



DATA CLEANING 2 A brief intro about the

Fixing incorrect, corrupted, duplicate and incomplete data in the database

CONCLUSIONS 4

Summary of facts and Observations recorded from analysis 3 ANALYSIS & VISUALIZATION

Analysing and visualization of various variables

5 REFERENCES



TERRORISM



Terrorism is defined in the Oxford Dictionary as "the unlawful use of violence and intimidation, especially against civilians, in the pursuit of political aims."

In our overview of terrorism, we try to understand how the number of terrorist acts varies around the world and how it has changed over time.

What is EDA?



Exploratory Data Analysis, or EDA, is an important step in any Data Analysis or Data Science project. EDA is the process of investigating the dataset to discover patterns, and anomalies (outliers), and form hypotheses based on our understanding of the dataset. EDA involves generating summary statistics for numerical data in the dataset and creating various graphical representations to understand the data better.

Data Pre-Processing



The Data mining is the process of extracting and

discovering patterns in large data sets.

Data Cleaning is one of the most integral part of analysis, in order miss to get something meaningful out of a huge dataset we need to figure out how to extract a generalized, smaller

and more skewed sub-dataFrame out of it.



DATA VISUALIZATION A

Data Visualization is the process of analyzing data in the form of graphs or maps, making it a lot easier to understand the trends or patterns in the data.

What's on the Database?



- 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.
- In this project we are going to analyze the Global Terrorism Data in order to find out valuable insights.

Extracting meaningful chunks out of the file by dropping unnecessary columns and NaN values.

Al

	terrorism_data.head(10)															
Ď		eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region		addnotes	scite1	scite2	
	0	197000000001	1970	7	2	NaN	0	NaN	58	Dominican Republic	2		NaN	NaN	NaN	Ι.
	1	197000000002	1970	0	0	NaN	0	NaN	130	Mexico	1		NaN	NaN	NaN	
	2	197001000001	1970	1	0	NaN	0	NaN	160	Philippines	5		NaN	NaN	NaN	
	3	197001000002	1970	1	0	NaN	0	NaN	78	Greece	8		NaN	NaN	NaN	
	4	197001000003	1970	1	0	NaN	0	NaN	101	Japan	4		NaN	NaN	NaN	
	5	197001010002	1970	1	1	NaN	0	NaN	217	United States	1		The Cairo Chief of Police, William Petersen, r	"Police Chief Quits," Washington Post, January	"Cairo Police Chief Quits; Decries Local 'Mili	V
	6	197001020001	1970	1	2	NaN	0	NaN	218	Uruguay	3		NaN	NaN	NaN	
(1	7	197001020002	1970	1	2	NaN	0	NaN	217	United States	1		Damages were estimated to be between \$20,000-\$	Committee on Government Operations United Stat	Christopher Hewitt, "Political Violence and Te	
()	(181691, 135)															



Result of dropping the unnecessary columns and NaN values

terrorism_data_new.head(5)

	year	month	day	country	region	state	city	latitude	longitude	success	attack_type	target_type	gang_name	weapon_typ
0	1970	7	2	Dominican Republic	Central America & Caribbean	unknown	Santo Domingo	18.456792	-69.951164	1	Assassination	Private Citizens & Property	MANO-D	Unknow
1	1970	0	0	Mexico	North America	Federal	Mexico city	19.371887	-99.086624	1	Hostage Taking (Kidnapping)	Government (Diplomatic)	23rd of September Communist League	Unknow
2	1970	1	0	Philippines	Southeast Asia	Tarlac	Unknown	15.478598	120.599741	1	Assassination	Journalists & Media	Unknown	Unknow
3	1970	1	0	Greece	Western Europe	Attica	Athens	37.997490	23.762728	1	Bombing/Explosion	Government (Diplomatic)	Unknown	Explosive
4	1970	1	0	Japan	East Asia	Fukouka	Fukouka	33.580412	130.396361	1	Facility/Infrastructure Attack	Government (Diplomatic)	Unknown	Incendiar

(181691, 16)

The Count of NaN values BEFORE processing.

```
terrorism_data_new.isnull().sum()
iyear
imonth
iday
country txt
region txt
provstate
                     421
city
                     434
latitude
                    4556
longitude
                    4557
SUCCESS
attacktype1_txt
targtype1_txt
gname
weaptype1_txt
nkill
                   10313
nwound
                   16311
dtype: int64
```

terrorism_data_new.isnull().sum()



```
year
month
day
country
region
state
city
latitude
               4556
longitude
               4557
success
attack_type
target_type
gang_name
weapon_type
no_of_kills
wounds
dtype: int64
```

The Count of NaN values AFTER processing.

