

Capstone Project

Global Terrorism Exploratory Data Analysis

Team Members

Yash Patil

Mohd. Zahid Ansari

Pritam Pawar

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 & 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.

Handling Missing Data

- ✓ [6] #This column shows us total no. of casualties in a particular attack
`gt_df['Casualties'] = gt_df['Killed'] + gt_df['Wounded']`
- ✓ [7] #Dropping datapoints/rows which have null values in latitude and longitude column for further visualisation.
`gt_df.dropna(subset=['latitude', 'longitude'], inplace=True)`
- ✓ [8] #Replacing -99 value for Damages column to drop rows for better visualisation.
`gt_df['Damages'].replace([-99.0, np.nan], 0.0, inplace=True)`
`gt_df['Damage_txt'].replace('Unknown', np.nan, inplace=True)`
- ✓ [9] # We have to remove Unknown values in Weapon_type column
`gt_df.drop(gt_df.index[gt_df['Weapon_type'] == 'Unknown'], inplace=True)`
`gt_df.drop(gt_df.index[gt_df['City'] == 'Unknown'], inplace=True)`

Data Wrangling

- ❖ Performed Feature extraction on the dataset to filter out useful data only.
- ❖ Renamed the columns for better understanding.

▼ Data Pre-Processing

Raw data is often unrecognized and contains a lot of information which is irrelevant to the project requirements. Data preprocessing methodology helps in converting this raw data into a more meaningful, focused, interpretable and readable format.

```
✓ ▶ #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', 'r  
[5] #Keep Relevant columns only  
gt_df = gt_df[['eventid', 'Year', 'Month', 'Day', 'Country', 'Region', 'latitude', 'longitude', 'Attack', 'Target', 'Killed', 'Wounded', 'Summary', 'Group_name', 'Target_type', 'Weapon_type', 'Moti
```


Major Visualization Aspects

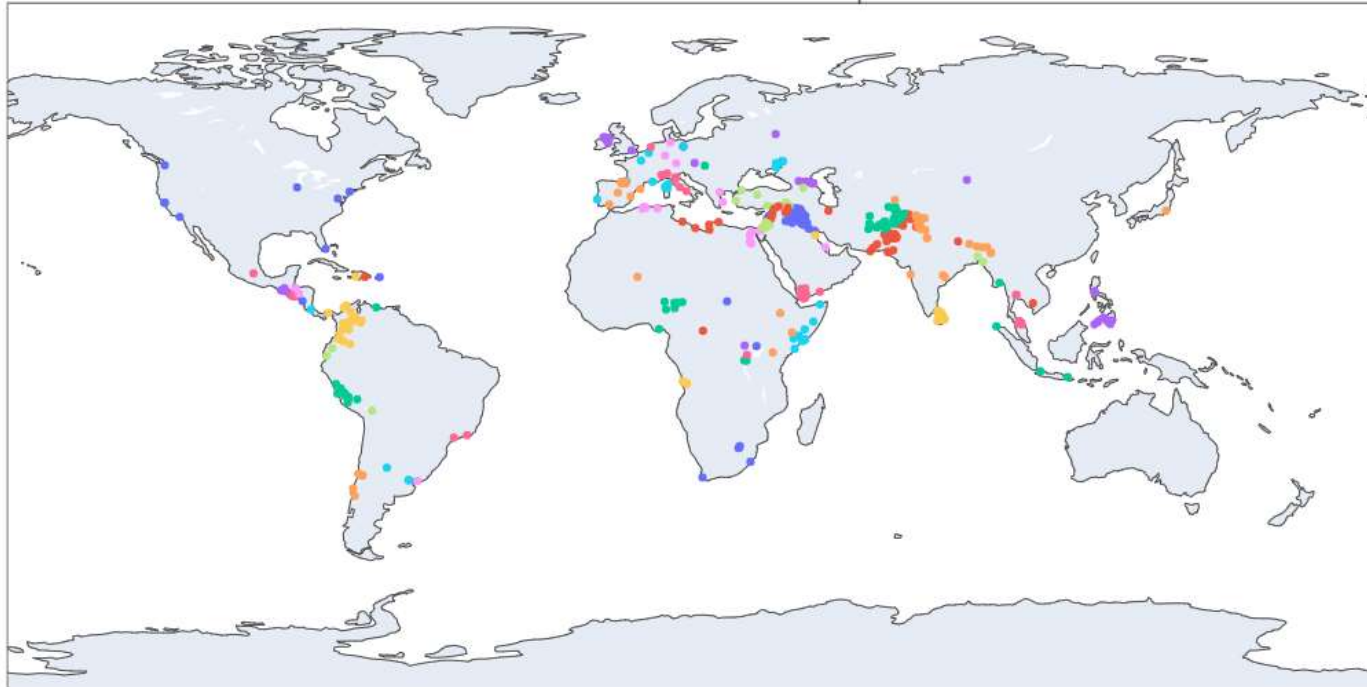
We divided this analysis on 4 major aspects with respect to features in datasets and they are:

- ❖ Regions / Countries
- ❖ Weapon Type
- ❖ Terrorist Groups
- ❖ Targeted Organizations.

EDA Explanation

- ❖ We analyzed and preprocessed data of features like latitude and longitude to get the coordinates of the terrorist attacks on the worldmap.
- ❖ Plotting the attack coordinates on world map we can figure out by densely populated marker, which country is highly prone to terrorist attacks!
- ❖ We can see that countries like Pakistan, Afghanistan, Colombia, Somalia, Iraq and etc. have high number of Terrorist Attacks.

