**Capstone Project Submission**

**Name - Kailash Sahu**

**Email -** [**kailashsahu13@gmail.com**](mailto:kailashsahu13@gmail.com)

**Github - https://github.com/Kailash-13011992/Global-Terrorism-Data-Set-.git**

**Chapters:**

1. Data Wrangling**:**

* Dealing with missing values
* Reshaping the data
* Eliminating unknown values
* Filtering data for India

2. Number of countries affected by terrorist attacks

3. Success rate of terrorist attacks

4. Terrorist attacks per year worldwide

5. Terrorist attacks per country and per region

6. Favorite targets of terrorist groups

7. Number of killings happened in every in India

8. Most targeted city in India

9. Most active terrorist groups in India

Github Link:-<https://github.com/Kailash-13011992/Kailash-13011992>

**Short summary of your Capstone project and its components,problem statement,approaches**

**& Conclusions .**

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. The dataset set huge having 135 columns so we decided to divide our project work.

**SUMMARY:**

Initially we started looking at database available to us for the GTD project. We went through various columns available within GTD database and tried to make sense of what all data is available to us. We took useful columns and understood that it contained details of each and every terrorist attacks worldwide including India. So we quickly understood that we have some common statistics pertaining to terrorism.

We first decided to do the analysis worldwide but before we start there were some hurdles like as we imported the csv file it was big so we encoded to parse it and there were 135 columns in the Data frame so the first task was to remove null columns. Then calculated the percentage of null values in each column and dropped off the columns having more than 50% null values and there were a lot missing values as well so to be accurate at analysis we filtered those values. For the sake of proper analysis we renamed columns to a better relevant name according to the dataset.

During the analysis data of year 1993 was missing because of some unknown reasons. For the better understanding of the data analysis we used different graphs to visualize the data and chose the graphs according to the data. We analysed that from the year 1970 to 2017 the average number of attacks were around 990 attacks each year and the increase in the attacks was drastic that was 15 times and whenever an attack took place it's probability of getting successful was 89% which very high.

And these attacks were covering 205 countries around the world that is approximately 95% of the world so we can say that terrorist have taken place all around the world that not a great news to hear. The most active terrorist group was Taliban around the world which make them the most dangerous terrorist group and the most attacked country was Iraq which was around 25000 and the most attacked region was Middle East and North Africa followed by South Asia.

We found out some of the statistics such as Cities with the most number of killings by terrorist in India and also tried to find which organization areas mostly involve into terrorist activity. We quickly understood some statistics related to attacks and found that the most active terrorist groups are Sikh Extremists followed by CPI-Maoist causing the maximum number of deaths in India.

For the better understanding of the data analysis we used different graphs to visualize the data and chose the graphs according to the data. We analysed that in the year 1992 there were highest number of deaths in India and for better understanding we used scatter plot for better visualisation.

IN year 1992, 1150 people lost their lives. We can see an exponential growth in the number of attacks and it's the maximum in the year 2016 with 1025 attacks and this is peak year with respect to attack. However, Srinagar witness highest number kills almost 540 people loss their life followed by Mumbai and New Delhi.

**Conclusions:**

1. The success rate of these terrorist attacks is huge that is 89%, this is a very big issue of concern for the security agencies around the world

2. 205 countries in the world are affected by terrorism according to the Dataset.

3. The most number of attacks took place till 2017 are in Iraq approximately 25000 which is huge in last 50 years.

4. Number of terrorist attacks are more in Iraq, Pakistan and Afghanistan and region wise it is more in Middle East and North Africa regions.

5. The most targeted assets of these organizations are Private Citizens & Property followed by Military and Police.

6. Taliban is the terrorist organization that targeted Afghanistan the most.

7. Srinagar was the most frequently attacked city in India which caused 540 deaths.

8. The most active terrorist groups are Sikh Extremists followed by CPI-Maoist causing the maximum number of deaths in India.