STATISTICS PROJECT

ON

EXPLORATORY DATA ANALYSIS

ON THE GLOBAL TERRORISM DATABASE (GTD)

Mentor: Dr. Sayantani Roy Choudhury

Submitted to –

Praxis Business School

Submitted by,

Shramana Bhattacharya – C22022

Saswata Pradhan – C22020

PGPDS – Jan 2022

Exploratory Data Analysis on Global Terrorism Database (GTD)

Submitted by,

Shramana Bhattacharya (C22022)

Saswata Pradhan (C22020)

Abstract:

This project explores the Global Terrorism database and tries to understand which are the worst affected countries, the types of terrorist attacks launched by the prominent terrorist groups, and the types of weapons used by them across the world. Year-wise analysis has been done to observe the increase in terrorism over the years from 1970 to 2019. We have tried to bring forward the main causes, weapons, injuries, and other details of terrorist attacks on those countries through our analysis. In our overview of the analysis of terrorism, we studied how the number of terrorist acts varied around the globe and how they changed over time.

Exploratory data analysis has been performed on the data obtained from the GTD website to understand the current global terrorism situation focusing on the top 5 countries where the maximum number of terrorism activities have occurred since 1970 and then concentrating on India, we found out the major affected states and the causes of the attacks there.

INTRODUCTION

Terrorism is an act of violence that seeks to create fear and harms or disrupts life. There are various definitions of the word 'terrorism'. Terrorism has been practiced by political organizations, nationalists and religious groups, revolutionaries, and state institutions such as armies, intelligence services, and police. In the 1790s, to refer to the terror used during the French revolution, the word was first coined by the revolutionaries against their opponents. Since the 20th century, the term has been mostly used to refer to any violence aimed at the government to influence any policy or topple an existing regime. Strength of arms is the principal means of victory used by the conventional military forces against their foe. Likewise, guerrilla forces often rely on violence to obtain military victory. Some social scientists have mentioned terrorism as the "weapon of the weakest."

Individuals and groups choose terrorism as a tactic because it can:

- Act as a form of asymmetric warfare in order to directly force a government to agree to demands
- Intimidate a group of people into capitulating to the demands in order to avoid future injury
- Get attention and thus political support for a cause
- Directly inspire more people to the cause (such as revolutionary acts) – propaganda of the deed
- Indirectly inspire more people to the cause by provoking a hostile response or over-reaction from enemies to the cause

The GTD (Global Terrorism Database) defines terrorism as: "The threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation."
The data collection team uses a series of inclusion criteria to systematically identify events for inclusion in the database. More information about the data collection process can be found in the GTD Codebook.
As of May 2021, the list extended through 2019 recording over 200,000 incidents, although data from 1993 is excluded. The database is maintained by the National Consortium for the Study of Terrorism and Responses to Terrorism (START) at the University of Maryland, College Park in the United States. It is also the basis for other terrorism-related measures, such as the Global Terrorism Index (GTI) published by the Institute for Economics and Peace.

The GTD was formally introduced in a paper in *Terrorism and Political Violence* by Gary LaFree and Laura Dugan of START, published in 2007. In 2018, the GTD suffered a lapse in funding that caused projects and updates to be temporarily put on hold. The Department of Defense Combating Terrorism and Technical Support Office and the German Federal Foreign Office were able to provide short-term funds which enabled the GTD to put out their report for 2018, but the Global Terrorism Database continues to seek funding for the long term.

While calling the Global Terrorism Database a treasure trove of information, a 2013 *Washington Post* fact-checking article criticized its use by government officials to hype the threat of terrorism around the world, given its use of a definition of terrorism conflicting with Congressionally required law.

In our project, we have used the information provided in the GTD to investigate the global scenario of terrorism, and its trend over time, focusing on the top 5 countries which include India, and then studied the situation of terrorism in Indian states.

DATA DESCRIPTION

The Global Terrorism Database (GTD) describes itself as the "most comprehensive unclassified database on terrorist events in the world" and includes over 200,000 terrorist attacks in its 2021 version. The link to the website from where the data has been downloaded is https://www.start.umd.edu/gtd/access/.

The entire database (about 80 MB excel file) includes more than 20,000 assassinations and more than 15,000 kidnappings.

The current GTD is the product of several phases of data collection efforts, each relying on publicly available, unclassified source materials. These include media articles and electronic news archives, and to a lesser extent, existing data sets, secondary source materials such as books and journals, and legal documents.

There are 135 attributes (columns) and 2,01,184 rows of data in the GTD database. We have only considered 21 columns that have less than 60% null values.

The columns included in our analysis are described below in brief: -

1. iyear: This field contains the year in which the incident occurred.

2. imonth: This field contains the number of the month in which the incident occurred.

3. country_txt: This field contains the name of the countries.

4. region_txt: It tells us about the regions or parts of different countries.

5. provstate: It tells us about the states of the countries.

6. Crit1: The act must be aimed at attaining a political, economic, religious, or social goal.
   Crit2:There must be evidence of an intention to coerce, intimidate, or convey some other message to a larger audience (or audiences) than the immediate victims.
   Crit3: The action must be outside the context of legitimate warfare activities.

7. attacktype1_txt: It informs us about the type of attacks i.e. kidnapping, explosion etc.

8. targtype1_txt: It informs us about the targets like Private, Governments, Journalist etc.

9. gname: It describes the name of the terrorist groups involved in the attack.

10. suicide: This variable is coded “Yes” in those cases where there is evidence that the perpetrator did not intend to escape from the attack alive

11. weaptype1_txt: It describes the type of weapons used by the groups

12. nkill: It describes the number of people killed

13. nwound: It describes the number of people wounded

14. nwoundte: It describes the number of terrorist wounded

15. latitude: It describes the location in coordinates

16. longitude: It describes the location in coordinates

17. nhostkid: It describes the number of hostages or kidnapping victims.

18. nhours: The duration of the incident is recorded either in this field or in the next field (ndays) depending on whether the incident lasted a matter of hours or days

19. ndays: The duration of the kidnapping / hostage incident last for more than 24 hours, this field records the duration of the incident in days

20. ransomamt: If a ransom was demanded, then the amount (in U.S. dollars) is in this field.

21. nreleased: If the “Attack Type” is “Hostage Taking (Kidnapping),” “Hostage Taking (Barricade Incident),” or a successful “Hijacking,” then this field will apply

EXPLORATORY DATA ANALYSIS

The critical process of data investigation in order to discover patterns, detect any anomaly, obtain the summary statistics and graphically representing the data is termed as Exploratory Data Analysis (EDA).

We have used data visualization methods and descriptive statistics to understand the data variables and the relationships between them (if any).

EDA helps in bringing out interesting characteristics from the dataset and eventually helps in model building and other data analysis from the dataset.

In our EDA, we have taken each of the attributes and tried to find out information on them.

DATA PREPROCESSING

Our dataset has 21 columns and 201183 Rows.
Each row in the dataset represents one particular terrorist attack and all the information corresponding to that particular terrorist attack are entered in the respective attributes.

Table 1: Descriptive Statistics on the numerical columns of the dataset

	nkill	nwound	nkillter	nwoundte	nhostkid	nhours	ndays	ransomamt	nreleased
count	201183	201183	201183	201183	201183	201183	201183	2.01E+05	201183
mean	2.267831	2.831964	0.355746	0.08198	0.985545	0.017994	0.940079	2.11E+04	0.27806
std	10.927856	39.728346	3.476231	1.25035	52.099804	0.498084	23.527725	2.48E+06	10.018089
min	0	0	0	0	0	0	0	0.00E+00	0
25%	0	0	0	0	0	0	0	0.00E+00	0
50%	0	0	0	0	0	0	0	0.00E+00	0
75%	2	2	0	0	0	0	0	0.00E+00	0
max	1570	10878	500	200	17000	72	2676	1.00E+09	2958

The above table shows the descriptive statistics for all the numeric columns. we can see that there is high variation among the data points in features that are denoted by high standard deviation. The median value is 0 for all of these features. We have replaced the missing values with the median.

Treating the missing values:

We have replaced the missing values with the medians of the columns.

Grouping of numerical features year-wise

The grouping of the numerical attributes was also possible to be done country wise, but since here, we are considering the global scenario of terrorism, we are more interested to know the distributions of data year-wise which will give us an idea of the global terrorism situation across the years in all the countries.

For the attributes: nkill, nwound, nkillter, nwoundte and nreleased; we have referred to the sum of these attributes across each year. Sum has been considered to understand the total number of persons who got affected by all the terrorist activities throughout the globe each year.

For the other attributes that are, nhours, ndays and ransomamt; we have considered the average values of them per year, that is, the average number of hours and days a person kidnapped in a terrorist activity was detained per year and also ransomamt refers to the average amount of money demanded by the terrorists for release.

Table 2: Displays the Descriptive Statistics of the numeric columns after grouping them year-wise.

