**RAMNIRANJAN JHUNJHUNWALA COLLEGE**

**Department of DSAI**

**Ghatkopar (West), Mumbai - 86**



**2021-2022**

**Project Report  
On**

**Global Terrorism Exploratory Data Analysis**

**In partial fulfillment of M.Sc. (DSAI)**

**By**

**Mr. Madhusudan Tiwari**

**Project Guide**

**Prof. Bharati Bhole**

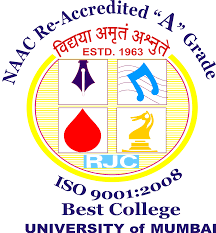
**RAMNIRANJAN JHUNJHUNWALA COLLEGE**

**(AUTONOMOUS)**

**(Affiliated to University of Mumbai)**

**GHATKOPAR(WEST), 400086.**

**Certificate**



This is to certify that the Project entitled, “Global Terrorism Exploratory Data Analysis” is bonafide work of **Mr. Madhusudan Tiwari** bearing **Seat No: - 38** submitted in partial fulfilment of the requirements for the award of Degree Master of Science in DSAI,

Signature of Internal Guide Sign of Co-Ordinator

Examiner

Date: College Seal

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **SR.NO** | **TOPIC** | PAGE NO |
| 1 | ACKNOWLEDGEMENT | 4 |
| 2 | ABSTRACT | 5 |
| 2 | **INTRODUCTION** | 6 |
| 3 | OBJECTIVE & PROPOSED SYSTEM | 7 |
| 4 | **EDA AND DESCRIPT** | 8 |
| 5 | **GLOBAL TERRORISM DATABASE** | 8 |
| 6 | WHAT IS TERRORISM? | 9 |
| 7 | GANTT CHART | 10 |
| 8 | FUNCTIONALITY AND SCOPE | 11 |
| 9 | SRS | 12 |
| 10 | DETAILS OF THE GTD | 13 |
| 11 | **AUTOMATED PROCESSES** | 14 |
| 12 | DATA PREPROCESSING | 14 |
| 13 | RESEARCH METHODOLOGY | 15 |
| 14 | DETAILS OF GLOBAL TERRORISM DATASET | 16 |
| 15 | PLOT TERRORISM BASED ON DIFFERENT CATEGORIES | 17 |
| 16 | DATA VISUALIZATION USING POWR BI | 23 |
| 17 | POWER BI DASHBOARD | 24 |
| 18 | INSIGHTS & ANALYSIS  PROJECT DETAILS | 26 |
| 19 | CONCLUSION & SUGGESTION | 27 |
| 20 | REFERENCE | 28 |
| 21 | PROJECT CODE LINK | 28 |

**ACKNOWLEDGEMENT**

Before we get into thick of things, we would like to add a few heartfelt words for the people who were part of Global Terrorism Exploratory Data Analysis.

project in numerous ways, people who gave unending support right from the stage the project idea was conceived. A project report is such a comprehensive coverage; it would not have been materialized without the help of many.

The four things that go on to make a successful endeavour are dedication, hard work, patience and correct guidance. Able and timely guidance not only helps in making an effort fruitful but also transforms the whole process of learning and implementing into an enjoyable experience.

In particular, I would like to thank our principal Dr. **Himanshu Dawda** & R.J college.

I would like to give a very special honour and respect to our teacher, **Prof.** **BHARATI BHOLE** who took keen interest in checking the minute details of the project work and guided us throughout the same. A sincere quote of thanks to the non-teaching staff for providing us software & their time.

**ABSTRACT**

This project addresses the problem of global terrorism is Compared to most types of criminal violence terrorism poses special data collection challenges. In response, there has been growing interest in open-source terrorist event data bases.

One of the major problems with these data bases in the past is that they have been limited to international events—those involving a national or group of nationals from one country attacking targets physically located in another country. Past research shows that domestic incidents greatly outnumber international incidents. In this paper we describe a previously unavailable open-source data base that includes some 70,000 domestic and international incidents since 1970.

We began the Global Terrorism Database (GTD) by computerizing data originally collected by the Pinkerton Global Intelligence Service (PGIS).

Following computerization, our research team has been working for the past two years to validate and extend the data to real time. In this paper, we describe our data collection efforts, the strengths and weaknesses of open-source data in general and the GTD in particular, and provide descriptive statistics on the contents of this new resource.

**INTRODUCTION**

The Global Terrorism Database (GTD) documents more than 190,000 international and domestic terrorist attacks that occurred worldwide since 1970. With details on various dimensions of each attack, the GTD familiarizes analysts, policymakers, scholars, and journalists with patterns of terrorism.

The GTD defines terrorist attacks 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. For each GTD incident, information is available on the date and location of the incident, the weapons used and nature of the target, the number of casualties, and — when identifiable — the identity of the perpetrator.

Compared to most types of criminal violence, terrorism poses special data collection challenges.

In response, there has been growing interest in open-source terrorist event data bases. One of the major problems with these data bases in the past is that they have been limited to international events — those involving a national or group of nationals from one country attacking targets physically located in another country.

Past research shows that domestic incidents greatly outnumber international incidents.

**PROPOSED SYSTEM**

1. In the proposed system, we are going to use the Global Terrorism Database to analyze and derive insights on the various terrorism acts that have taken place. We will pre-process the database to get reliable and accurate patterns and insights and use descriptive statistics to draw conclusions such as — major causes of terrorism, major groups behind terrorism, categories of the targets, locations of terrorist attacks.