CORRELATION HEATMAP MATRIX

year	1	0.00014	0.018	0.17	0.0039	-0.083	0.021	0.017
ШОШС	0.00014	1	0.0055	-0.016	-0.0039	-0.0028	0.004	0.0032
day	0.018	0.0055	1	0.0034	-0.0023	-0.012	-0.0032	-0.0011
annina	0.17	-0.016	0.0034	1	0.0015	-0.074	-0.013	0.016
annification	0.0039	-0.0039	-0.0023	0.0015	1	-0.00086	-0.00059	0.0002
SACCESS	-0.083	-0.0028	-0.012	-0.074	-0.00086	1	0.05	0.023
SIIIN IO OI	0.021	0.004	-0.0032	-0.013	-0.00059	0.05	1	0.44
woulds	0.017	0.0032	-0.0011	0.016	0.0002	0.023	0.44	1
	_{veat}	month	804	latitude	ngitude	success no	of kills	wed under

The matrix shows the relation between various

The further away

correlation coefficient is

from zero (0), the stronger

the relationship between

the

- 0.25

- 0.00

columns.

the Columns.

- 0.75

- 0.50

--0.:

- -0.75

What we are working on



- Geographic Heat map of the total terrorism attacks.
- Total number of attacks every year by region.
- Analyzing the trend of attacks across the decades.
- Analyzing the attacks in the year 2014 on the basis of region.
- Top 10 Countries which were most attacked.
- Top 10 States under attack across the year.
- Which attack type is the most common among the terrorist groups.
- Most Active Gangs across the years.
- Analyzing which attack type is used by the deadliest gang 'Taliban'.
- What was the most common target of the terrorists.
- Which year had the most number of casualties.
- Comparing the number of casualties and wounded people across every region.
- Relation between total, successful attacks and failures.



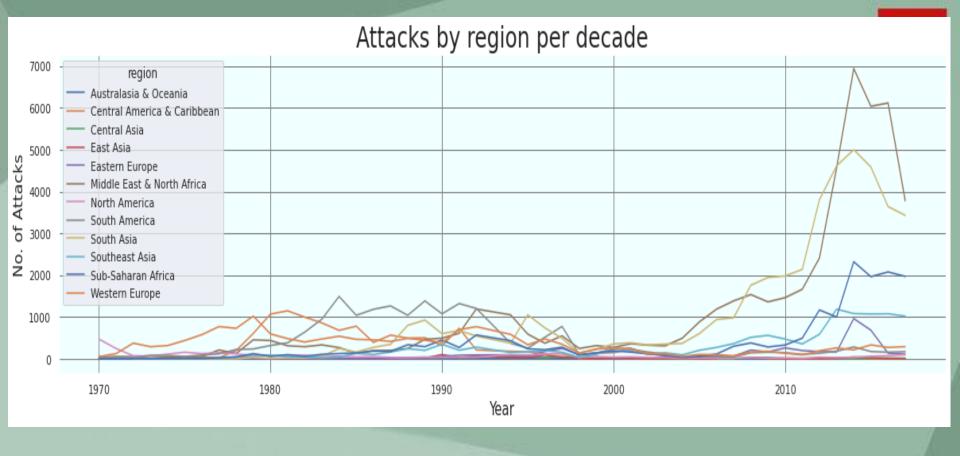


Word Cloud is a data visualization technique used for representing text data in which the size of each word indicates its frequency or importance.

