**EXPORATORY DATA ANALYSIS ON GLOBAL TERRORISM**

**Sachin Gawande ,Vishal Pagare Sandesh Arsud,**

**Rushikesh Sharma Data science trainees,**

**AlmaBetter, Bangalore**

* **Abstract:**
* Introduction data set
* Dataset Preparation
* Attributes in Global Terrorism Analysis
* Number Of Attacks happening in Year and analyze total Number of people killed count
* Most active terrorist group
* Number of attacks in each state
* Number of Attacks by Region, since 1970
* Total Number of people killed in Jammu and Kashmir , Since 1970 to 2017
* The Most attacks by Year
* **Introduction**
* In this data set contain about terrorism activity from 1970 to 2017
* and consists of information of more than 180,000 terrorist attacks
* In global terrorism .and is also information about where people killed
* and his location, target type, motive of terrorist organization, and
* most active terrorist group in glob ,attack type and wounded and so more
* In this data set consist of year wise ,state wise ,city wise data
* This data shows where and which type of attack on city and state.

# In this data set we are going to perform exploratory data analysis on global terrorism Such as to find out hot zones of terrorism. exploratory data analysis is nothing but Analyzing given data and finding the trends ,patterns and making sum assumption In this data base we have so many unused columns that are not required for our data analysis so we kept only important columns .5. How Surge pricing works

* **DATA PREPARATION :-**

**Importing Libraries :** Importing all the required libraries like Pandas, Numpy, Matplotlib and Seaborn.

**Importing The Dataset :** Importing the global terrorism analysis dataset from google drive to our colab notebook

**Reading The Data :** Reading the csv file and its text encoding option to deal with different format of files ( read\_csv('file', encoding = "ISO-8859-1"))

**Cleaning the Data :-** Renaming the columns name and selecting the important columns.

**Exploratory Data Analysis :-** Analyzing the data set and creating insights from it.

**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.

**Matplotlib**: It is a 2D based plotting package that provides required modules and

**functions**. A developer can customize font properties, styles, axes properties, etc.

**NumPy**: It provides structures for multiple dimensional array objects and tools for related operations. NumPy is usually used for high performance scientific computational tasks

* **DATA PREPROCESSING:-**
* 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 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
* **Exploratory Data Analysis :-**

After loading the dataset we performed this method by comparing our target variable that is **GLOBLE TERRORISM** with other independent variables. This process helped us figuring out various aspects and relationships among the target and the independent variables. It gave us a better idea of which feature behaves in which manner compared to the target variable.

* **Null values Treatment:-**

Our dataset contains a large number of null values which might tend to disturb our accuracy hence we dropped them at the beginning of our project in order to get a better result.

* **Encoding of categorical columns :-**

We used One Hot Encoding to produce binary integers of 0 and 1 to encode our categorical features because categorical features that are in string format cannot be understood by the machine and needs to be converted to numerical format.

**Preparation of attribute:-**

**Year :** This column contains Calendar details of the event. **Month :** This column contains the month of terror activities.

**Day :** This column contains the day of terror activities.

**Country :** This column contains the country of terror activities.

**Region :** This column contains name of the region where the attack happened. Region consists values like East Asia, South Asia, Western Europe, etc

**State :** This column contains the state of terror activities

**City :** This column contains the city of the terrorist activities

**Attack Type :** This Columnconsists of categories like explosion, armed assault, assassination, kidnapping, unarmed assaults.

**Target Type :** This column contains consists of categorical values like private citizens, military, police, government officials, transportation, education, religious institution, airports, etc.

**Group Name :** This column contains terrorists groups in global terrorism.

**Weapon Type :** This column contains consists used weapons like Explosives, Firearms, Unknown, Incendiary, Melee in terrorist activities.

**Killed :** This column contains Number of people killed in any event

**Wounded :** This Columnconsists Number of people wounded in any event

**Loss/Casualty :** This Columnconsists casualty/loss in any event

* **Number Of Attacks happening in Year and Analyse total Number of people killed count**
* For above barplot as we can see the lowest number of attacks happening in year 1992,2008 & 2010 and killed above 4000 peoples
* The most number of attacks happening in year 2014,2015 & 2016 and killed 12000 to 16000 peoples

**Most active terrorist group:-**

* Taliban is most active terrorist group
* Taliban and ISIL are the major contributors in the rise in attacks
* One of the reasons for this observation could be the resistance terrorist groups have faced from multiple counter-terrorism forces over time.

**Total Number of people killed in Jammu and Kashmir , Since 1970 to 2017**

* **As** we can see in this graph Attacks were more during 1990 to 2000 in between most people Killed. When compared to the last 10 years scored a maximum Killed onwards, 1998 There is huge spike in Jammu and Kashmir.
* we can see the from 2000 onwards count started decreasing.
* **CONCLUSION**

**Summary**

The goal of this project was to build a tool which helps users to understand and

interpret the nature of terrorism. Users can perceive the START dataset through visual

designs. A visualization which can be used to calculate the total number of attacks, total

kill counts and location based on the selected region and year provides interactive interface

to explore this dataset. Users can understand various patterns, trends and correlation in

terrorism through visual interpretation and its provided explanation. Users can also explore

START dataset and other terrorism related sources for additional research purposes

provided in this tool. 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.