Capstone Project

Global Terrorism Exploratory Data Analysis

By

Pritam Pawar

Index

Summary

- 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.
- As the first step for handling such huge dataset we started with data cleaning and data preprocessing. Performed data wrangling on dataset to get insights and understanding from the features for visualizing the data comprehensively.
- We divided this analysis on 4 major aspects with respect to features in datasets and they are, Regions / Countries, Weapon Type, Terrorist Groups and Targeted Organizations.

Summary (Continued)

- In Region and Countries visualization on the basis of coordinates given in the form of latitude and longitude in the datasets, we plotted the points on world map to see which region and countries is highly prone to terrorist attacks.
- The goal of this project was to understand and interpret the nature of terrorism efficiently and comprehensively with the use of data visualizations.
- Users can understand various patterns, trends and correlation in terrorism through visual interpretation and its provided explanation.
- This work can be used by curious civilians, security related policy-makers, international organizations hosting worldwide events, foreign investors and academic researchers for the purpose of understanding terrorism and its nature.

Features in EDA

- Eventid : Unique Id assigned to a terrorist attack.
- Year : Year of the attack
- Month : Month of the attack.
- Day : Day of the attack.
- Country : Country in which attack took place.
- Region : Region in which attack took place.
- Latitude : Latitude co-ordinate w.r.t to world map.
- Longitude : Longitude co-ordinate w.r.t to world map.
- Attack : Type of attack.
- Target : Targeted facility of the attack.
- Killed : Number of people killed in this attack.
- Wounded : Number of people wounded in this attack.
- Summary : Attack description in short.
- Groupname : Terrorist Group name.

Features in EDA

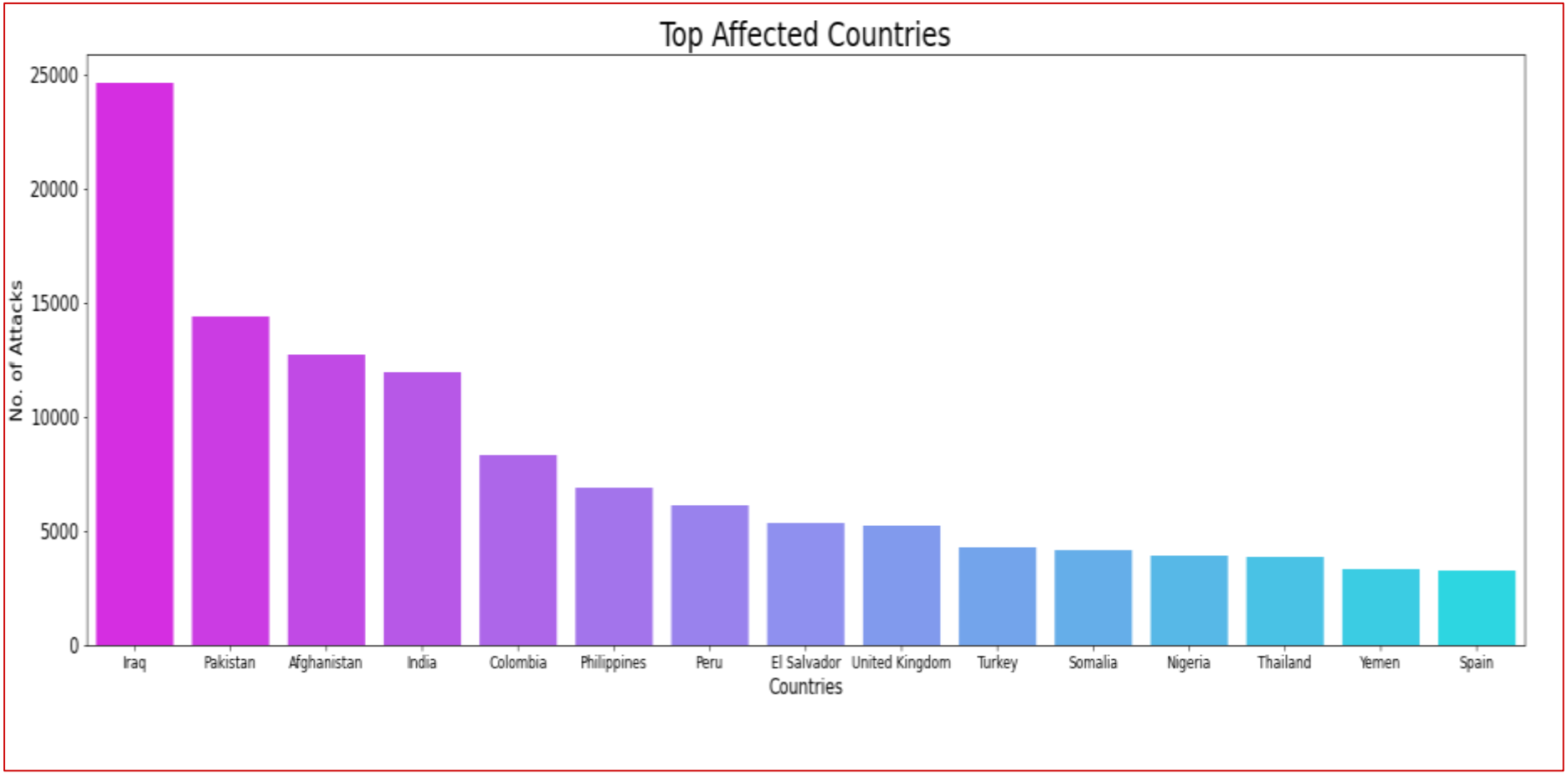
- Target_type : Name of the Specific entity suffered by the attack.
- Weapon_type : Weapon type used by the terrorists.
- Motive : Reason behind the attack.
- Damages : Damages incurred in dollars($).
- Damage_txt : Scale of damage done(Minor, Major, Catastrophic)
- Suicide : Suicide number of terrorists.
- City : City in which attack took place.