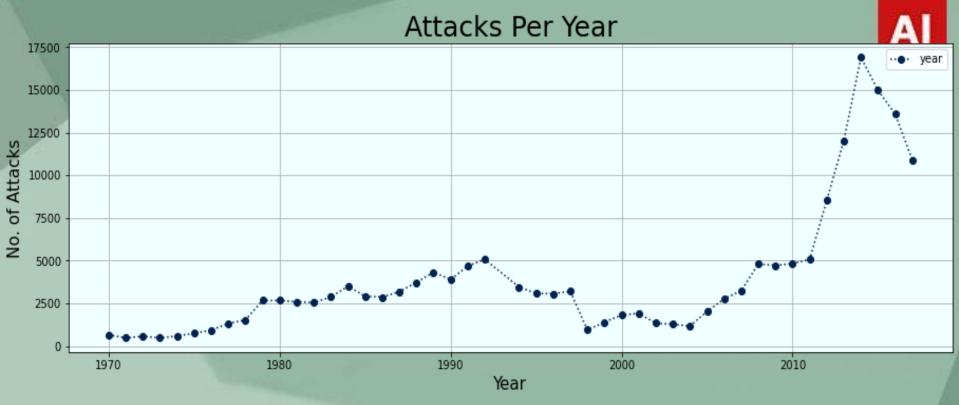
Iraq

52 Years of Terrorism

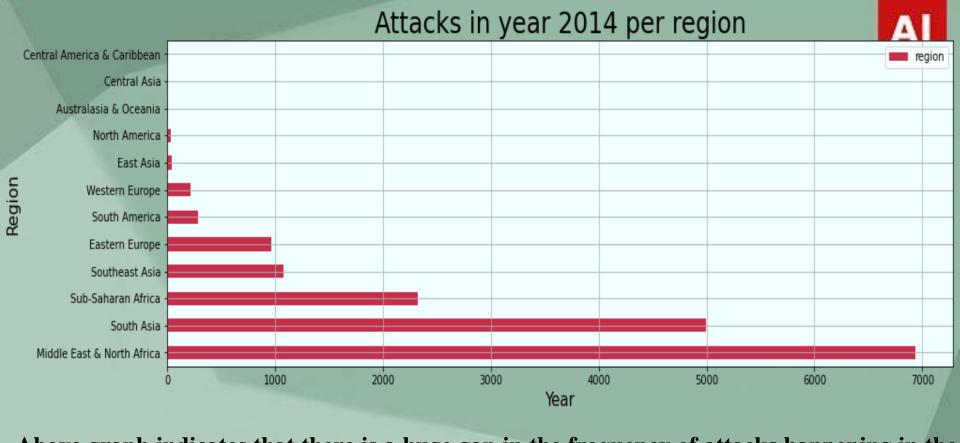




By looking at the graph we analyzed that attacks were taking place at a nearly constant rate across the world for 4 decades, but in 2010 a huge spike came in the Middle East and South Asia region.

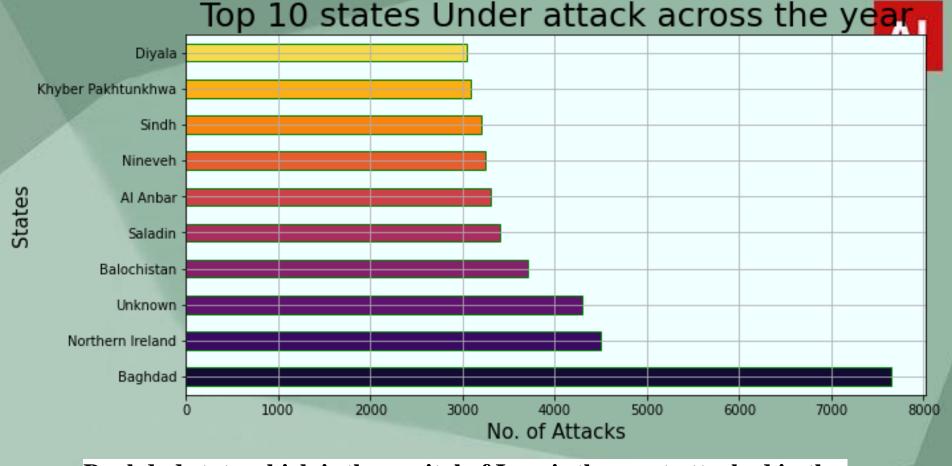


The graph shows that attacks were slowly rising from 1970 to 1990 and then a dip took place till 1998 and then a sudden spike in attacks were to be seen after 2010 across the world.

