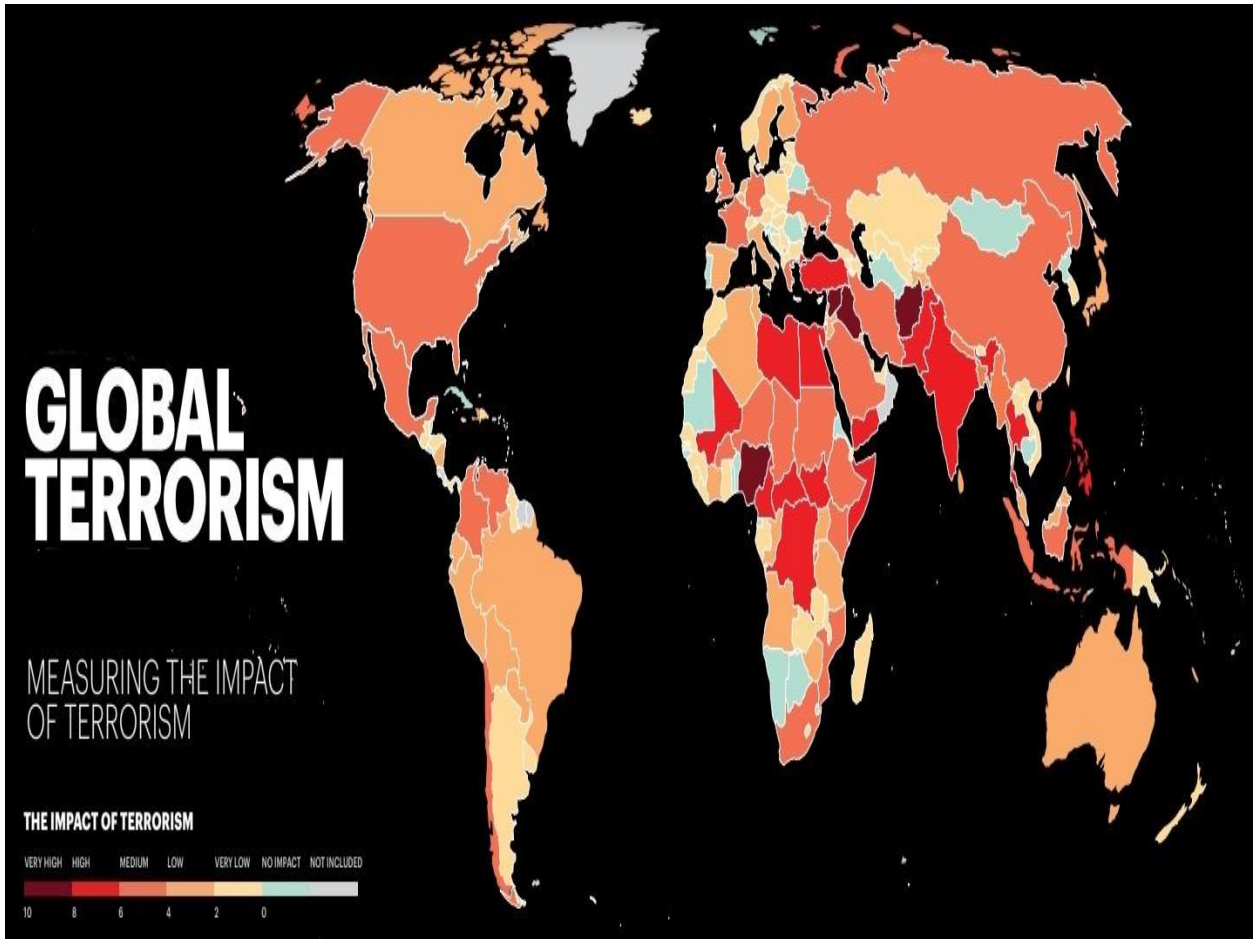
Exploratory Data Analysis on Global Terrorism Data



Submitted by

Kanike Lakshmi Narayana

Abstract

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 and now includes more than 180,000 attacks. The database is maintained by researchers at the National Consortium for the Study of Terrorism and Responses to Terrorism (START), headquartered at the University of Maryland. Explore and analyse the data to discover key findings of terrorist activities.

Problem Statement

The Global Terrorism Database (GTD) is the most comprehensive unclassified database of terrorist attacks in the world. The National Consortium for the Study of Terrorism and Responses to Terrorism (START) makes the GTD available via this site to improve understanding of terrorist violence so that it can be more readily studied and defeated.

Introduction

Terrorism, as we know is a global menace that has been present across the globe for a long time as a means of coercion.

Global Terrorism Database is an extensive collection of such acts around the world from 1970 to 2019. The dataset used here has been updated until 2017. This dataset has been digitised and maintained by The National Consortium for Study of Terrorism and Response to Terrorism.

The GTD defines a terrorist attack 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.

Objective

The main objective is to explore and analyse the data to discover key findings of terrorist activities.

Libraries imported

Libraries that are used in the analysis.

- Pandas: Mostly used to load the dataset and for cleaning the dataset.
- Numpy: Used for warping the dataset.

- Seaborn: Used for the visualization of the data.
- Matplotlib.pyplot: Used for the visualization of the data.
- Warnings: To stop showing warnings in the output.
- Plotly.express: Used for visualization.
- WordCloud: Used for visualization.

Handling of the Dataset

The given Global Terrorism Dataset contains many columns, the greatest challenge I faced is which columns are to be used and when to move on, in which I used only 16 columns for analysis.

- The dataset contains the data of terrorist attacks from 1970 to 2017 containing over 180000 observations.
- Renaming the columns and making a separate data frame with 16 columns.
- Made a heatmap for checking the presence of NaN values in the dataset.
- Replaced the NaN values with Unknown and “0” in Categorical and numerical variables, respectively.
- Comparing the incidents within regions and countries.

Columns used in the analysis

- eventid: consists of a unique number to all the incidents.
- iyear: renamed as year, contains the year in which the incident took place.
- imonth: renamed as month, contains the month number in which the incident took place.
- iday: renamed as day, contains the day in which the incident took place.
- country_txt: renamed as country, contains the name of the country in which the incident occurred.

- region_txt: renamed as region, contains the name of the region in which the incident occurred.
- provstate: renamed as state, contains the name of the state in which the incident occurred.
- city: contains the name of the city in which the incident occurred.
- latitude: states the latitude point of the place of incident.
- longitude: states the longitude point of the place of incident.
- success: informs that the incident is successful or not.
- suicide: states that the person who handled the event did commit suicide or not.
- attacktype1_txt: renamed as attack_type, contains the type of attack used by terrorists for the incident.
- targettype1_txt: renamed as target_type, contains the main target of the group.
- targetsubtype1_txt: represents the sub-target during the incident.
- claimed: indicates the group that caused the incident has claimed the responsibility of the attack or not.
- motive: consists of the motive behind the attack.
- gname: renamed as an organization, mentions the name of the group responsible for the event.
- doubtterr: renamed as doubt_terror, represents the consequences recognised as might be a terrorist attack.
- weapontype1_txt: renamed as weapon_type, renamed as weapon_type, represents the type of weapon used in the incident.
- nkill: renamed as perpetrator_killed, represents the number of people who died due to the incident.
- nwound: renamed as perpetrator_wounded, represents the number of people who got injured due to the incident.
- ransomamt: renamed as ransom_amount, represents the ransom amount demanded by the group in the incident.

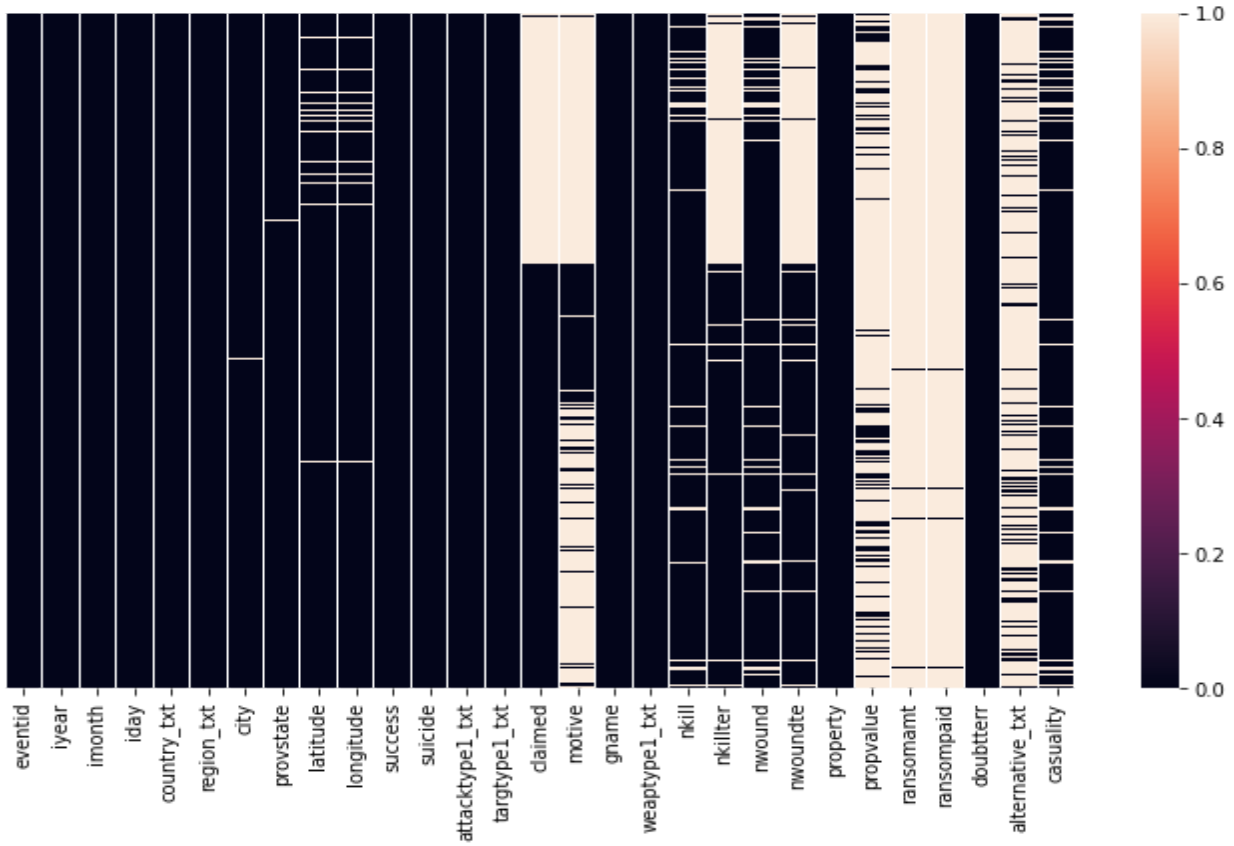
- ransompaid: renamed as ransom_paid, represents the ransom paid against the ransom amount demanded by the group in the incident.
- property: states that the event took place in any public property like malls, markets.
- propvalue: if the property column contains yes then the equal value in terms of us. dollars will be mentioned.

Approach

To give the scope of the project we have dived into the EDA with the dataset. It is presented in the tabular format and also in the form of visualization with all the information of the dataset.