2. Using EDA, we will show the trend of the growth of terrorism in the world and how it spread to various parts of the world. Using graphs, we study how one factor behind terrorism is interrelated to the other.

3. We also do predictive analytics as to what the magnitude of the terrorism acts can be in the future years.

**OBJECTIVE OF ANALYSIS AND STUDY**

The GLOBAL TERRORISM INDEX defines **Terrorism** 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”.

**The objectives of this project are:**

* To study and analyze the different terrorism acts that have taken place in various parts of the world, at different times.
* The various causes behind terrorism acts.
* Different types of terrorism acts.
* Categories of terrorism acts.
* Distribution of terrorism acts.

**EDA & DESCRIPTIVE ANALYTICS**

**Descriptive analytics** is the interpretation of historical data to better understand changes that have occurred in a business. **Descriptive analytics** describes the use of a range of historic data to draw comparisons. A statistical method that is used to search and summarize historical data in order to identify patterns or meaning. Preliminary stage of data processing that creates a summary.

**Exploratory data analysis**(**EDA**) is an approach to analyzing data sets to summarize their main characteristics, often with visual methods. EDA is for seeing what the data can tell us beyond the formal modeling or hypothesis testing task. It’s where the researcher takes a bird’s eye view of the data and tries to make some sense of it.

**GLOBAL TERRORISM DATABASE**

**About the dataset**

The Global Terrorism Database (GTD) is an open-source database including information on domestic and international terrorist attacks around the world from 1970 through 2017, and now includes more than 181,691 cases. For each event, information is available on the date and location of the incident, the weapons used and nature of the target, the number of casualties, and–when identifiable–the group or individual responsible. There are 18,1691 records and 135 columns including date, time, location, number of hostages, killed, wounded, if there was a ransom, the outcome, if there was a suicide attack, claims, weapons used. Unit of analysis: Attack. The number of variables is > 100 on location, tactics, perpetrators, targets, and outcomes.

**WHAT IS TERRORISM?**

1. Terrorism is defined in the Oxford Dictionary as “the unlawful use of violence and intimidation, especially against civilians, in the pursuit of political aims.” The key problem is that terrorism is difficult to distinguish from other forms of political violence and violent crime, such as state-based armed conflict, non-state conflict, one-sided violence, hate crime, and homicide.

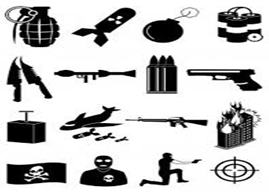
2. In our coverage of terrorism, we rely strongly on data from the Global Terrorism Database (GTD), which defines terrorism as “acts of violence by non-state actors, perpetrated against civilian populations, intended to cause fear, in order to achieve a political objective.”

3. To be considered an act of terrorism, an action must be violent, or threaten violence. As such, political dissent, activism, and nonviolent resistance do not constitute terrorism. There are, however, many instances around the world of authorities restricting individuals’ freedom of expression under the pretext of counter-terrorism measures. Human rights groups, such as Amnesty International and Human Rights Watch, publish reports on such cases of censorship.

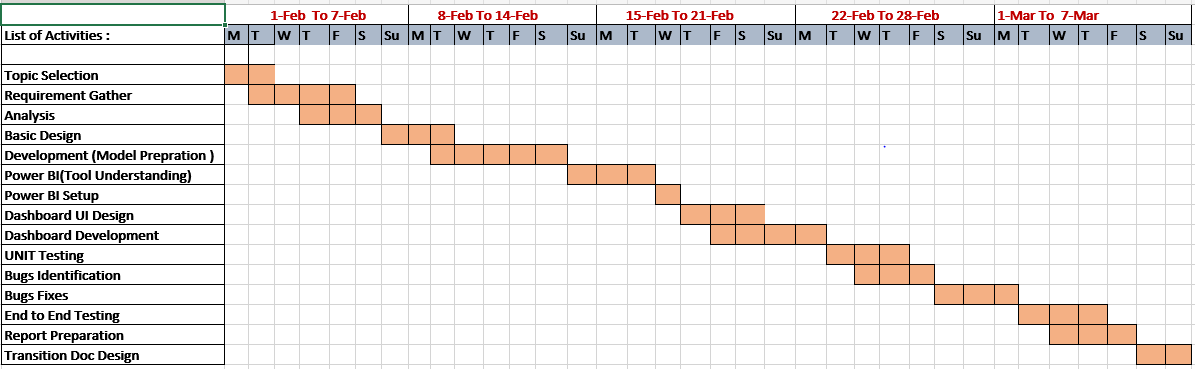
4. The inclusion of damage to private and public property in the definition of terrorism is a point of contention, but it is generally accepted in legal and statistical contexts.

5. An action must also be carried out for political, economic, religious, or social purposes to count as terrorism. For example, the terrorist organisation Islamic State of Iraq and the Levant (ISIL) has clearly stated its political goal to establish itself as a caliphate. Likewise, attacks perpetrated by white extremists have discernible socio-political motivations, and so are considered acts of terrorism. By contrast, violent acts committed without a political, economic, religious or social goal are not classified as terrorism, but instead as ‘violent crimes’

6. To be classified as terrorism, actions must be designed to have far-reaching psychological repercussions beyond the immediate victim or target. In other words, an action must aim to create terror through “it’s shocking brutality, lack of discrimination, dramatic or symbolic quality and disregard of the rules of warfare”.



