# PREDICTIVE ANALYSIS OF GLOBAL TERRORISM

## PROBLEM STATEMENT

Terrorism is one of the biggest concerns among all nations in the world. It threatens the lives of people by frequently doing attacks which involve violence, resulting in loss of lives and property. It is becoming very hard for the governments to predict these attacks. Hence, we collected data on previous attacks that happened between 1970 and 2017 across various regions in the world and analyzed it to provide governments with insights about the most likely groups that can may pose threat to nations in those regions, thereby helping governments in taking necessary steps to eradicate terrorism and mitigate the damage.

This project aims to use statistical methods and visualization techniques to identify the terror hotspots in the world based on the number of terrorist attacks that happened over the past few decades and predict which terrorist groups are highly likely to attack in these hotspots. Besides, we also performed Test of Hypothesis to prove/disprove the claim of recent terrorism statistics report which states that 77.2% of the attacks in Iraq, which ranked first in the Global Terrorism Index 2016, are done by using explosives.

## DATA TO BE COLLECTED

In order to get statistical information about the terrorist attacks all over the world, we used the readily available global terrorism dataset (GTD). This dataset contains 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 includes more than 180,000 records.

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. In practice this means in order to consider an incident for inclusion in the GTD,all three of the following attributes must be present:
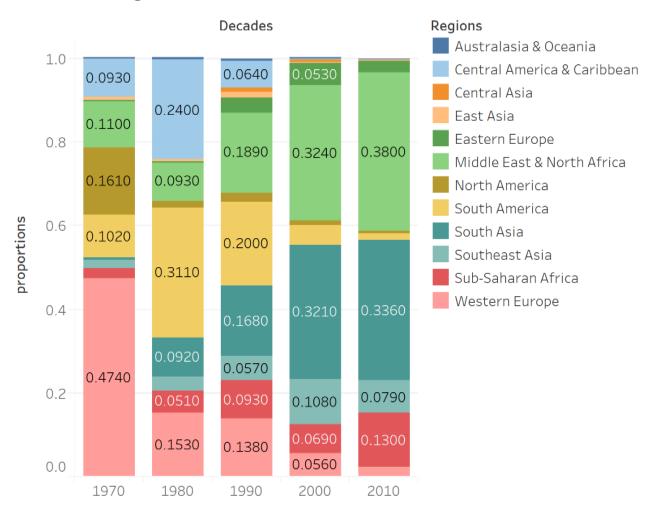
- The incident must be intentional – the result of a conscious calculation on the part of a perpetrator.

- The incident must entail some level of violence or immediate threat of violence - including property violence, as well as violence against people.

- The perpetrators of the incidents must be sub-national actors. The database does not include acts of state terrorism.
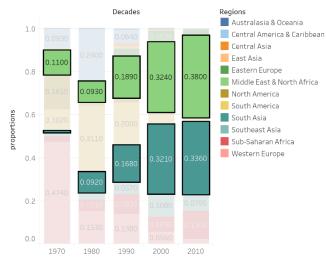
# DESCRIPTIVE STATISTICS

We exported the global dataset as .csv file into R studio and divided the years in the dataset into decades and calculated the proportion of incidents that occurred all over the world decade wise. The same data has been exported to Tableau and visualized as segmented bar graphs shown below.

## Proportions of incidents reported in different regions over decades



Sum of Proportions for each Decades. Color shows details about Regions. The marks are labeled by sum of Proportions. The view is filtered on Exclusions (Decades,Regions), which keeps 58 members.

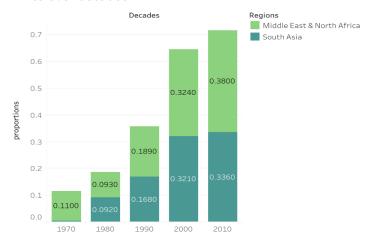Proportions of incidents reported in different regions over decades

Sum of Proportions for each Decades. Color shows details about Regions. The marks are labeled by sum of Proportions. The view is filtered on Exclusions (Decades,Regions), which keeps 58 members.



Proportions of incidents reported in South Asia, Middle East & North Africa over decades

Sum of Proportions for each Decades. Color shows details about Regions. The marks are labeled by sum of Proportions. The view is filtered on Exclusions (Decades,Regions) and Regions. The Exclusions (Decades,Regions) filter keeps 58 members. The Regions filter keeps Middle East & North Africa and South Asia.