- Global at glance
- South Asia at a glance
- India at glance

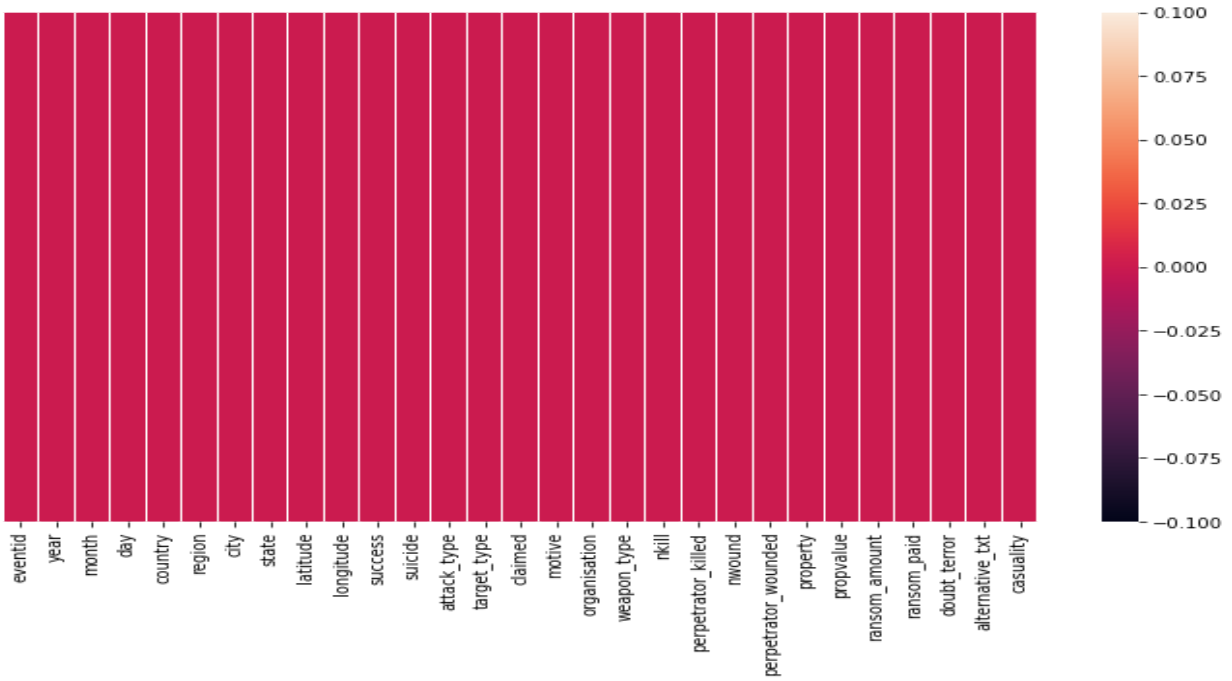
Data Cleaning



The above figure clearly shows the gaps in between the heat map which means the presence of NaN values in the dataset.

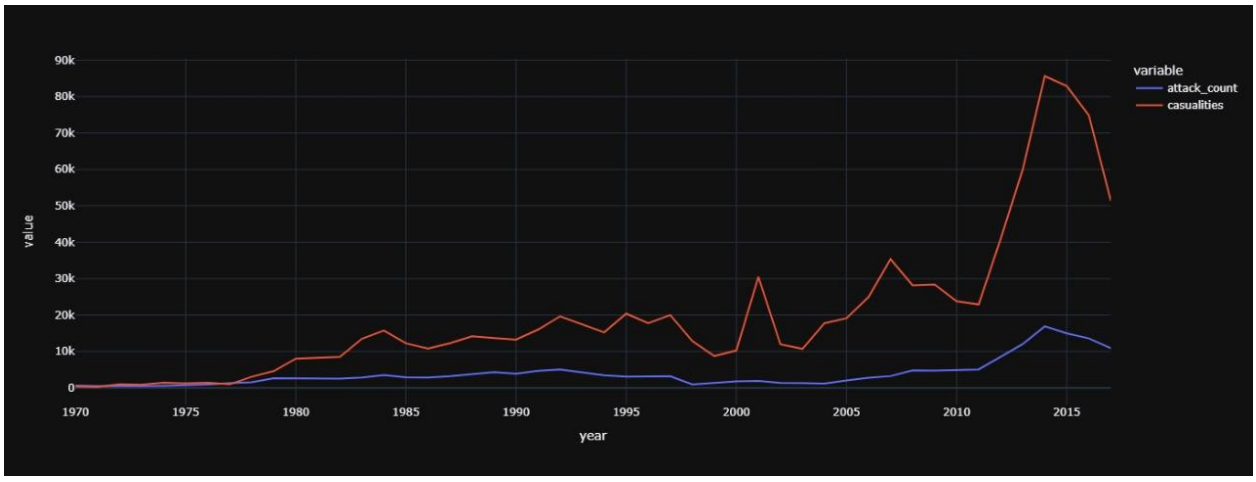
Filling the nan values

- For categorical variables, the nan values are replaced with unknown.
- For numerical variables, the nan values are replaced with zero(0).



The above figure shows the null values are filled and the names of the columns are renamed for analysis purposes.

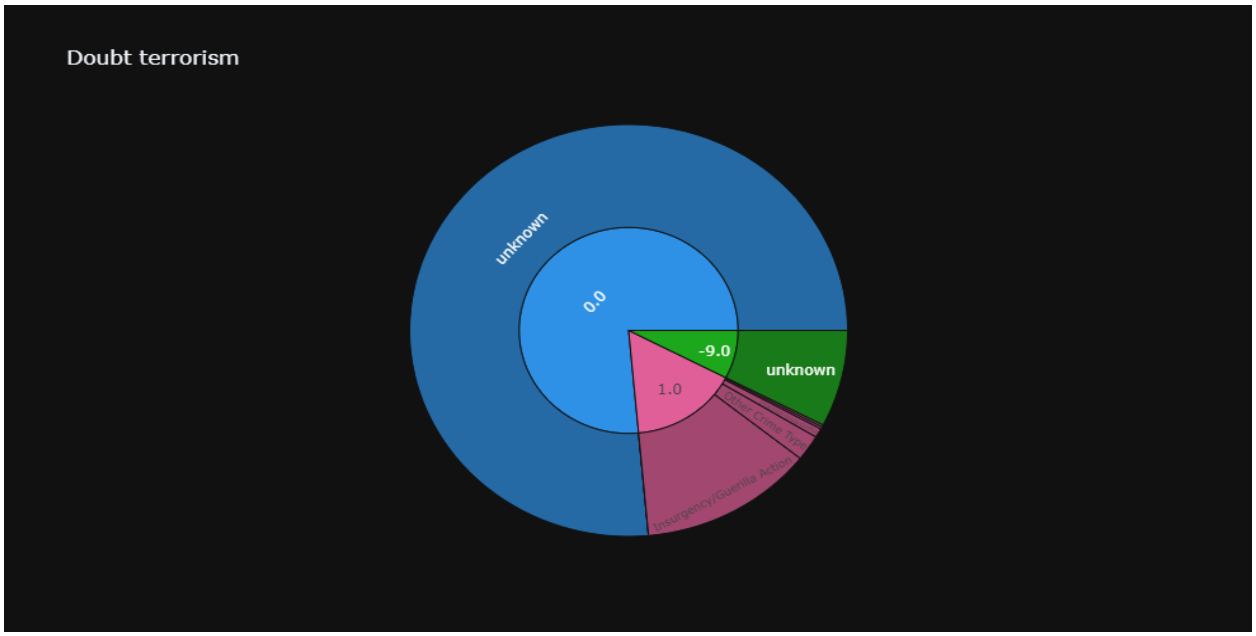
Year of Attacks



It's evident from the figure that the terrorist attacks are quite concentrated in a handful of countries like Afghanistan, Pakistan, Iraq, India etc. The Middle East & North Africa is the region with the most casualties, followed by South Asia, Sub-Saharan Africa, South America and so on. Iraq is the country the most affected by terrorism. Casualties: 213,279. The attack frequency over the world peaked in the year 2014: 16903 attacks The casualties due to terrorism peaked in the same year: 85618 casualties.

There was a slight downfall in the year 1998 which had attacks of 934 and started increasing in the count of consequences. After reaching a high range during 2014 there is again the downfall in the count of incidents.

Doubt Terror



There are instances when some incidents do not match any two of the 3 criterions set by START to qualify as a terrorist attack. Such incidents are marked 1 in the 'doubtterr' field in the data set, implying that there is some amount of doubt as to whether this was an act of terrorism. These attacks are further classified in the field alternative_txt.

- 138,906 attacks are such that there is no essential doubt in them being terrorist attacks.
- 29001 attacks are such that they can't be classified as terrorist attacks under START criterions. They are classified into follows:
- Insurgency/ Guerilla Action: 23,408

- Lack of intentionality: 319
- State Actors: 321
- Intra/Intergroup conflict: 1296
- Another crime type: 3657

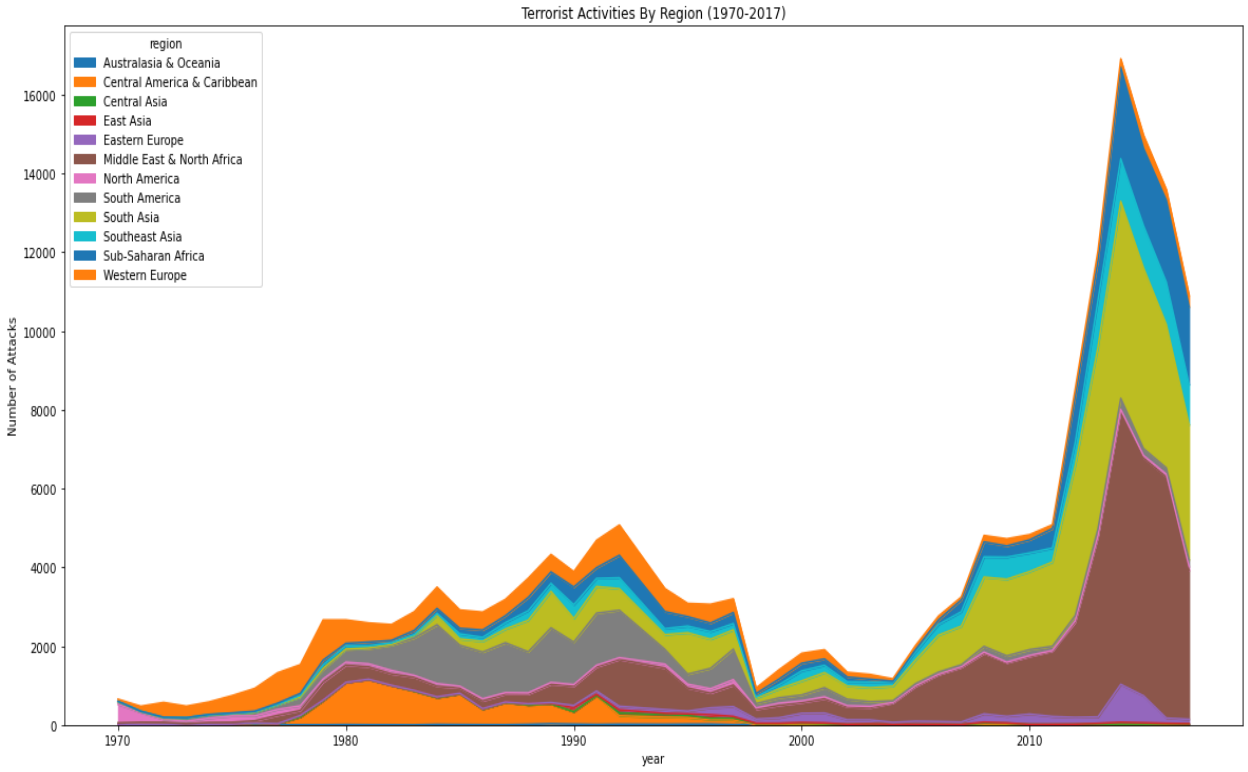
Organizations responsible for events

The Unknown faction contributes the most to the number of attacks over the world. These could be the attacks, not claimed by any organizations, work of individuals or groups of individuals who do not belong to any formal organization. Taliban has the upper hand in terms of an absolute number of attacks all over the world.