EDA Visualization



Country

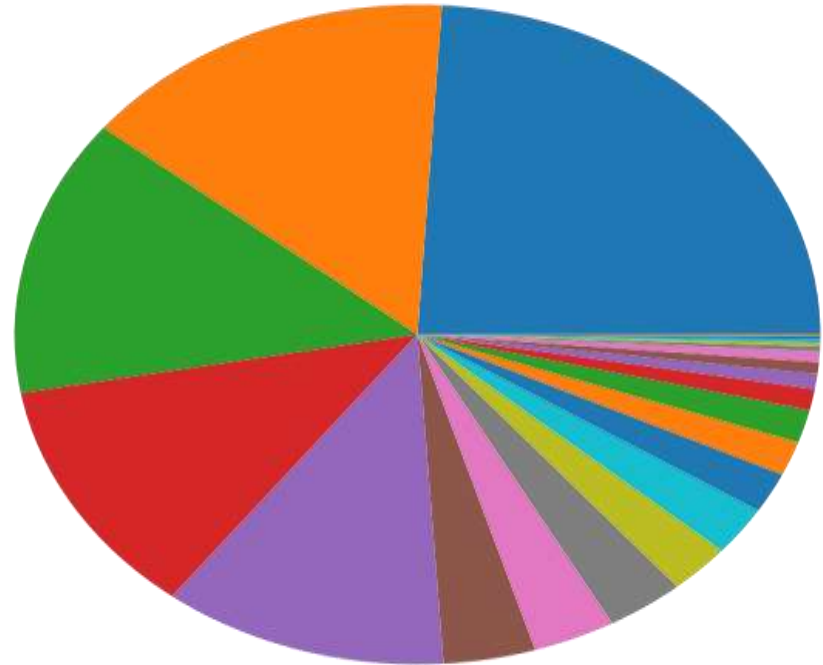
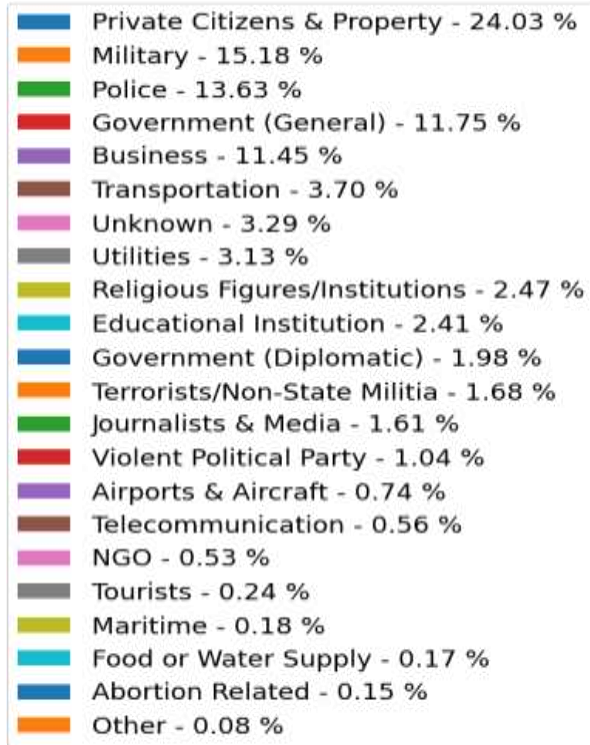
- Iraq
- Pakistan
- Peru
- United Kingdom
- Chile
- Somalia
- El Salvador
- Turkey
- Greece
- Colombia
- Lebanon
- Libya
- Afghanistan
- Guatemala
- India
- France
- Italy
- Bangladesh
- Egypt
- Israel
- United States
- Syria
- Nigeria
- Philippines
- Spain

EDA Explanation

- ❖ Analyzing the features through data wrangling we also considered the Targeted Organizations by Terrorist from 1970 to 2017.
- ❖ Understanding the features and visualizing them we got to know that Targeted Organizations mainly includes publicly held institutions or organizations and government facilities.
- ❖ Highest terrorists attack are done on Military, Private Citizens & infrastructure and Government (in General).

EDA Visualization

Types of targets attacked by terrorists.

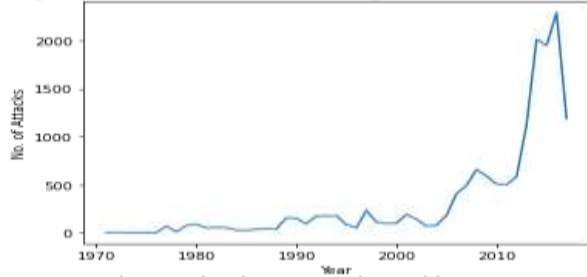


EDA Explanation

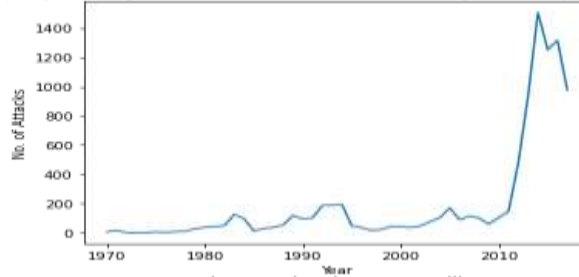
- ❖ In this Visualization we are analyzing the trend of terrorist attacks from 1970 to 2017.
- ❖ This visualization also includes top Regions affected versus the number of attacks happened in that region.
- ❖ We can understand from Subplot(1,1) that in 1900's the terrorist attacks where much lower than what can we see after the 2000's.
- ❖ We can also deduce that the terrorist attacks where on Private Citizens and Property and also on their Military Facilities.

EDA Visualization

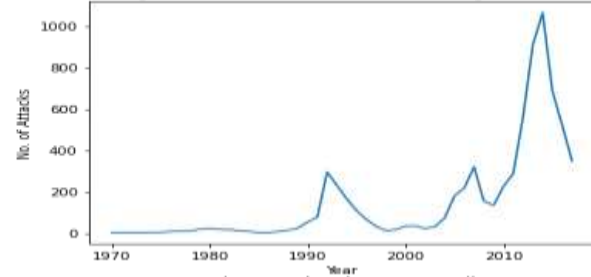
Region: Middle East & North Africa, Target: Private Citizens & Property



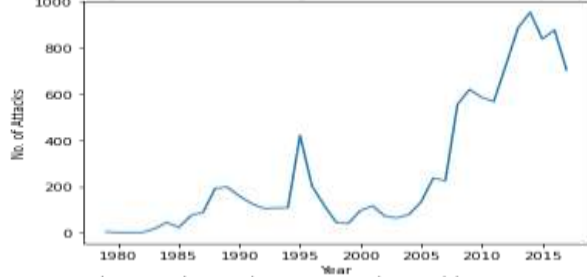
Region: Middle East & North Africa, Target: Military



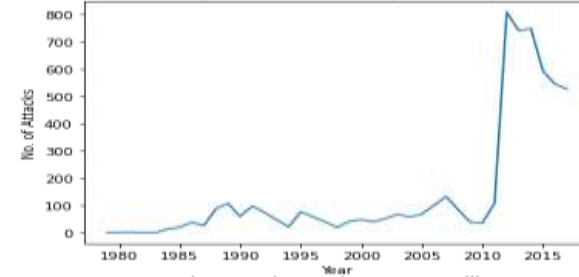
Region: Middle East & North Africa, Target: Police



