

# **Capstone Project**

## **Global Terrorsim Analysis**

**By - Shourya Chandra Sai**

# Project Agenda

- **Data Cleaning**
  - Features Selection.
  - Dealing with null values.
- **Data Analysis**
  - Preliminary Analysis.
  - India Terrorism Vs Iraq Terrorism.
- **Conclusions**

# Data Cleaning

## Part I – Features Selection.

There are 135 columns in the dataset. It doesn't make sense to analyse all the columns so I have narrowed them down to 18 columns.

### *Approach used:*

- A lot of columns in the dataset had a ton of null values. So I dropped those columns that had a ton of null values
- I also dropped those columns which spoke the same information.

### *Data Summary:*

- year = The year in which attack has taken place.
- month = The month(represented in numbers) in which the attack has taken place.
- day = The day in which the attack has taken place.

- **extended** = If the attack has extended for more than one day then value 1 is assigned and if it isn't then value 0 is assigned.
- **country** = The country in which the incident took place.
- **city** = The city in which the incident took place.
- **latitude** = This field represents the latitude.
- **longitude** = This field represents the longitude.
- **success** = This field informs us whether the attack was a success or a failure(0 = failure, 1 = Success).
- **suicide** = This field informs us whether the attack was carried out as a suicide mission or not (0 = not a suicide mission, 1 = suicide mission)
- **attack\_type** = The type of attack.
- **target** = The type of target/victim.
- **organisation** = Terrorist Organisation that has claimed the attack.
- **motive** = The motive behind the attack.
- **weapon\_type** = The type of weapon used in the attack.
- **killed** = The number of people killed in that attack.
- **wounded** = The number of people wounded in that attack.

## Part II - Dealing with null values

Four columns in subsetting data set have null values namely *motive*, *killed*, *wounded*, *city*. *motive* and *city* were categorical variables and *killed* and *wounded* were continuous variables.

Approach used for categorical variables:

Using the `value_counts()` method I have got an understanding of the value that is being used if the data is not present. String 'unknown' was being used. So I have replaced all the null values with string 'unknown'.

Approach used for continuous variables:

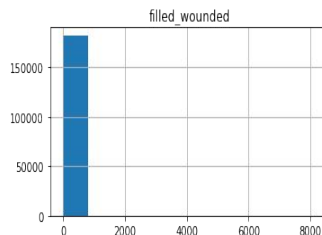
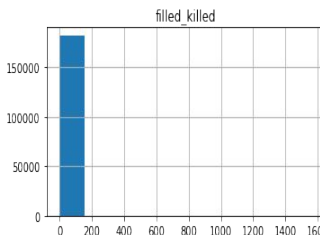
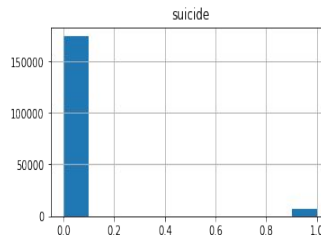
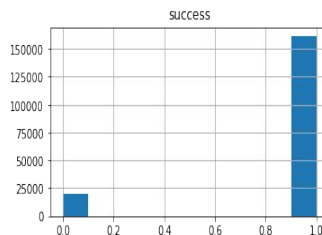
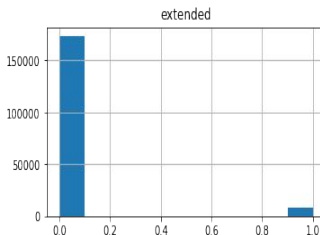
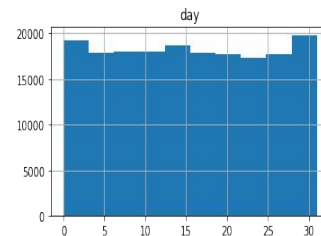
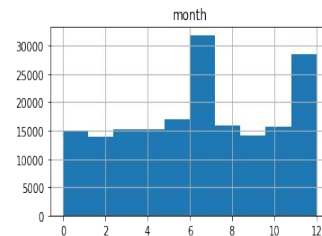
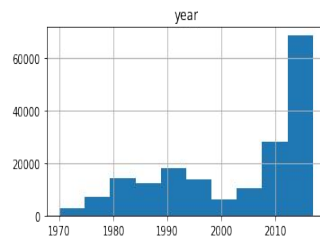
For both, the *killed* and the *wounded* columns, I have replaced the null values with their respective means grouped them by their countries.