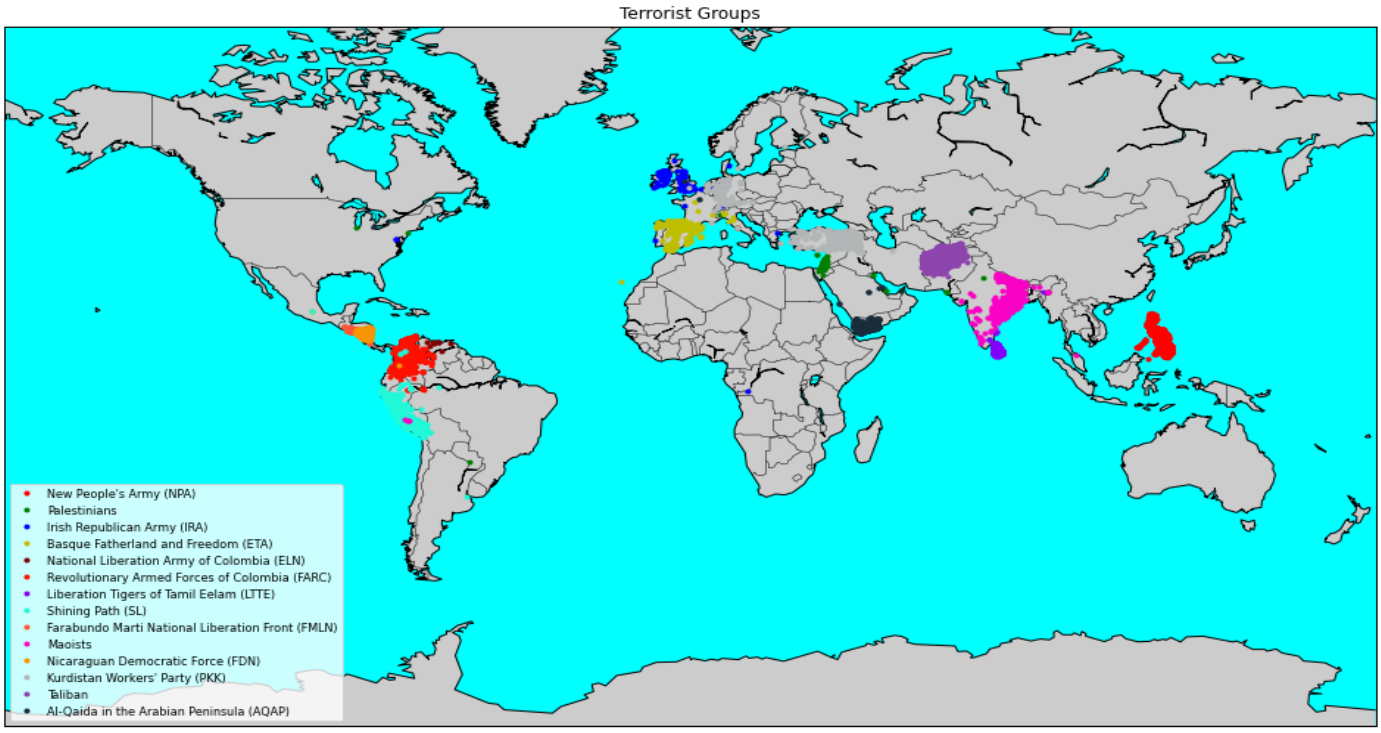
	nkill	nwound	nkillter	nwoundte	nhostkid	nhours	ndays	ransomamt	nreleased	suicide
count	49	49	49	49	49	49	49	49	49	49
mean	9311.204082	11627.40816	1460.612245	336.591837	4046.428571	0.022041	0.826735	37251.24367	1141.653061	148.346939
std	9866.017997	11511.86116	3092.165603	778.902725	6036.225666	0.017909	0.605611	107576.158	1404.460091	258.272833
min	173	82	4	0	60	0	0.02	634.39	45	0
25%	4391	3637	36	4	1084	0.01	0.37	3460.81	322	1
50%	7036	7384	206	16	1934	0.02	0.62	10933.01	645	15
75%	9444	15953	666	109	3944	0.03	1.18	32095.07	1228	184
max	44524	44204	11817	2970	34295	0.07	2.25	744494.76	6807	985

Adding another column:

Another column named 'Cause' is formed where it stores the criteria corresponding to each attack. The 'crit1','crit2','crit3' columns only have values 1's or 0's depending on whether the criteria is relevant for the attack or not. Hence, this new column is a categorical column which stores the corresponding criteria types for each attack.

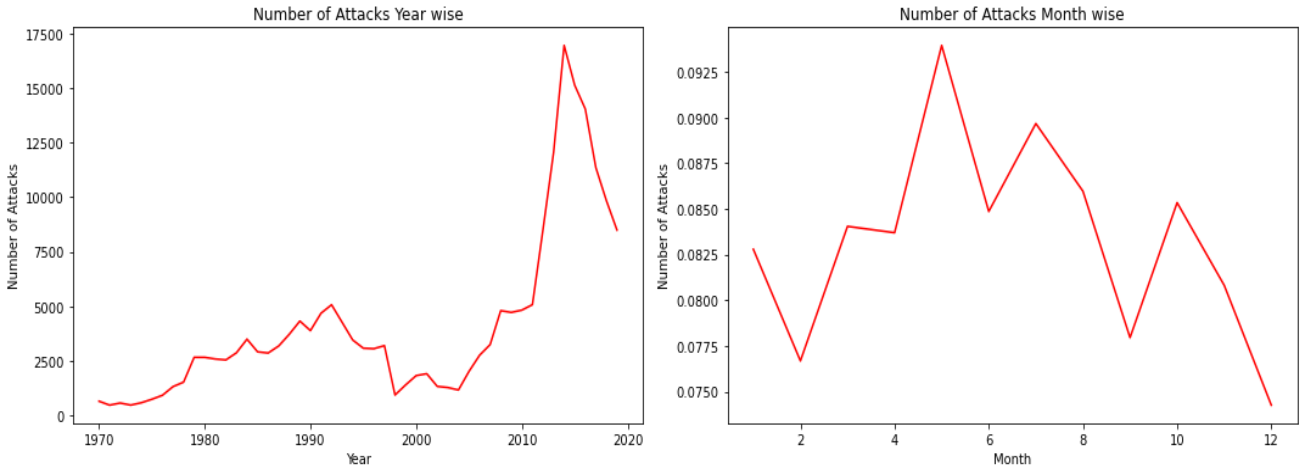
EDA Part 1: On Global Database (considering all countries)

The world map shows the global scenario of active terrorist groups.



We can see the concentration of certain terrorist groups in countries like Iraq, Pakistan, Afghanistan, India and so on.

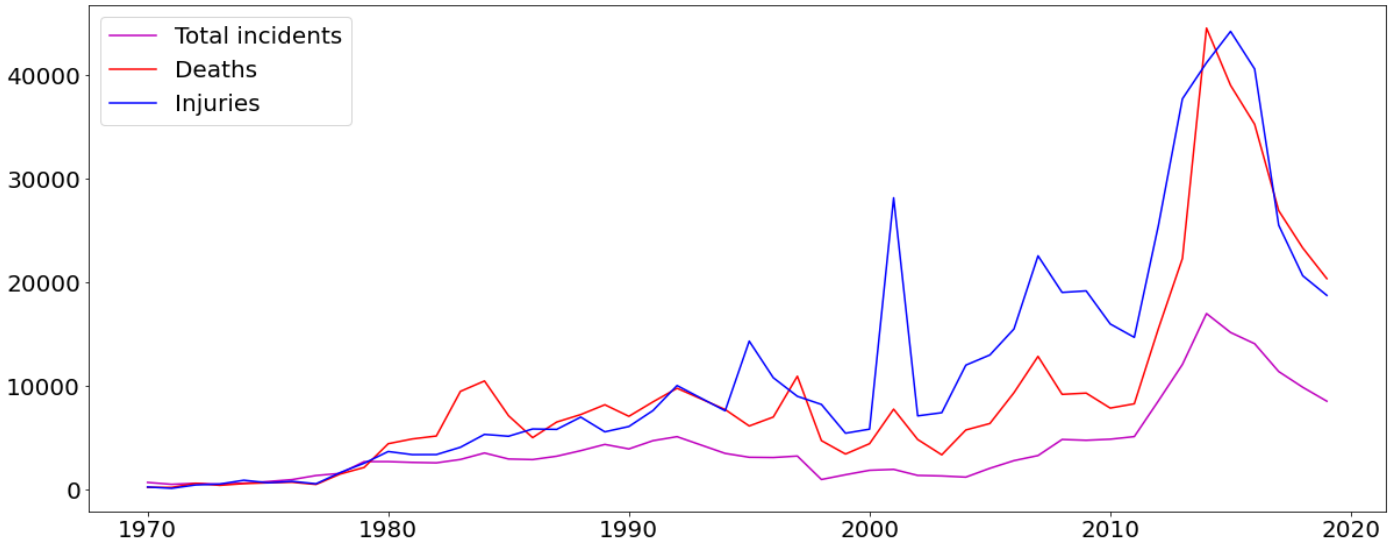
Figure 1: Displays the number of attacks year-wise and month-wise



Insights from Figure 1:

- The above plot shows that the number of global terrorist attacks increased gradually till around the year 1990 and then decreased slightly with minimum in the year of 1998.
- Then the count increased sharply after 2010 to a great extent and reached its peak in 2014 and decreased sharply after that. The data is available from 1970-2019.
- Month wise average number of attacks is fluctuating a lot across the months. The count has become maximum at month 5 (May).