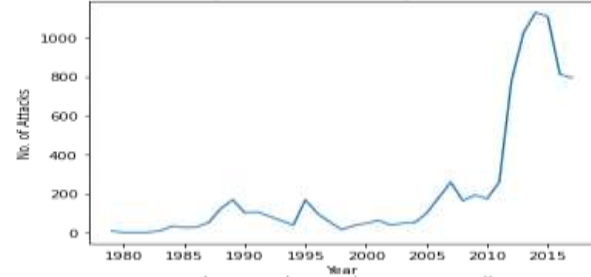
Region: South Asia, Target: Private Citizens & Property



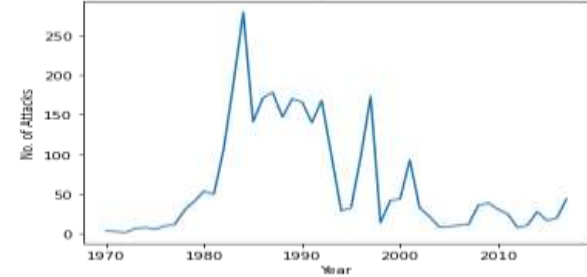
Region: South Asia, Target: Military



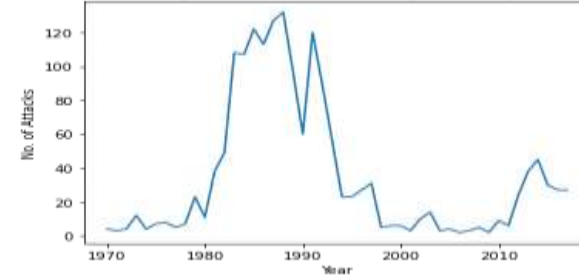
Region: South Asia, Target: Police



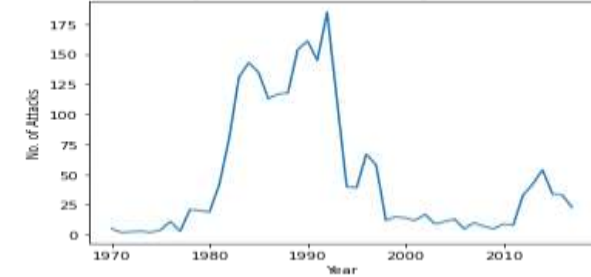
Region: South America, Target: Private Citizens & Property



Region: South America, Target: Military

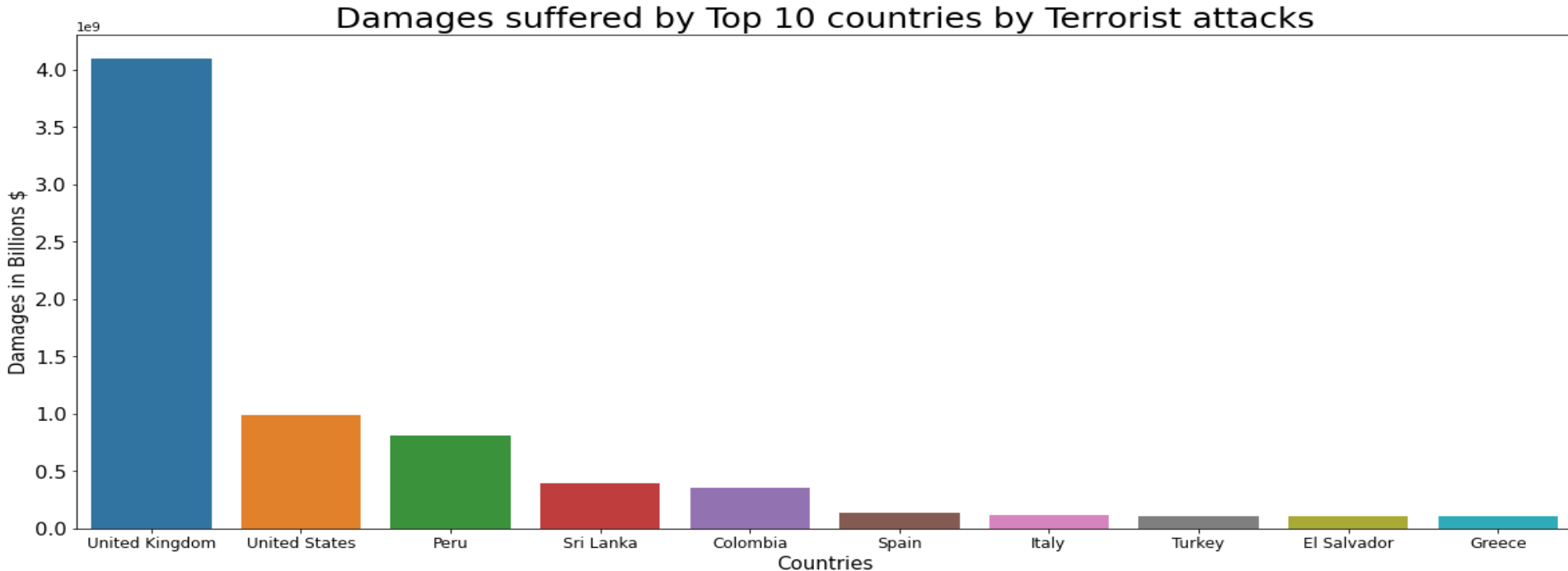


Region: South America, Target: Police



EDA Visualization and Explanation

❖ In this Visualization we are analyzing the damages incurred by which Countries from Terrorist attacks over the years.