# Data Analysis

## Part I – Preliminary Analysis

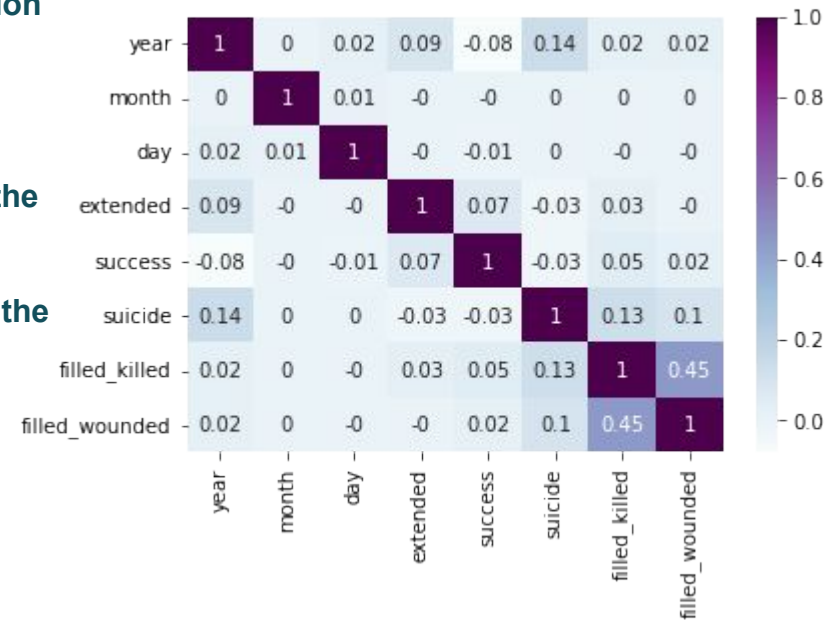
*Distribution of all the Numeric Features in the dataframe:*

- Columns extended, success, suicide have a bernoulli distribution.
- Column year is skewed towards the left. this indicates that a lot of data is present from the year 2010.
- Column month is evenly distributed.
- Column month has a lot of data for month July and December.
- Column filled\_killed has a lot of data between 0 and 200, meaning for most of the attacks, deaths were between 0 and 200.
- Column filled\_wounded has a lot of data between 0 and 1000, meaning for most of the attacks, wounded were between 0 and 1000.



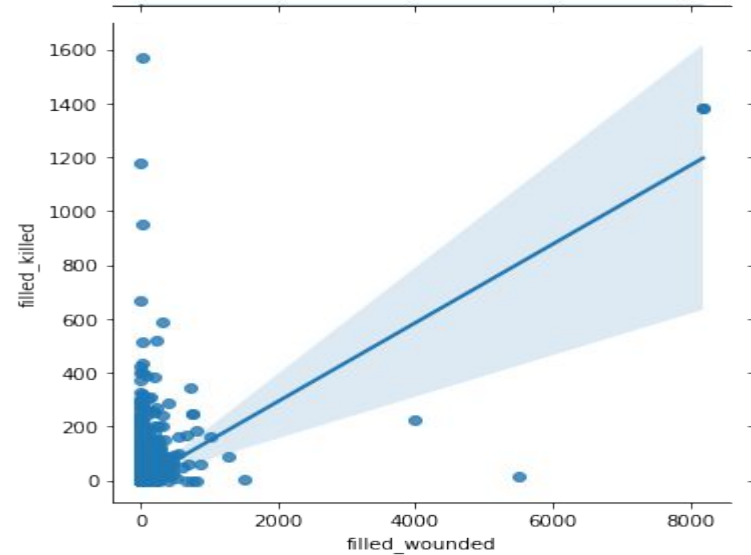
## Understanding how close the variables are.

- To understand the intimacy of the variables a correlation table is plotted.
- A correlation coefficient is between -1 and 1.
- If the coefficient between the variables is closer to 1, the more positively related they are.
- If the coefficient between the variables is closer to -1, the more negatively related they are.
- If the coefficient is equal to 0 then there is no relation.
- Only filled\_killed and filled\_wounded are a tad related with a coefficient of 0.45.