EDA Data Wrangling & Handling Missing Data

- Imported the Global Terrorism csv file.
- Identified and replaced null values
- Identified the Data Format of all the features under the dataset.

```
[4] #Rename and extract useful columns for better understanding before getting in dataframe.
    gt_df.rename(columns ={'iyear':'Year', 'imonth':'Month','iday':'Day','country_txt':'Country','region_txt':'Region','attacktype1_txt':'Attack','target1':'Target','nkill':'Killed','nwound':'Wounded','summary
    #Keep only Relevant columns in the Dataframe.
    gt_df = gt_df[['eventid','Year','Month','Day','Country','Region','Attack','Target','Killed','Wounded','Summary','Group_name','Target_type','Weapon_type','Motive','latitude','longitude','Suicide','City']]
    #This column shows us total no.of casualties in a particular attack
    gt_df['Casualties'] = gt_df['Killed'] + gt_df['Wounded']
    #Removed Unknown Values in City Column
    gt_df.drop(gt_df.index[gt_df['City'] == 'Unknown'], inplace= True)
```

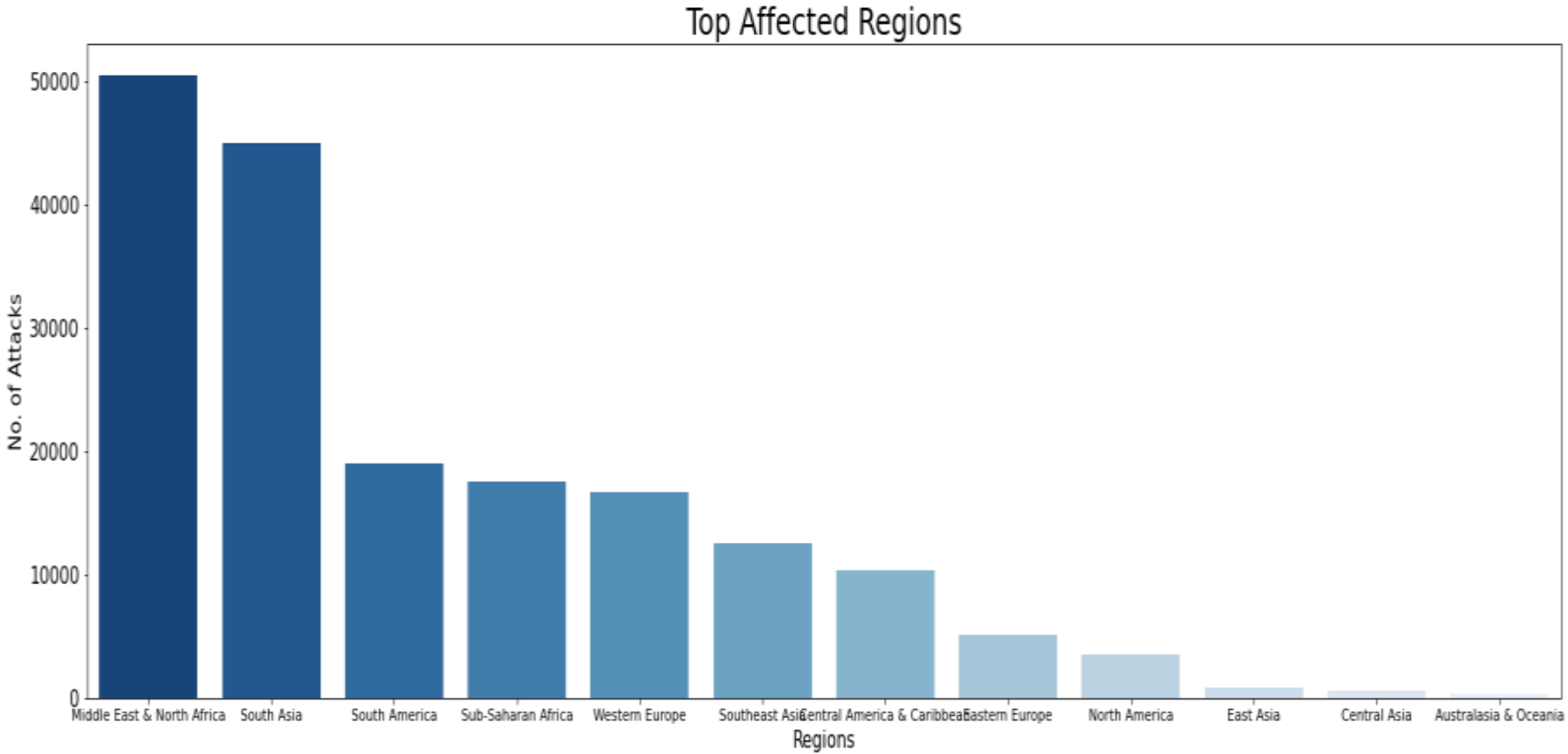
EDA Visualization



EDA Explanation

- Top affected Countries Visualization shows some of the most affected countries are Iraq, Pakistan, Afghanistan, and India based on the total number of attacks.
- But graph does explain how some countries are prone to violent actions and difference in an ideology which can lead to extreme terrorism.