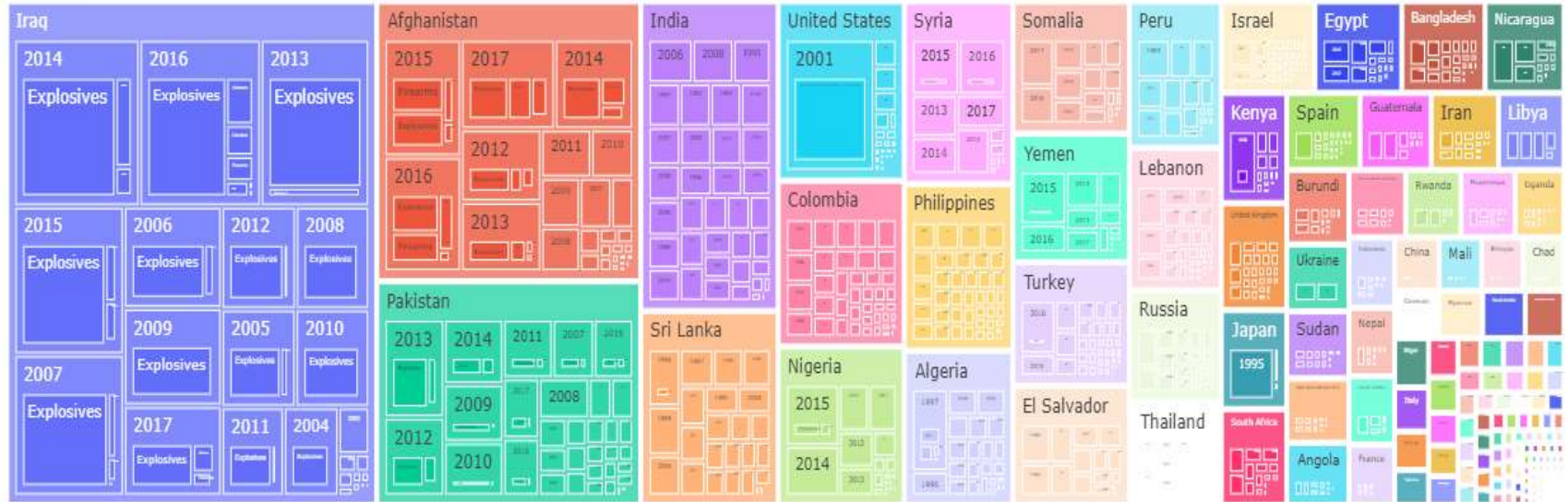
EDA Explanation

- ❖ In this Visualization we are analyzing the Causalities caused by different types of Weapons over the years using a Treemap.
- ❖ We can see that weapons like High Grade Explosives and Firearms cause the greatest number of casualties.
- ❖ Looking into the metrics we find that Explosives caused over 5 lakh+ casualties and Firearms caused over 2 lakh+ casualties.

EDA Visualization

Casualties by different types of weapons.

Weapons

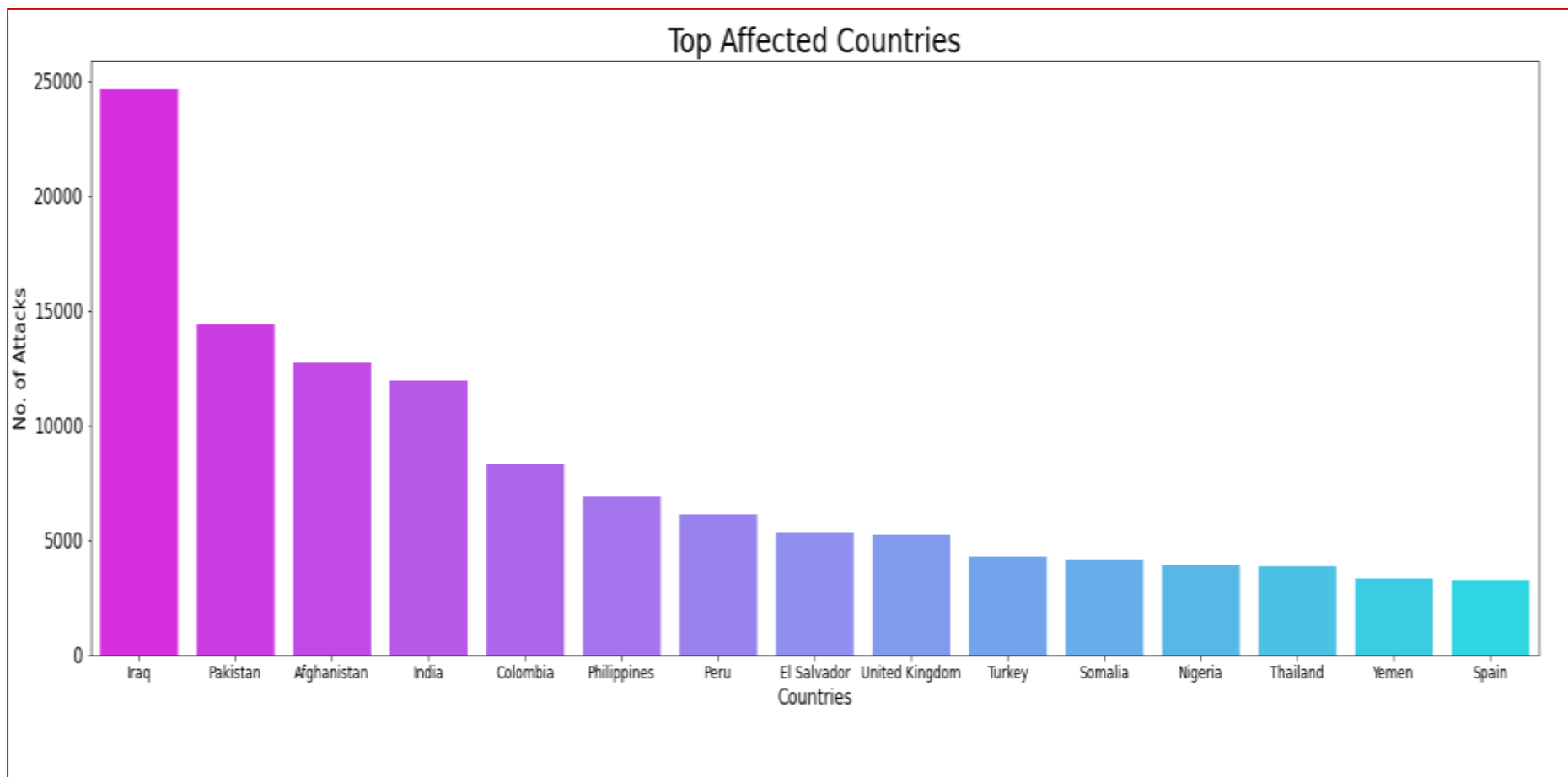


EDA Explanation

- ❖ 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.