Mostly affected States

It is evident from the wordcloud that Baghdad in Iraq is the most affected state in the world. Ireland, Balochistan, Saladin, Khyber, Nineveh, Jammu and Kashmir etc are a few other notable names that pop up.

Attacks Vs Region Vs Year

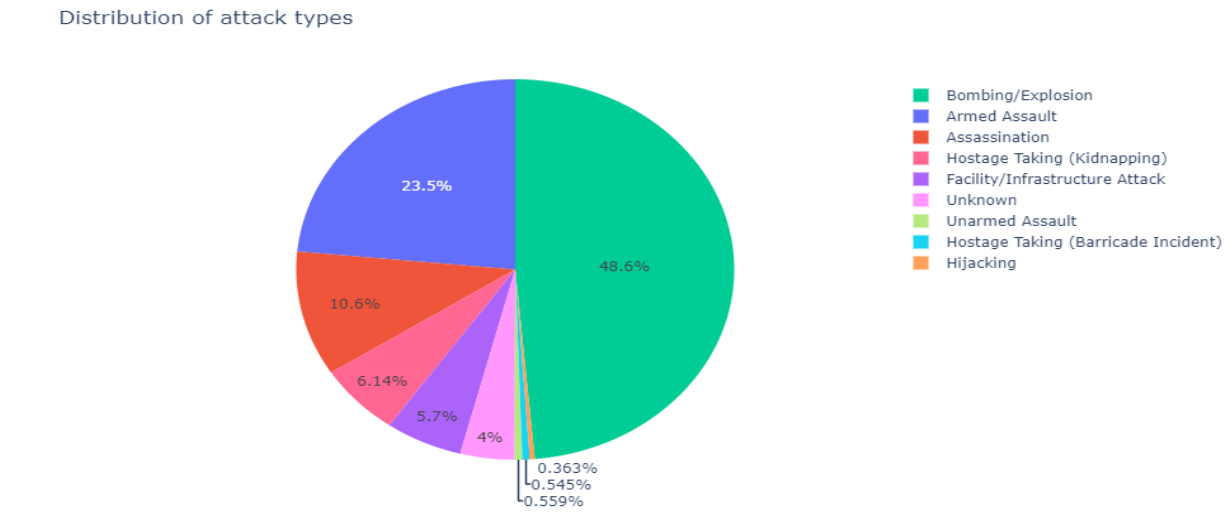


region	total_deaths	perpetrator_killed	total_wounded	perpetrator_woundec	Attacks
Middle East & North Africa	137642	19788	214308	2020	50474
South Asia	101319	21342	141360	8161	44974
South America	28849	789	16704	69	18978
Sub-Saharan Africa	78386	12970	52857	736	17550
Western Europe	6694	250	18332	63	16639
Southeast Asia	15637	1233	26259	430	12485
Central America & Caribbean	28708	246	8991	21	10344
Eastern Europe	7415	1268	12045	474	5144
North America	4916	85	21531	37	3456
East Asia	1152	252	9213	34	802
Central Asia	1000	55	2009	9	563
Australasia & Oceania	150	13	260	7	282

As confirmed before, the Middle East and North Africa followed by South Asia, South America are the most attacked regions. Central and North Americas lead the

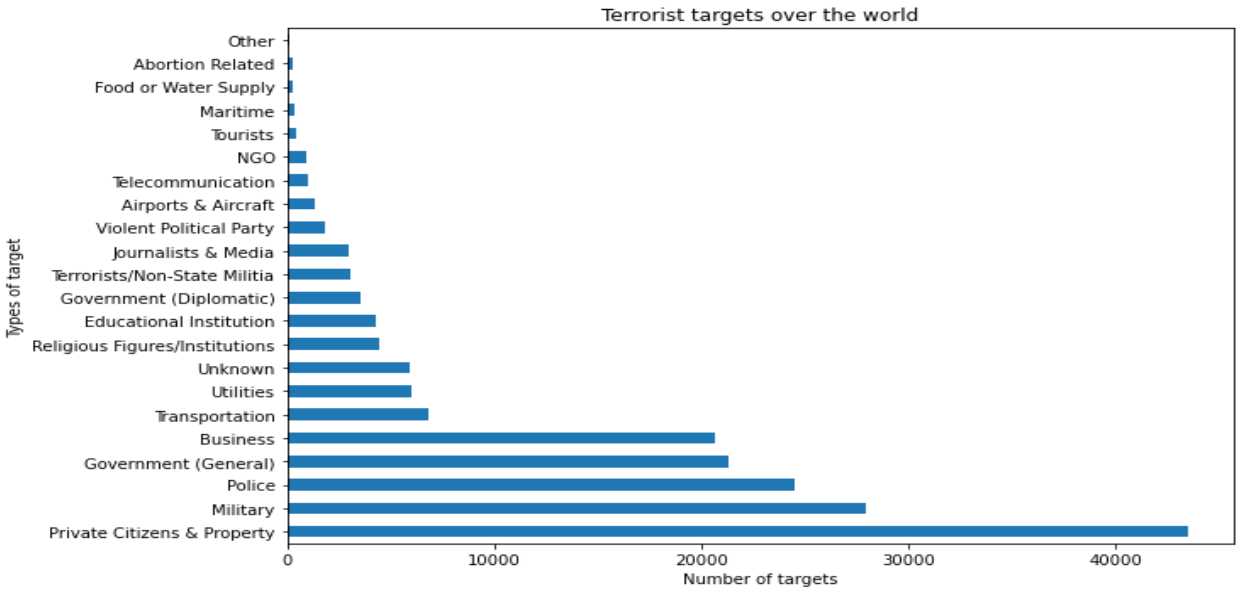
chart in the 70s and quickly lost traction. The Middle East started gaining traction around 2000 and went on to become the most affected region. The table clearly shows the number of deaths, people killed and injured attacks on a region-wise basis.

Type of Attacks



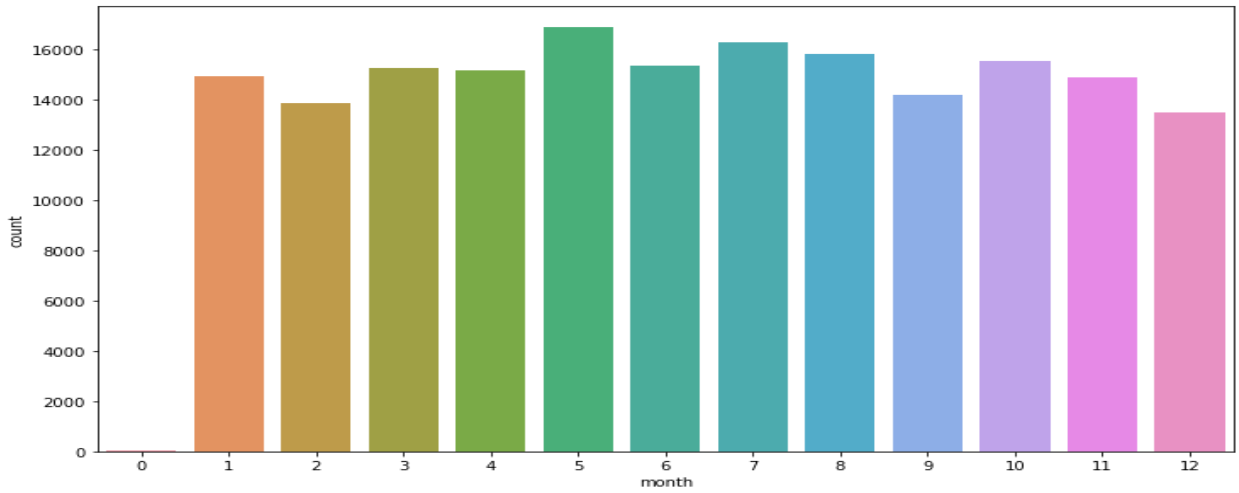
Most of the attacks are taken over by Bombing/Explosion and the least is by Hijacking of 659. 7296 of the attacks are unrecorded and considered as unknown. The next highest attacks are done by Unarmed Assault.

Targets and their frequency



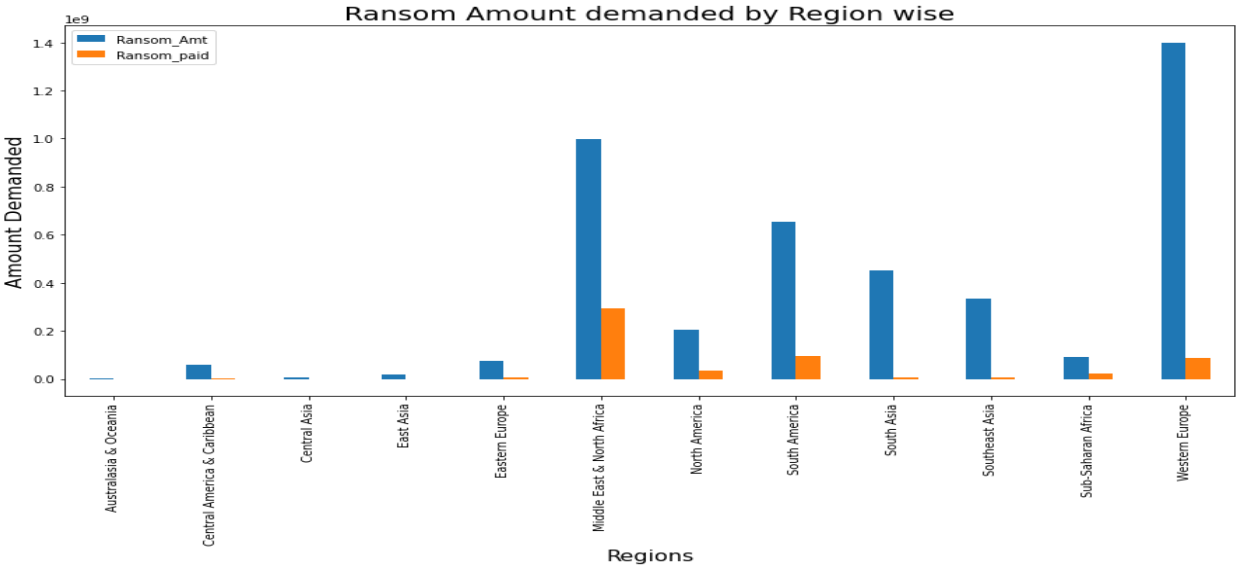
The private citizens and property are mostly targeted for attacks and property means malls, markets and parks, where crowds will be more and many people can be killed easily there may be more loss to the nations.

