

Capstone Project Submission

Instructions:

- i) Please fill in all the required information.
- ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

Name: Soma Pavan Kumar

Email: spkumar1998@gmail.com

Contribution:

1. Data Importing
2. Data Cleaning
3. Getting the Statistics of Data
4. Data Visualizations
 1. Number of Attacks according to the year
 2. Number of attacks with their respective countries
 3. Number of attacks according to their months
 4. Found out what type of attack is used by most of the terrorists
 5. Also Found out what kind of population is more susceptible for terror attacks
5. Conclusion

Please paste the GitHub Repo link.

Github Link:- <https://github.com/PavanKumar181098/Global-Terrorism-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)

Write here the short summary

The Global Terrorism Database (GTD) is an open-source database that contains information on terrorist attacks around the world from 1970 through 2017. The GTD includes systematic data on domestic as well as international terrorist attacks that occurred during this time period and 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)

For the First step, the data is imported from a csv file and converted to a Pandas DataFrame. Pandas is an in-built library in python that is used to handle and manipulate DataFrames. A DataFrame is basically a collection of Rows and Columns where the Columns represent the various features of the dataset and the Rows represent the terrorist attacks that took place.

It was found that there were many columns(features) that were either empty or did not contain any important information. Hence, the Data is trimmed of all the columns that contain either null values or which are irrelevant to our analysis.

Once the data is clean, we check for the statistics of the cleaned data. The statistics of the data tells us the mean, median (basic descriptive data) and the distribution of the data and some more information of the columns of the dataset.

After the data is cleaned, different Visualizations can be done on the data and many inferences can be made from the visualizations.

One such visualization is to find the trend of the number of terror attacks per year. We find out that there is a sharp rise in attacks after 2007 till 2014. After 2014 the number of terror attacks falls. We find out that 2014 was the year where most of the attacks happened, with over 20,000 attacks that year.

Another such visualization is to find out the top 5 countries that where most of the terrorist attacks happened and we find out that Iraq, Pakistan, Afghanistan, India and Columbia are the top 5 attacked countries with almost 25000 attacks in Iraq.

Similarly, we can plot the frequency of attacks according to their months and we found that there is no correlation between the month of the year and attacks.

We plotted the weapon and type of attack that were most used by the Terrorists in their attacks and found that Bombing/Explosion was the most used type of attacks. Taliban is the organization responsible for most of the attacks.

At last, by plotting the various targets on which the terrorist attack, we find that Private People and private property are the most attacked targets which was obvious as they are unarmed and most susceptible and last on their targets is Hijacking/Kidnapping.

In the end, It was a good Experience working on the terrorist Data, But one downside with exploring these datasets is, the reason why these attacks happen are mostly political and neither can be explained nor be predicted by the most of Machine learning Algorithms. But many conclusions can be made from the various visualizations.