**Gantt Chart**

****

**Functionality and Scope**

* We can get the different types of python libraries which are as follows:
* **Pandas**
* **NumPy:**
* Matplot
* Seaborn
* Pyplot
* Line interpolation / smoothing
* Bi-polar bar chart
* Simple pie chart
* Pie chart with custom labels
* Animating a Donut with SVG animate
* We can also get data which is available on Kaggle.
* We will able to find out in which area the tweets are more trending with help of world map.
* We can visually represent all insightful data with the help of creating dashboard.

**SRS (Software Requirement Specification)**

**Internal Interface Requirement:**

The recent explosion in data pertaining to users on social media has created a great interest in performing data analysis on this data using Dataset and Machine Learning principles to understand terrorism across all world. This project intends to perform the same tasks. The difference between this project and other analysis tools is that, it will perform real time analysis of data and create attractive dashboard. Describe the context and origin of the product being specified in this SRS. For example, state whether which terrorism rate is higher in which country or state. If the SRS defines a component of a larger system, relate the requirements of the larger system to the functionality of this software and identify interfaces between the two. A simple diagram that shows the major components of the overall system, subsystem interconnections, and external interfaces can be helpful. The data functions are:

* Collect data from Kaggle website
* Remove redundant information from these collected data
* Do the development on Jupyter notebook
* Perform exploratory data Analysis on the global terrorism dataset to classify their nature viz. positive, negative and so on.

**External Interface Requirement:**

We classify External Interface in 3 types, those are:

**User Interface:** Describe the logical characteristics of each interface between the software product and the users. This include sample screen images, GUI standards or product family style guides that are to be followed, screen layout constraints, standard buttons and functions that will appear on every screen, keyboard shortcuts, error message display standards, and so on. Define the software components for which a user interface is needed.

**Hardware interface:** Describe the logical and physical characteristics of each interface between the software product and the hardware components of the system. This include the supported device types, the nature of the data and control interactions between the software and the hardware, and communication protocols to be used.

**Software Interface:** Describe the connections between this product and other specific software components (name and version), including databases, operating systems, tools, libraries, and integrated commercial components. Identify the data items or messages coming into the system and going out and describe the purpose of each. Describe the services needed and the nature of communications. If the data sharing mechanism must be implemented in a specific way (for example, use of a global data area in a multitasking operating system), specify this as an implementation constraint.

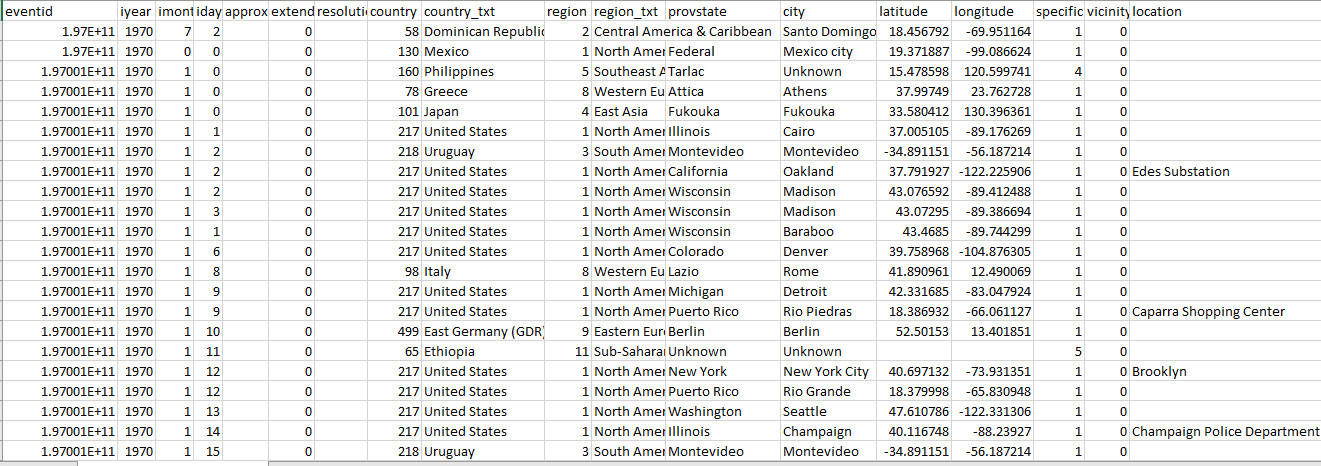
Python (Pandas Libraries) – Backend process

Power Bi– Data visualization

Dataset – Kaggle

**DETAILS OF THE GTD**

* Contains information on over 181,691 terrorist attacks
* Currently the most comprehensive unclassified database on terrorist attacks in the world
* Includes information on more than 91,000 bombings, 20,000 assassinations, and 13,000 kidnappings and hostage events since 1970
* Includes information on at least 45 variables for each case, with more recent incidents including information on more than 120 variables
* More than 4,000,000 news articles and 25,000 news sources were reviewed to collect incident data from 1998 to 2017 alone



**Automated Processes**

* Initial Boolean filters of articles.
* Remove duplicate articles
* Location identification
* Review relevant source articles; apply GTD inclusion criteria to identify unique terrorist attacks.
* Populate database with attack characteristics according to established coding rules.
* Present analysts with high-validity, topically clustered source articles.