Months Vs Attacks



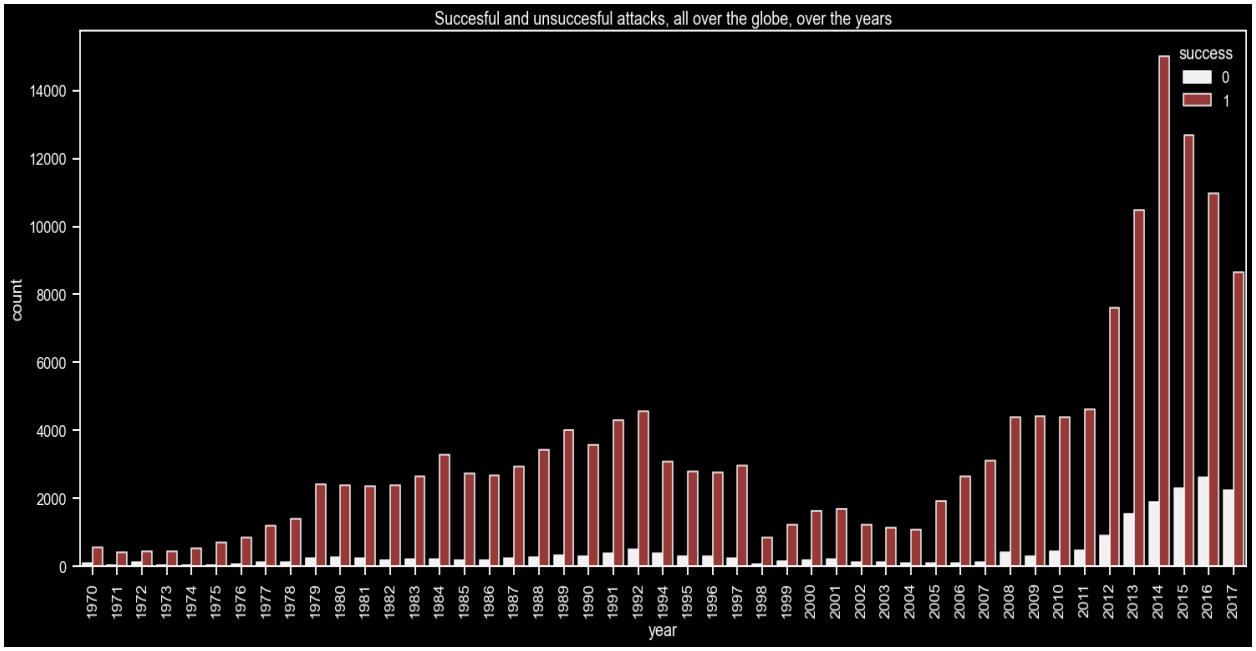
The above figure shows that there are more incidents recorded in May. We also have 20 null values and they are considered as zero(0) for the analysis purpose. But no month, in particular, can be said to be more likely to have a terrorist attack, than any other.

Ransom Amount demanded and paid vs Region



It can be observed in all the regions that when a situation involving ransom arises, it is rarely paid or is recorded to be paid officially. Western Europe is the region with most cases of ransom being demanded and Australasia & Oceania the least. The Middle East and North Africa is the region where the victims are most likely to pay the demanded ransom.

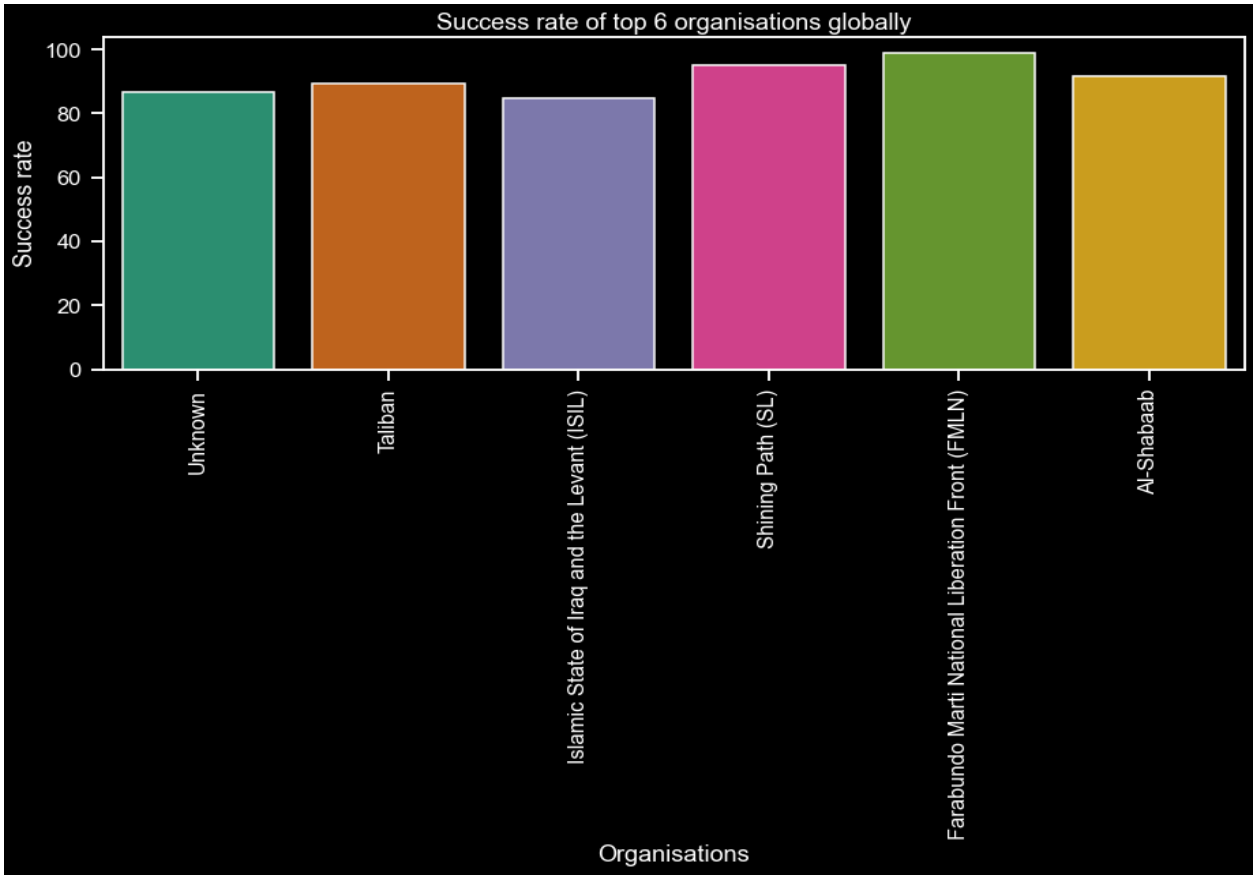
The Success and Unsuccessful attacks




```
global_org_success.head()
```

	organisation	success	attacks	succes_rate_percentage
585	1 May	9	10	90.0
905	14 K Triad	4	5	80.0
2683	14 March Coalition	1	1	100.0
1256	14th of December Command	3	3	100.0
2682	15th of September Liberation Legion	1	1	100.0

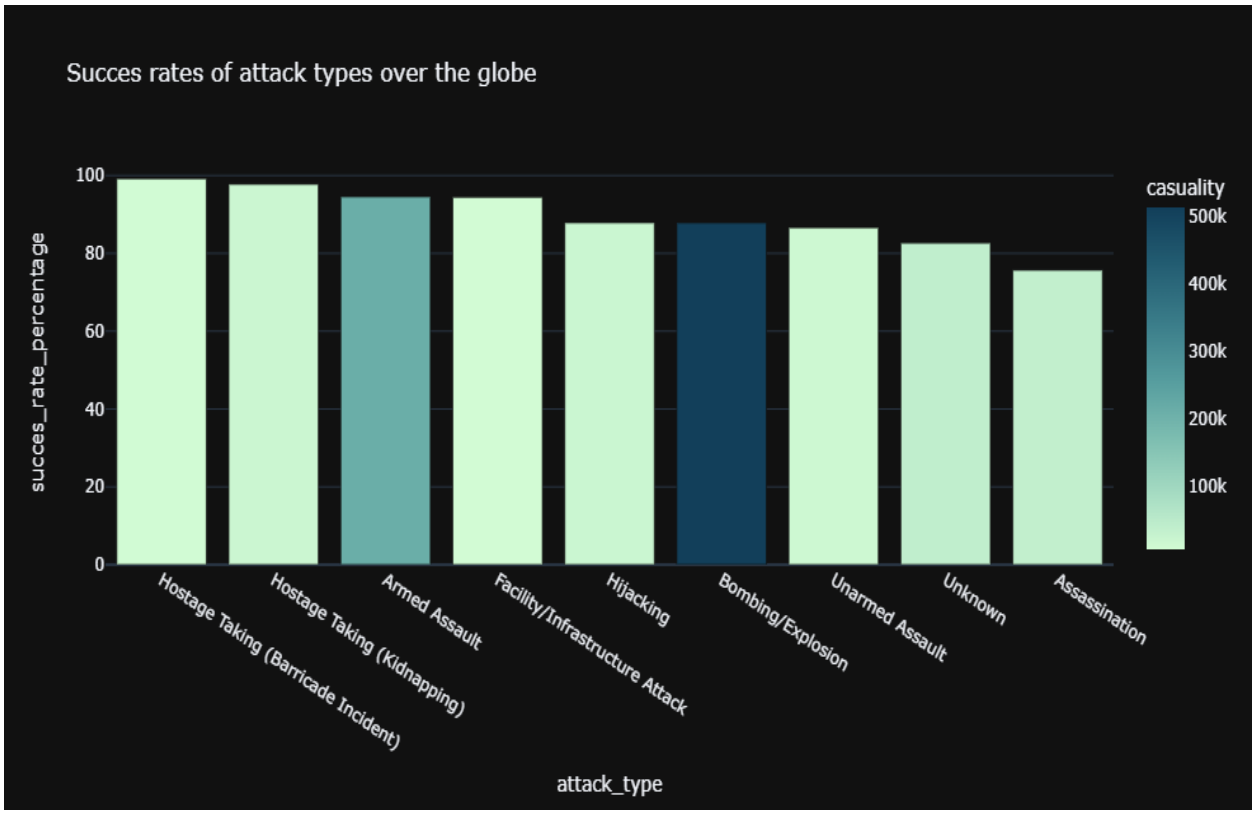
The success rate of attacks of different groups



Groups with highest attack frequency over the globe, and their success rate

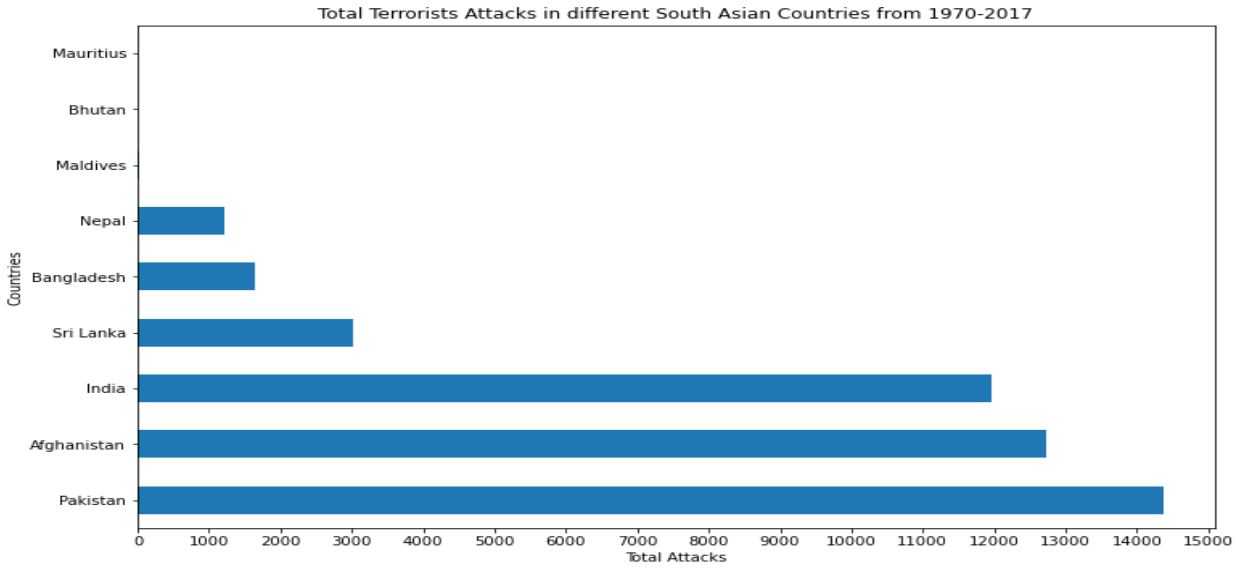
	organisation	success	attacks	succes_rate_percentage
0	Unknown	71748	82782	86.67
1	Taliban	6680	7478	89.33
2	Islamic State of Iraq and the Levant (ISIL)	4759	5613	84.79
3	Shining Path (SL)	4337	4555	95.21
4	Farabundo Marti National Liberation Front (FMLN)	3317	3351	98.99
5	Al-Shabaab	3016	3288	91.73