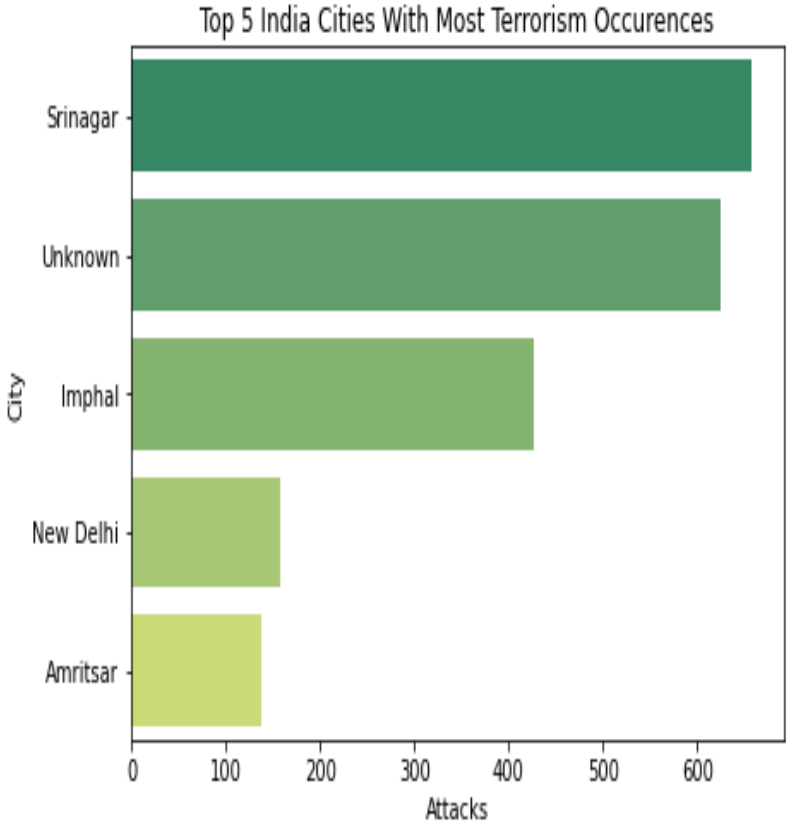
EDA Visualization

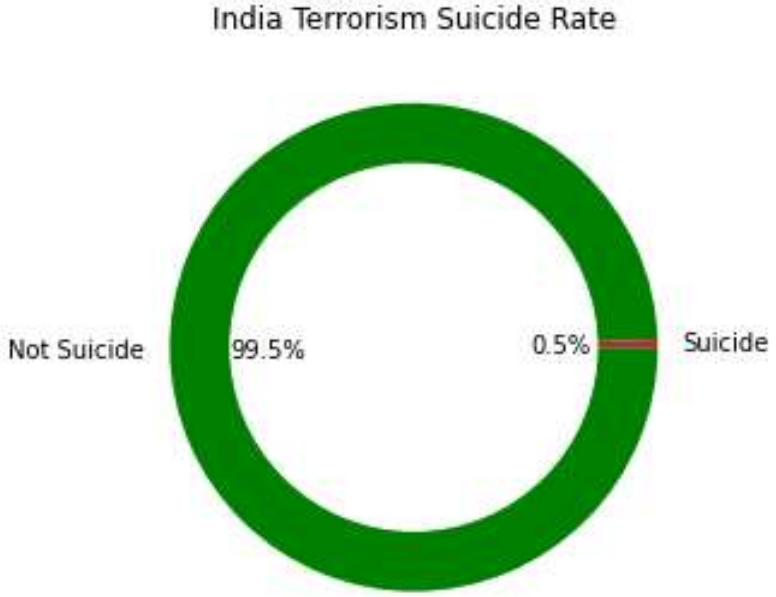


EDA Explanation

- Based on the geographic location of countries, they have been subcategorized into twelve regions to compare the rate of terrorism in each one of them as shown in Visualization
- Middle east and north Africa have the highest number of attacks followed by South Asia and South America. Terrorism here does not show an equal distribution among all regions.
- As a result, based on the number of attacks, different level of attention is required for each individual region.