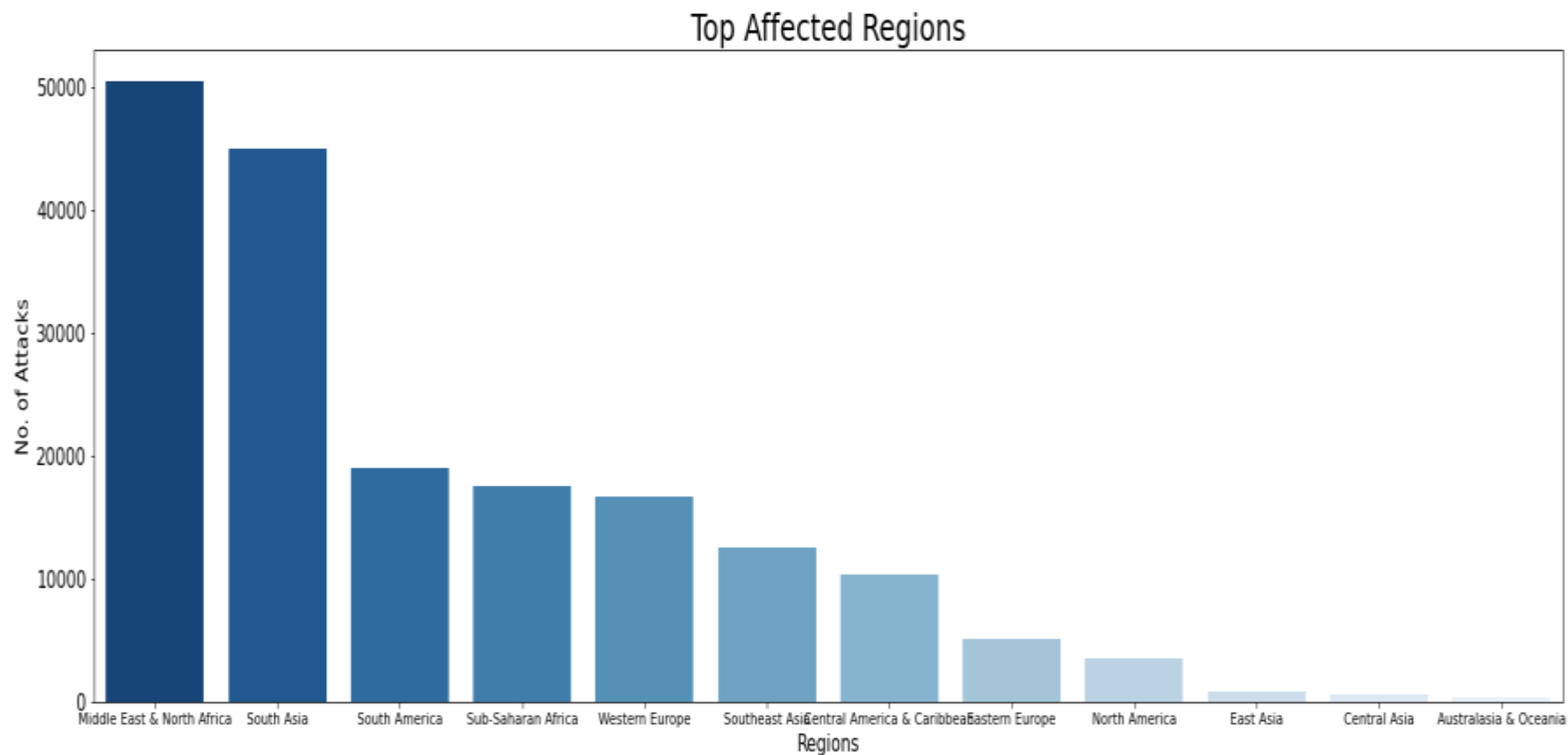
EDA Visualization



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

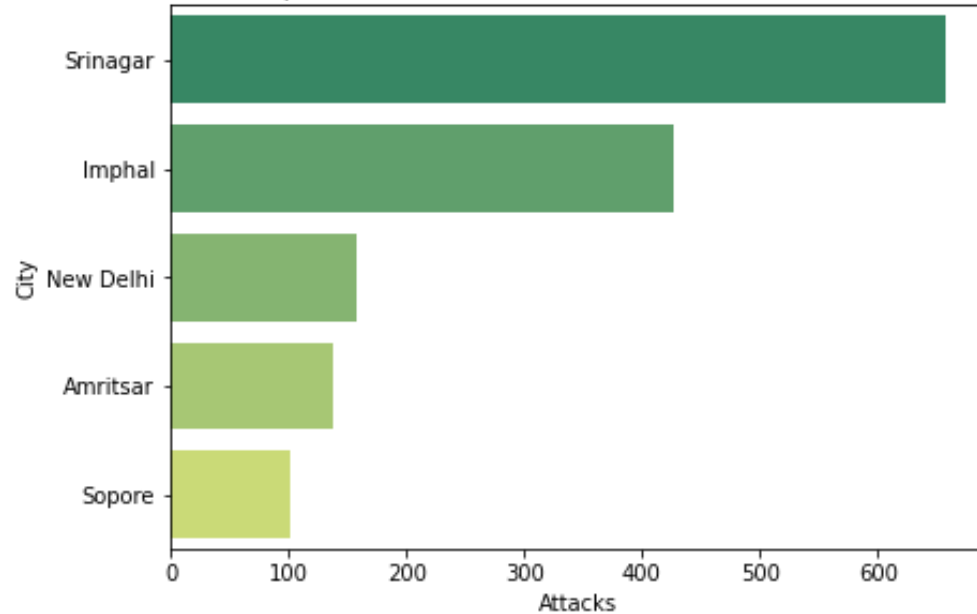


EDA Explanation

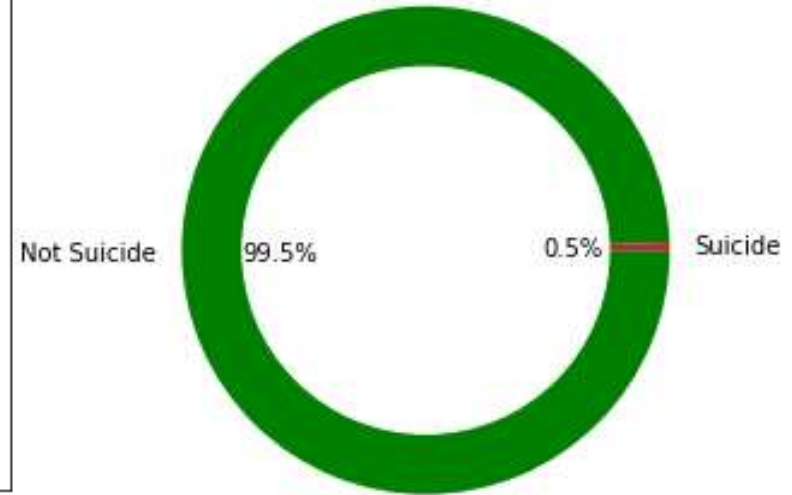
- ❖ Terrorism Analysis in India Between 1970 to 2017
analysis shows us that Srinagar is a Hotspot of terror attacks with 600+ attacks are from here.
- ❖ In subsequent years, Imphal had been caught in a spiral of violence and have more than 450+ terror attacks happened here, as various militant outfits 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