Attacks vs Success rate



South Asia at Glance

Mostly affected South Asian Countries

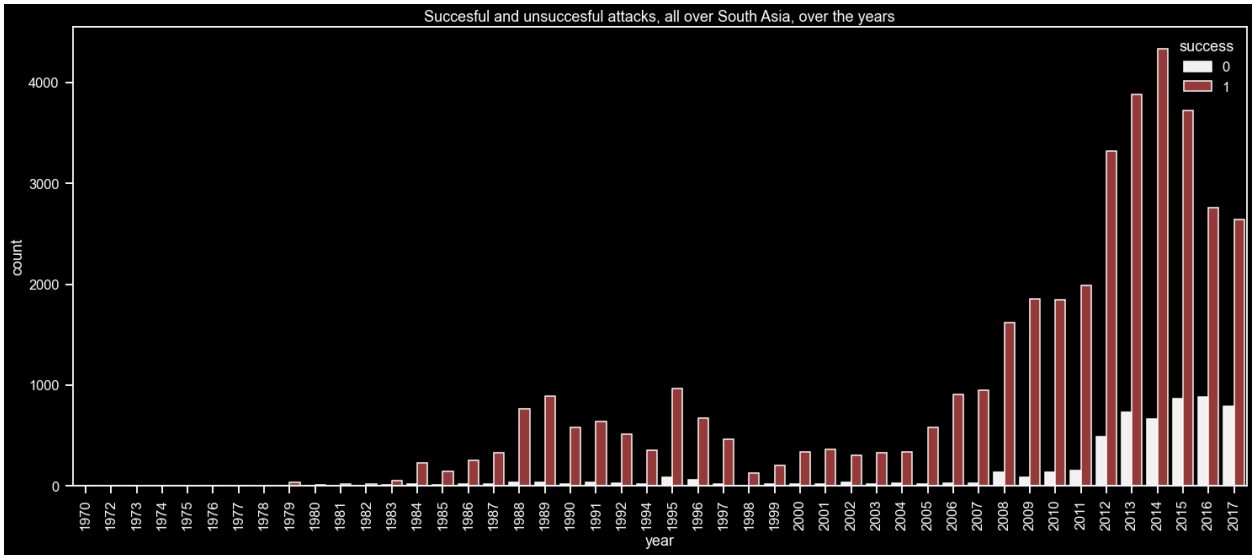


Pakistan is the most affected country in the South Asian region, followed by Afghanistan and India. In India, nearly 12000 attacks took place from 1970-2017.

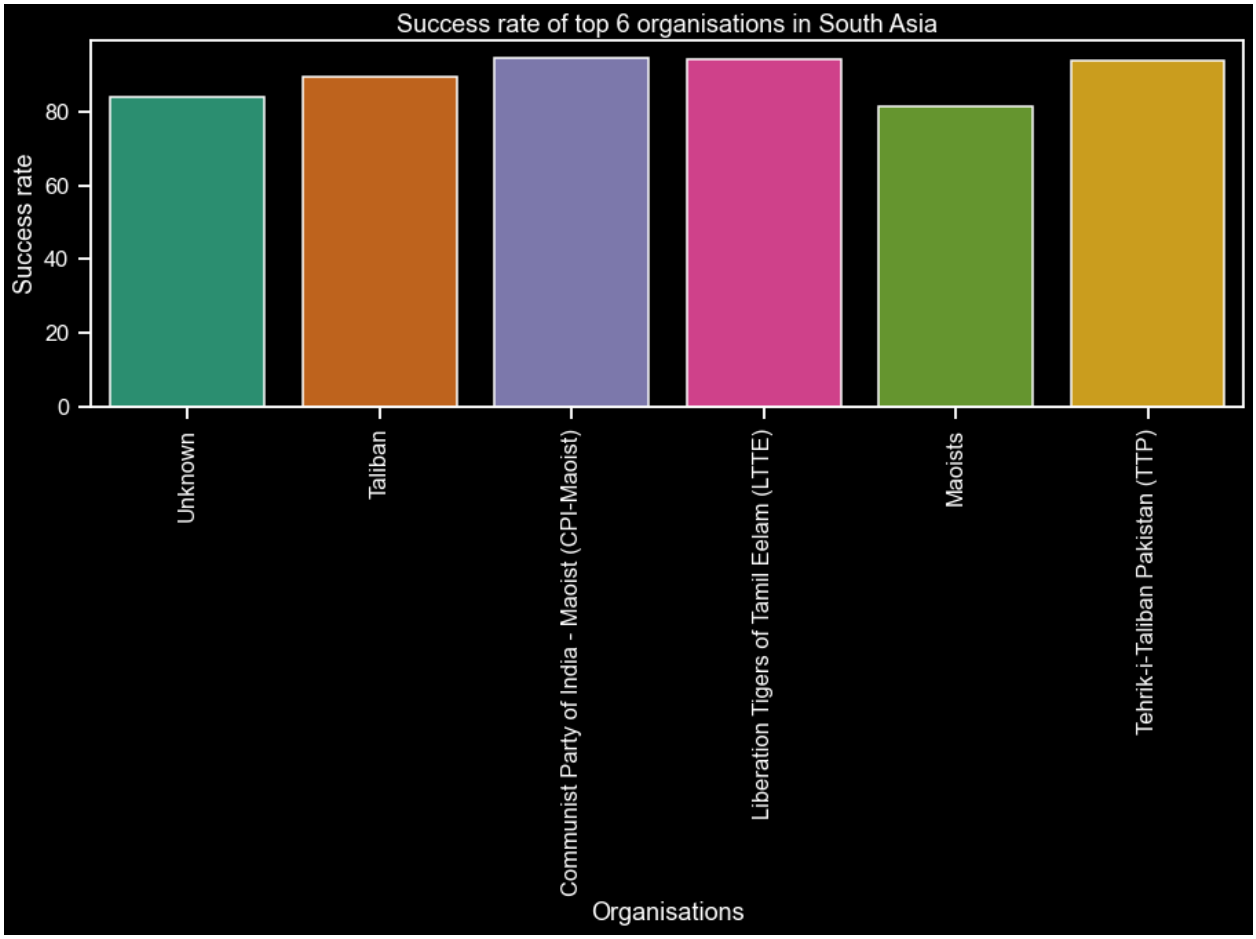
Terrorist Organisations in South Asia

Unknown faction again has the highest contribution to the number of attacks in the South Asian Region. Taliban again, comes up at top of the chart in South Asia, followed by CPI-Maoists, LTTE, TTP etc.

Success and Unsuccessful attacks



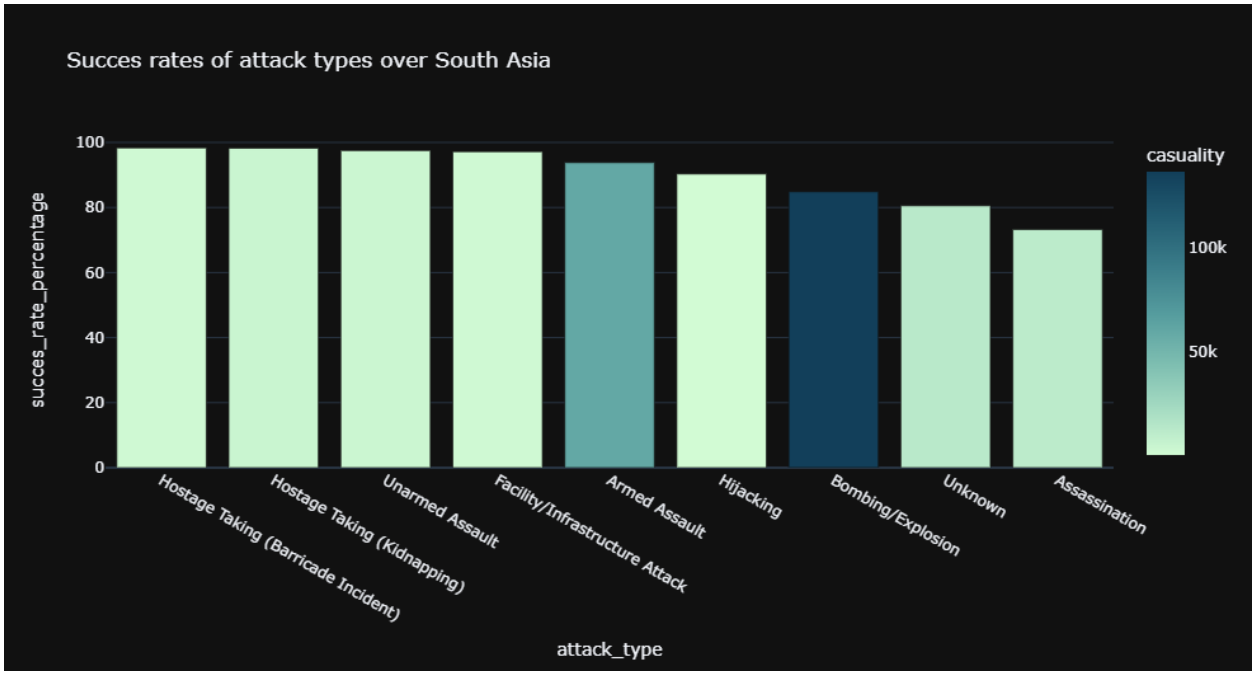
The success rate of attacks of different groups



Groups with highest attack frequency over the globe, and their success rate

	organisation	success	attacks	succes_rate_percentage
0	Unknown	71748	82782	86.67
1	Taliban	6680	7478	89.33
2	Islamic State of Iraq and the Levant (ISIL)	4759	5613	84.79
3	Shining Path (SL)	4337	4555	95.21
4	Farabundo Marti National Liberation Front (FMLN)	3317	3351	98.99
5	Al-Shabaab	3016	3288	91.73