**Data pre-processing**

The Data contains a total of 135 columns, which are both numeric and text. Now the since the project contains different types of Graphs and different aspects of it have been explored so different type of data pre-processing has been done before plotting the respective graph, which could be making of a new column or a new data frame itself.

**In order to calculate the casualties in the dataset I introduced:**



**RESEARCH METHODOLOGY**

**Exploratory Data Analysis**

Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

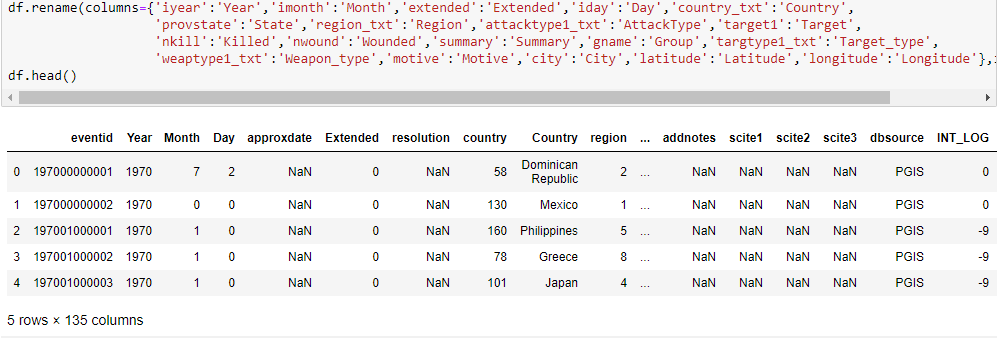
Exploratory Data Analysis (EDA) is an approach/philosophy for data analysis that employs a variety of techniques (mostly graphical) to

* maximize insight into a data set
* uncover underlying structure
* extract important variables
* develop parsimonious models
* determine optimal factor settings.

**Details of Global Terrorism Dataset**

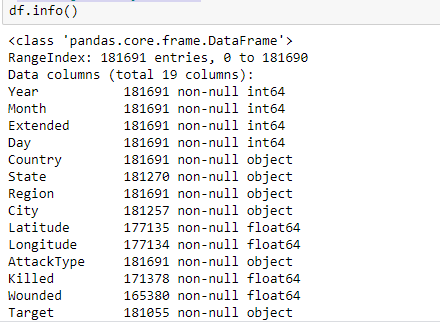
**Renaming Columns for our understanding and Deleting unwanted columns.**

rename():- Pandas **rename()** method is used to rename any index, column or row. Renaming of column can also be done by dataframe.columns = [#list]. But in the above case, there isn’t much freedom. Even if one column has to be changed, full column list has to be passed. Also, the above method is not applicable on index labels.



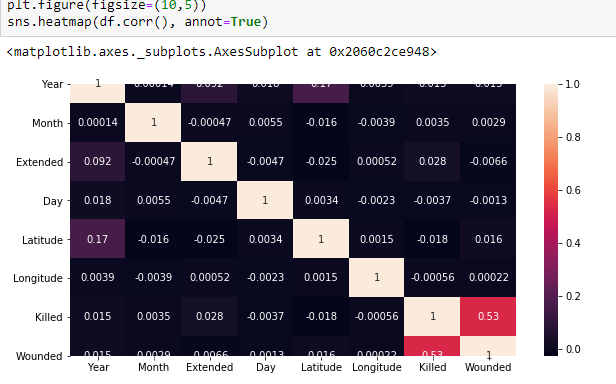
**Checking the information**

Info() :- The info() function is used to print a concise summary of a Data Frame. This method prints information about a Data Frame including the index dtype and column types, non-null values and memory usage. Whether to print the full summary. By default, the setting in pandas.

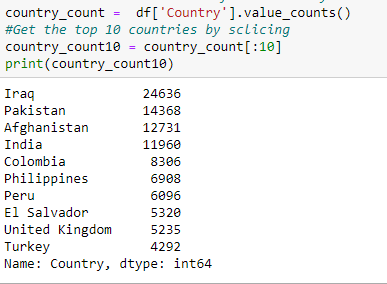


**PLOT TERRORISM BASED ON DIFFERENT CATEGORIES**

* **Plotting heatmap**

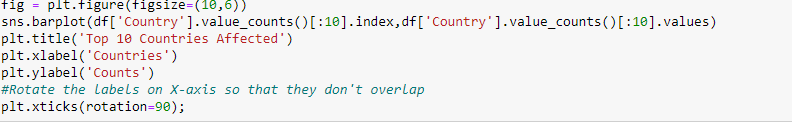


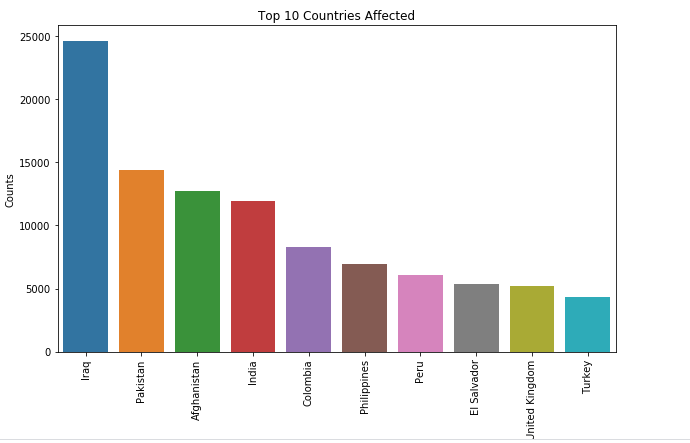
* **Countries which are affected most.**