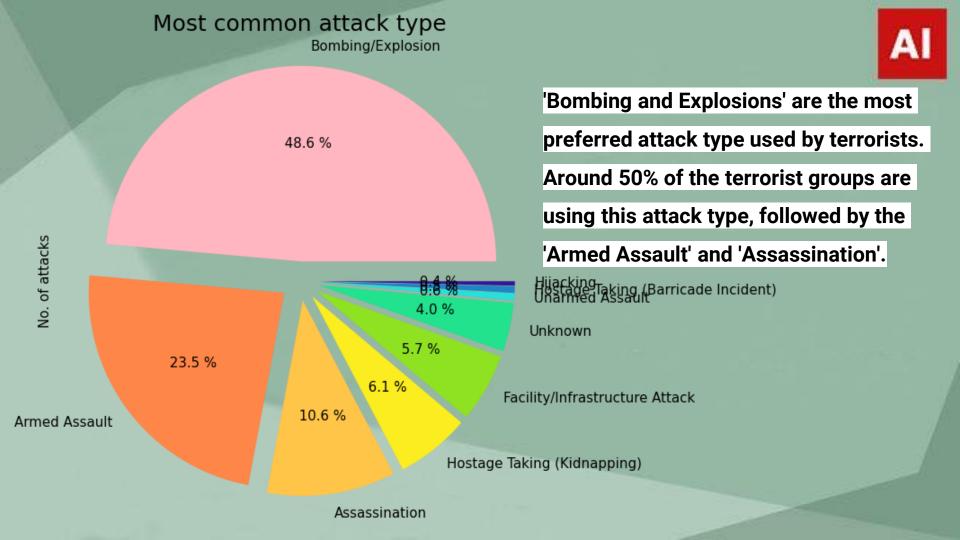


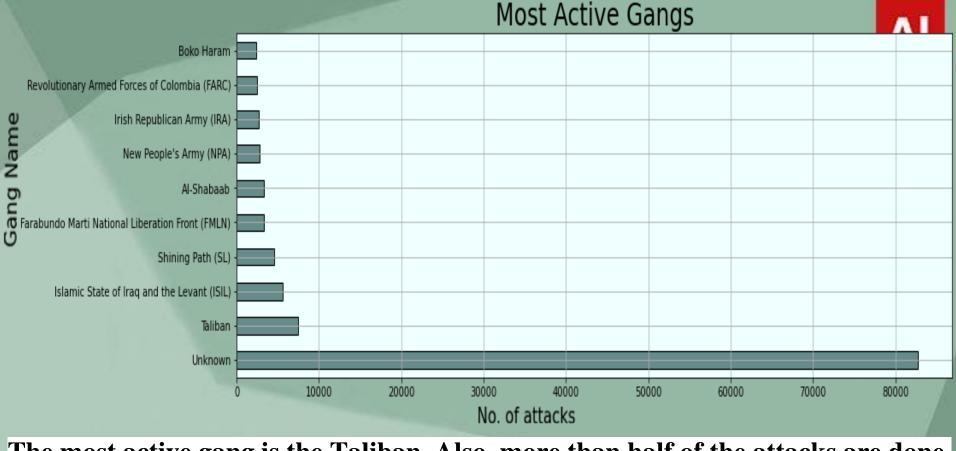
Above graph indicates that there is a huge gap in the frequency of attacks happening in the least and most vulnerable regions across the world. Also, it shows that the Middle East and Asian part is the most vulnerable.

Top 10 Countries which were most attacked Pakistan Iraq Iraq was the most 14.4 % 24.7 % attacked country across Afghanistan the world with 24636 12.7 % Countries attacks followed by 4.3 % Pakistan, Afghanistan, Turkey 5.2 % 12.0 % and India. 5.3 % United Kingdom 6.9 % 6.1 % India 8.3 % El Salvador Colombia Peru Philippines



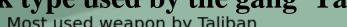
Baghdad state which is the capital of Iraq is the most attacked in the world with 7645 attacks followed by Northern Ireland.

