no. Of columns and rows a dataset contains.

**Type():** Used in knowing the data structure of the dataset.

**df.dtypes:** Used in understanding the data types of the dataset.

**IsNull():** used in checking if any null values are available.

**Sum():** used in giving sum of the values.

**value\_counts() :** Used in providing counts of particular values present in data.

**idmax():** function returns index of first occurrence of maximum over requested axis.

**rename():** used to rename a file or directory or data columns.

**corr():** aggregate function returns a coefficient of correlation between two numbers.

**unique():** function is used to find the unique elements of an array.

## Steps in exploratory data analysis

- Setting directory/path.
- Loading dataset.
- Data cleaning.
- Data wrangling.
- Visualizations.

- Data pre-processing- feature selection.

## Setting directory/path

Before we proceed with python to understand the data, it is important to let python understand first that where our data available is so that, we can perform the operation on data using python. Setting directory is similar to that and helps in understanding the presence of the data.

**os.chdir(your path)**

## Loading dataset

Once directory is set now we need to load the dataset which is commonly done using:  
**Pd.read\_csv('file\_name')**

## Data cleaning

Now once we have our dataset loaded it is important that we look for data in good format and it is not containing anything which is not required for our analysis or that contains values which make no sense.

Data which has null values in this step of cleaning the data and treating the null values or empty columns is called data cleaning.

Essentially data cleaning have below operations to perform:

1. *Dropping Columns in a DataFrame.*
2. *Changing the Index of a DataFrame.*
3. *Tidying up Fields in the Data.*
4. *Cleaning the Entire Dataset in term of null values.*
5. *Renaming columns and skipping rows.*

## Data wrangling

Process of cleaning and unifying messy and complex data sets for easy access and analysis. Converting the format of data to suitable format.

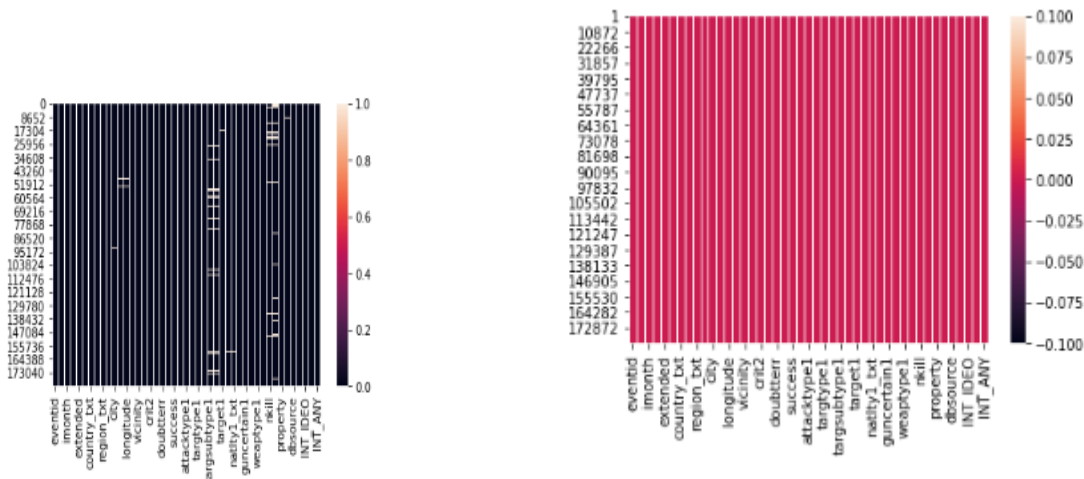
We can rename our variables using the .rename() function put your columns inside a list of a dictionary in case columns are more and which requires renaming to a more sensible form.

This step involves finding manually the data and making suitable alteration for that it make more sense.

## Data Visualization

In this step of visualization it involves checking the data in pictorial form. We can see outliers explore null values and also we can understand the variables which are correlated to each other and how data is distributed.