EDA Visualization

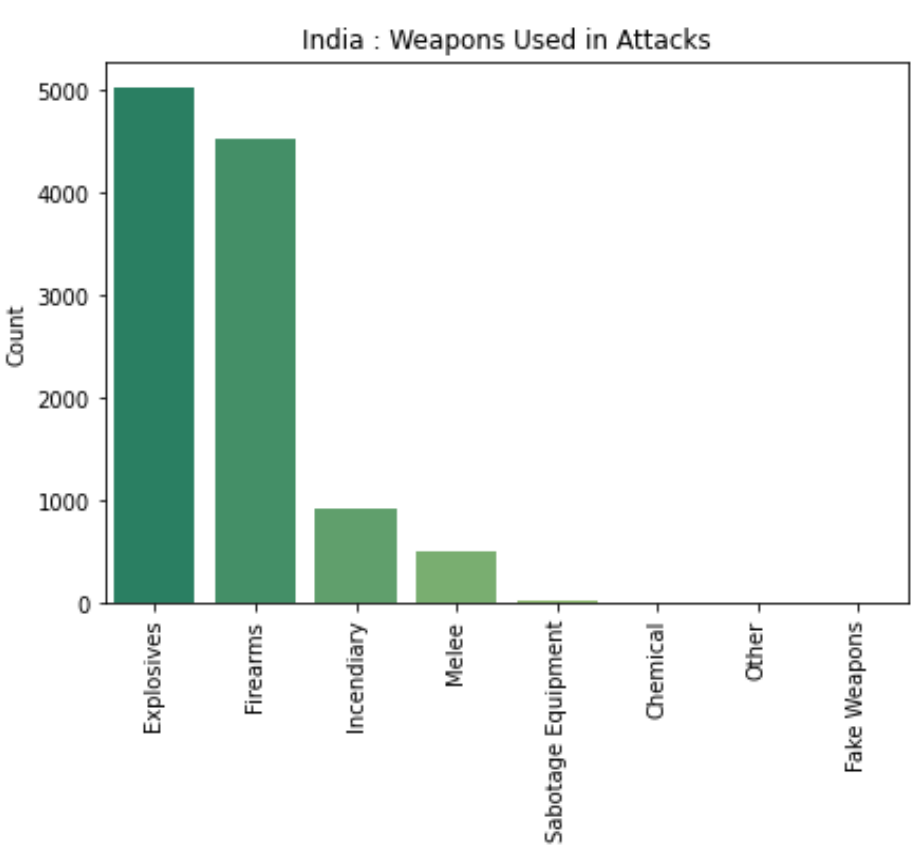


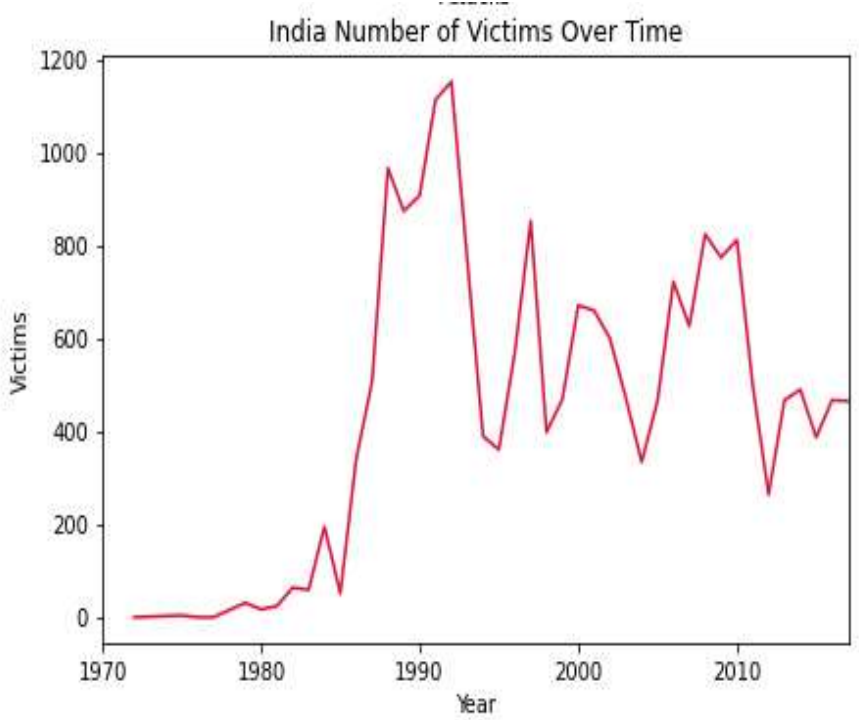


EDA Explanation

- Terrorism Analysis in India Between 1970 to 2017 analysis shows us that Shrinagar is a Hotspot of terror attacks with 600+ attacks are from here.
- In subsequent years, Imphal had been caught in a spiral of violence and have more than 450+ terror attacks happened here, as various militant outfits carried out deadly attacks on security forces or engaged in factional clashes.
- After that New Delhi and Amritsar come up with 100+ attacks on each.
- 2nd visualization shows that suicide rate of India is negligible through all years.