Top 5 India Cities With Most Terrorism Occurences



India Terrorism Suicide Rate

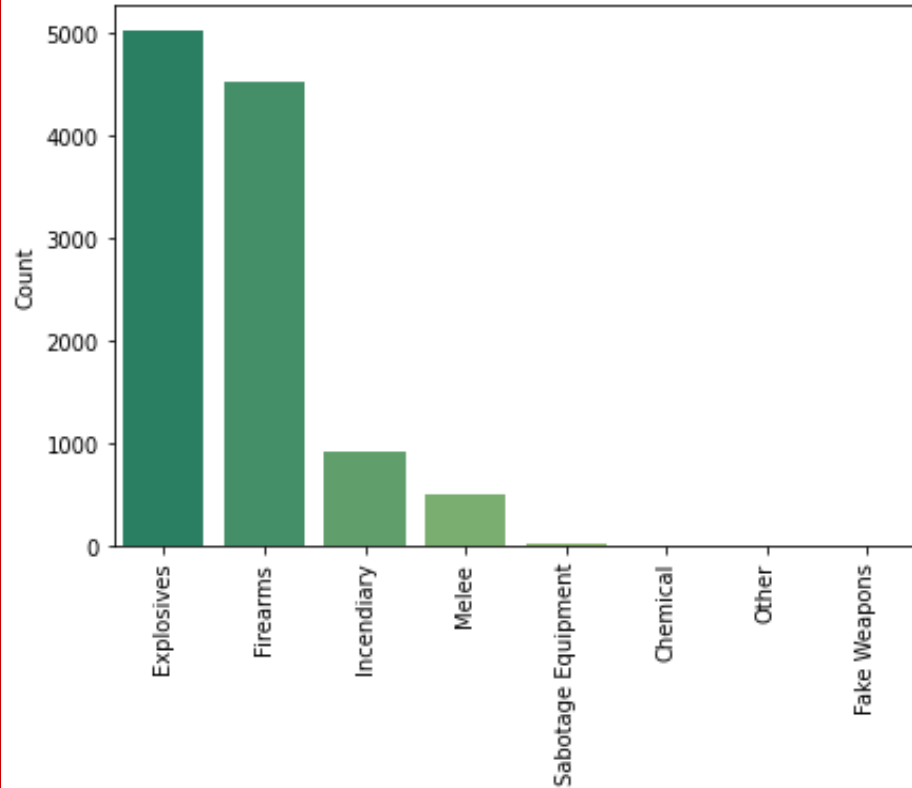


EDA Explanation

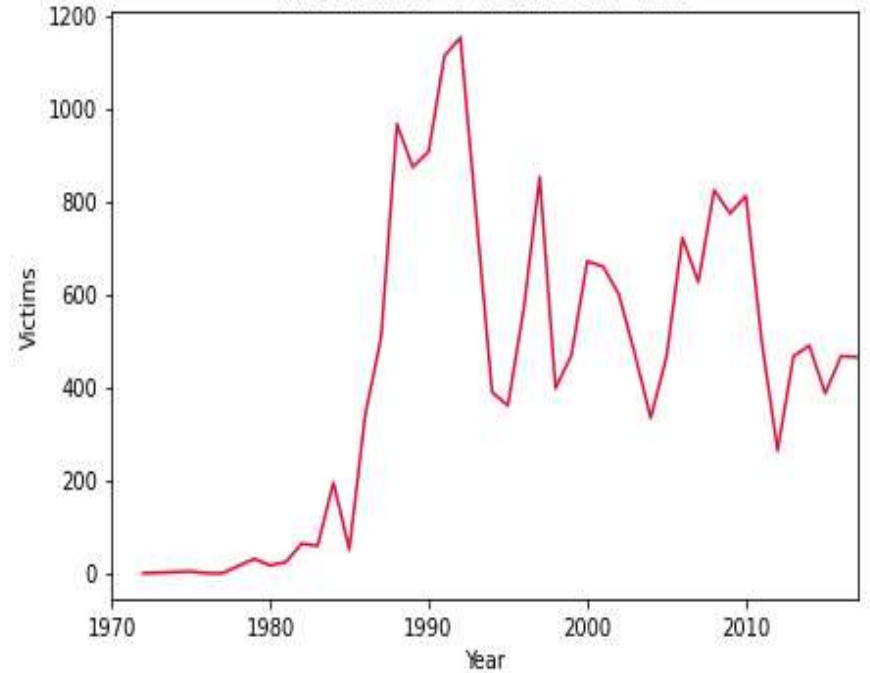
- ❖ 1st Visualization shows that in India different types of weapons and methods have been used by attackers. There are 8 categorical values for the defined attack type.
- ❖ They are bombing/explosion, firearms, incendiary, melee, sabotage equipment, chemical, fake weapons and others.
- ❖ Most attacks 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 india but from 1980 the lineplot is going upward.