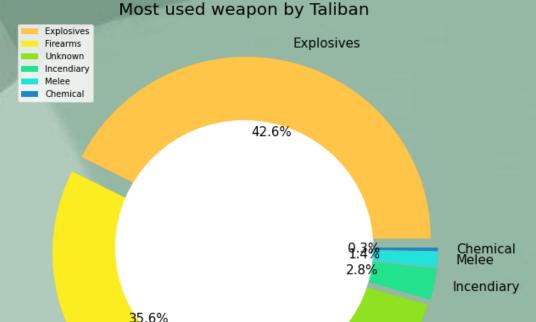




The most active gang is the Taliban. Also, more than half of the attacks are done by unknown terrorist groups which seem to be a huge concern.

Attack type used by the gang 'Taliban'





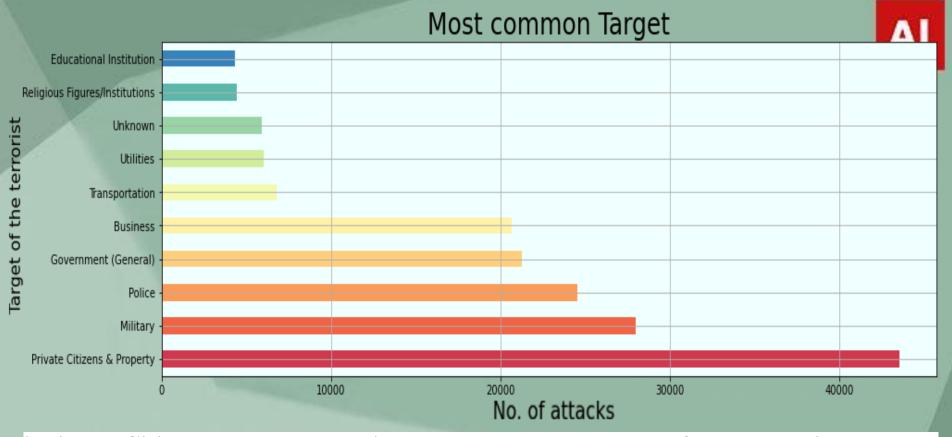
Firearms

17.4%

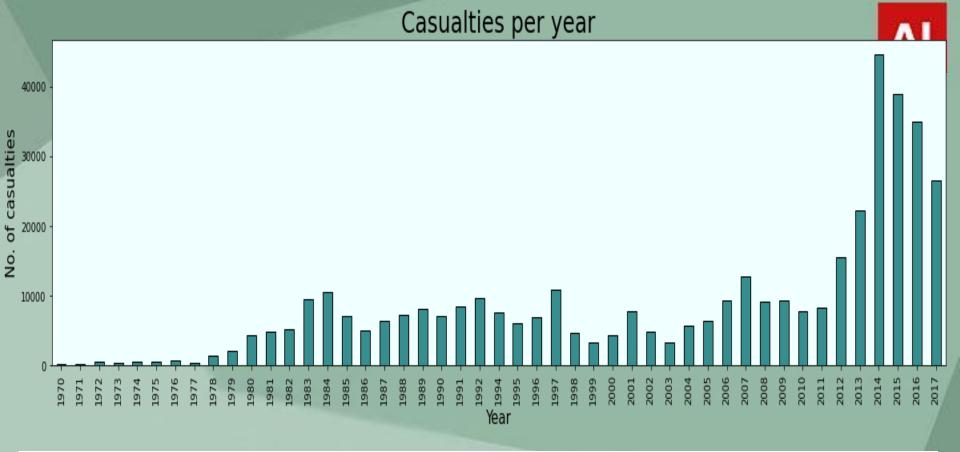
Unknown



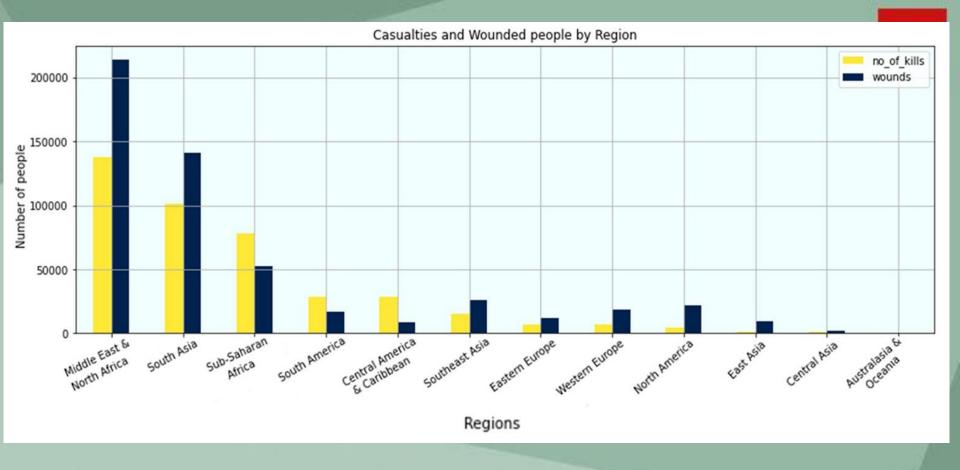
It is visible from the above donut chart that Firearms and Explosives are the most used attack types used by the deadliest gang Taliban. so, there