Figure 2: Displays the year-wise total number of terrorist attacks, deaths and injuries.



The above figure shows the global terrorism data on total number of attacks, injuries and fatalities year-wise. The trend is increasing over the years with fluctuations for all the three line-plots. Throughout, we can see that number of injuries and deaths have been quite high except the initial years; where number of injuries have been maximum across most of the years. Whenever, any global terrorist activity occurs, number of deaths and injuries are huge.

The major observations that are made:

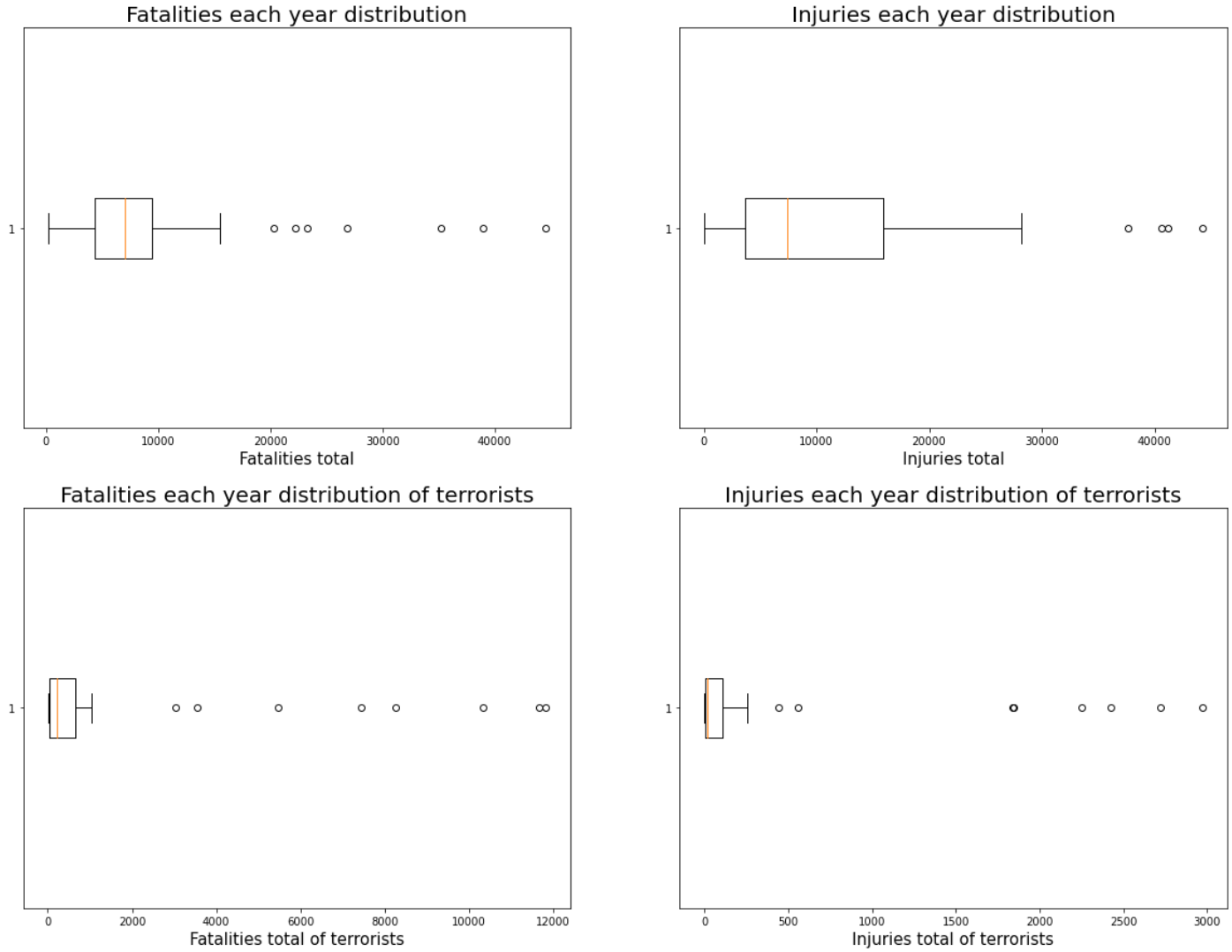
1. In 1984, the total number of deaths increased to a great extent as compared to the number of injuries and the number of incidents.

2. For the year 1995, the number of injuries and for 1997, the number of deaths have taken a sudden increase.
3. For the years 2001,2007,2013-2016, the counts of injuries and deaths have been significantly high.

Refer to the following table to get an idea on the significant years mentioned above.

Years	Attacks	Injuries	Deaths
1984	3495	5291	10450
1995	3081	14295	6103
1997	3198	8968	10906
2001	1912	28137	7727
2007	3247	22531	12825
2013	12045	37690	22280
2014	16959	41177	44524
2015	15133	44204	38993
2016	14046	40576	35236

Figure 3: Boxplots of fatalities and injuries of common people and fatalities and injuries of terrorists year-wise



All the above boxplots are positively skewed and many outliers with the following observations: -

- Fatalities each year distribution:

Mean= 9311.20, median= 7036.0, SD= 9866.0, Range= 44524.0 – 173.0 =44351, IQR= 5053

- Injuries each year distribution :

Mean= 11627.41, median= 7384.0, SD= 11511.86, Range= 44122-82 =44122, IQR= 12316

- Fatalities each year distribution of terrorists:

Mean= 1460.6, median= 206.0, SD= 3092.16, Range= 11817-4 =11813, IQR= 630

- Injuries each year distribution of terrorists:

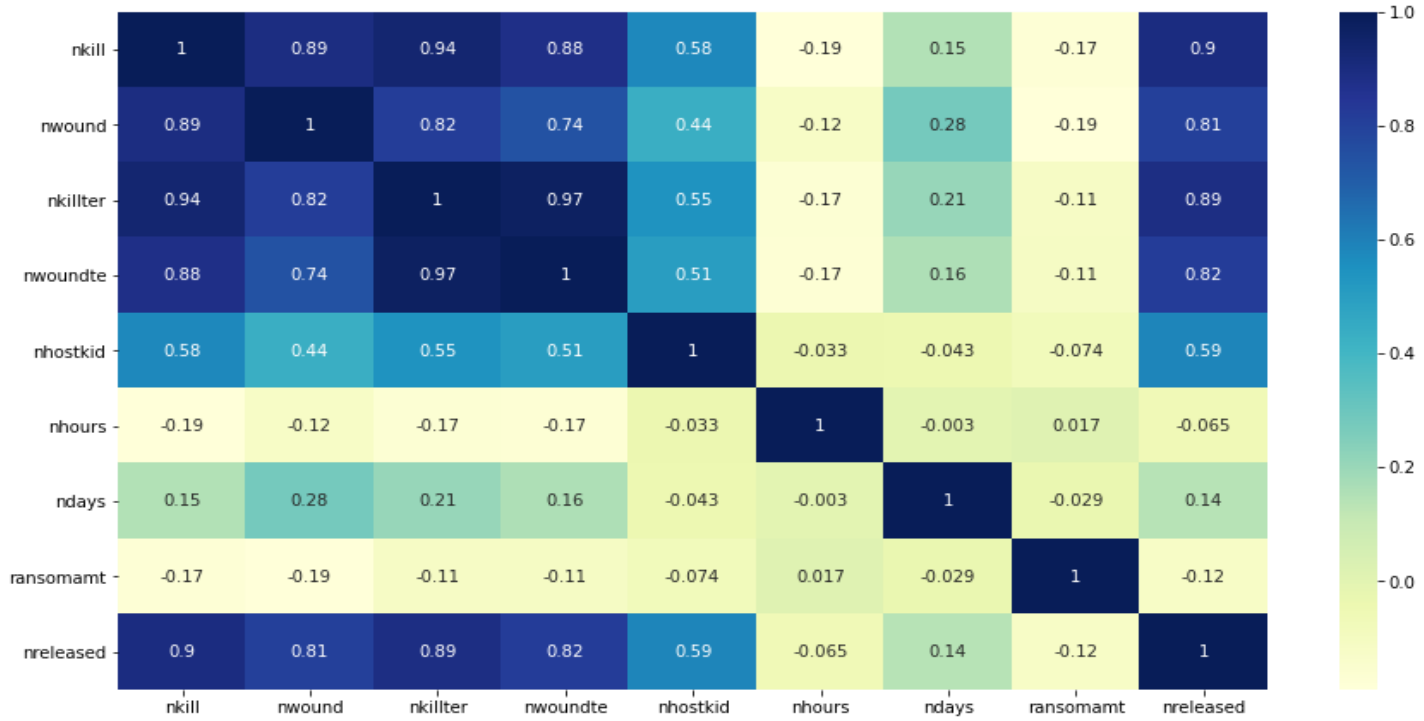
Mean= 336.59, median= 16.0, SD= 778.90, Range= 2970-0 =2970, IQR= 105

From the above box plots we can clearly see that the mean fatalities of common people is much higher than the mean fatalities of the terrorists. Each of these attacks takes around 9400 lives of innocent people. The standard deviations are quite high indicating that there is a huge spread of the fatalities and injuries distribution of common people over the years.