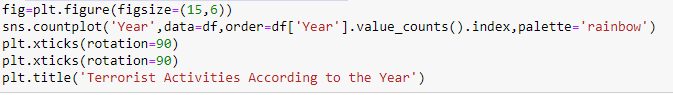
* **Create Figure (empty canvas)**

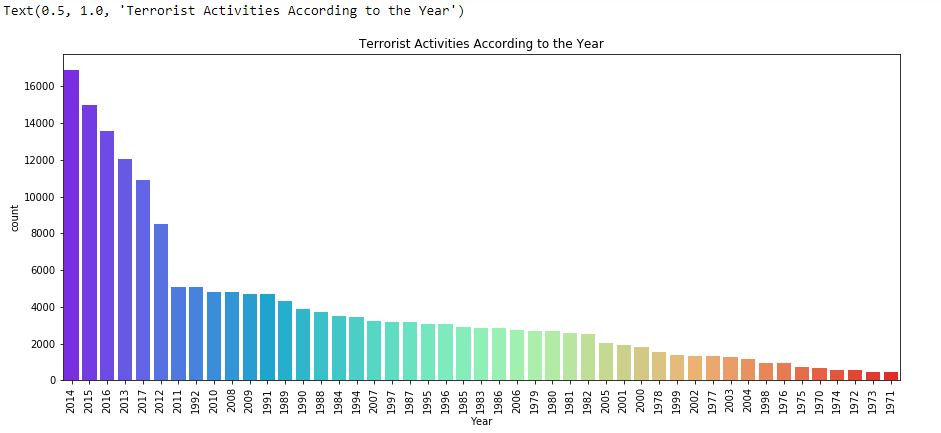
**Top 10 Countries Affected Most in Graph.**



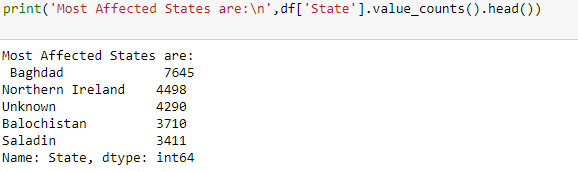


* **Terrorist Activities According to the Year**



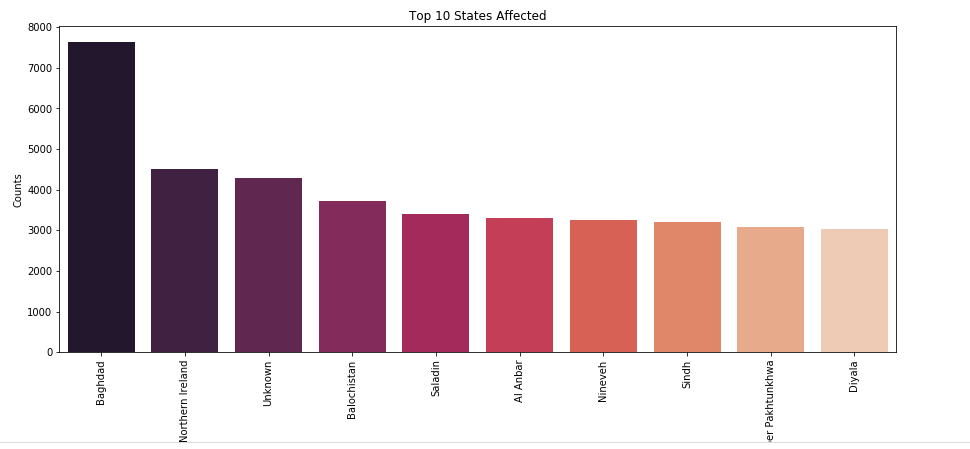


* **Most Affected State**

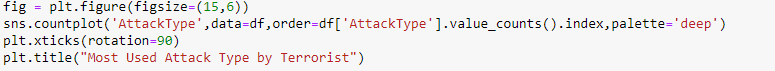


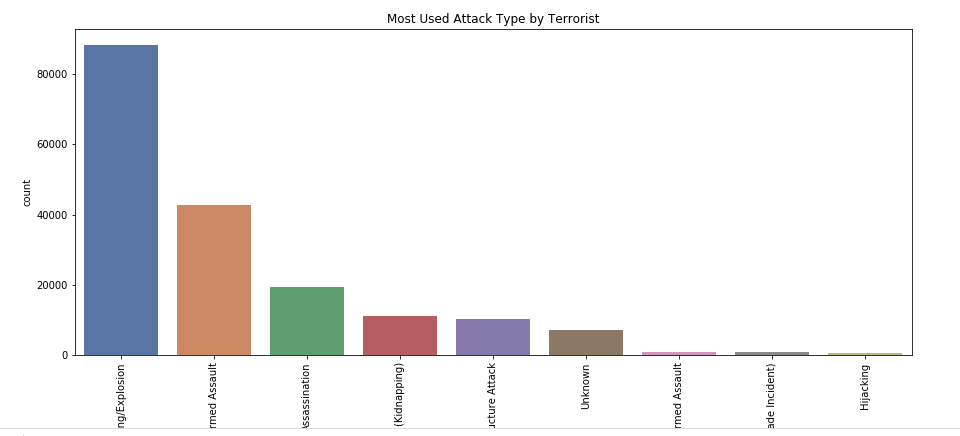
* **Top 10 State Affected Most in Graph.**