## *Understanding the relationship between filled\_killed & filled\_wounded.*

- From the scatter plot, we can conclude that filled\_wounded and filled\_killed are positively correlated.
- With the increase in filled\_wounded, there will be an increase in filled\_killed.

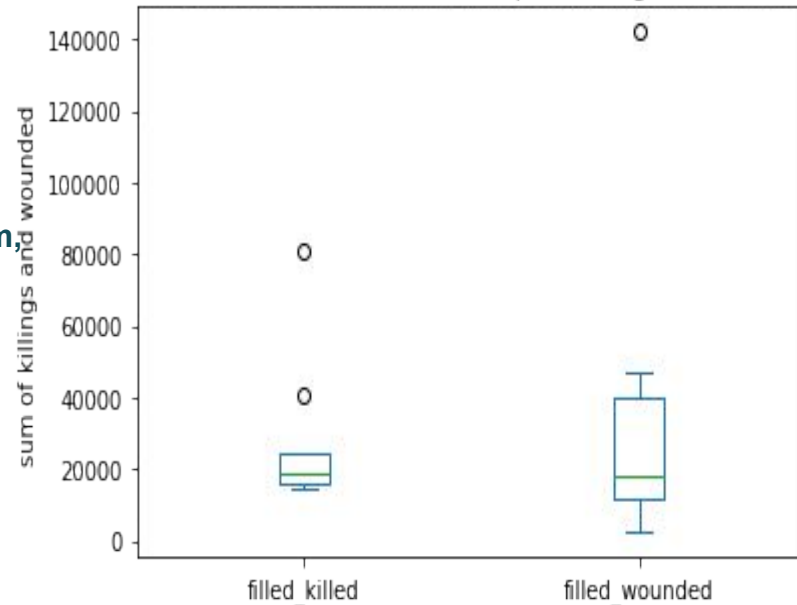




## *Analysing the Five number Summary for filled\_killed & filled\_wounded.*

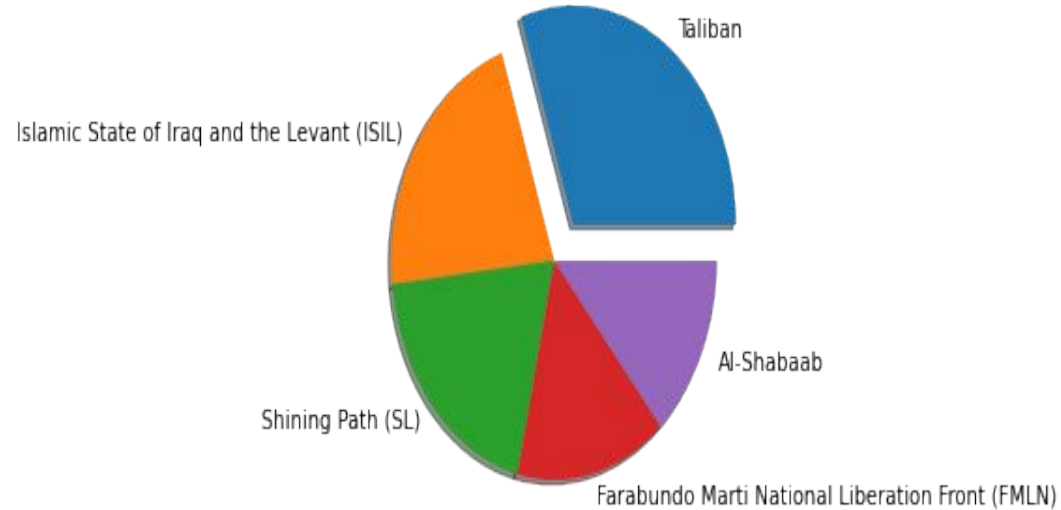
- The five number summary includes stats Like minimum, first quartile, median, third quartile, and maximum.
- For both, filled\_killed & filled\_wounded, the median looks the same.
- But all the other stats i.e., minimum, maximum, First quartile, third quartile are different.
- Filled\_killed has 2 outliers whereas Filled\_wounded has only one outlier.