50% of the fatalities of common people lie in the interquartile range of 5053 which is much higher than the IQR of terrorists' fatalities which is 630.

Average injuries and IQRs are also very high for common people than the terrorists.

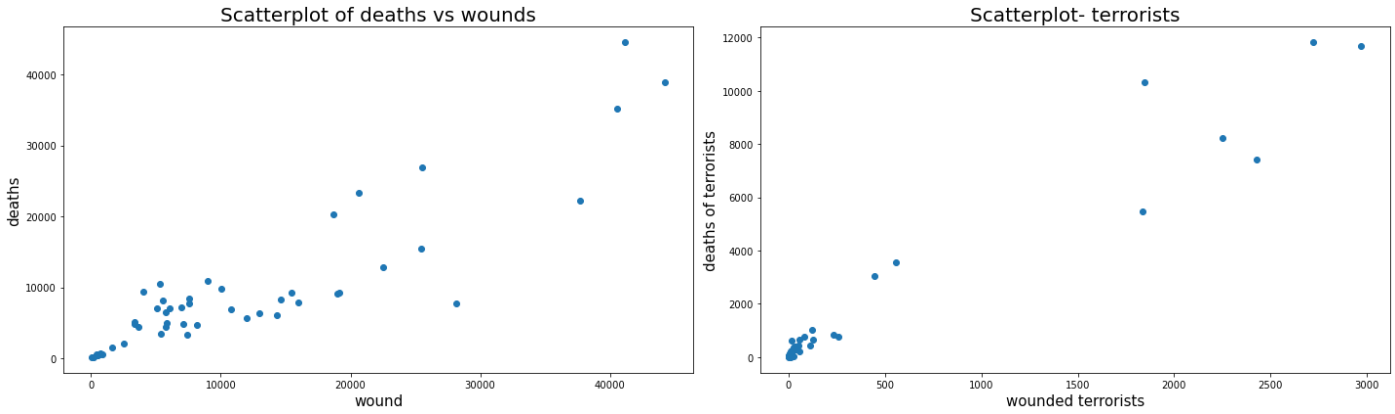
Figure 4: Correlation Matrix (Heatmap)



Insights:

- Number of people killed and number of wounded have high correlation. Also, number of terrorists killed and wounded and number of common people killed are highly correlated.
- Number of common people killed are also highly correlated with the number of people released on being kidnapped.
- A similar high correlation is observed with number of people wounded and number of dead terrorists and number of kidnapped released.
- Number of terrorists killed and the number of terrorists wounded are highly correlated. The number of terrorists killed is highly associated with the number of kidnapped people released.

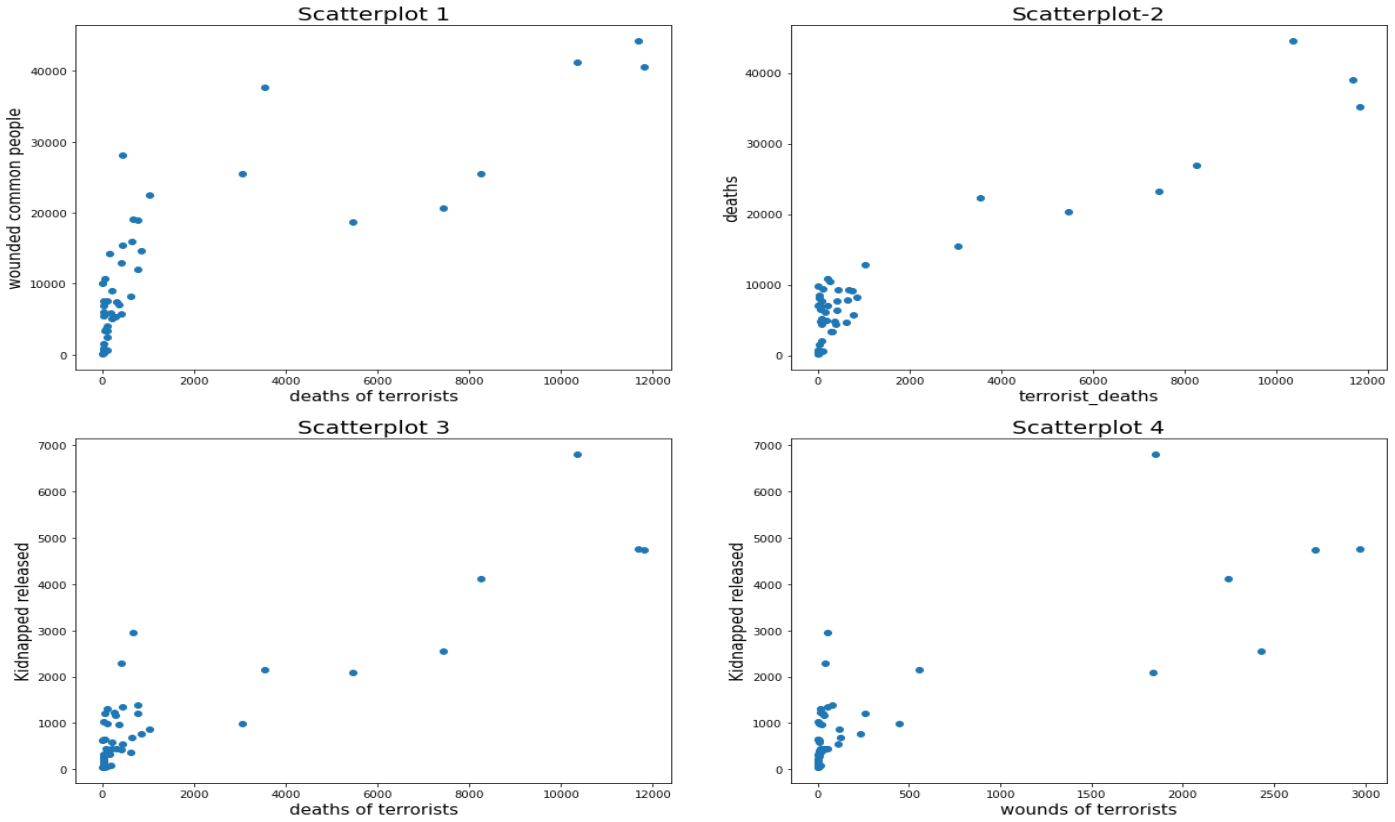
Figure 5: Scatterplots…1



There are outliers in each of the scatterplot.

i. From the first scatterplot we can see the pattern for deaths and wounded. There is a positive association between the two variables.
ii. There is a positive relationship between deaths of terrorists and wounded terrorists.

Figure 6: Scatterplots…2



All the four scatterplots show positive association between the x and y variables. Each scatterplot depicts presence of outliers.

We can infer that, wounded people and terrorist deaths are positively related and so are the terrorist deaths and common people deaths. As more terrorists are wounded or died, they released more kidnapped people.

Inference from adjacent Pie chart:

Figure 7: Weapon-Type: Pie Chart….1

Mostly the weapons used were explosives, firearms, incendiary and Melee (A melee or pell-mell is disorganized hand-to-hand combat in battles fought at abnormally close range with little central control once it starts.)

Since the values apart from the Explosives, Firearms, Incendiary, and Melee are very small, they have been combined together into "Others" category and a pie plot has been obtained for the same.

“Others” include weapon types like chemical, biological weapons, etc.