Apart from this, visualization helps in easy understanding of the data and representation of most of the data information in right and quick fashion.



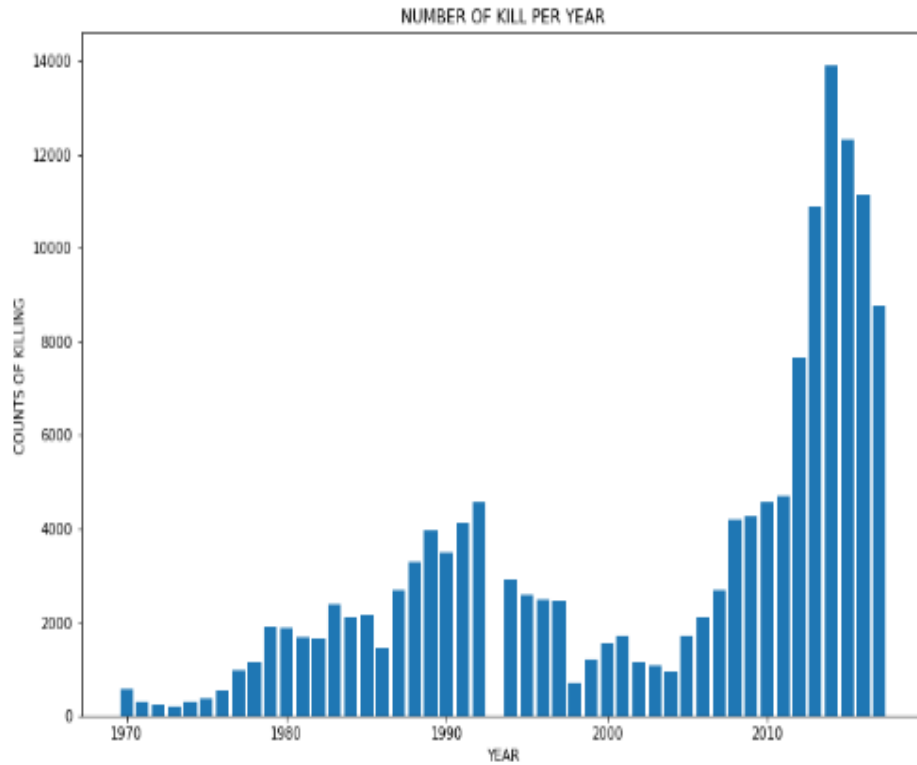
Looks like there are still many null values available which still need treatment.

From above fig we can say that yes null has been completely removed!