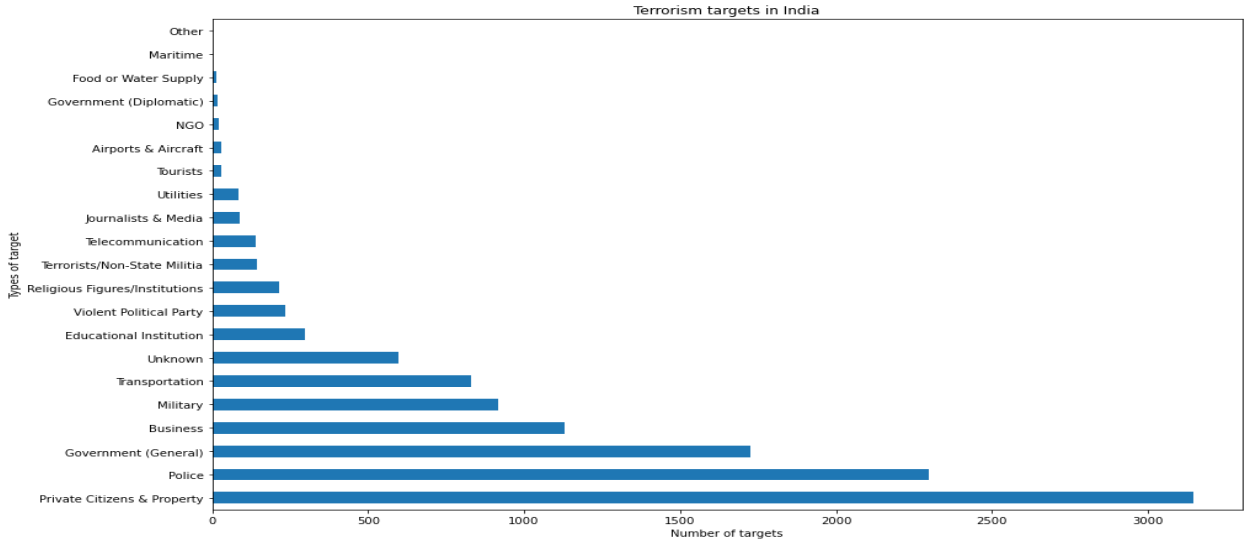
Attacks vs Success rate



India at Glance

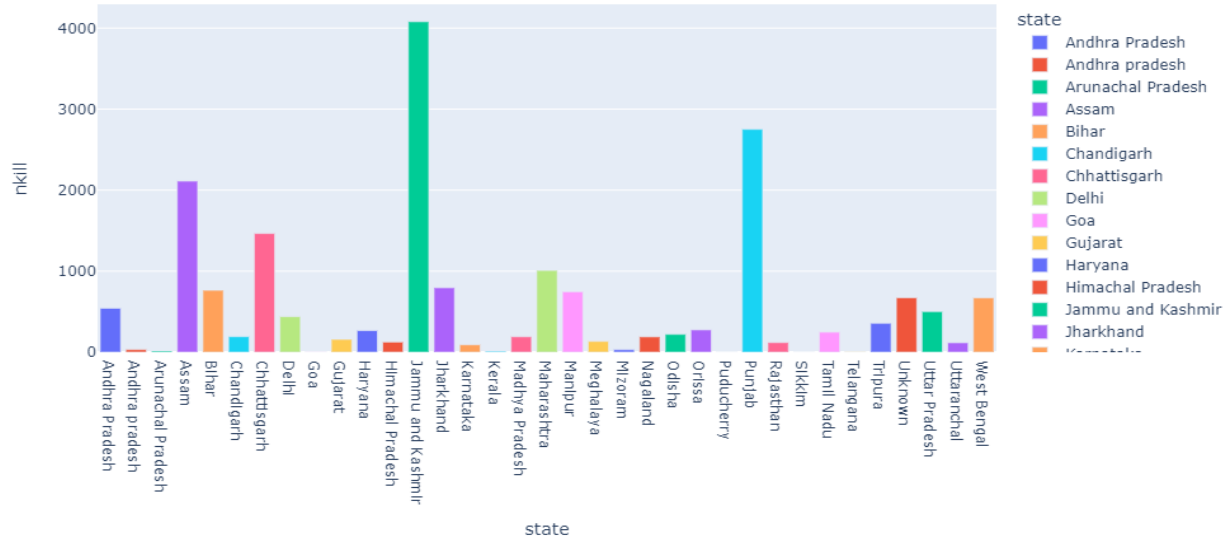
Communist Party of India(Maoists) is the most dominant group in India followed by Sikh Extremists, Maoists etc.

Targets in India



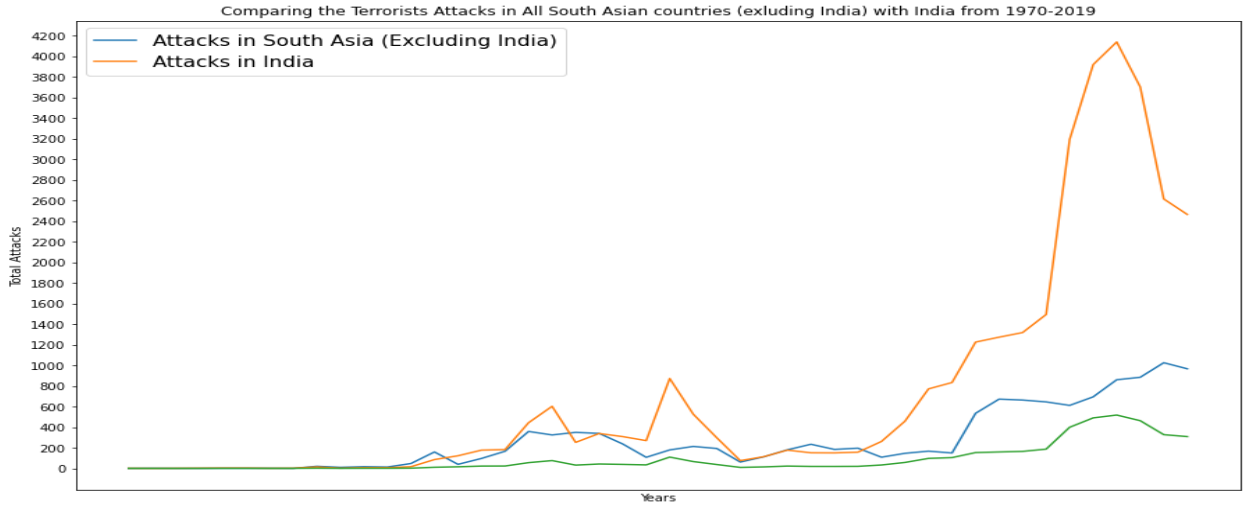
Private citizens followed by Police, Government, Businesses and Military are the most sought after targets by the terrorists in India.

States in India



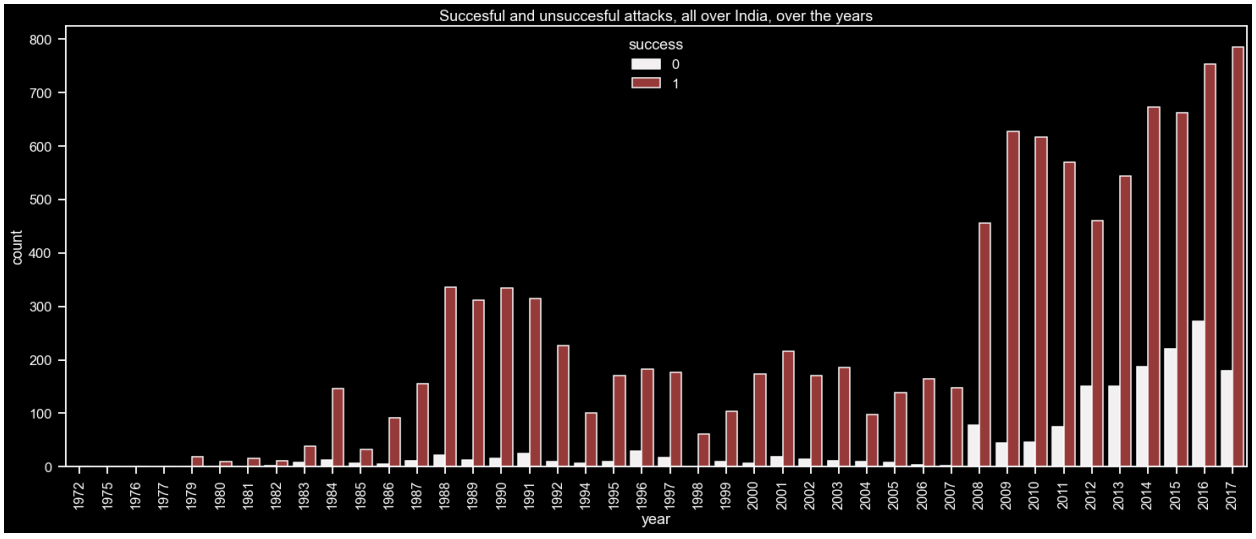
From the above figure we can say that Jammu and Kashmir, Punjab and Assam are the most affected states in India. And Jammu and Kashmir is one of the topmost affected states in the world.

India vs South Asia



The above figure visualizes the graph of India vs other countries in South Asia in the number of attacks from 1970-2017. It can be observed that there were quite a few instances in the late 80s and 90s, where the number of attacks in India alone was higher than the rest of South Asia Combined. India has always had more attacks than average attacks per country in the rest of South Asia.

Success and Unsuccessful attacks



A common trend can be observed at all three levels that the number of successful attacks has been on a gradual increase over time. But the unsuccessful attacks, or in other words our ability to render those attacks unsuccessful has not been able to keep

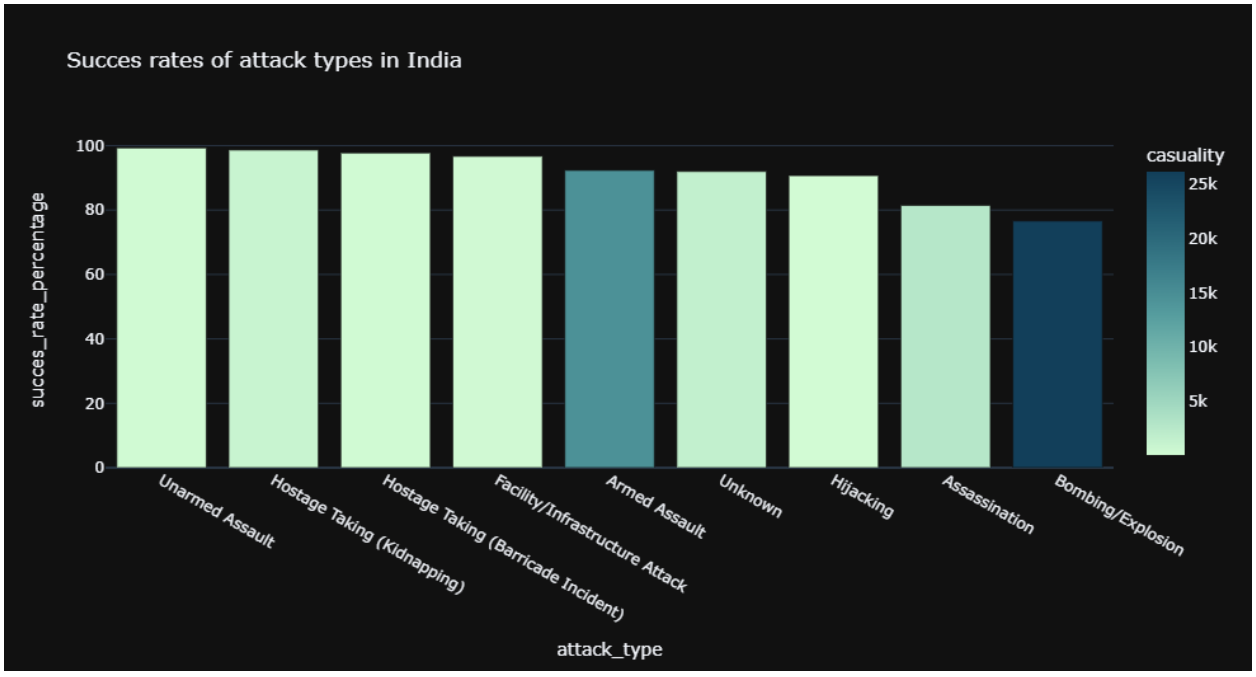
up with the rise in terrorism. In short, one could say, terrorism has been winning over the years.