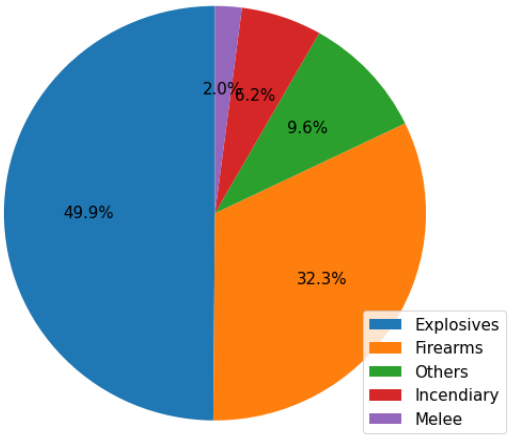
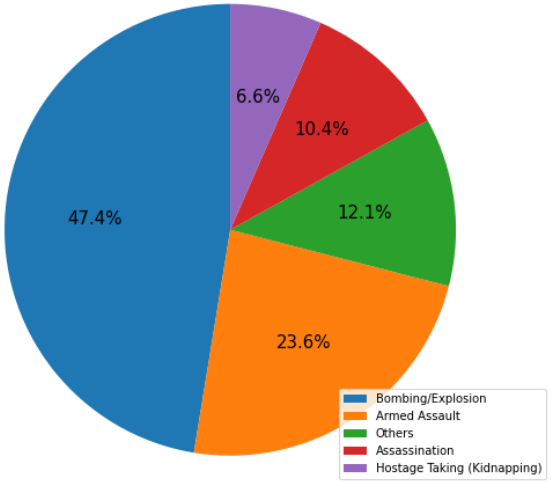


Figure 8: Attack-Type: Pie Chart ….2

Inference from adjacent Pie Chart:

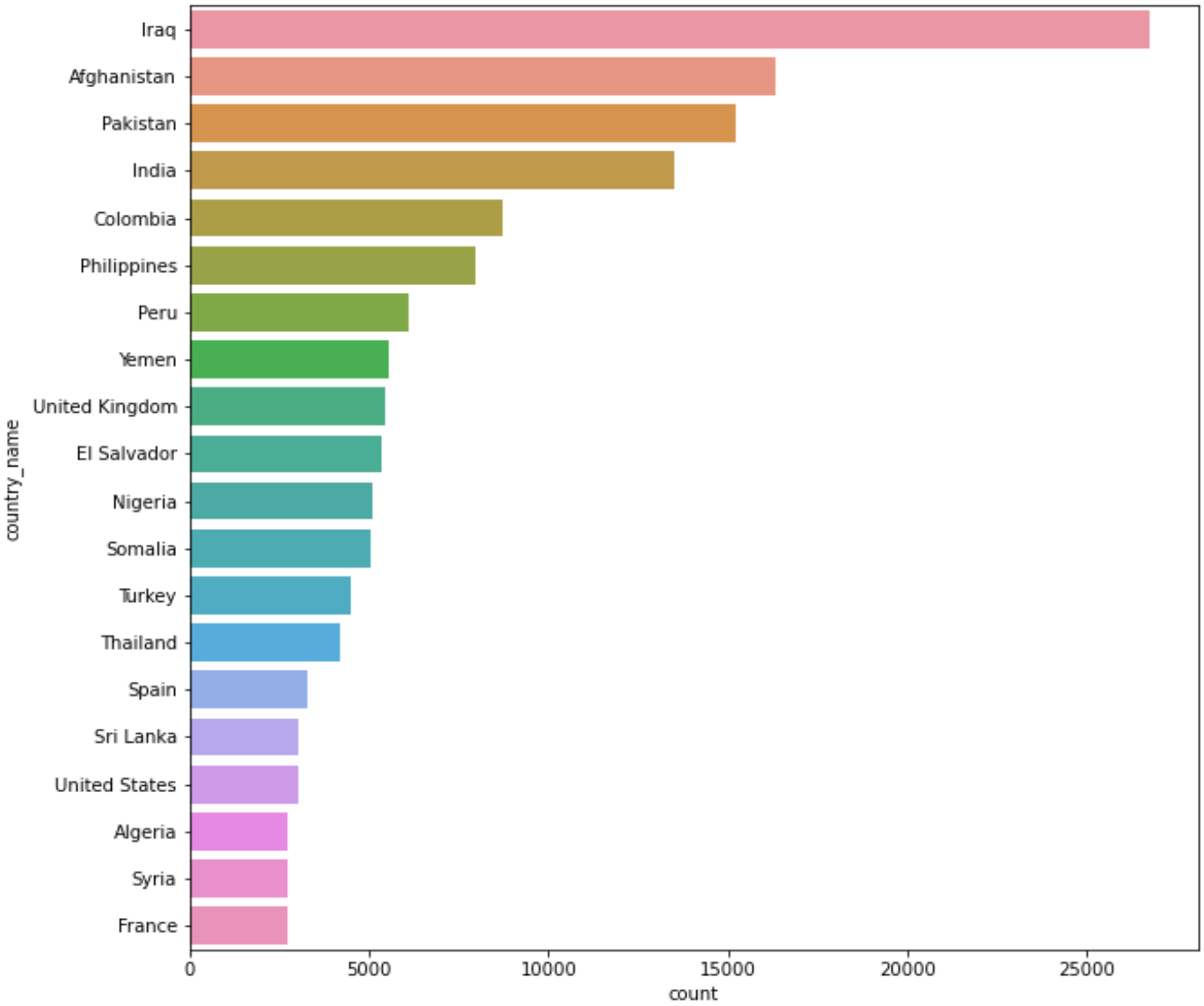
A similar kind of analysis has been done to understand the nature of attacks that were launched by the terrorists. From the pie chart, it is clear that mostly bombing/explosion is used, followed by armed assault and so on.

“Others” include attack types like unarmed assault, hijacking, etc.



EDA Part 2: Considering the top 5 countries

Figure 9: Bar Chart showing the number of terrorist attacks country wise

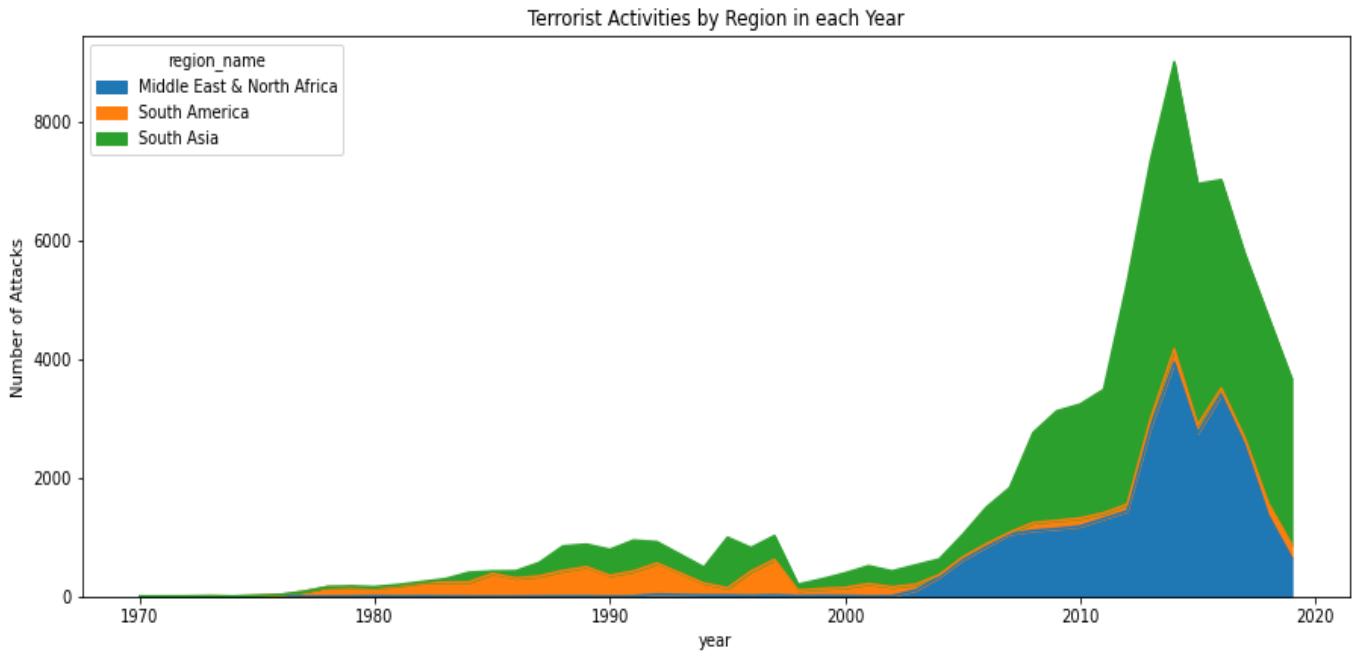


Country-wise the maximum number of attacks have occurred in Iraq, followed by Afghanistan, Pakistan, India, Columbia, Philippines, Peru and so on.

The bar plot shows the total number of attacks across all the years from 1970 to 2019. The minimum number of attacks (that is equal to 1) are faced by some countries like the Vatican City, South Vietnam, New Hebrides, etc.
Let us consider the top 5 countries that have had the maximum number of terrorist attacks from 1970 to 2019.

