Capstone Project Submission

Team Member's Name, Email and Contribution:

Team Member's Name	Email id
Syed Aquib	syedaquib153@gmail.com
Ganesh Subramanian	gsubramanian5@gmail.com
Mayur Chulbhare	chulbharemayur@gmail.com
Abhishek Kirar	abhishekkirar27@gmail.com
Ajinkya Shingote	shingoteajinkya65@gmail.com

Contribution:

Syed Aquib :
- Data Wrangling
    Loading and Pre-processing
    Structuring data
    Enriching data
- Data Mining
- Data Analysis
- Visualizations
    Bar graphs and Pie charts
- Observation
- Summarization
- Conclusions
- Technical Document
- Power Point Presentation

Ganesh Subramanian :
- Data Wrangling
    Loading and Pre-processing
    Structuring data
    Enriching data
- Data Mining
- Data Analysis
- Visualizations
    Bar graphs and Pie charts
- Observation
- Summarization
- Conclusions
- Technical Document
- Power Point Presentation

 Mayur Chulbhare :
- Data Wrangling
    Loading and Pre-processing
    Structuring data
    Enriching data
- Data Mining
- Data Analysis

- Visualizations
  - Bar graphs and Pie charts
- Observation
- Summarization
- Conclusions
- Technical Document
- Power Point Presentation

Abhishek Kirar :
- Data Wrangling
  - Loading and Pre-processing
  - Structuring data
  - Enriching data
- Data Mining
- Data Analysis
- Visualizations
  - Bar graphs and Pie charts
- Observation
- Summarization
- Conclusions
- Technical Document
- Power Point Presentation

Ajinkya Shingote :
- Data Wrangling
  - Loading and Pre-processing
  - Structuring data
  - Enriching data
- Data Mining
- Data Analysis
- Visualizations
  - Bar graphs and Pie charts
- Observation
- Summarization
- Conclusions
- Technical Document
- Power Point Presentation

Please paste the GitHub Repo link.

Github Link:-

https://github.com/syedaquib153/GTD-data-analysis

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

Project Name: The Global Terrorism Database (GTD).

Problem Statement:

Global Terrorism Dataset provided to us is in unformatted manner, uneven data, and duplicate data and also some data columns in it is irrelevant, since it's a piled-up data coming from various different countries. For doing the analysis on the data the data needs to be in correct format and well organized formed. The main objective of the analysis is to obtain the meaning full insights and facts from the given huge datasets.

Approach:

As a team we read the data present in the file and gone through the details in each and every column. The data set was huge in which some of the data was not required for the analysis so the data was cleaned by dropping some unwanted columns and created new data frame, with the columns we required for the analysis. The first problem we faced was the name of the columns which was not proper and the nan values present in the data. We renamed the columns by using dictionary format and replaced all the nan values to zero in int dtype and unknown in object dtype by using replace syntax. Each and every column were compared to gain the insights about the data. We worked individually to gain some insights by doing the exploratory data analysis using python. Cleaning the dataset, analysing the data and visualizing the data by plotting the data into different graph and charts so that the trend and relationship between the various indicators can be understand easily.

Conclusion:

- ❖ The year in which maximum attacks took place was during the year 2014.
- ❖ The attacks were seen gradually increasing from 2005 to 2014
- ❖ The country which was attacked the most are Iraq, Pakistan, Afghanistan and India
- ❖ Different kind of attacks were done the major was through Bombing/Explosion as it gives higher damage and Armed Assault
- ❖ The Middle East and North Africa was the most affected region
- ❖ The most affected city in the world is Baghdad and most affected Indian City is Srinagar
- ❖ We observed that attack mostly takes place on weekdays(i.e Mon, Tue, Wed, Thu) compared to weekends (i.e Fri, Sat, Sun)
- ❖ When we analysed the attack monthwise we observed that every month has equal contribution and when analysed date wise most of the attack took place on 1st and 15th of the month and minimum attack at the end of the month
- ❖ Overall 171378 people were killed during the attack and Iraq country has the highest count of the people killed.
- ❖ The motive behind the attacks was some unknown means only reason behind attack is to spread terrorism worldwide and some of them are "Protest the Vietnam War and the R.O.T.C" & "protest the War in Vietnam and sabotage the draft".
- ❖ The most common Targets are Private Citizens & Property followed by Military   sectors   and Police
- ❖ Taliban is the most notorious group