Distribution of killed and wounded in the top 10 killings and wounded countries



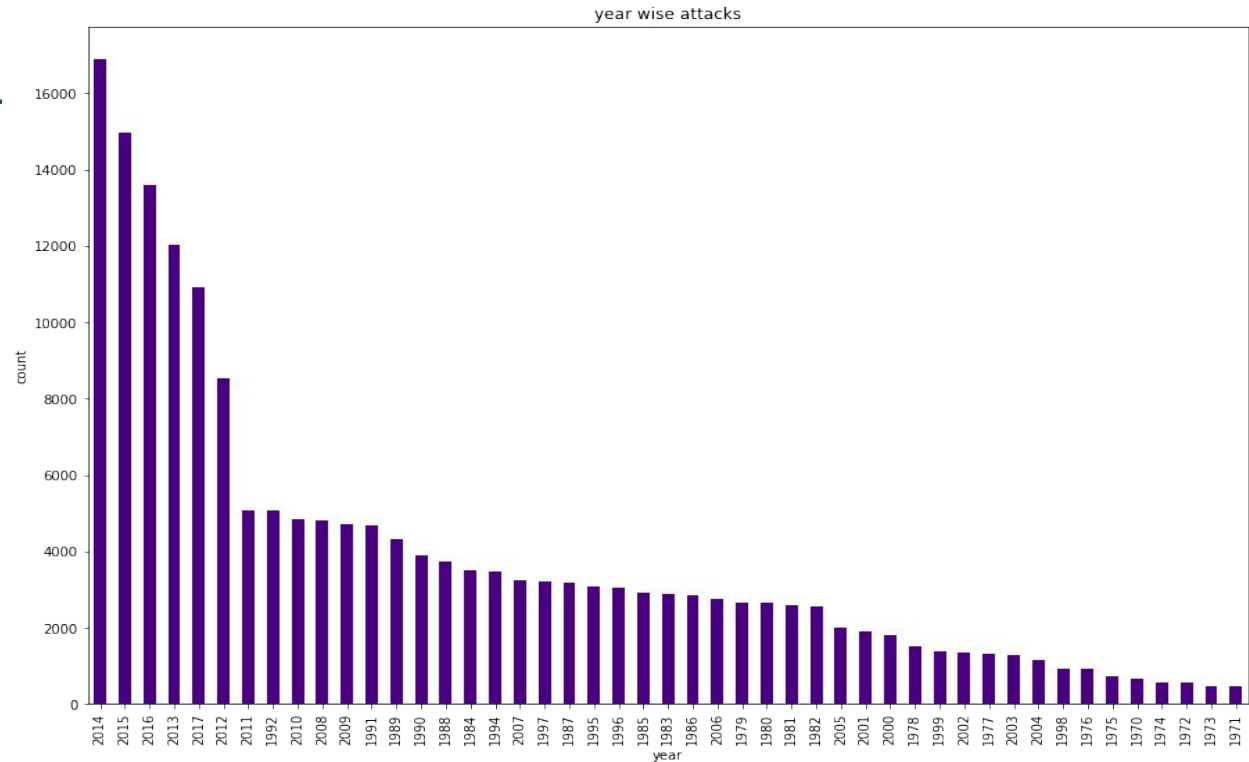
## *The Most Successful Terrorist Organisation.*

- From the Pie Chart, we can conclude that Taliban terrorist organisation is the most successful.
- The Islamic state of Iraq and the levant (ISIL) are not far behind though.