The success rate of attacks of different groups

```
Groups with top 5 highest attack frequency in India, and their success rate
```

	organisation	success	attacks	succes_rate_percentage
0	Unknown	3337	4263	78.28
1	Communist Party of India - Maoist (CPI-Maoist)	1773	1877	94.46
2	Maoists	1083	1396	77.58
3	Sikh Extremists	667	710	93.94
4	United Liberation Front of Assam (ULFA)	327	357	91.60
5	Hizbul Mujahideen (HM)	182	199	91.46

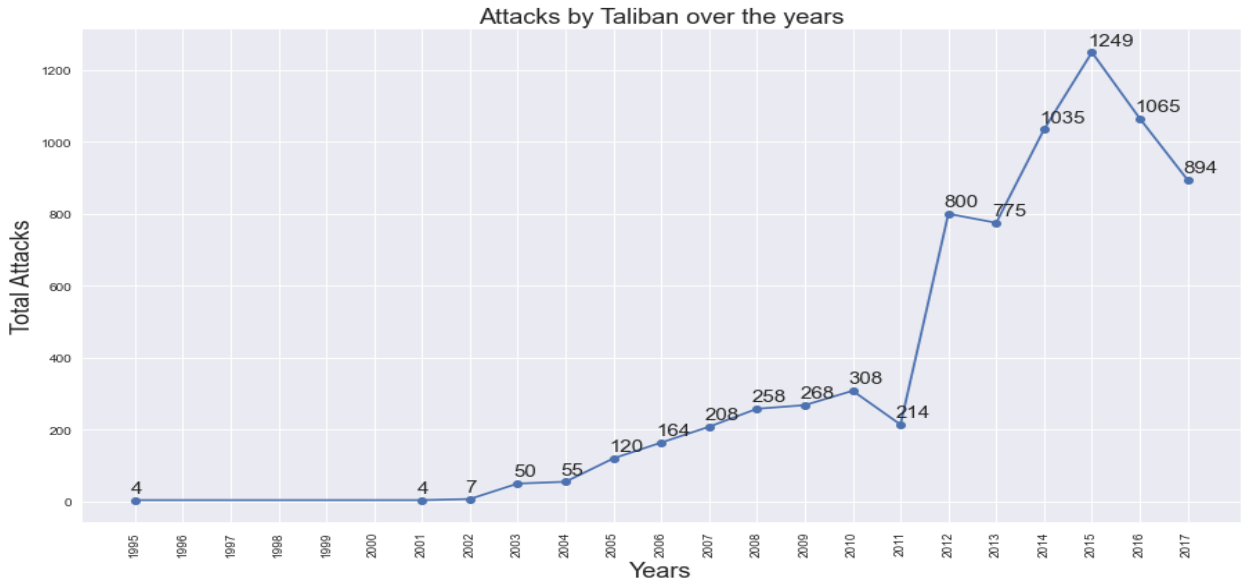
The success Rate vs Attacks



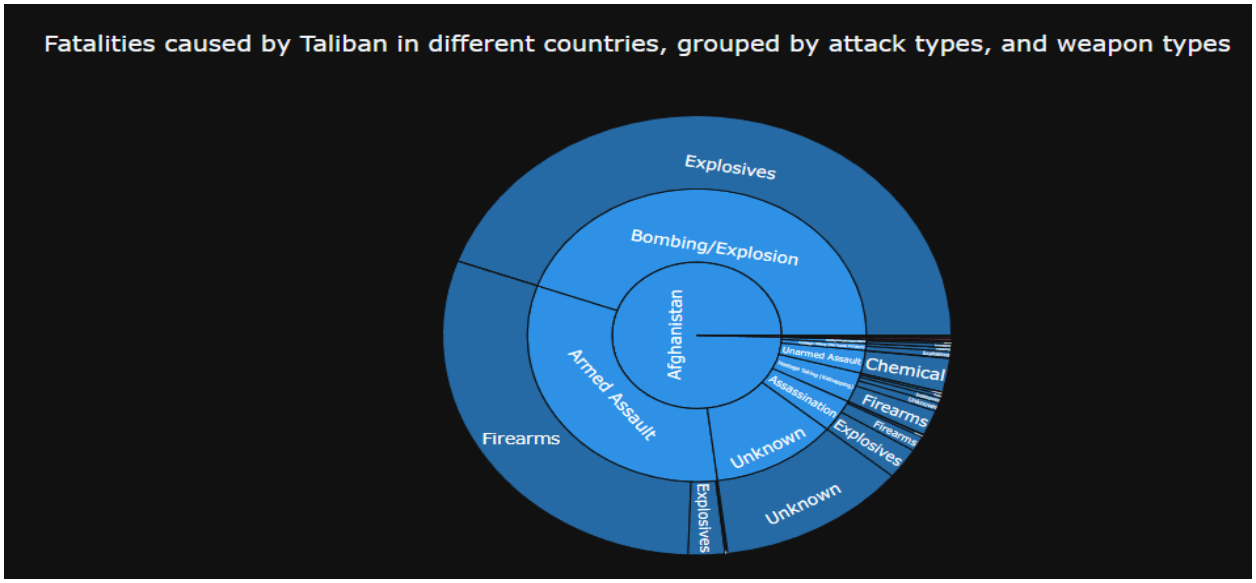
Hostage Taking(Barricade Incident), Hostage Taking(Kidnapping) has proved to be the most successful attack type at both global and South Asian level and result in comparatively lower casualties than other attack types. This can be attributed to the very nature of the attack type. Since the terrorist has the leverage of hostages, it

ressurises the concerned authority to meet the demands, hence resulting in high success rates. Bombing and explosion, which is amongst the most used attack types, and accounts for most casualties, has yielded low success rates at all three levels. The success of a bombing/explosion attack is determined by whether the explosive detonated or not.

Taliban attacks



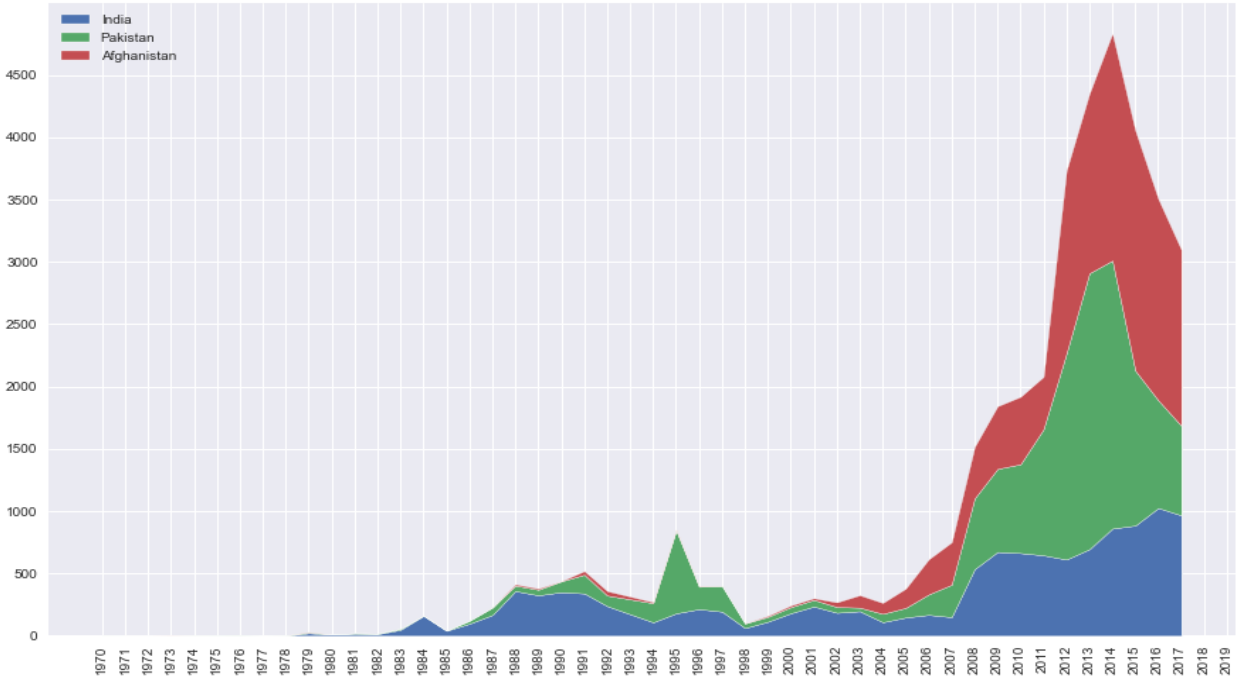
Taliban are more active after the year 2011 and as a result, there are more attacks over the world after 2011, mostly in 2014 and 2015 years.



In the above chart, the innermost circle represents the countries the organisation have caused casualties in. The area is determined by the number of casualties

It's interesting to note that, even though the Taliban has the highest contribution to the total number of attacks in the world, they have caused casualties in a total of just 3 countries: Afghanistan(57140 casualties), Pakistan (197 casualties) and Turkmenistan (5 casualties). Taliban has caused the most casualties using Bombing/Explosions (24650) followed by Armed Assaults (18901) and unknown attack types(7149). Explosives (24639) and Firearms(17466) are the most common weapon types used by the Taliban.

India vs Pakistan vs Afghanistan

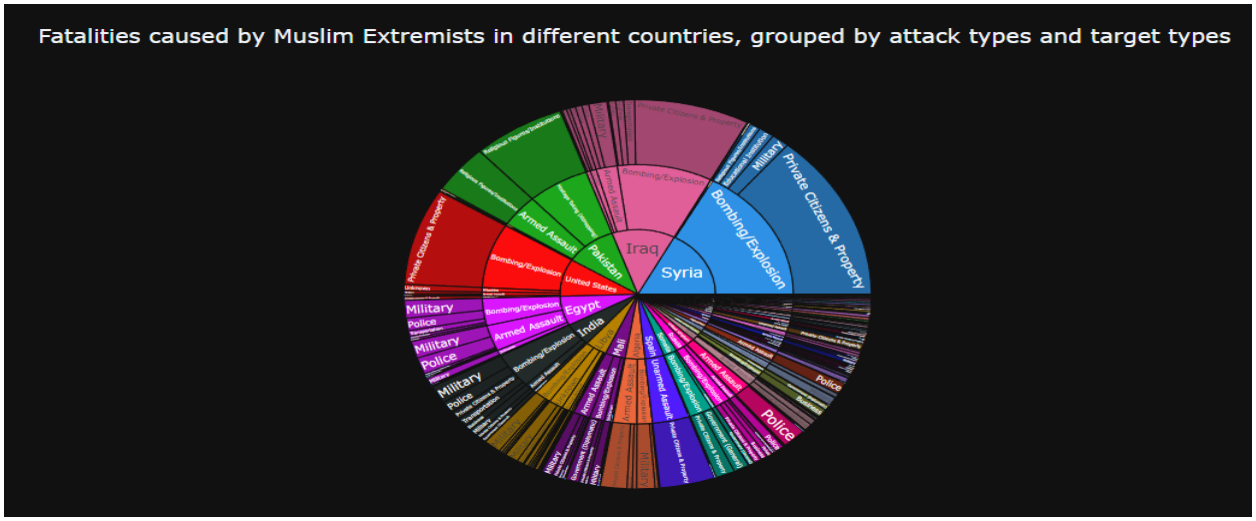


Terrorism is in India, Pakistan and Afghanistan appear to be connected. The fall and rise in terrorism activities in these countries are observed to follow similar patterns. As per the dataset these regions (India, Pakistan and Afghanistan) were relatively safe earlier, but in the last 2 decades, there has been a continuous rise in terrorism.