We have focussed our analysis on the top 5 vulnerable countries and then looked at the condition of India since it is in the 4th position from the top which is of course a matter of concern for all Indians.

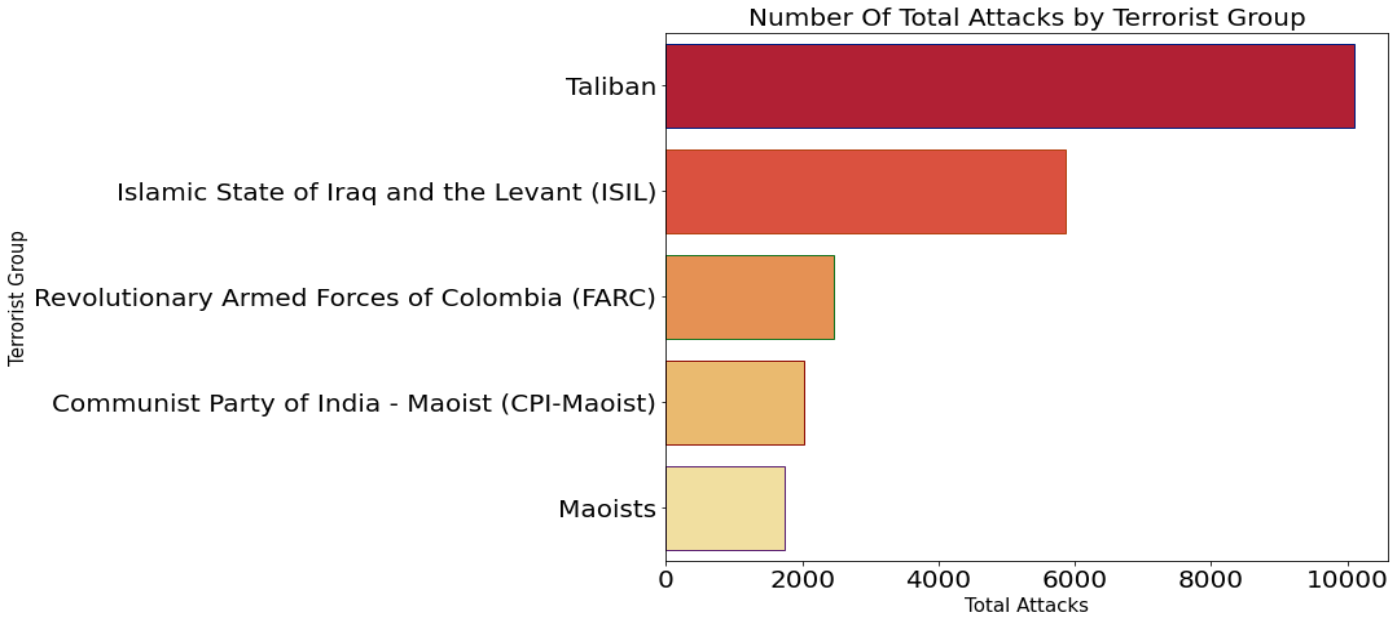
Figure 10: Stacked Area Chart



The stacked area plot shows the trend across the years which is an increasing one with certain fluctuations for different regions along with the total shown in form of the area occupied by each region.

We have Afghanistan, Pakistan and India in the region of South Asia which occupies the maximum area depicting those maximum attacks have occurred in this region. Next comes Iraq for the Middle East and North African region followed by Colombia in the South American region.

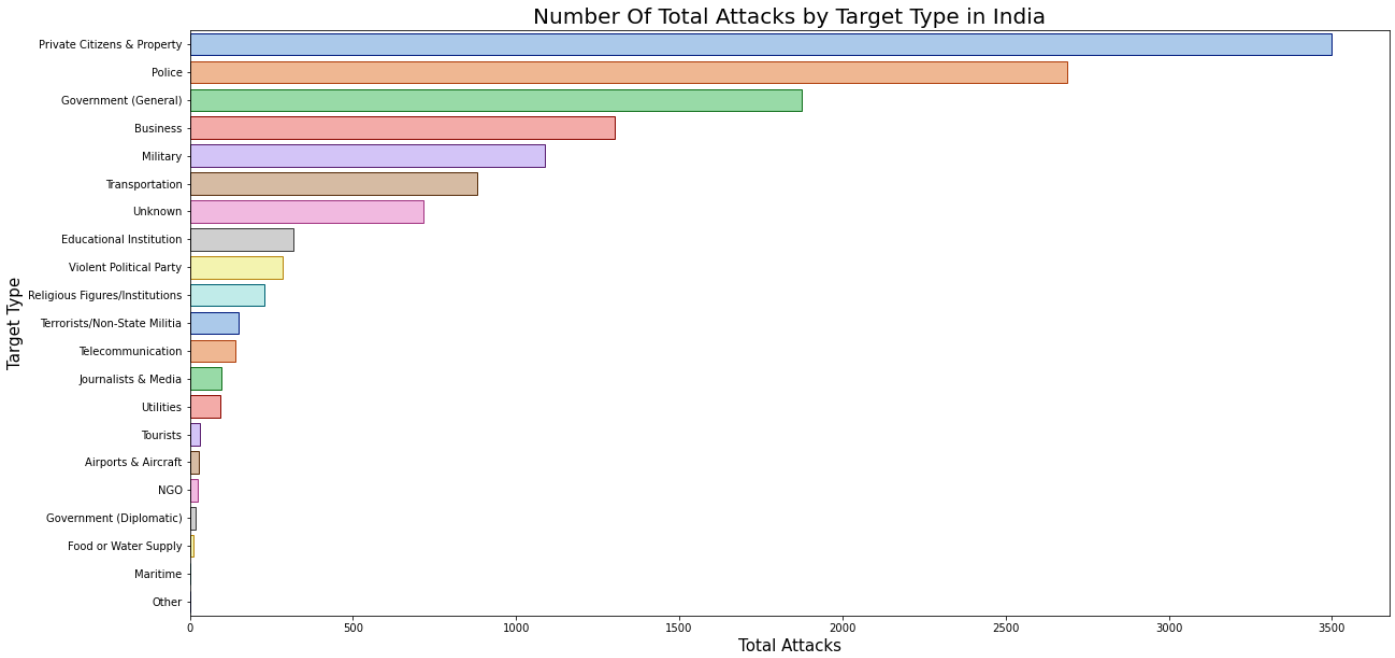
Figure 11: Bar Chart showing the top terrorist groups



The bar plot displays that among all the other terrorist groups, Taliban has launched the maximum number of attacks.

EDA Part 3: Analysing the situation in India

Figure 13: Bar Chart depicting the Target Types attacked in India



The above bar plot depicts the target types of the attacks launched in India. Mostly the attacks were on Private Citizens and Property, police, government, business and so on.

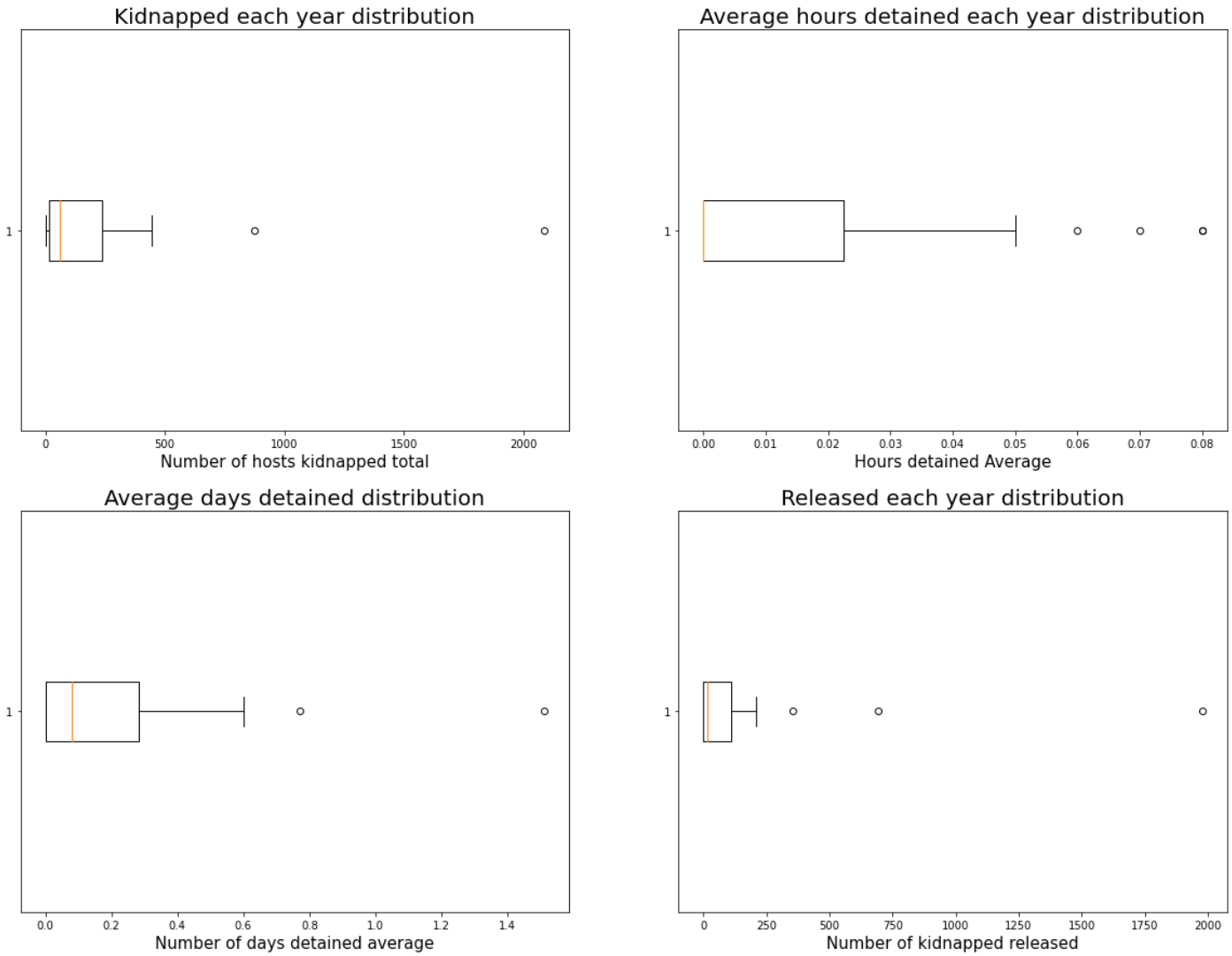
There are many terror groups active across the India.