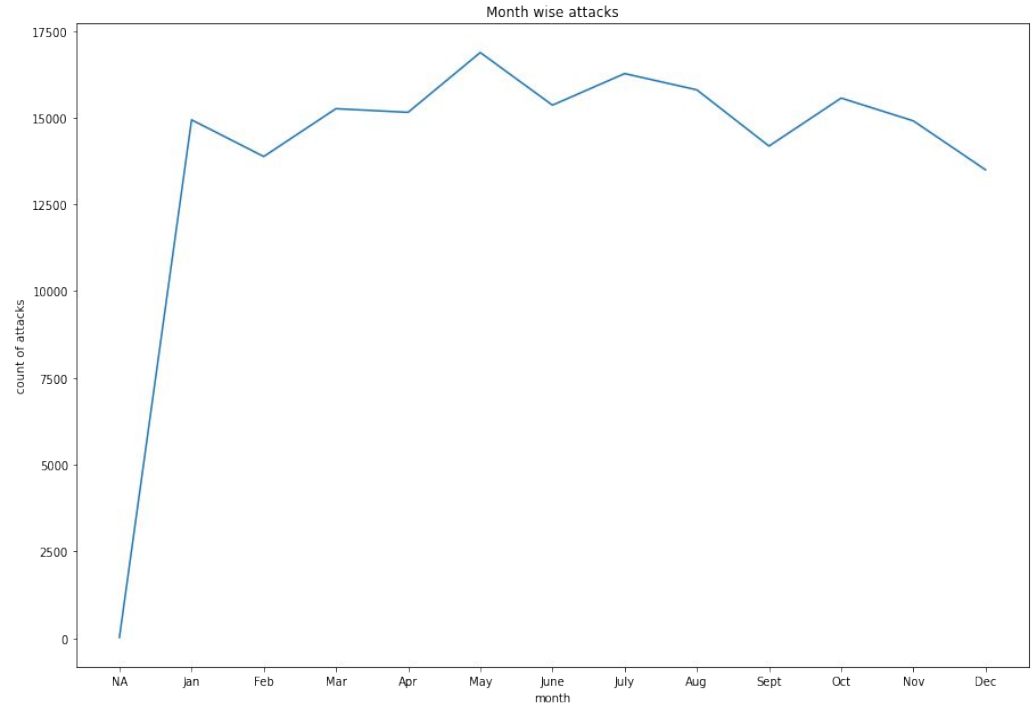
## *Most Deadliest year.*

- We can see that 2014, was the most attacked year.
- and 1971 is the least attacked year.



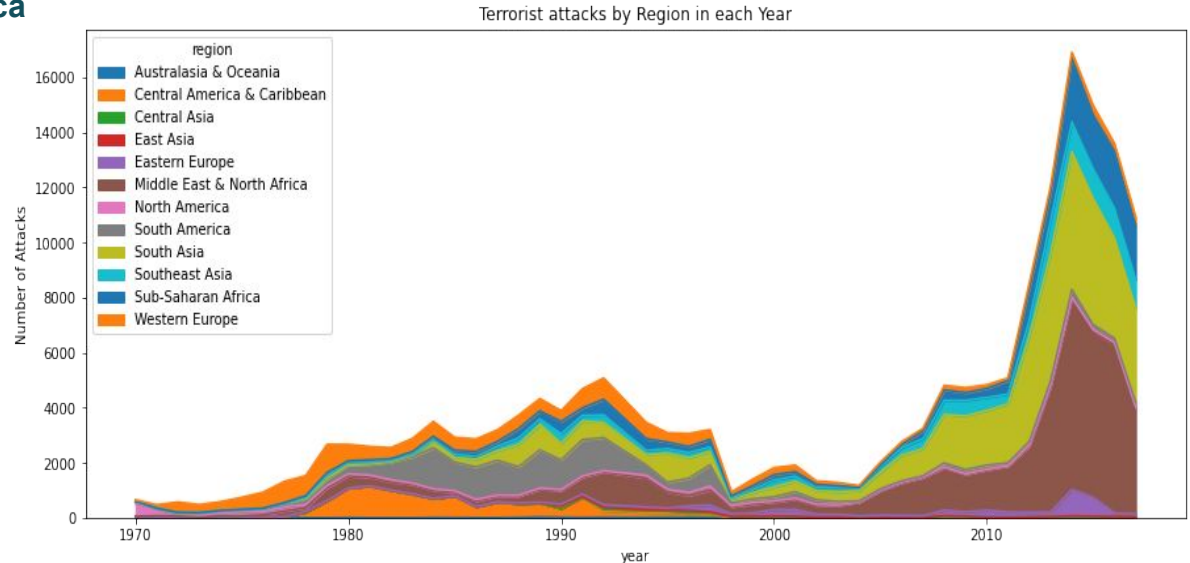
## *Understanding the relationship between attacks and months of year.*

- There is no clear relationship between the number of attacks and the months in which the attacks have taken place.
- There were constant attacks throughout all the months of a year.
- “NA” in the month axis represents the unknown month for which the attack has taken place.