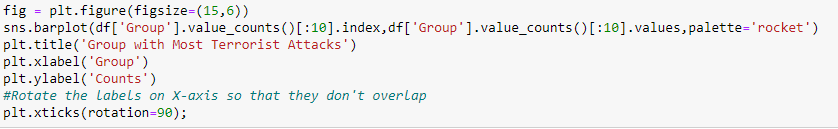


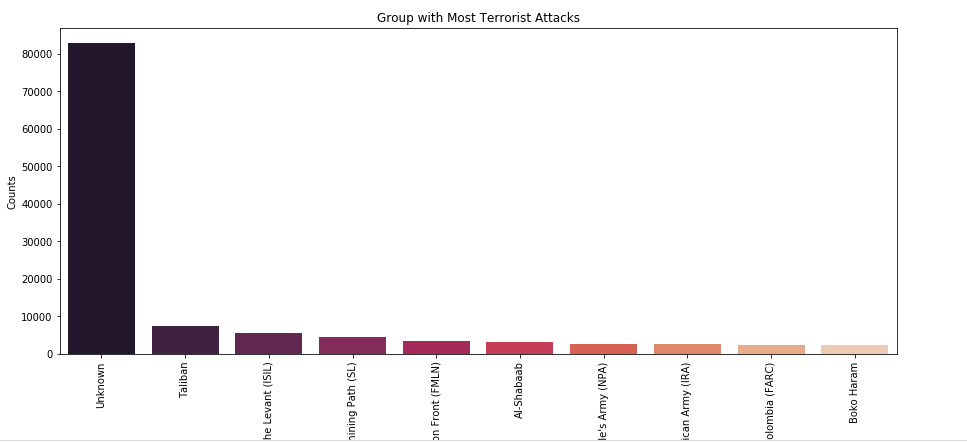
* **Most Used Attack Type by Terrorist**



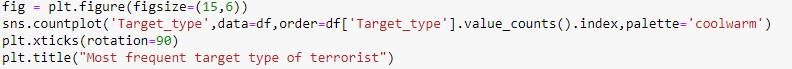


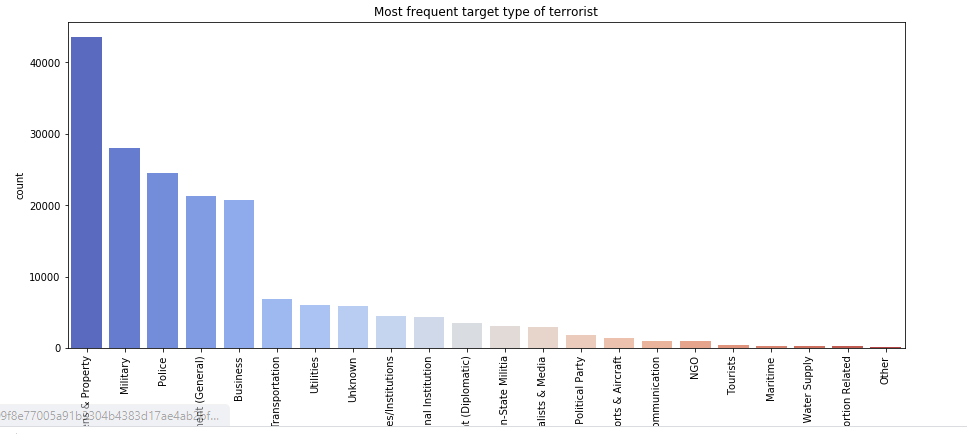
* **Groups with most terrorist attacks**



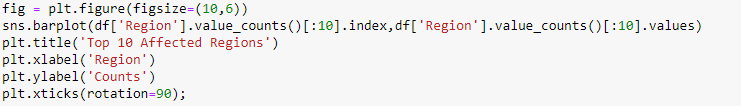


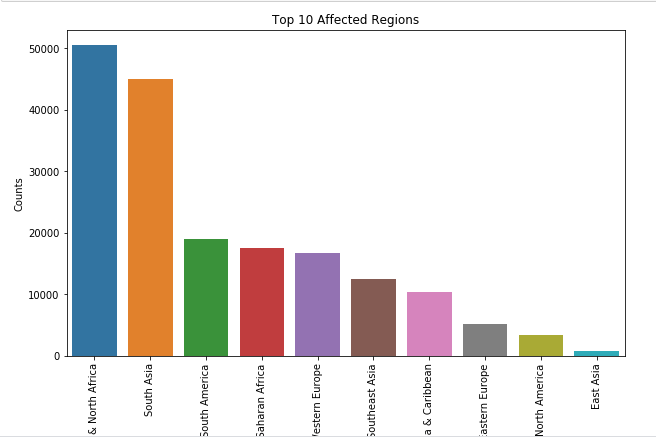
* **Target type of terrorist**





* **Most affected region**





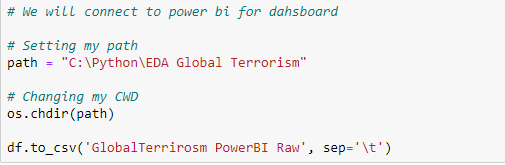
**DATA VISUALIZATION USING POWR BI**

**What is Power BI?**

Power BI is a business analytics service by Microsoft. It aims to provide interactive visualizations and business intelligence capabilities with an interface simple enough for end users to create their own reports and dashboards.