Some of the famous among them are CPI-Maoist,ULFA,Hizbul Mujahideen,Sikh Extremist etc.

Table 3: Terrorist Groups

Terrorist Group	Total Attacks
Communist Party of India - Maoist (CPI-Maoist)	2036
Maoists	1745
Sikh Extremists	710
United Liberation Front of Assam (ULFA)	379
Hizbul Mujahideen (HM)	269

Figure 14: Boxplots of kidnapped each year, average hours detained, average days detained and number of kidnapped people released distribution each year in India

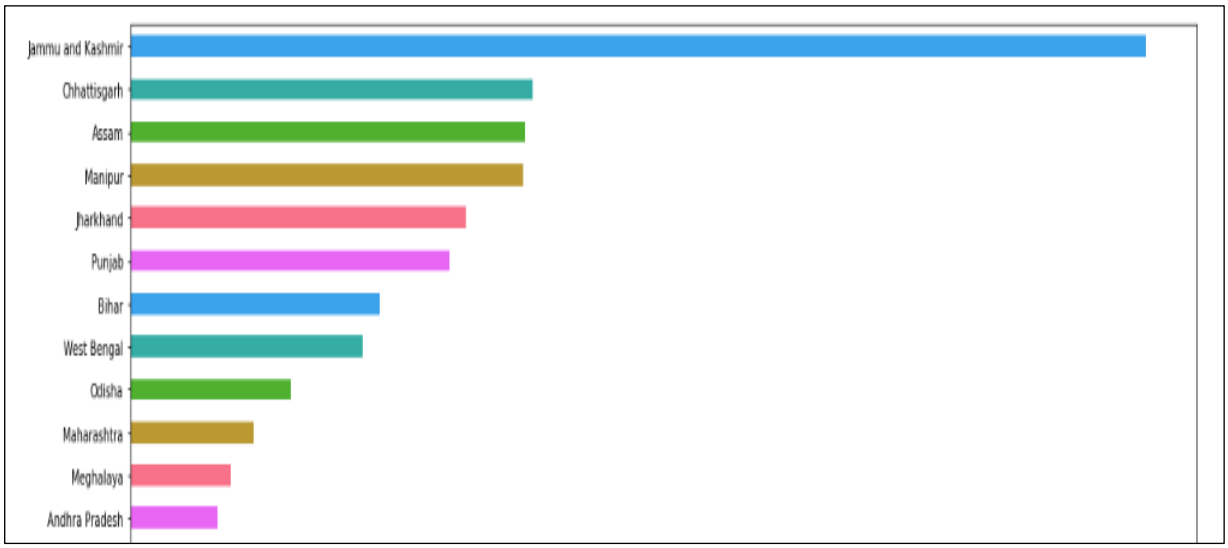


All the above boxplots are positively skewed and many outliers with the following observations:-

- Kidnapped each year distribution:
  Mean= 180.568182, median= 60.5, SD= 339.482260, Range= 2086-0 =2086, IQR= 221.75
- Average hours detained each year distribution:
  Mean= 0.014, median= 0, SD= 0.023, Range= 0.08-0=0.08, IQR= 0.0225
- Average days detained each year distribution:
  Mean= 0.18, median= 0.08, SD = 0.28, Range= 1.51-0=1.51, IQR= 0.28
- Number of kidnapped released each year distribution:
  Mean= 114.93, median= 14.00, SD= 313.06, Range= 1976-0=108.5, IQR= 108.5

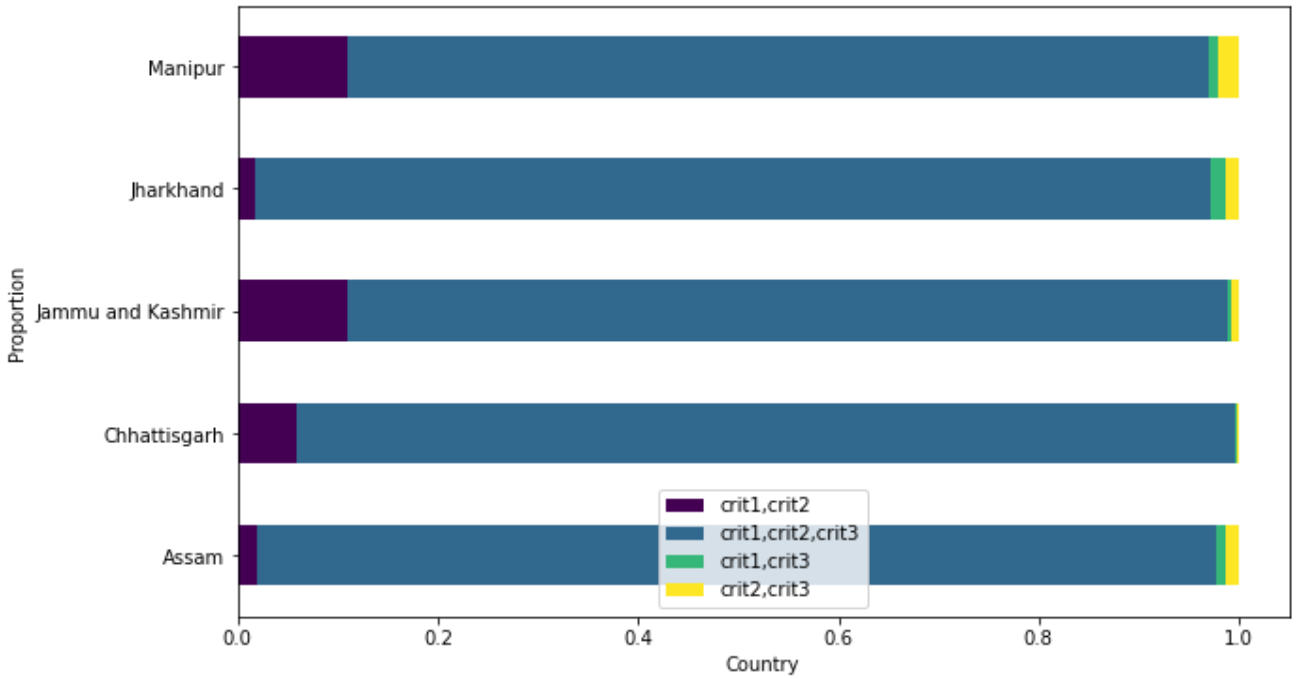
On an average, 181 people were kidnapped each year as a result of terrorist attacks in different states of India and they were detained on an average of less than 1 day (0.18). On an average, 115 people kidnapped were released by the terrorists.

Figure 15: Top Indian States Attacked shown using Bar Chart



Major terrorist activities have occurred in the States Jammu and Kashmir, Chhattisgarh, Assam, Manipur, Jharkhand and so on. For Jammu and Kashmir, the number of terrorist activities are very high in comparison to the other states.

