Capstone Project

Global Terrorism Exploratory Data Analysis

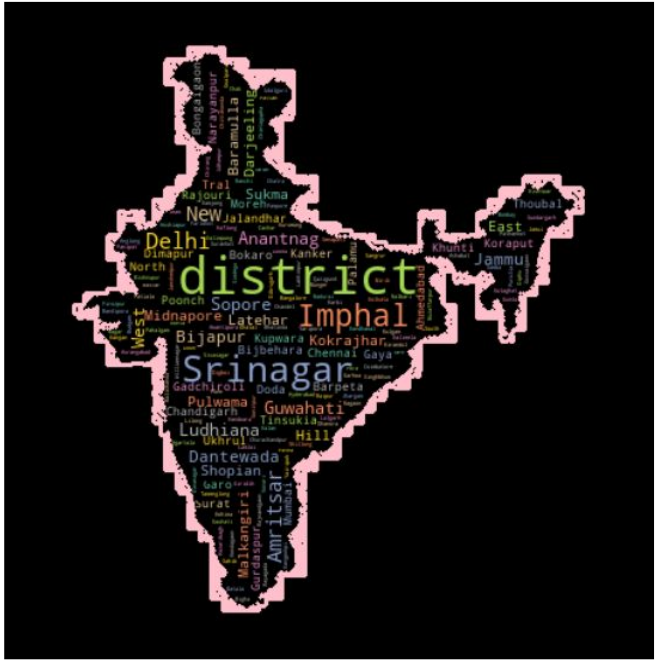
CREATED BY

Mohd. Zahid Ansari

Index

1. Summary
2. Features in EDA
3. EDA & Handling Missing Data
4. Data Wrangling
5. Major Visualization Aspects
6. EDA Visualization with their Explanation.
7. Conclusion
8. Challenges
9. Question and Answers

Most Occurence Terrorist Attacks Cities of India



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']

     #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.