should be stricter rules to prevent the movement of firearms and explosives into and from the countries.



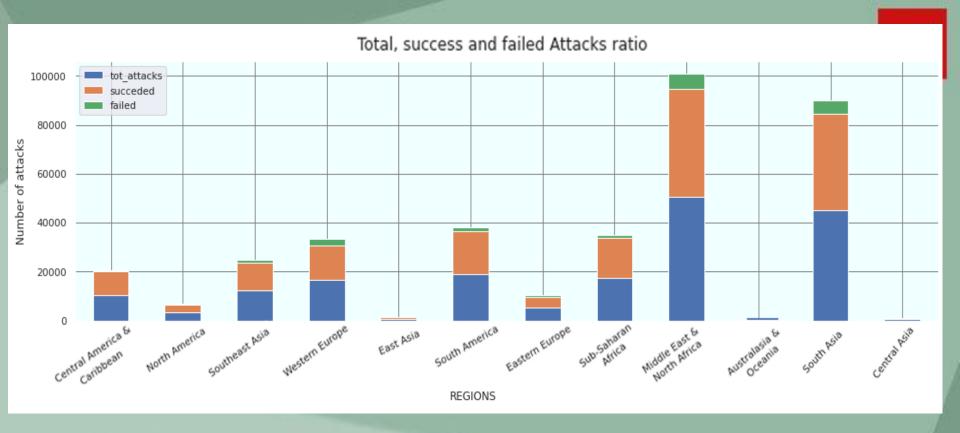
'Private Citizens and Property' are the constant target of the terrorist groups, tighter security and surveillance should be provided, especially in densely populated regions.



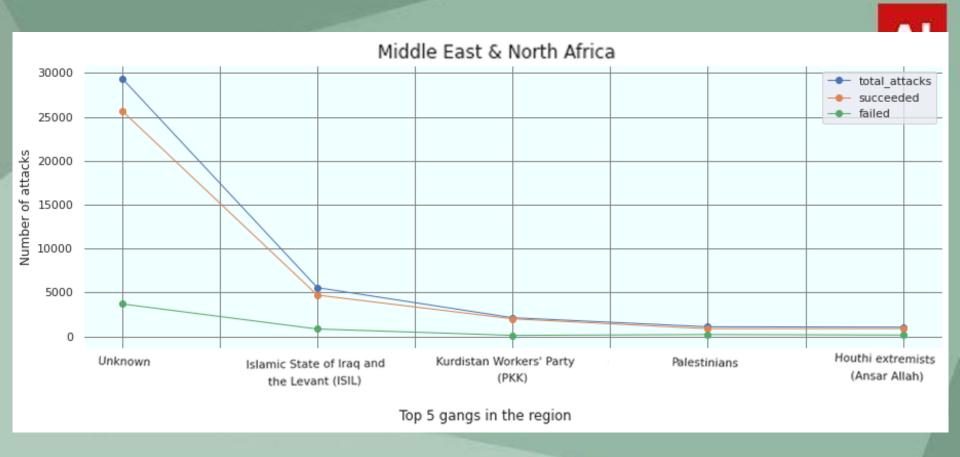
As clearly visible from the above graph, the year 2014 had the most number of casualties, with the most number of attacks followed by Year 2015 & 2016.