Figure 16: 100% Stacked Bar Chart showing the Causes of attacks in the Indian States



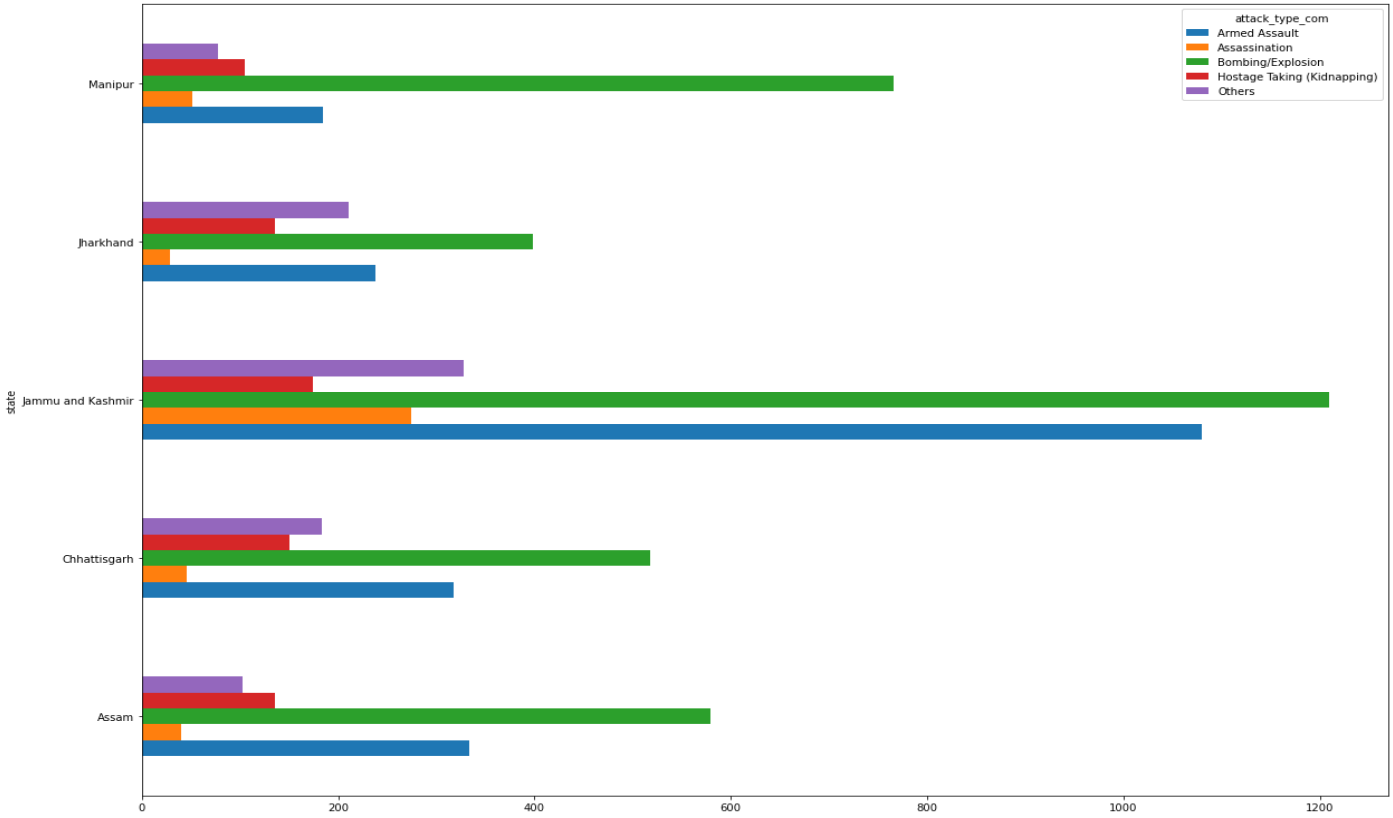
Criterion 1: POLITICAL, ECONOMIC, RELIGIOUS, OR SOCIAL GOAL (CRIT1): The violent act must be aimed at attaining a political, economic, religious, or social goal.

Criterion 2: INTENTION TO COERCE, INTIMIDATE OR PUBLICIZE TO LARGER AUDIENCE(S) (CRIT2): To satisfy this criterion there must be evidence of an intention to coerce, intimidate, or convey some other message to a larger audience (or audiences) than the immediate victims.

Criterion 3: OUTSIDE INTERNATIONAL HUMANITARIAN LAW (CRIT3): The action is outside the context of legitimate warfare activities, insofar as it targets non-combatants.

In the top 5 Indian States as well, we can observe that most of the attacks have occurred as a result of all the three criteria.

Figure 17: Composite Bar Chart showing the attack types



Weapon type used was mostly explosives for all states except Jammu and Kashmir where both firearms and explosives dominate.

Next, let us see which are the top cities in Jammu and Kashmir that has experienced nearly 22% of the total terrorist attacks in India from 1970 to 2019.

Table 4: Cities in Jammu & Kashmir mostly

Srinagar	23.82
Sopore	4.00
Anantnag	2.89
Jammu	2.08
Tral	1.88

The mostly affected cities in the state of Jammu and Kashmir are Srinagar, Sopore, Anantnag, Jammu and Tral. The descriptive statistics for the suicides in the state is displayed in the table below.

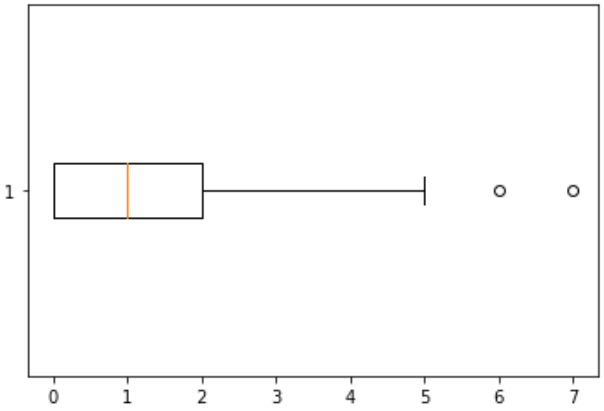
Table 5: Descriptive Statistics of suicide attacks year-wise

mean	1.48
std	1.98
min	0.00
25%	0.00
50%	1.00
75%	2.00
max	7.00

Table 6: Descriptive Statistics of suicide attacks year-wise in India

mean	1.48
std	2.13
min	0.00
25%	0.00
50%	0.00
75%	2.25
max	8.00

Figure 18: Boxplot of suicide attacks in Jammu and Kashmir



The above box plot and table shows the distribution of the year-wise suicidal attacks in Jammu and Kashmir. Table 6 also shows the descriptive statistics for suicide attacks in India as a whole. We can compare the two and see that the major suicide attacks in India have occurred in Jammu and Kashmir. The mean number of suicide attacks is around 2 with the maximum number going to 7 for Jammu and Kashmir as can be seen from the box plot.

CONCLUSION

Some key insights from our analysis include:-

- Global terrorism expanded rapidly post (2001), and the center of terrorism has shifted from the West to the Middle East and in South Asia, it has sharply increased.
- The motives of terrorists are political, and emanate from groups who desire political or religious change or revenge, yet lack the political power to enforce the change through any means other than violence. The number of attacks reached its peak in the year 2014.
- The number of killings, and wounded has increased significantly after 2010. Throughout, we can see that number of injuries and deaths has been quite high whenever any global terrorist activity occurs.
- Most of the time explosive has been used as weapons and attacks type was bombing/explosion.
- The top four countries facing terrorism are Iraq, Afghanistan, Pakistan, and India and the major terror groups are Taliban, ISIL, FARC, Maoists, etc.
- Private citizens were not always the target of terrorists worldwide — in some cases, police, government, military targets, and infrastructure were also vulnerable.
- In India major terrorist activities have occurred in the States of Jammu and Kashmir, Chhattisgarh, Assam, Manipur, and Jharkhand. People were kidnapped with an average number of 181 year-wise as a result of any terrorist activity.
- For Jammu and Kashmir we can see firearms were used more than explosives as compared to the other 5 states where explosives were mostly used.
- The highly affected cities in the state of Jammu and Kashmir are Srinagar, Sopore, Anantnag, Jammu, and Tral and suicidal attacks are also prevalent in the state with the highest number reaching 7 year-wise and that shows us the increase in radicalization.

SOME THOUGHTS

We believe that terrorism cannot be the solution to any problem but only leads to meaningless loss of life, damage to property, mental trauma and many more. In a world, where we are still fighting major life-taking diseases like COVID-19, cancer, etc.; and where there are other serious issues to deal with which take millions of lives, it absolutely makes no sense for humankind to take to arms and fight against each other. We consider ourselves to be the most educated and the smartest species on Earth, however, when incidents like terrorism occur, it completely shatters hope for the good. Through this analysis, we could study the major factors leading to terrorism, the year-wise terrorist incidents, the injuries, targeted violence, etc. Unfortunately, we as aspiring data scientists, have to deal with facts and data and cannot possibly come up with an immediate solution to stop terrorism. We can only hope to establish peace and stop terrorism which would create a peaceful society for us and for the generations to come.

REFERENCES

- https://www.start.umd.edu/gtd/
- https://ourworldindata.org/terrorism
- https://reliefweb.int/report/world/global-terrorism-index-2022