From the above figures, we found out that in two regions, MENA (Middle East & North Africa) and South Asia, the number of terrorist attacks increased over the decades, while it decreased in the other regions. So, we wanted to focus our study in these two regions.
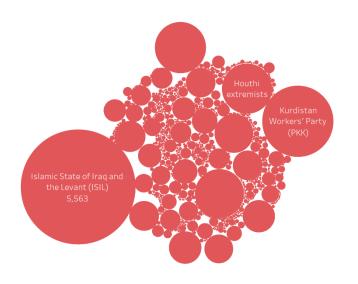
**SAMPLING METHOD**

We divided the population into clusters based on geographical location. Among those clusters, we chose two regions where the number of terrorist attacks is the highest. They are MENA (Middle East & North Africa) and South Asia.
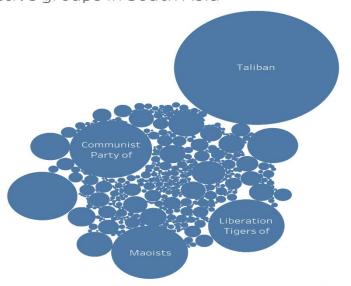
# ADVANCED ANALYTICS RESULTS

X is a random variable of the number of terrorist attacks made by a particular terrorist group. From the global dataset, we filtered the countries that fall in Middle East & North Africa (MENA) and South Asia regions along with the group name and the number of attacks made by them in that particular-country with the help of R studio. The same data has been exported to Tableau as .csv file and the following analysis was made on the regions.

Active groups in Middle East &North Africa
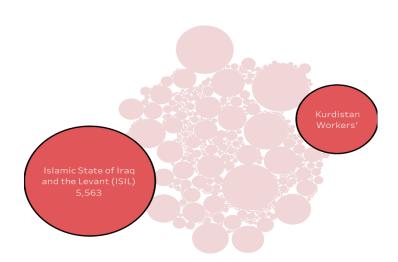
Top two active groups in Middle East &North Africa
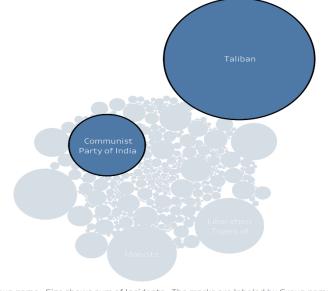


Houthi extremists

Kurdistan Workers' Party (PKK)

Islamic State of Iraq and the Levant (ISIL) 5,563

Islamic State of Iraq and the Levant (ISIL) 5,563

Kurdistan Workers'

Active groups in South Asia

Top two active groups in South Asia

Taliban

Communist Party of

Liberation Tigers of

Maoists

Taliban

Communist Party of India

Liberation Tigers of

Maoists

Group name. Size shows sum of Incidents. The marks are labeled by Group name.

Group name. Size shows sum of Incidents. The marks are labeled by Group name.

From the above graphs, it can be inferred that:

- **Islamic State of Iraq and the Levant (ISIL)** & **Kurdistan Workers' Party (PKK)** are the top two active terrorist groups in MENA region.

- **Taliban** & **Communist Party of India (CPI)** are the top two active terrorist groups in South Asia.

After this, we went on to find out which country in these regions are at high risk of attack by the above groups using conditional probability.

## Calculations

Events are defined as follows:

- 'A' be an event that a terrorist incident happened in a country in MENA and South Asia.

- 'I' be an event that the country was attacked by ISIL

- 'K' be an event that the country was attacked by Kurdistan Workers' Party

- 'T' be an event that the country was attacked by Taliban

- 'C' be an event that the country was attacked by Communist Party of India.

$P(I|A)$ is defined as the probability of finding that an attack was made by ISIL given that it occurred in country A

$P(K|A)$ is defined as the probability of finding that an attack was made by PKK given that it occurred in country A

$P(T|A)$ is defined as the probability of finding that an attack was made by Taliban given that it occurred in country A

$P(C|A)$ is defined as the probability of finding that an attack was made by CPI given that it occurred in country A

The following calculations are performed using MS Excel.

| Country (MENA) | ISIL | PKK | P(I\|A) | P(K\|A) |
|---|---|---|---|---|
| Bahrain | 2 | 0 | 0.00036 | 0 |
| Egypt | 3 | | 0.000539 | 0 |
| Iran | 2 | 2 | 0.00036 | 0.000936 |
| Iraq | 4797 | 23 | 0.862305 | 0.010763 |
| Israel | 1 | 0 | 0.00018 | 0 |
| Jordan | 6 | 0 | 0.001079 | 0 |
| Lebanon | 50 | 0 | 0.008988 | 0 |
| Libya | 2 | 0 | 0.00036 | 0 |
| Saudi Arabia | 15 | 0 | 0.002696 | 0 |
| Syria | 594 | 3 | 0.106777 | 0.001404 |
| Tunisia | 5 | 0 | 0.000899 | 0 |
| Turkey | 82 | 2109 | 0.01474 | 0.986898 |
| West Bank and Gaza Strip | 3 | 0 | 0.000539 | 0 |
| Yemen | 1 | 0 | 0.00018 | 0 |
| | | | | |
| Total number of attacks made | 5563 | 2137 | | |

| Country(South Asia) | Taliban | Communist Party of India | P(T\|A) | P(C\|A) |
|---|---|---|---|---|
| Afghanistan | 7423 | 0 | 0.992778 | 0 |
| India | 0 | 1877 | 0 | 0.999468 |
| Nepal | 0 | 1 | 0 | 0.000532 |
| Pakistan | 54 | 0 | 0.007222 | 0 |
| | | | | |
| Total number of attacks made | 7477 | 1878 | | |

From the above tables, it can be inferred that:

- If any terrorist incident occurred in **Iraq** there is 86.23% chances that it was done by ISIL.