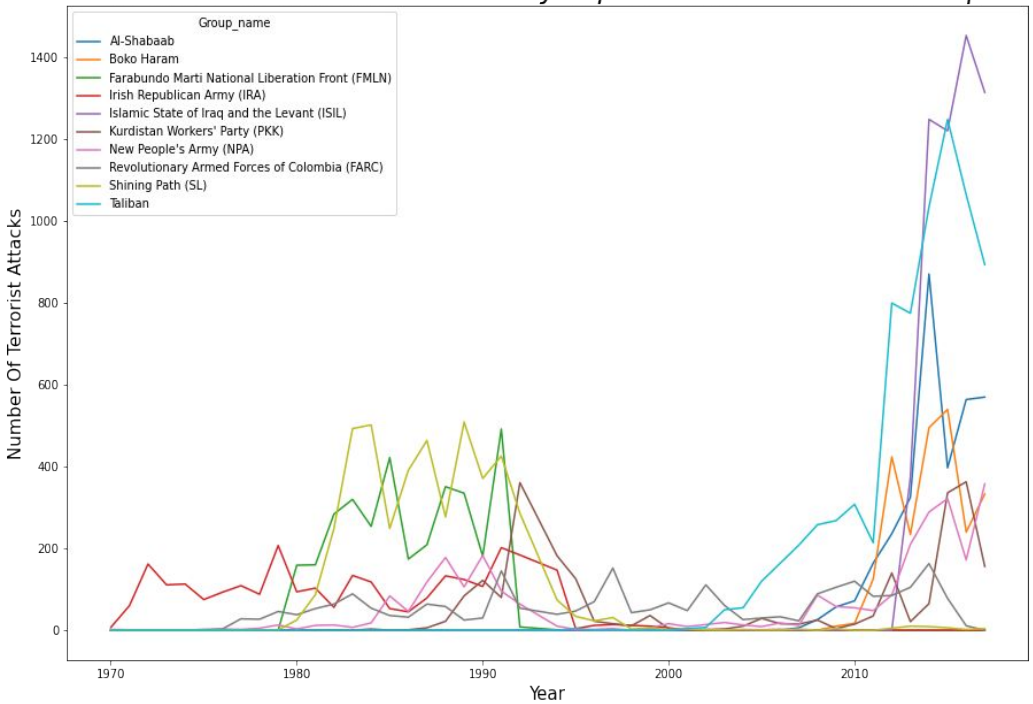
Major Visualization Aspects

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

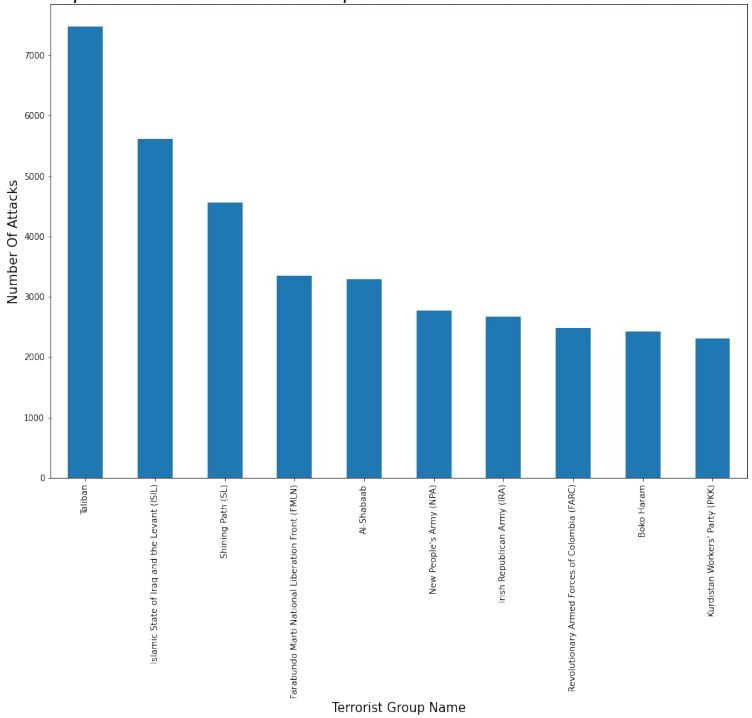
- Regions / Countries
- Weapon Type
- Terrorist Groups
- Targeted Organizations.

EDA Visualization

Number Of Attacks Per Year By Top 10 Active Terrorist Group



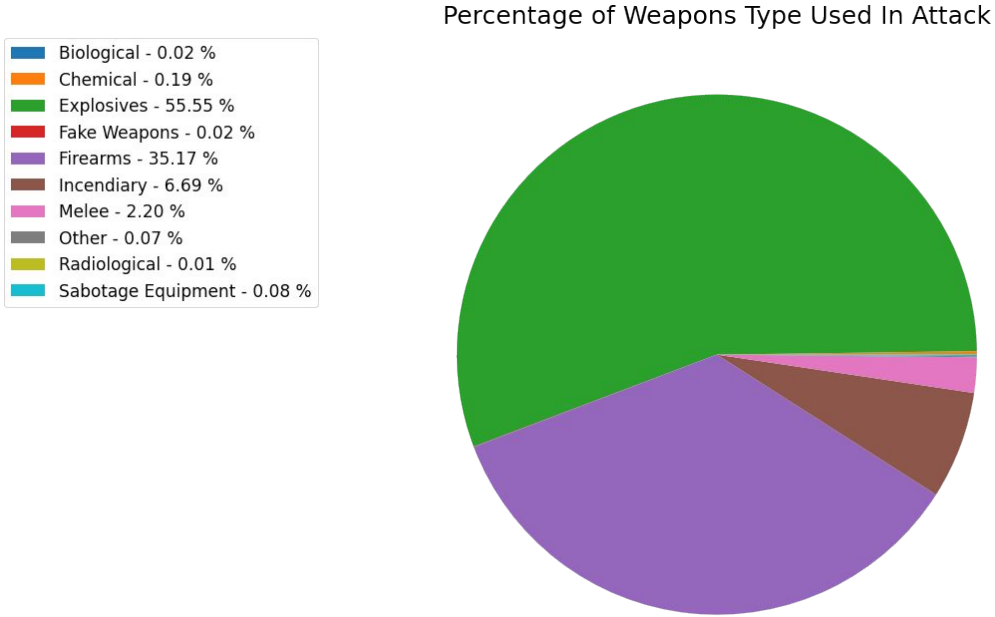
Top 10 Terrorist Attack Groups And Their Number Of Terror Attacks



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.