There are several instances when attacks in country from (India, Pakistan and Afghanistan) were higher compared to other two e.g. from 1982 to the mid-1993 and 1998 to 2004 terrorist activities in India were higher, from 1994 to 1996 and from

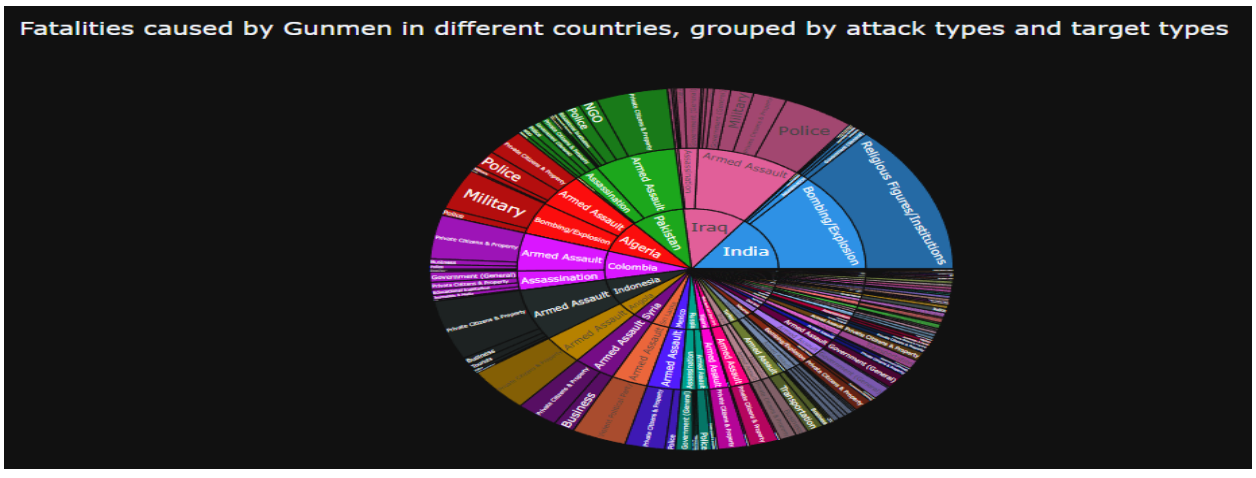
2009 to 2013 terrorist activities in Pakistan were higher. Similarly, in the case of Afghanistan, we can say that it was relatively a quiet place as compared to India and Pakistan from 1970 to 2001 but after 2001 there was a sudden rise in terrorist activities surpassing both the other countries in terms of several terrorist attacks in mid of 2014 till 2017(last record).

Attacks by Muslim Extremists



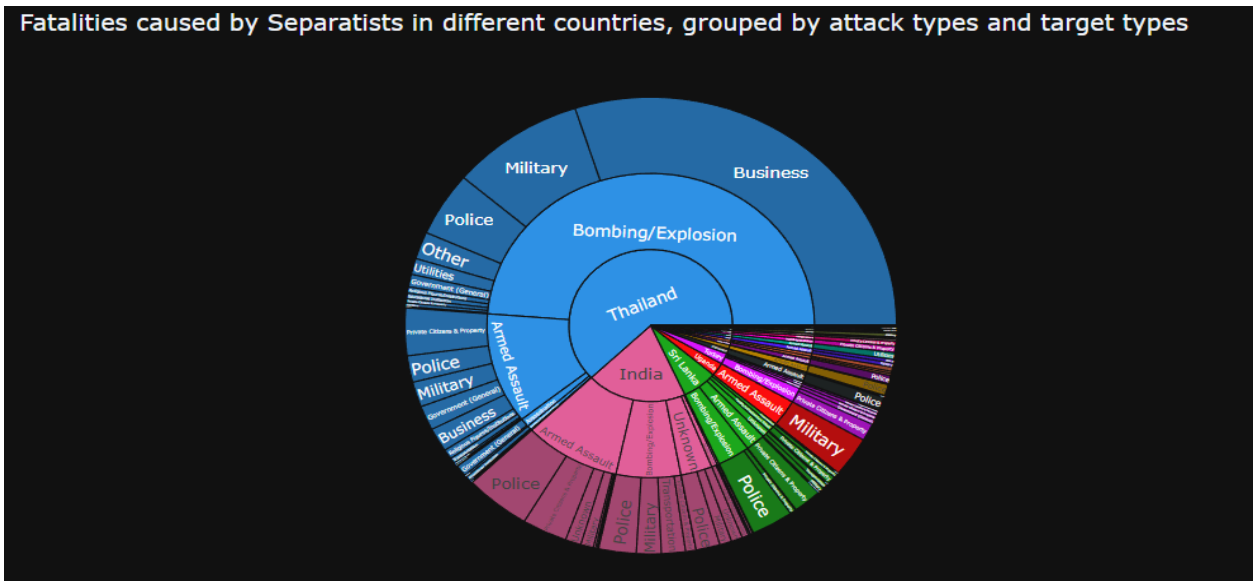
- Countries caused casualties is: 51
- Country caused most casualties in Syria; 600 casualties
- Preferred Attack type: Bombing Explosions
- Preferred weapon type: Explosives

Attacks by Gunmen



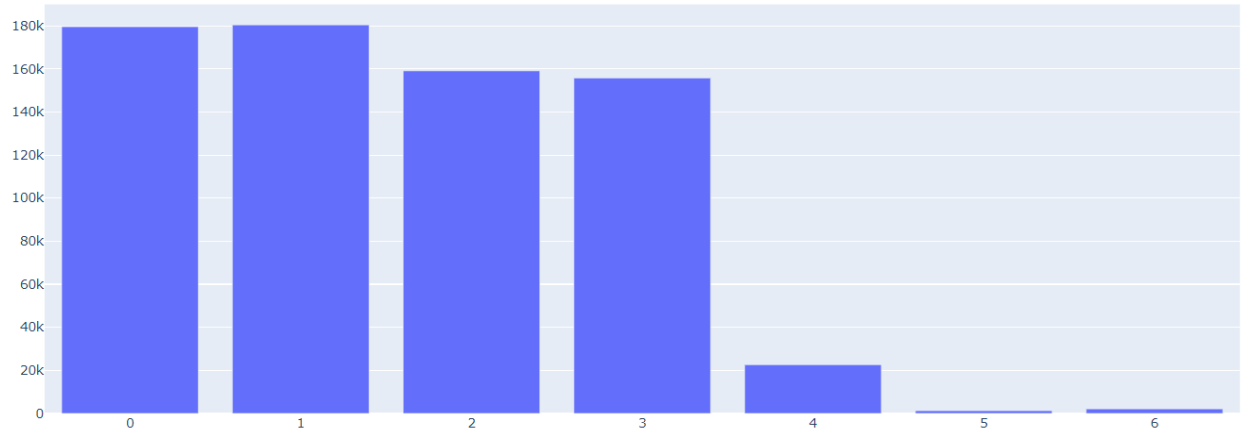
- Countries caused casualties is: 50
- Country caused most casualties in India; 113 Casualties
- Preferred Attack type: Bombing Explosions
- Preferred weapon type: Explosives

Attacks by Separatists



- Countries caused casualties is: 30
- Country caused most casualties in Thailand; 1256 Casualties
- Preferred Attack type: Bombing Explosions
- Preferred weapon type: Explosives

Analysis on Criterion

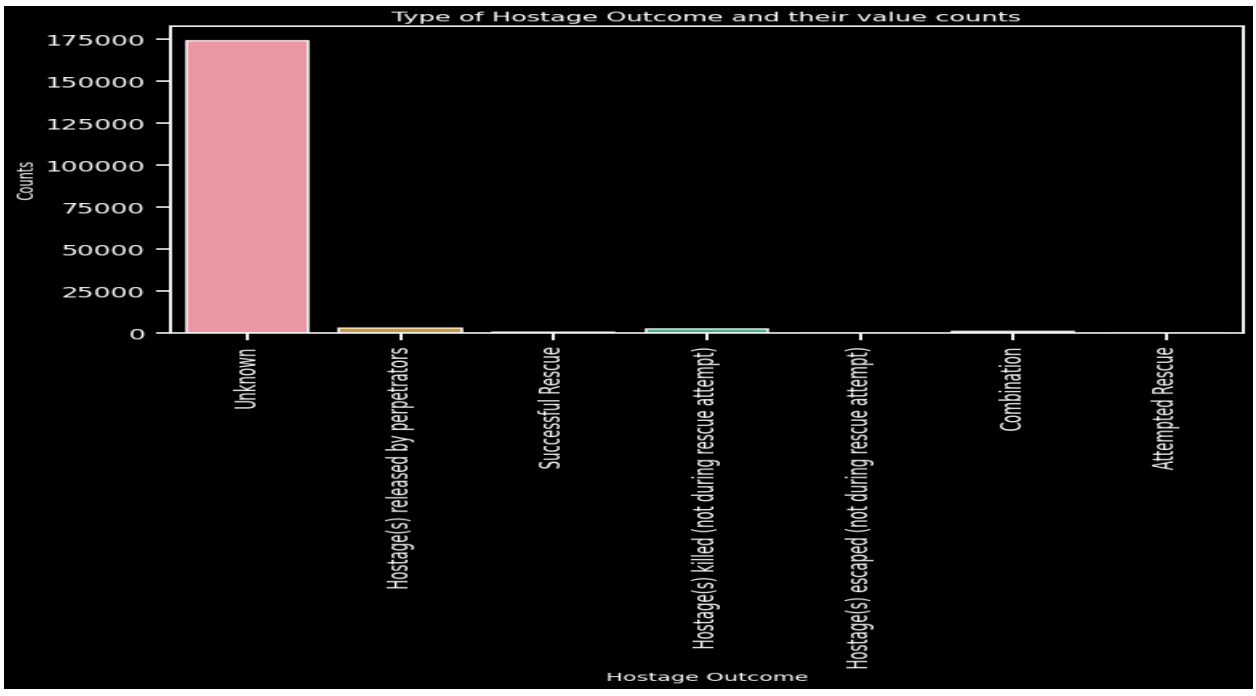


Following is the distribution of attacks according to criterions. It's also the case at times when a single attack meets multiple criterions. The same has been recorded as well.

- Criterion 1: 179.6 K
- Criterion 2: 180.43 K
- Criterion 3: 159.10 K
- All the Criterions: 155.762 k
- Criterion 1 and 2 only: 22.59 k
- Criterion 1 and 3 only: 1255
- Criterion 2 and 3 only: 2084

It can be observed that attaining some political/ sociological/ economical / Social goal and publicizing their message to a larger audience is almost always the goal behind these attacks.

Analysis of Hostage Outcomes



- It can be observed in most cases, that the outcome of these hostage situations is unknown.
- The hostages were released by perpetrators in over 3124 cases.

- But the cases of hostages getting killed is also considerable, with a count of 2522.
- The cases of hostages being rescued (542) and hostages escaping the situation (165) are quite low.

Conclusion

- We have selected only 25 out of 136 fields from the given dataset and made some useful visualizations.
- Divided the data into main parts like global level, regional level, south Asia and India.
- Attack frequency, casualties, major groups, affected regions, comparison between regions etc, were the major components of this analysis.
- Along with the above, we also made some analysis on the success rate of attacks, groups involved, the number of deaths and injured, ransom demanded and paid, etc.
- The project is limited only to the extent of exploring and analyzing the data.
- Future scopes of the project may include prediction and prevention of such attacks to a good extent by integration of Machine Learning and Artificial Intelligence elements.