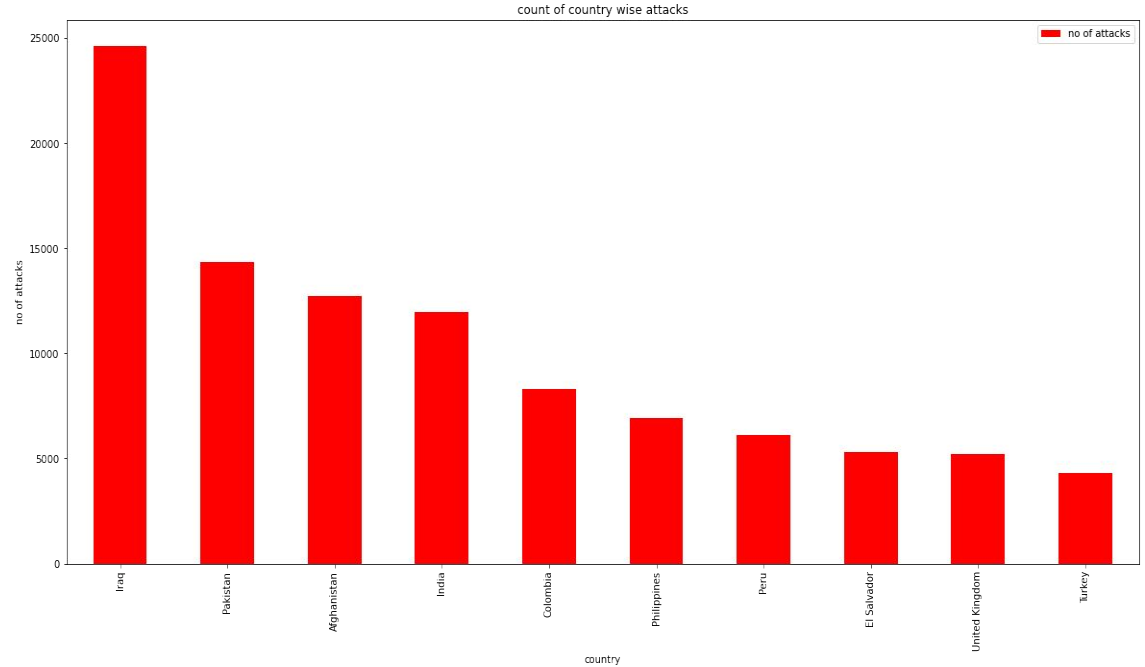
## *Terrorist attacks by region.*

- The Middle East and North Africa regions have had the most number of attacks in recent times.
- It wasn't the case between the years 1970 to 1990. The attacks have picked up only after 1990s.
- Between 1970 and 1990, South America and Western Europe were hotspots for attacks.



## *Most Terrorized Country.*

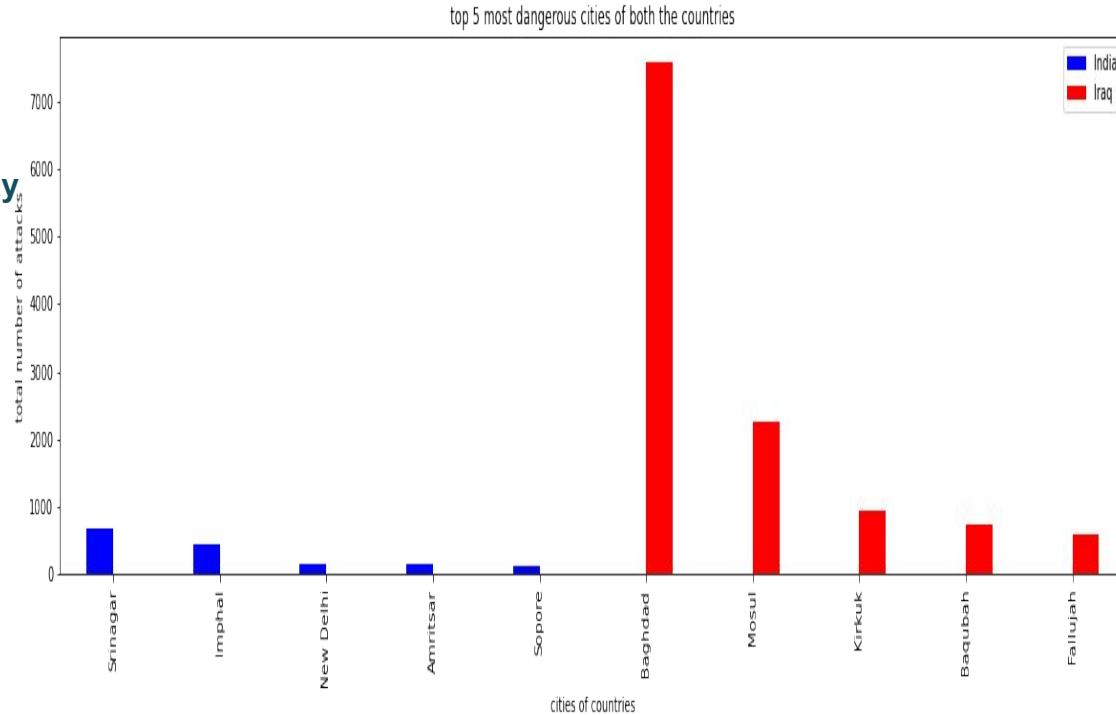
- From the bar graph, we can see That Iraq is by far the most attacked country with close to 25,000 attacks.



