**Global Terrorism Dataset**

**Ruchira Umesh Shrawagi**

**Abstract**

The Global Terrorism Database is an open-source database including information on terrorist attacks around the world from 1970 to 2017. Global Terrorism Database includes systematic data includes domestic and international terrorist incidence that have occur during this time period and now includes 180000 attacks. The database is maintained by researchers at the National Consortium for the study of terrorist and response to terrorism headquarters at the University of Maryland.

**Problem Statements:**

The objective of the report is to understand the terrorist events around the world. By making use of interactive charts and animations we have tried to make the exploration easy and more informative.

The problem statement here is to build a tool that can present processed information in the form of intuitive visual representation of analysed data. Implementation of this project involves system design, backend design, visual design, and user interface.

Visual design mostly consists of analyses and visualization techniques to construct different graphics representing the end results in an easy-to-interpret format.

**Objective:**

* An elaborate exploratory data analysis on the dataset from the Global Terrorism Dataset.
* The analysis comprised of multiple questions which helped in deriving insights about terrorism hot spots and security questions.
* Prepare the data for processing and analysis
* Perform data pre-processing to manipulate the data for analysis and draft out targeted tables for analysis
* Analyse the data more deeply and extract insights
* Visualize the data using Python.

**Dataset key attributes:**

This dataset is an unclassified, open-sourced, freely available dataset for anyone to use and has the most comprehensive collection of terrorist events among all other available datasets.

Dataset contains data of more than 180000 terrorist events happened since 1970 and has over 135 variables describing each attack. Some of the key attributes consisting those variables which are used for this project are listed below:

|  |  |
| --- | --- |
| **Attributes** | **Description** |
| Day, month, year | Calendar details of the event. |
| Region\_txt | Name of the region where the attack happened. Region\_txt consists values like East Asia, South Asia, Western Europe, etc. |
| Longitude | Longitude of the location |
| Latitude | Latitude of the location |
| Attacktype1\_txt | The type of attack happened. Attacktype1\_txt consists of categories like explosion, armed assault, assassination, kidnapping, unarmed assaults. |
| Target1\_txt | Type of target involved in the attack. Target1\_txt consists of categorical values like private citizens, military, police, government officials, transportation, education, religious institution, airports, etc. |
| Nkill | Number of people killed in any event |
| Nwonded | Number of people wounded in any event |
| Motive | Known motive of the attacker. |

**Work Flow:**

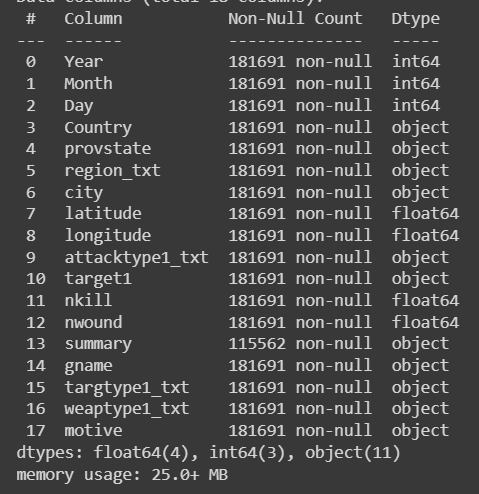
**Data Processing:**

Data processing is the first step to be done after collecting data. 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 processing methodology helps in converting this raw data into a more meaningful, focused, interpretable and readable format.

We ended up working on 18 columns from 135. See below screen print:



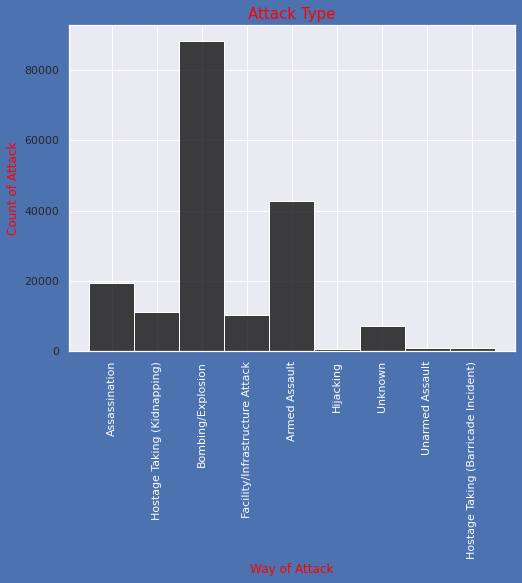
Fill NA Screen print



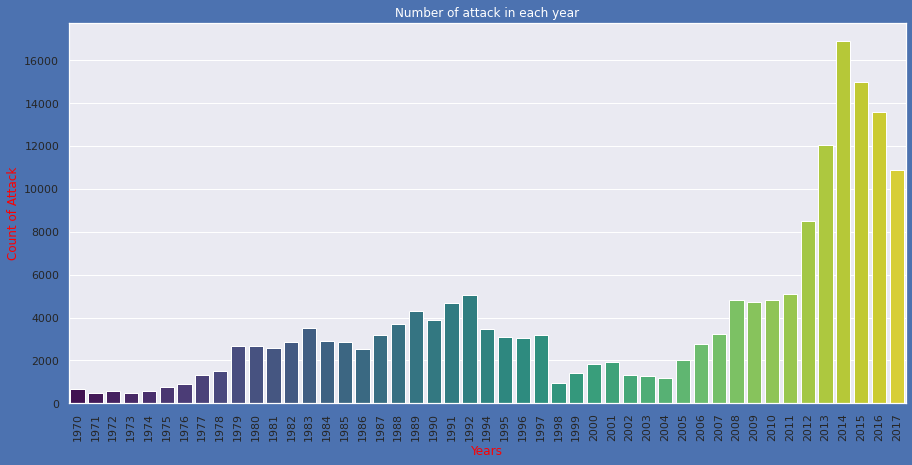
**Visualization:**

This section consists of details regarding the visual results for the Dataset.

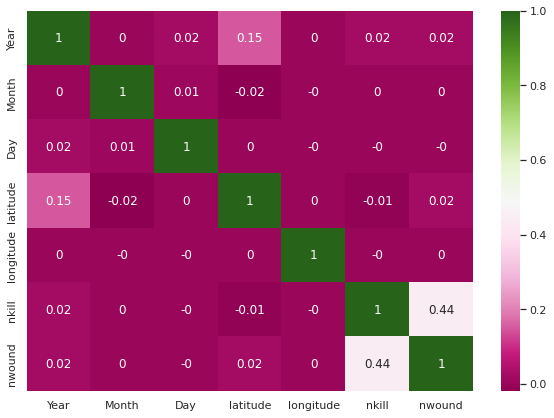
1. Below screen print shows the maximum and minimum terrorist attack type:



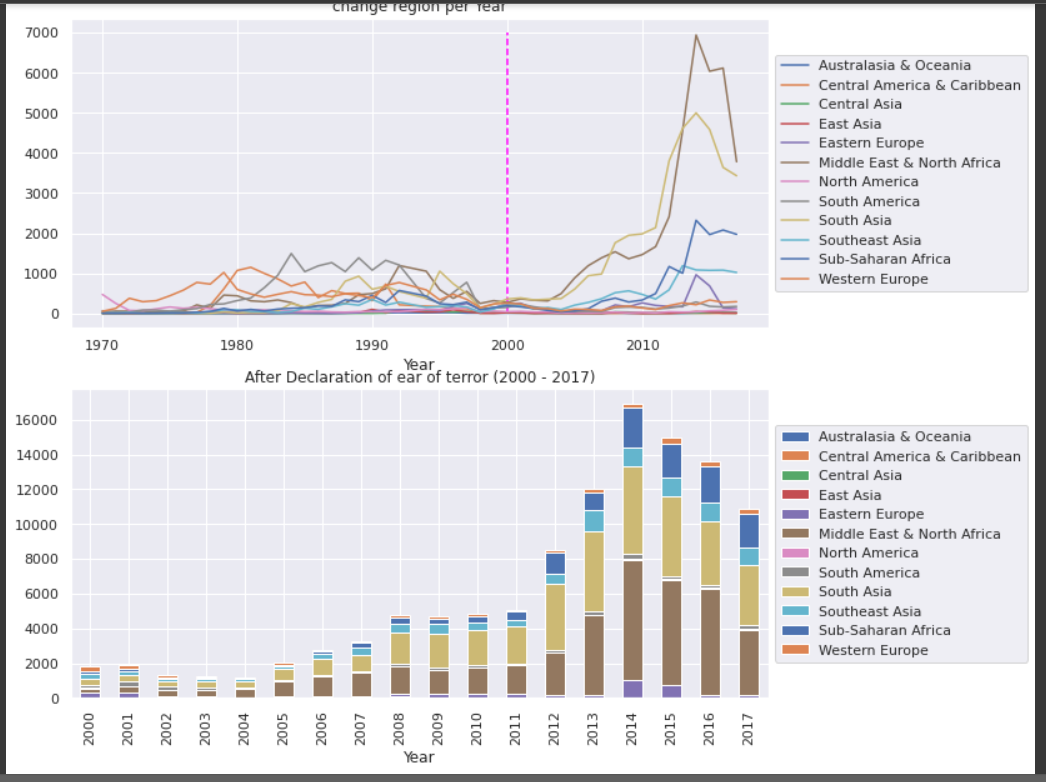
1. Below figure shows the number of attack in Each year:



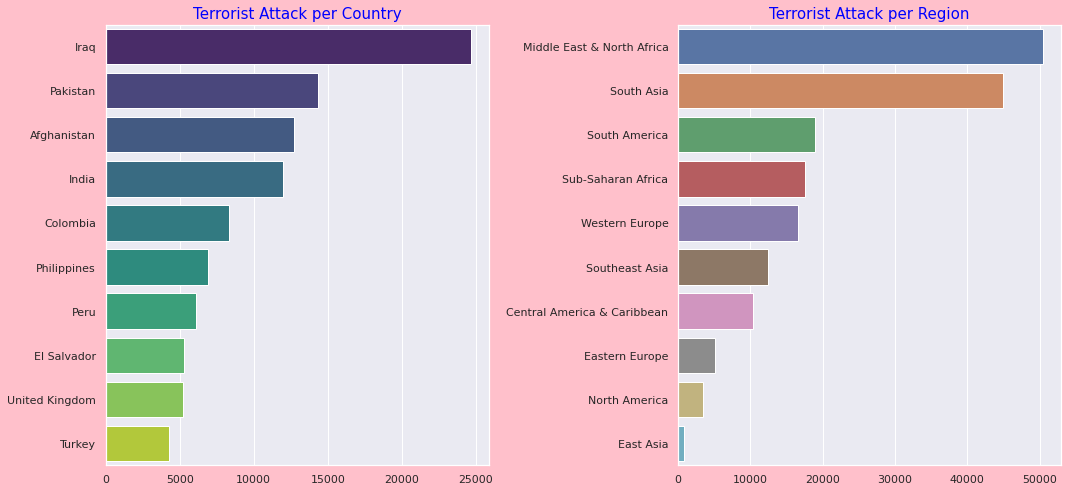
1. This figure helps to find dependencies among the various parameters in the dataset can reveal a key pattern about the nature of terrorism.



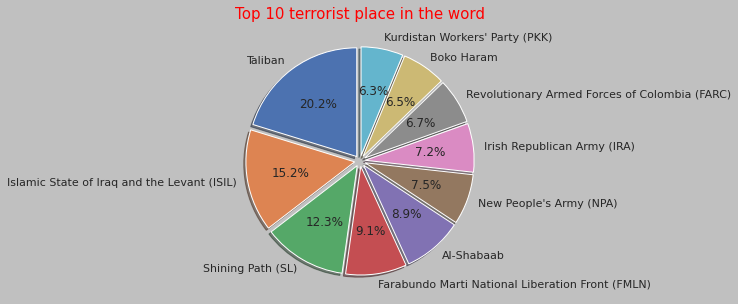
1. Below figure shows the trend before and after the war on terror.



1. Below sub plot shows the minimum and maximum terrorist attack happen by country and region wise:



1. Now let’s see top 10 terrorist group placed in which country around the world



**Technologies Used:**

Python is a high-level interpreted language that supports different platforms like Windows, Linux, Mac, Rasberry Pi, etc. Python can be used for creating web applications, database systems, handle big data, perform complex mathematical calculations. Python can be treated in an object-oriented, functional or procedural way.

**Conclusion:**

The Conslusion of this project was to build a tool which helps users to understand and interpret the nature of terrorism. A visualization which can be used to calculate the total number of attacks, total kill counts and location based on the selected region, attack type, terrorist place and year provides interactive interface to explore this dataset. We can understand various patterns, trends and correlation in terrorism through visual interpretation and its provided explanation. This work can be used by curious civilians [28], security related policy-makers, international organizations hosting worldwide events, foreign investors and academic researchers for the purpose of understanding terrorism and its nature.