- If any terrorist incident occurred in **Turkey,** there is 98.69% chances that it was done by Kurdistan workers' party.

- If any terrorist incident occurred in **Afghanistan** there is 99.3% chances that it was done by Talibans.

- If any terrorist incident occurred in **India** there is 99.9% chances that it was done by CPI.
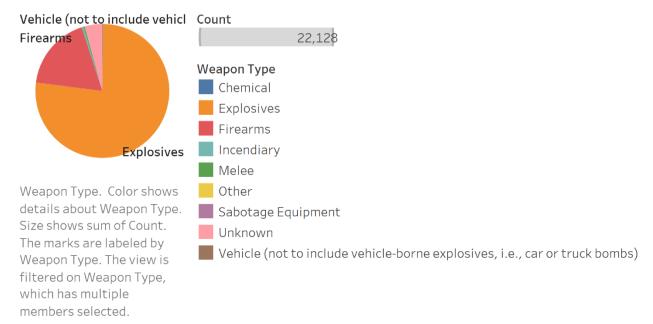
Therefore, Iraq and Turkey are the two countries which are at high risk of terrorist attack in Middle East and North Africa & Afghanistan and India are the two countries which are at high risk of terror attack in South Asia.

## STATISTICAL ANALYSIS

It is also very crucial to know what type of weapon is used by these groups to attack so that government agencies can take safety measures accordingly and strengthen the security forces in these regions. For this purpose, we found out the distribution of weapons used using Tableau as shown below.

We took a claim made by recent report on global terrorism that the percentage of attacks



## Distribution of weapon type

Vehicle (not to include vehicl Count
Firearms
22,128

Weapon Type
- Chemical
- Explosives
- Firearms
- Incendiary
- Melee
- Other
- Sabotage Equipment
- Unknown
- Vehicle (not to include vehicle-borne explosives, i.e., car or truck bombs)

Weapon Type. Color shows details about Weapon Type. Size shows sum of Count. The marks are labeled by Weapon Type. The view is filtered on Weapon Type, which has multiple members selected.
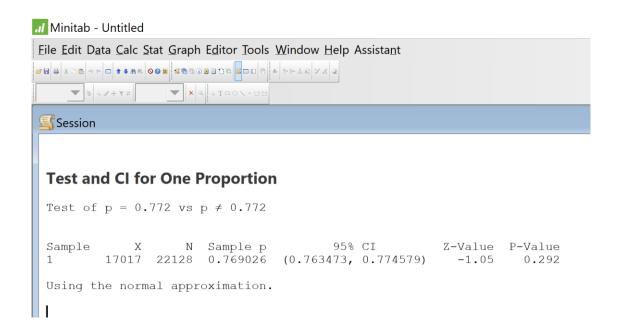
by the terrorists using explosives as weapon is 77.2% of the total number of attacks in Iraq between 1970 and 2016 and tested the same using hypothesis test.

## Results for confidence interval and hypothesis test

The total number of attacks in Iraq during that period is 22128 and we used 95% confidence interval to determine if hypothesis is true or not. The hypothesis is stated as :

{H0: P0 = 0.772, H1: P0 != 0.772

Clearly, this is a two-tailed hypothesis. The results were calculated using Minitab as shown below.

**Test and CI for One Proportion**

```
Test of p = 0.772 vs p ≠ 0.772


Sample      X       N   Sample p          95% CI          Z-Value   P-Value
1       17017   22128   0.769026   (0.763473, 0.774579)    -1.05     0.292

Using the normal approximation.
```

As p-value =0.292>alpha value (0.05), we fail to reject the null hypothesis and hence conclude that 77.2% of the attacks in Iraq were done using explosives.

## CONCLUSIONS

As we can see from the inferential analysis conducted above, Iraq and Turkey are the two countries in Middle East and North Africa having high chances of terror attacks by ISIL and PKK respectively and Afghanistan and India are the two countries in South Asian region having high chances of terror attack by Taliban and CPI respectively. There is more likely chance that explosive type of weapon will be used by ISIL in Iraq which already ranks top one in Global Terrorism Index 2016. These insights could be used by the government agencies of the above countries to monitor the activities of the above four deadly terror groups, tighten the security check of people travelling to these countries to ensure the safety of people living here.

## REFERENCES

- https://www.kaggle.com/START-UMD/gtd