## Part II - India Terrorism vs Iraq Terrorism.

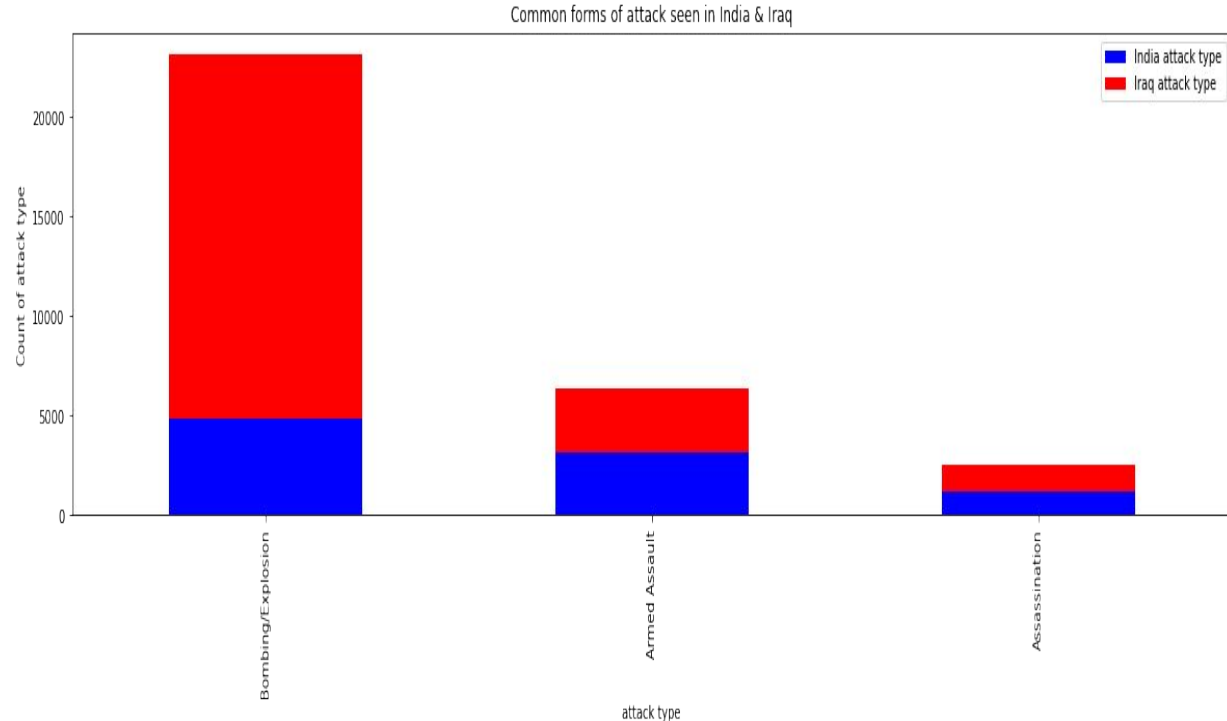
*Analyzing the top cities with the most  
Number of attacks:*

- The amount of threat faced by Iraq is way more than the threat faced by India.
- All the top 5 cities in Iraq face more threat than all the top 5 cities in India
- From this analysis we can safely conclude that India counters terror better than Iraq.