EDA Visualization

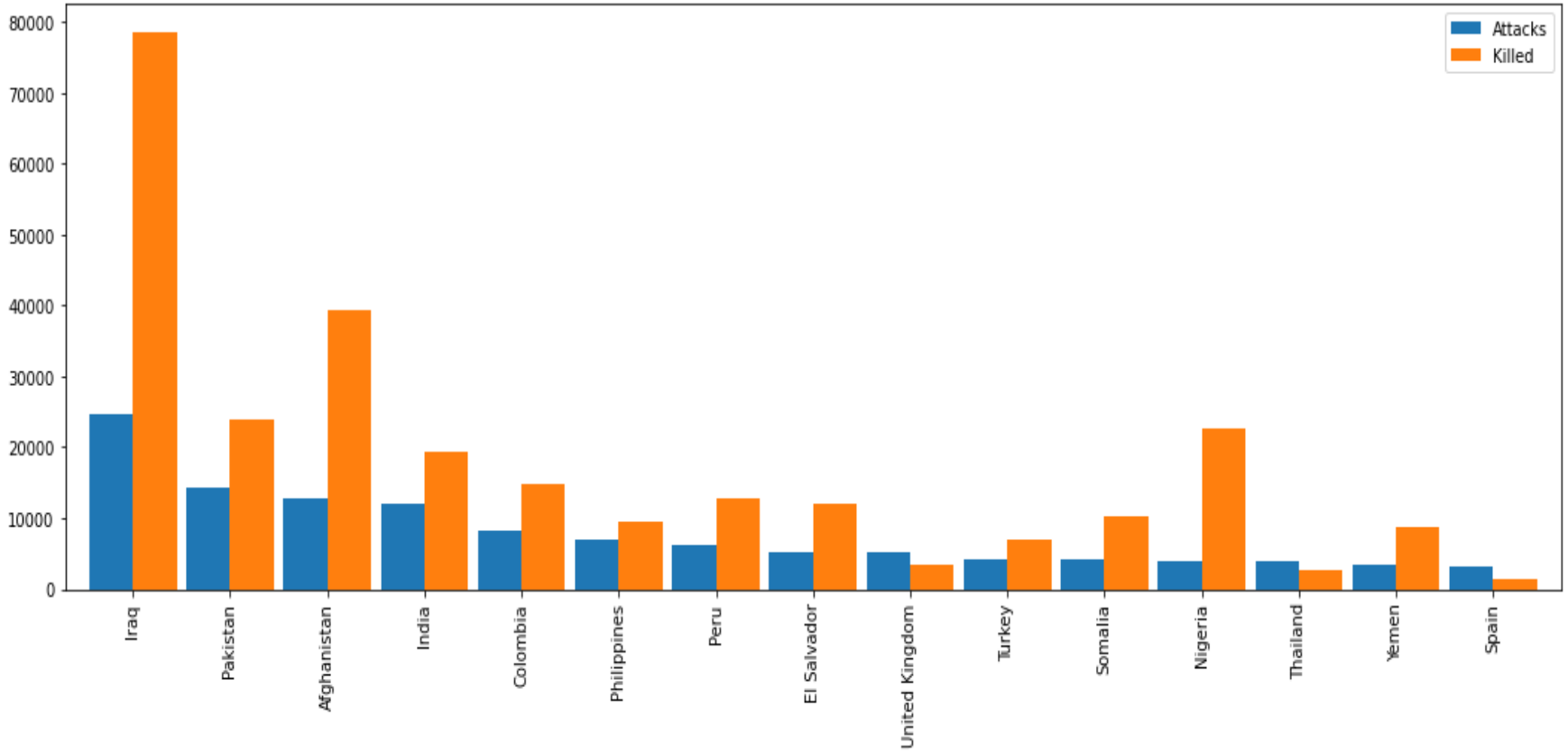




EDA Explanation

- 1st Visualization shows that in India different types of weapons and methods have been used by attackers. There are 8 categorical values for the defined attack type.
- They are bombing/explosion, firearms, incendiary, melee, sabotage equipment, chemical, fake weapons and others.
- Most attacks are by bombing/explosion hence the total number of casualties by explosive weapons is almost double than the next most attack which is firearms.
- 2nd visualization shows that number of victims i.e casualties of people appeared in the attack. The years 1970-80 are the peaceful years for inda but from 1980 the lineplot is going upward.

EDA Visualization



EDA Explanation

- Visualizations lists the most countries affected by terrorism based on the total number of attacks. Another bar along with each countries' attack count is the number of total victims killed in those attacks combined for that country.
- We can analyse kills to attack ratio for the most affected countries. For Iraq, that ratio is very high. The average kill for each attack is over 3. There are countries like Philippines, Peru and the United Kingdom which has faced an almost similar number of attacks but have a different number of kills.
- This comparison among countries can be taken into considerations while devising new tactics against terrorism.

Conclusion

- A visualization which can be used to analysed the India's major aspects related with terrorist attacks, number of victims over time, most used weapons and the global analysis to calculate the total number of attacks on countries and regions, kills to attack ratio, most attacks by the terror groups over years provides interactive interface to explore this dataset.
- Users can also explore START dataset and other terrorism related sources for additional research purposes provided in this tool. This work can be used by curious civilians, security related policy-makers, international organizations hosting worldwide events, Indian's securities and organizations, foreign investors and academic researchers for the purpose of understanding terrorism and its nature.