Here I have performed EDA (Exploratory Data analysis) using python script and extract insightful data in power bi to create the dashboard.

* **Python Script for connecting power BI instance.**

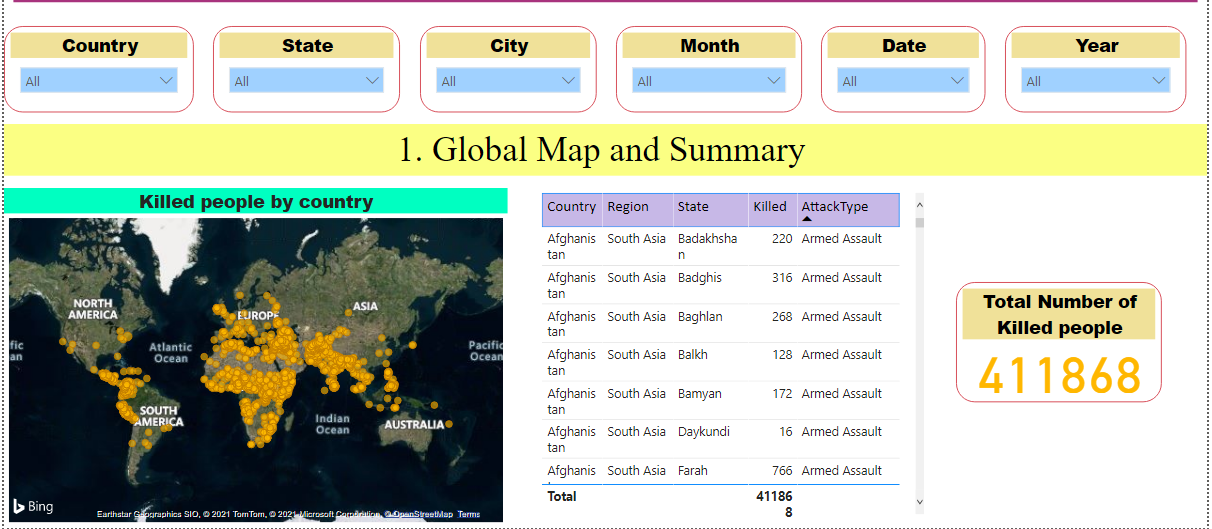


**POWER BI DASHBOARD**

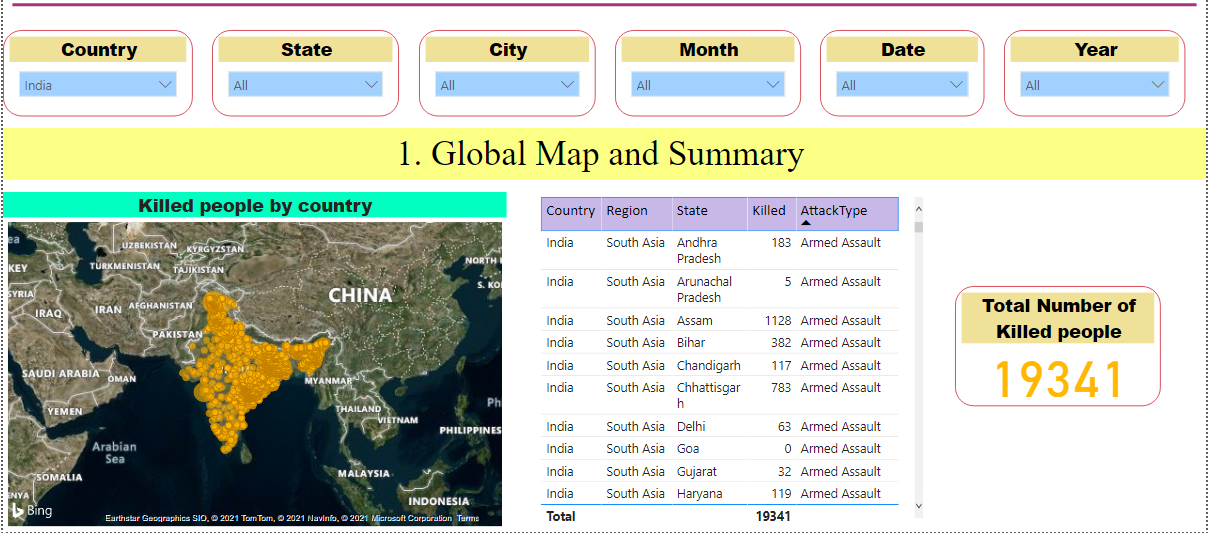
**Header**



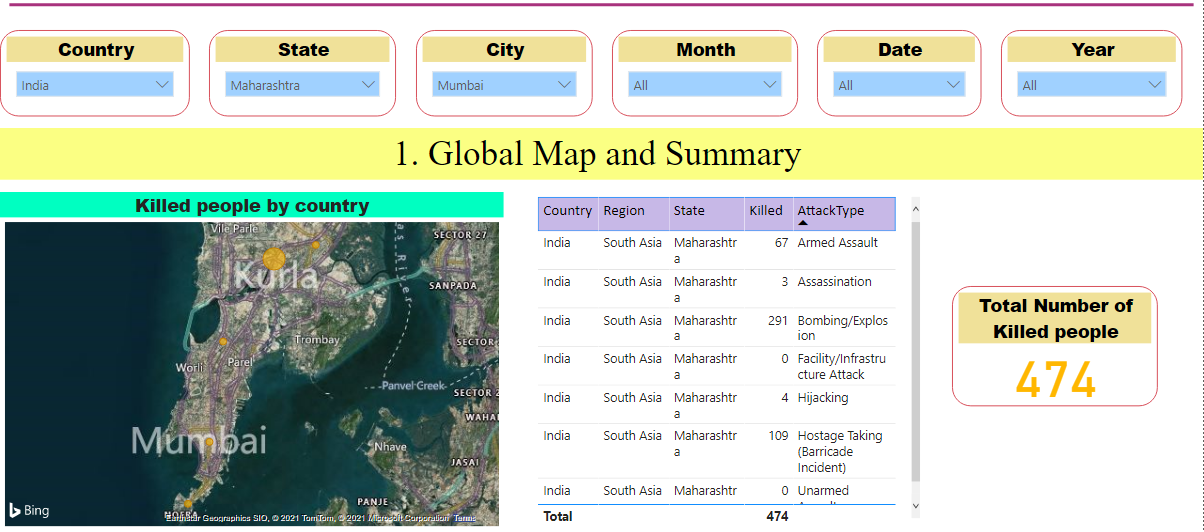
* **Here you can see the total number of killed people in Power Bi dashboard.**



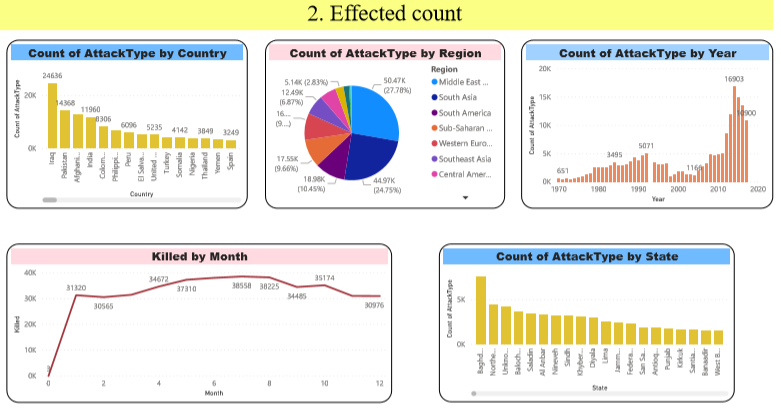
* **You can also search for particular country for which you want to see the insights. (For ex. India)**



* **Here you can also filter out content based on state, city, month, year and date. (For ex. Mumbai city)**



* **Now in below you can see effected count across the world.**



**Insights & Analysis**

* Most Affected Country is **Iraq**.
* The year 2014 has observed the greatest number of terrorist activities.
* Baghdad is the most affected state due to terrorism
* Bombing/Explosion is the most frequently used attack type by the terrorist.
* An Unknown group is the most active Terrorist organization followed by the Taliban.
* Private Citizens and Property the most frequent target type of the terrorist.
* Middle East and North Africa is the most affected region due to terrorism.

**Project Details**

|  |  |
| --- | --- |
| **TASK NAME** | **Assigned To** |
| **Analysis** | Madhusudan |
| **Design Document** | Madhusudan |
| **Development** | Madhusudan |
| **Development Review** | N/A |
| **Unit Testing** | Madhusudan |
| **UAT (User Acceptance Testing)** | N/A |
| **THB (Transition Hand Book)** | Madhusudan |
| **Project Work Model** | Agile |
| **Sprint Duration** | 1 Month |
| **Scrum Master** | Mrs. Bharti Bhole |
| **Product Owner** | RJ College |