## Analyzing the common forms of attack.

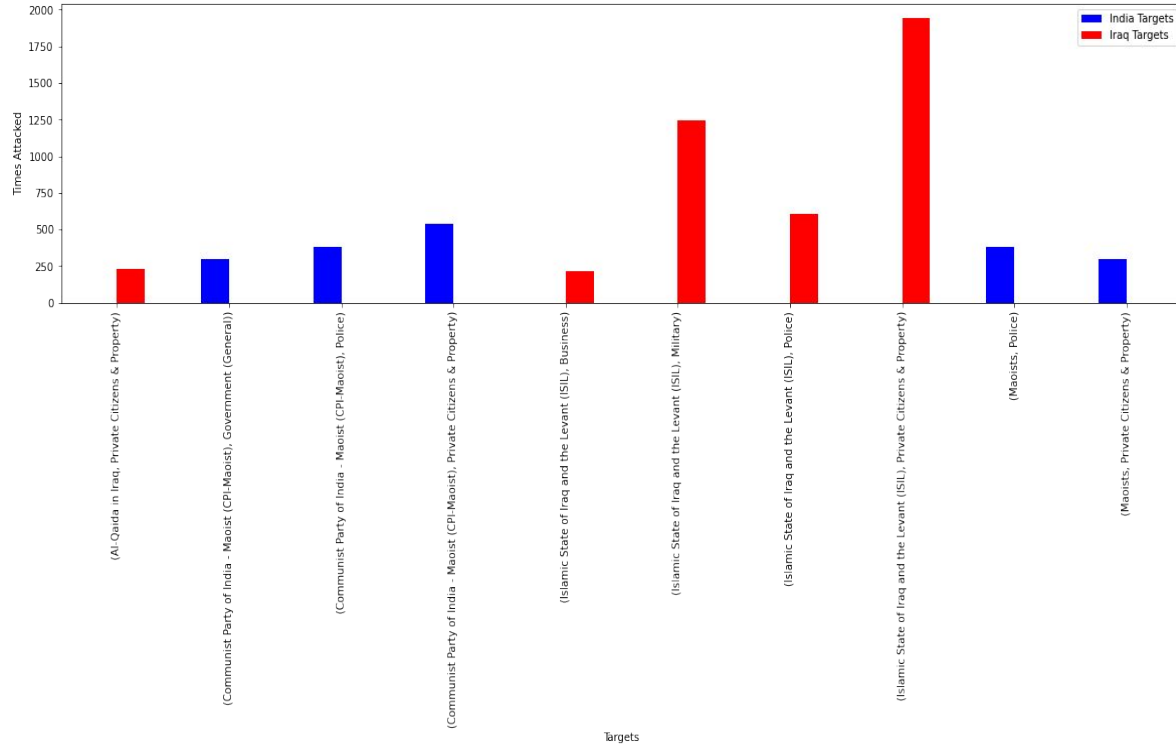
- Bombing/explosion is the most common form of attack in both the nations.
- Then comes, armed assault and assassination.
- But the magnitude of attacks in Iraq is way more than in India.





# Analysing the top terror organisations and what they target the most.

- The top terror organisation in India is CPI-Maoist, Maoists and they target Citizens, Police
- In Iraq, the top terror organisations are-  
Al-Qaida - targets citizens  
ISIL - targets Citizens, Military the most.



## Conclusions

- Killed and Wounded are the only 2 features that have some form of influence among each other and the relationship is a positive one.
- The medians of killed and wounded are approximately the same, but the First Quartile, Third Quartile, Minimum and Maximum differ for both.
- Taliban is the most successful terrorist organisation.
- The most,least number of attacks took place in the year 2014,1971 respectively.
- Terrorist attacks take place throughout all the months of a year.
- The most attacked region is middle-east and north africa.
- A large portion of terror attacks took place in Iraq.
- Conclusions from analysing Indian terrorism and Iraq Terrorism-
  - All the cities in India counter terrorism better than all the cities in Iraq.
  - The forms of attacks used in both the countries is similar.
  - In India,The top terror organisation CPI-Maoist,Maoists they target Citizens,Police whereas in Iraq Al-Qaida,ISIL terror organizations target the Citizens,Military the most.

Thank You.