**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| * **Name :** Vishal Rahul Pagare * **Email :** [pagarevr4103@gmail.com](file:///D:\Global%20terrerisom\pagarevr4103@gmail.com) * **Contribution :** Contributed innotebook   data cleaning, data manipulation, exploratory data analysis, finalizing the conclusion and visualization with help in ppt and technical documentation   * **Name :** Sachin Shivaji Gawande * **Email :** [sachingawande4599@gmail.com](mailto:sachingawande4599@gmail.com) * **Contribution :** Contributed in notebook in exploratory data analysis and visualization with help in technical documentation and ppt * **Name :** Rushikesh Sunil Sharma * **Email :** [rishi77714@gmail.com](file:///D:\Global%20terrerisom\rishi77714@gmail.com) * **Contribution :** Contributed in notebook exploratory data analysis and help in choosing the dataset * **Name :** Sandesh Sundarlal Arsud * **Email :** [arsudsandesh97@gmail.com](file:///D:\Global%20terrerisom\arsudsandesh97@gmail.com) * **Contribution :** Contributed in notebook exploratory data analysis and data visualization, finalizing observation and conclusion with making ppt all points covered, help in technical documentation |
| **Please paste the GitHub Repo link.** |
| Github Link:-  <https://github.com/vishal120700/Global_Terrorism_Analysis_Capstone_-Project_EDA> |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| **Exploratory Data Analysis on Global Terrorism :**  In 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 database is maintained by researchers at the National Consortium for the Study of Terrorism and Responses to Terrorism (START), headquartered at the University of Maryland.  We Explore and analyze the data and discover the terrorist activities. In this project, we focus on terrorism by analyzing the dataset . to explore meaningful patterns and statistics.  The problem statement here is to build a tool that can present processed information in the form of intuitive visual representation of analyzed data. Implementation of this project involves system design, backend design, visual design, and user interface.  System design includes the overall design plan of the whole project system which explains how each individual module is correlated with others.  Visual design mostly consists of analyses and visualization techniques to construct different graphics representing the end results in an easy-to-interpret format  Most of the operations on the dataset are done by python on google colab Notebook which is an open-source web-based application for python code development. Python is used for data preprocessing, data modeling, analyses, and visualization.  **Data preprocessing :**  Data preprocessing is the first step to be done after collecting data. It is a set of operations performed on the START (Study of Terrorism and Response to Terrorism) dataset to modify ambiguous data which can be a bottleneck to analytical results. Raw data is simply a collection of related information put together. Raw data is often unorganized and contains a lot of information which is irrelevant to the project requirements. Data preprocessing methodology helps in converting this raw data into a more meaningful, focused, interpretable and readable format.  Available START dataset from the Global Terrorism Database is incomplete, inconsistent, contains many errors, missing attributes values, contains outliers, incorrect  tags, and duplicate entries. Data preprocessing can help resolve these discrepancies.  **Python packages :**  Following are some of the python packages used in this project.  **Pandas:**  It is used for data analysis and manipulation. Pandas can convert data structures and dataset formats to data frames on which operations like loading data, rename attributes, mapping, crosstab, sub-data frames, plotting, etc. can be performed.  **NumPy :**  It provides structures for multiple dimensional array objects and tools for related operations. NumPy is usually used for high performance scientific computational tasks.  **Matplotlib :**  It is a 2D based plotting package that provides required modules and  functions. A developer can customize font properties, styles, axes properties, etc.  **Number Of Attacks happening in Year and Analyse total Number of people killed**  **Middle East & North Africa :**  Most attacked and the large number of people killed in this region  The Middle East & North Africa leads 1st among all the regions and then South Asia takes 2nd place  **Analyse the most attacks by Year ?**  in 2014 : The large number of attacks took place in this Year  in 1971 & 1973 : The less number of attacks happening in this Years  Attacks were more during 2014 and then in 2015. When compared to attacks from 1970 onwards, the last 6 years scored a maximum. But from 2014 onwards count started decreasing. |