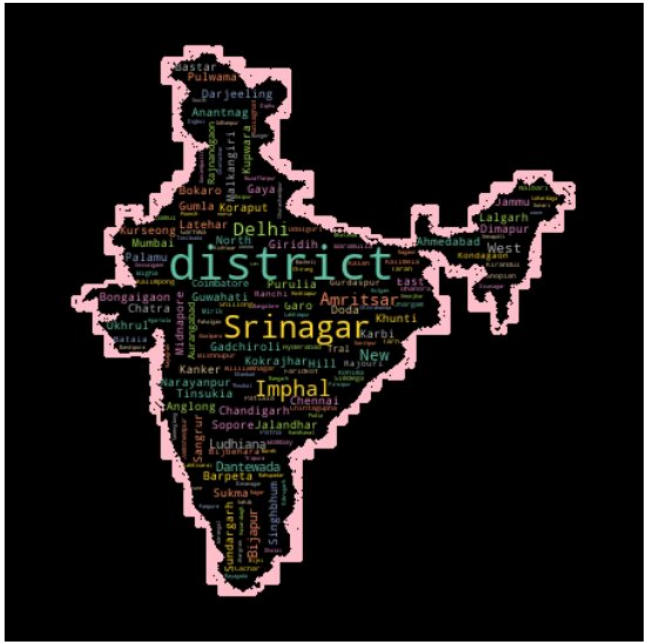
EDA Visualization and Explanation



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

EDA Visualization and Explanation

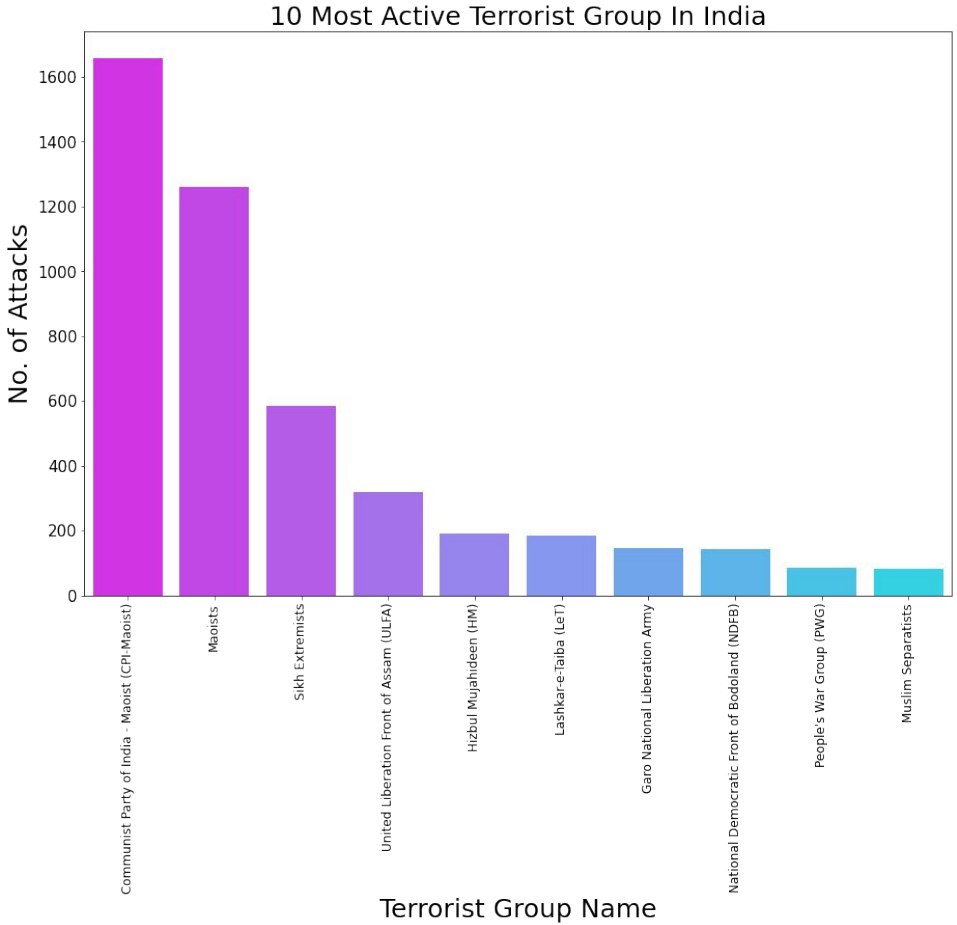
Most Occurrence Terrorist Attacks Cities of India



- From this word cloud of highly attacked cities of India we can infer that Srinagar, Imphal, Amristsar etc. are highly attacked cities of India those are in North India. So terrorists mostly attacks in the north side of India.

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 prioritizing 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 per year by which terrorist group, total attack counts on Indian cities based on 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 policymakers, 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