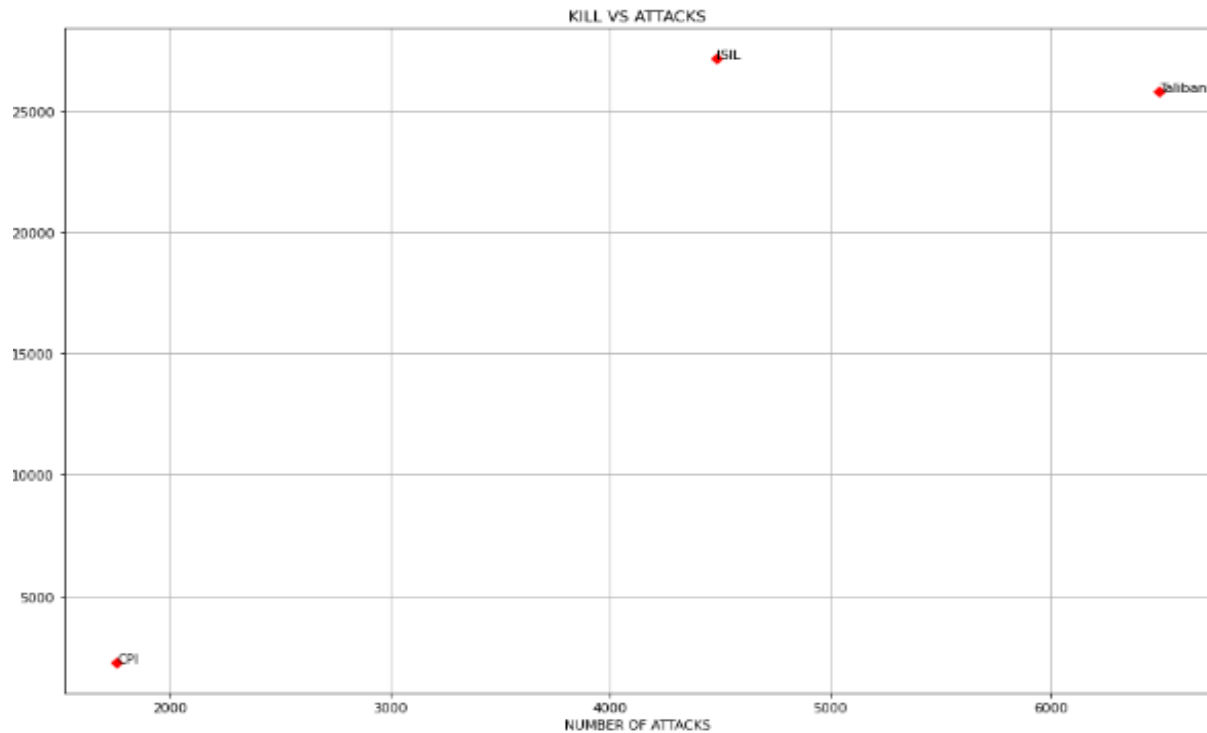


Map shows 4 location as per coordinates with most attacked country.