EDA Visualization

India : Weapons Used in Attacks

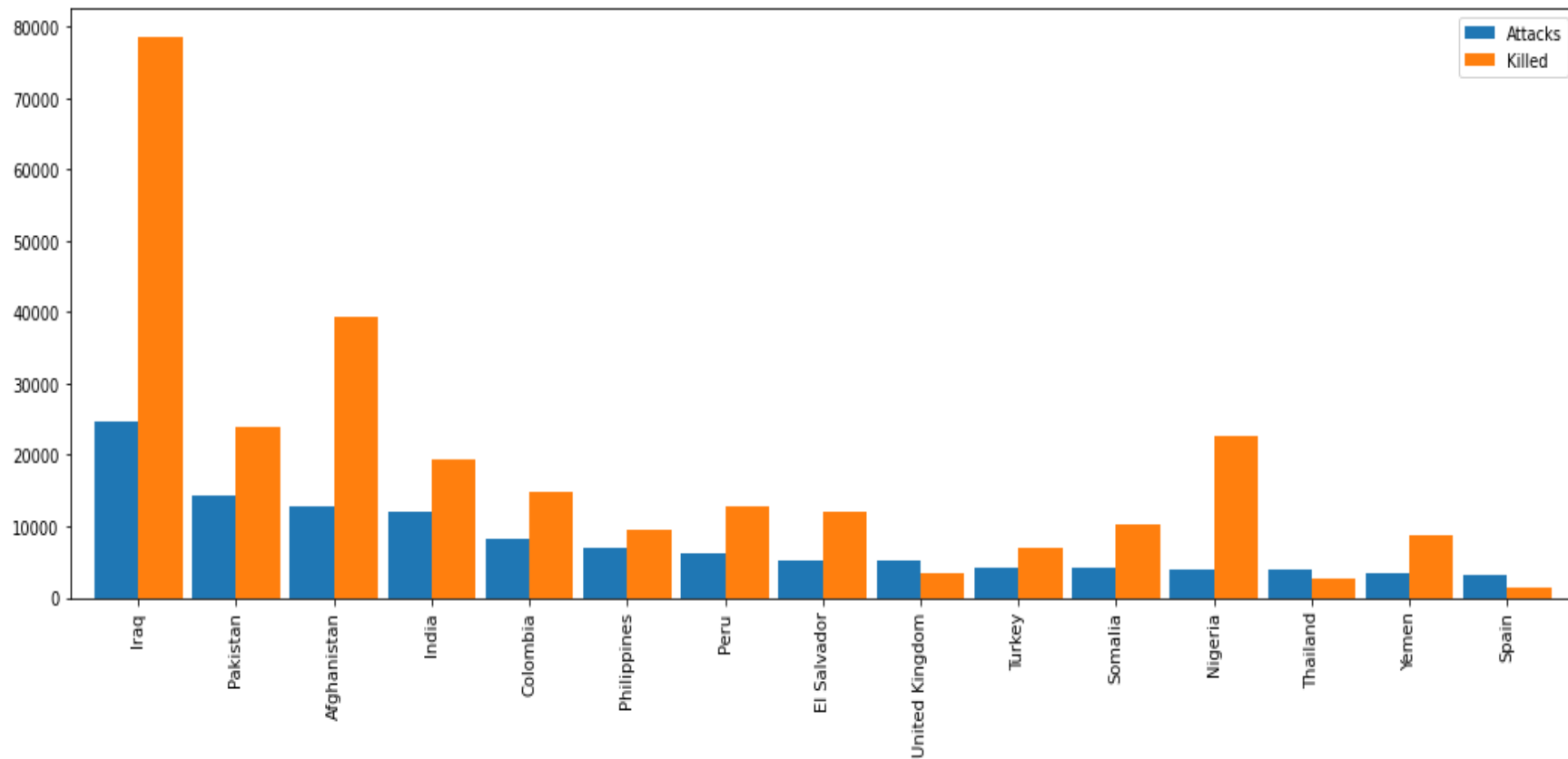


India Number of Victims Over Time



- ❖ 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.

EDA Visualization

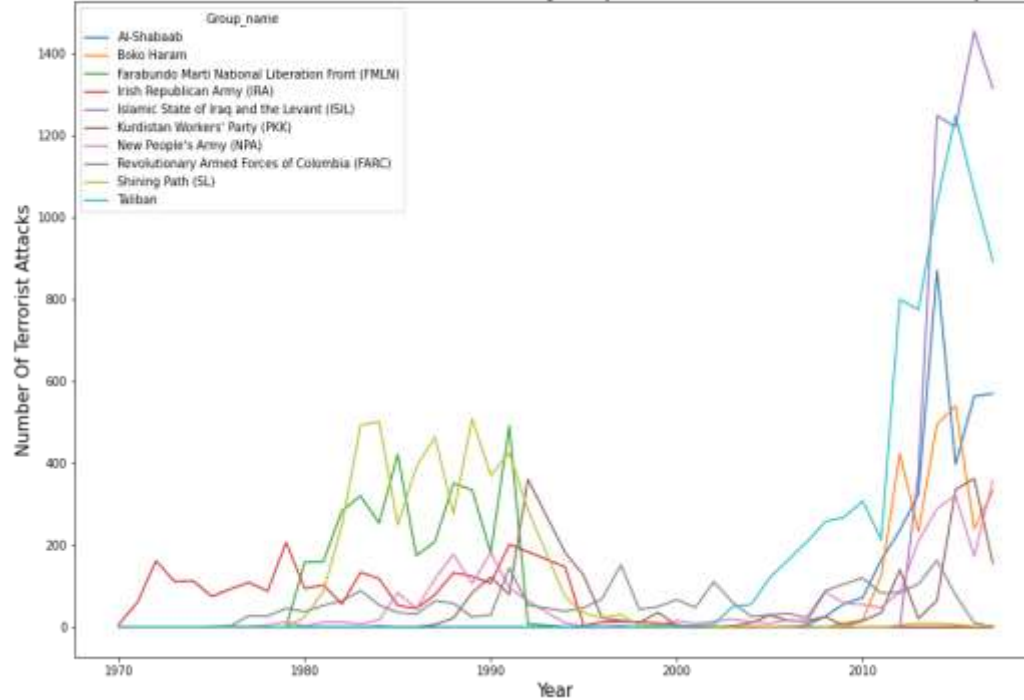


EDA Explanation

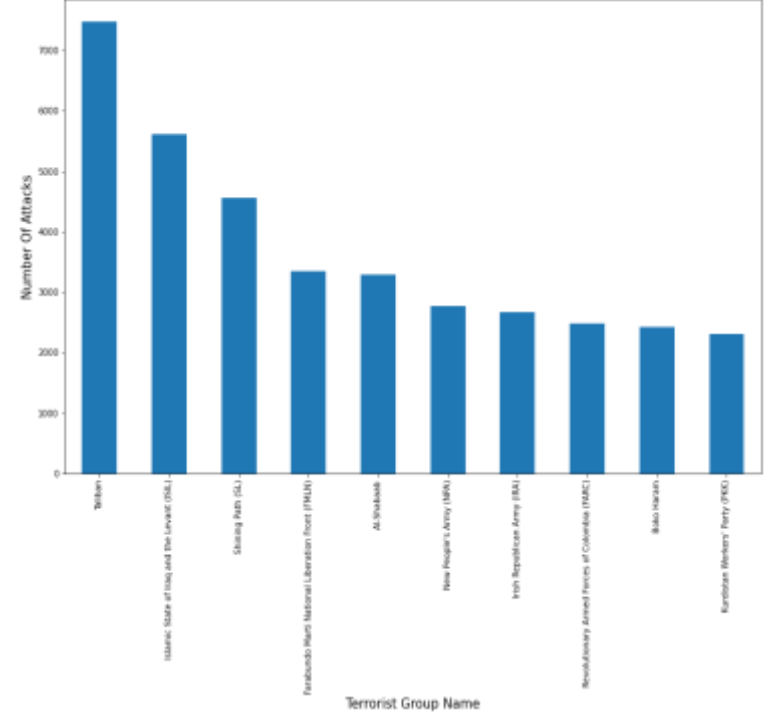
- ❖ From above line chart we can see the number of terror attacks by top 10 terrorist groups from 1970 to 2017.
- ❖ Here we can see how Taliban rapidly grow in terror activities from 2003 to 2017.
- ❖ Shining path terrorist group started their terror activities from 1979 and had been worked since 1995. After that their activities are not that much.