| **Story Point** | 8(T-shirt Sizing Method) |
| **Epic Name** | DSAI(SEM1) |

**CONCLUSION AND SUGGESTIONS**

We exploited the resources of the Python language and its libraries for our use and have managed to complete our main objective of Exploratory Data Analysis of the Global Terrorist Dataset.

The comprehensive analysis done for the world as well as specifically for India gives us enough information, which can help us to conclude and trace some very significant insights and conclusions.

The maps used are interactive in nature and dynamic, which make them more interesting but at no point have we compromised with the information in order to accommodate the aesthetics.

Any peculiar or not expected trend has been examined in order to make it more readable and understandable for the readers.

In 2017, an estimated 10,900 terrorist attacks were executed.

Maximum number of attacks have been in the Middle East Area and hence the maximum number of casualties suffered is in that region.

Now this alone might be insignificant but the in-depth analysis that we have done breaks it down to the core so as to get the maximum information available from the dataset and hence use those for Predictive and finally Prescriptive analysis which can help the Governments and Private organisations.

**REFRENCE**

1. <https://www.kaggle.com/search?q=global+terrorism+dataset>
2. <https://en.wikipedia.org/wiki/Global_Terrorism_Database>
3. <https://towardsdatascience.com/global-terror-and-the-eda-visualization-rabbit-hole-a953c998b7e9>
4. <https://www.youtube.com/watch?v=tzRPDQ4uayQ>

**PROJECT CODE LINK**

1. <https://github.com/Madhusudantiwari1/Task4-.git>
2. <https://www.linkedin.com/posts/madhusudan-tiwari-%F0%9F%87%AE%F0%9F%87%B3-64318319a_gripfeb21-activity-6768920736305315840-paqI>