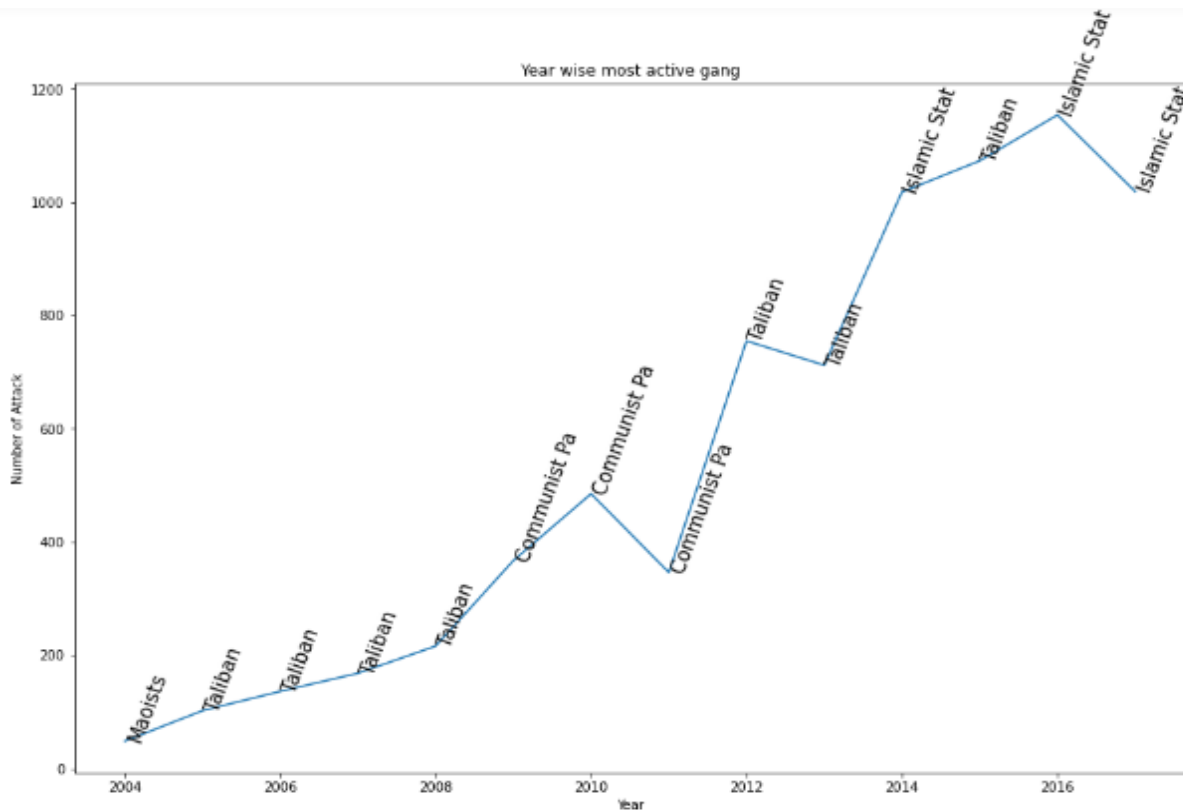
- Basra
- Mosul
- Kirkuk
- Baghdad



- According to above observation, it can be concluded that number of kills increased rapidly after 2010 and where 2014 being the most deadly year having maximum number of killing.\*



We can conclude that most attacks and most killings are performed by ISIL followed by Taliban.



From this plot we can conclude few points

- Most active gangs are Taliban, CPI-Maoist, ISIL

## Data pre-processing- feature selection.

This part in data pre-processing involves in selecting the variables or features which are useful and making sense for the analysis.

The features which are useless and makes nearly no sense in exploring would have to drop accordingly as early as possible in the process.

There will be instances where it is also required to rename the columns suitably.

For example a column is something written- what gender do you belongs to? This would require renaming to Gender.

Few more examples are below from my dataset.

```
{'eventid':'event_id'}
```

```
{'iyear':'year'}
```

```
{'imonth':'month'}
```

Column selection and dropping is important and we can confirm it with the help of null value percentages along with correlation.



Correlation approach involves in checking strong +ve and -ve correlation also weak correlation or columns which may not even be correlated.

Variables which are not correlated to any of the variables would have to drop.

Second approach if null values are way too high and keeping that column would not confer the much difference in our analysis so it can be dropped.

## Observations

- ***Most attacks***

- \*Country with most attacks: Iraq
- \* City with most attacks: Baghdad
- \*Region with the most attacks: Middle East & North Africa
- \*Year with the most attacks: 2014
- \* Month with the most attacks: 5
- \*Group with the most attacks: Taliban
- \*Most Attack Types: Bombing/Explosion

- ***Map shows 4 locations as per coordinates with most attacked country.***

- \* Basra
- \* Mosul
- \* Kirkuk
- \* Baghdad

- ***After plotting attacks for years.***

- \* Number of kills increased rapidly after 2010 and where 2014 being the most deadly year having maximum number of killings.

- ***These three groups have been more contagious.***

- \* Unknown 95399.0(number of kills).
- \*Islamic State of Iraq and the Levant (ISIL) 27174.0(number of kills).
- \* Taliban 25785.0(number of kills).

- ***Most killings in year.***

- \* Analysis tells that in the year 2001 most were killed with 1340 lives.
- \* Wounded were 8190 lives.
- \* Al- Qaida terrorist group involved and it was in Hijacking of Airlines.

- ***Few Observations about Al-Qaeda.***

- \* Al-Qaeda groups have done most killings.
- \* Weapon used for killing is only two: 1) Explosives and 2) vehicle-borne explosives.
- \* Year 2001 more than 1 attack by Al- Qaeda.
- \* Seems like USA and Al-Qaida were in war for some issue.

- ***Top active Terrorist groups.***

- \* Most active gangs are Taliban,CPI-Maoist,ISIL.
- \* These gangs are using mostly Bombing/Explosion and Armed Assault.
- \* We can conclude that most attacks and most killings are performed by ISIL followed by Taliban.