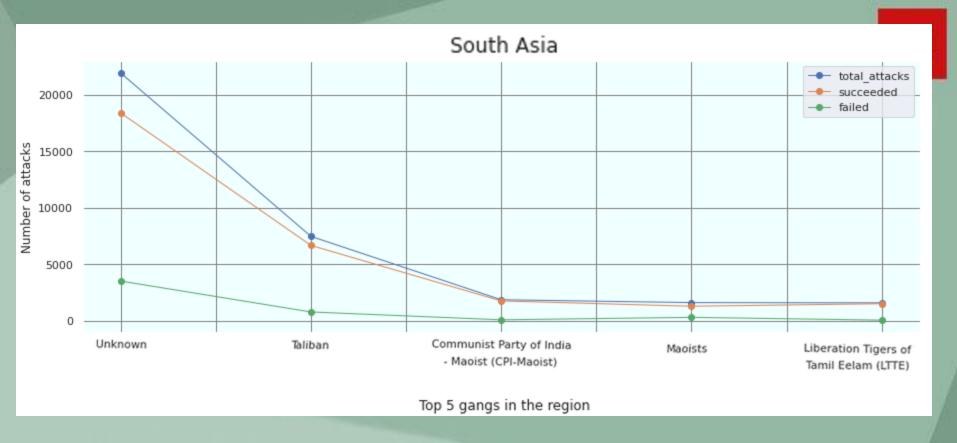
We can observe that there are more wounded people than casualties in all the regions except 'Sub-Saharan Africa', 'South America' and 'Central America & Caribbean', which have more casualties than wounded people.



The graph shows that the success rate of attacks is very high in all regions. Also, The highest number of attacks happened in the Middle East and North Africa followed by South Asia, whereas Central Asia and East Asia are the least attacked. Although in the Middle East and North Africa attacks that failed are high, the ratio of the success and failed attacks remains to be the same for other regions.



Above graph indicates that unknown groups are responsible for the most of the attacks and also success ratio seems to be very high as well.



South Asia which is the second most affected region also has unknown gangs active and highly successful followed by the Taliban, which also has a good success rate. Moreover, we found that no same group comes under the top 5 attackers when we compare data of the two most affected regions.

Al

CONCLUSION

- The year 2014 had the most number of terror attacks in the last decade. Approximately 17000 attacks in one year. This means that around 47 attacks were happening every single day during that year in multiple locations around the world.
- The Middle East & North Africa were the top affected region. 28.04 % of all events and a staggering 36.47 % of total casualties have been exclusively from these regions.
- Iraq has been the country with the highest number of attacks and 21.87% of all casualties have been from Iraq.
- Baghdad has been the most attacked city in the world, 7 of the cities are from Iraq and 4 of the cities are from Pakistan.



- Bombing/Explosion has been consistently the most popular method of attack over the last 5 decades with 47.7% of all attacks.
- Taliban has gained much prominence since 2012 and is now responsible for the most number of terror attacks, and the most common attack type used by them are firearms and explosives. This has been the same for most terrorist groups.
- Around 82000+ attacks were done by unknown groups of gangs, which is a major security concern.
- Private Citizens and Property are the most attacked targets, followed by the Military, Police, Government, Transportation, etc.



RECOMMENDATIONS

- Since Private Citizens and Property are being targeted consistently and explosives are the most favored weapon type of terror groups, tighter security and surveillance should be provided, especially in densely populated regions
- More surveillance is required especially in the Middle East & North African Regions.
- Stricter rules to prevent the movement of explosives into and from countries.
- Agencies need to be more active as around half of the attacks were done by unknown gangs and they need to be traced to stop them.
- Intelligence and Security needs to be updated as almost all attacks in the areas appear to be a very high in success rate.



REFERENCES

• Global Terrorism Analysis(GTA):

https://www.start.umd.edu/gtd/

• Python Pandas Documentation: https://pandas.pydata.org/pandas-docs/stable

• Python Numpy Documentation:

https://numpy.org/doc/

• Python MatPlotLib Documentation:

https://matplotlib.org/stable/index.html



THANKYOU