EDA Visualization

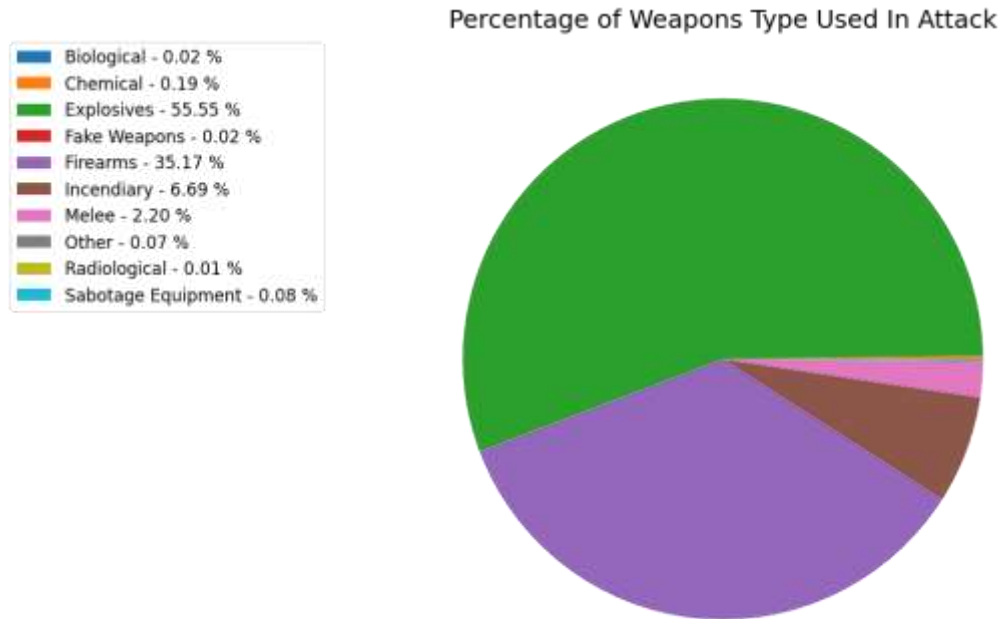
Number Of Attacks Per Year By Top 10 Active Terrorist Group



Top 10 Terrorist Attack Groups And Their Number Of Terror Attacks



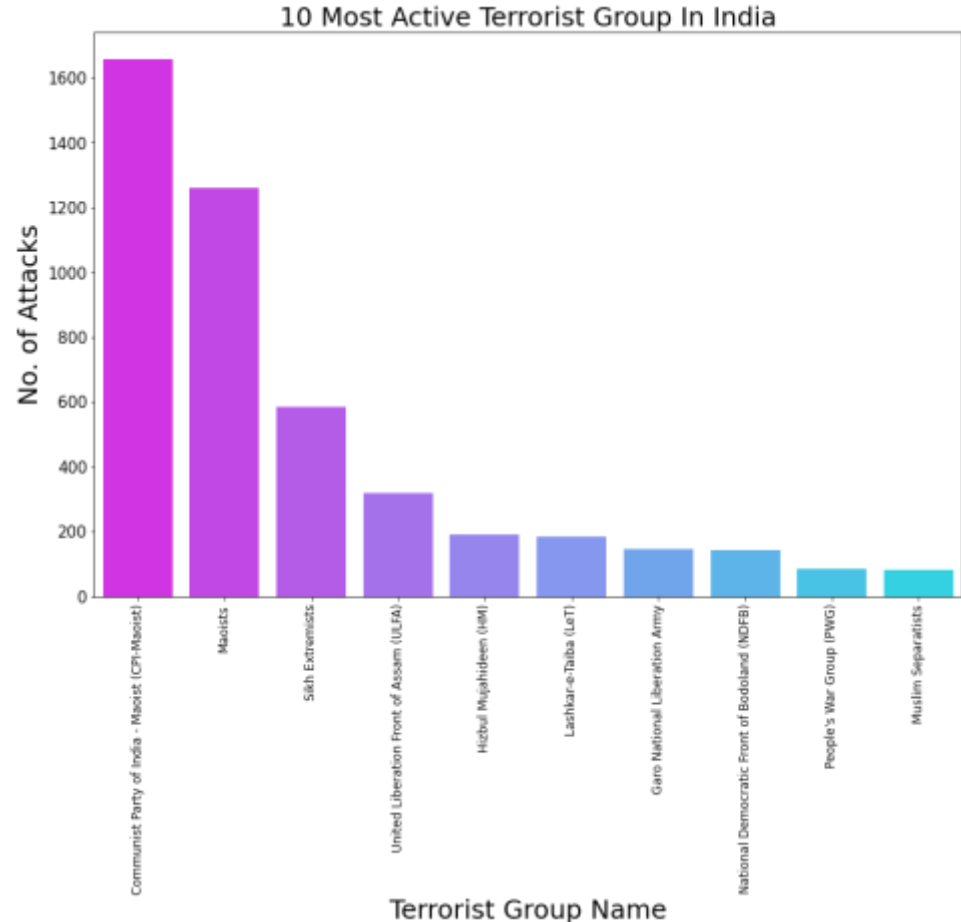
EDA Visualization and Explanation



- ❖ From above pie chart we can see that terrorists mostly use explosive, firearms and incendiary weapon type.

EDA Visualization and Explanation

- ❖ From this bar plot of 10 Most Active Terrorist Groups In India, we can infer that CPI-Maoist is the most active terrorist group in India.
- ❖ We can deduce that CPI is the most active Terrorist Group in India followed by Maoists and Sikh Extremists.



Conclusion

- ❖ Strategic intelligence gives an insight into terrorist intent, capability useful in prioritising risks and developing preventative measures
Also helpful in focusing on key vulnerabilities.
- ❖ A visualization which can be used to calculate the total number of attacks, total kill counts and location based on the selected region and year provides interactive interface to explore this dataset. 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.

Challenges

- ❖ Huge amount of data needed to be analyzed and understood to extract useful features from it.
- ❖ Handling sparse dataset like Global Terrorism was challenging because of many features and many null values included in the dataset.
- ❖ Pre-processed values in the features for better visualization.

Q & A