**Global Terrorism Exploratory Data Analysis**

**Mandar Pate, Shivam Jadhav,**

**Tejas Chaudhary, Pranav Gades**

**Data science trainees,**

**AlmaBetter, Bangalore**

**Abstract:**

Terrorist attacks are now being global issue both in developing and developed countries. There are more than 180,000 terrorist attacks in 1970-2017. Most of the countries attacked by terrorist through Bombings and firearms cause fatalities. Classification of terrorist attacks can be performed based on either the attack succeeds or not. Succeed attack is defined as an unavoided action that caused fatalities. There are seven attributes studied in this paper: year, attack type, terrorist group name, target attack, city, and

states, regions, weapon type uses to attack.

***Keywords: Analysis, Success rates,***

***Univariate analysis***

**1.Problem Statement**

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 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.

The main objective is to Analyze this model, which could help them in predicting and taking particular action proactively. This would in turn help the world facing less Terror attacks and fatalities

* **provstate**: State which affected by Terrorist attack
* **country**\_**txt**: Country which affected by terrorist attack
* **regiontype1\_txt**: Region in which attacked was took placed.
* **Attacktype1\_txt**: Terrorist groups preferred the type of attack for attacking
* **weaptype1\_txt**: Weapon through which attacked was performed.
* **gname**: Group name of attackers.
* **iyear**: year in which terror attack was took placed.
* **imonth**: year in which terror attack was took placed.
* **success**: ‘1’ if attack is successful,’0’ if attack is not successful.
* **suicide**: ‘1’ if attack is suicide,’0’ if attack is not suicide.

**2. Introduction**

### The terrorist’s groups doesn’t have any fix pattern of attacking, based on any situation they attacked in the whole world using different Attack types, target types, weapon types.

### Above features are mostly categorical features.

### features like country, states are affecting by their geographical and geopolitical locations,

Countries, States and Regions are taken from all around the world.

### The attack around the world increases suddenly after 2001.

### Our goal here is to build a Analytical model, which could help everyone in predicting the attack type proactively.

We will investigate Variables having various types such as Attack types, Weapon types,

Target types only.

## **3. Types of Attacks**

* Bombing /Explosion
* Armed Assault
* Assassination
* Armed Assault
* Infrastructure attack

### Attacks varies with the groups and materials required to lead the attack.

With the Weapon types as:

* Explosives
* Firearms

Mostly Targeting,

* Private citizens
* Military
* police

### Success rate also affecting attack types.

## **4. Reasons for High Attacks**

The reasons for high attacks are:

* Economic In stabilization
* Geographical location of state, country
* Religious Events
* Political Influence
* Publicity and Media attention.
* Irrationality or Insanity

Whenever rates are raised due to surge pricing, the app lets riders know. Some riders will choose to pay, while some will choose to wait a few minutes to see if the rates go back down.

**6. Steps involved:**

* **Data Exploration**

It is the first step of data analysis used to explore and visualize data to uncover insights from the start or identify areas or patterns to dig into more.

* **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.

* **Column Selection**

In these steps with the help of pandas and NumPy we selected only few columns to get insights from it

* **Finding Correlation**

Correlation is a statistical measure that expresses the extent to which two variables are linearly related

* **Data Visualization**

we tried various visualization like:

1. **Countplot**
2. **Pairplot**
3. **Pie chart**
4. **Histogram**
5. **KDE plot**
6. **Barplot**

**7.: Libraries & Methods**

* 1. **Pandas:**

that is made mainly for working with relational or labeled data both easily and intuitively. It provides various data structures and operations for manipulating numerical data and time series. This library is built on top of the NumPy library. Pandas is fast and it has high performance & productivity for users. Methods we have used are,

1. Info()
2. Describe()
3. Dropna()
4. Fillna()
5. Rename()
6. Groupby()
7. Value\_counts()
8. Idxmax(), idxmin()
9. Max(), min()

**7.2. Numpy:**

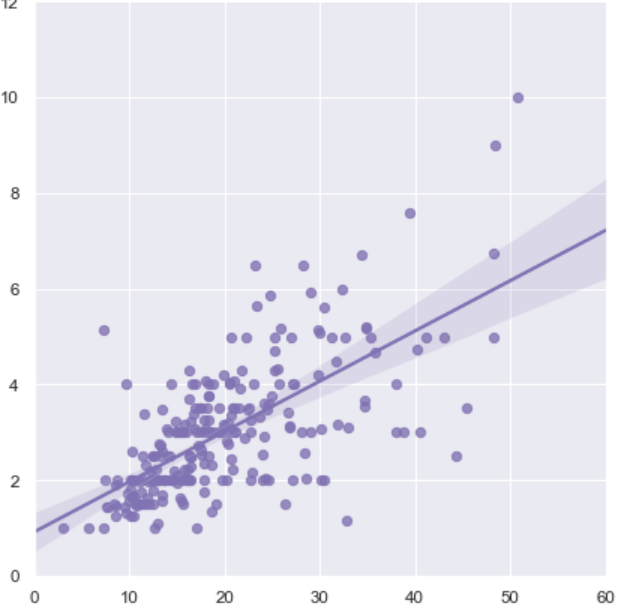
NumPy is a Python library used for working with arrays.

It also has functions for working in domain of linear algebra,

* 1. **Seaborn & Matplotlib:**

**Seaborn is a Python data visualization library based on**matplotlib**. It provides a high-level interface for drawing attractive and informative statistical graphics.**

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python

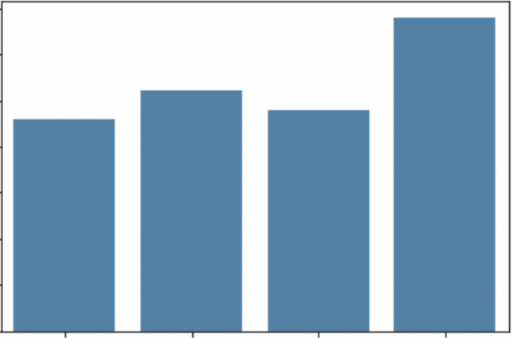


**Jointplot**

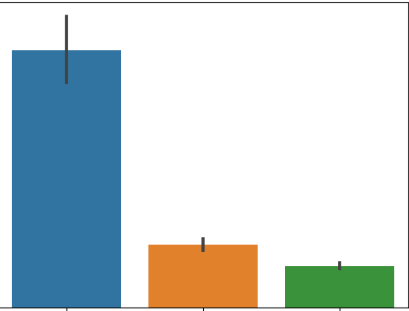
**Chart, histogram

Description automatically generated**

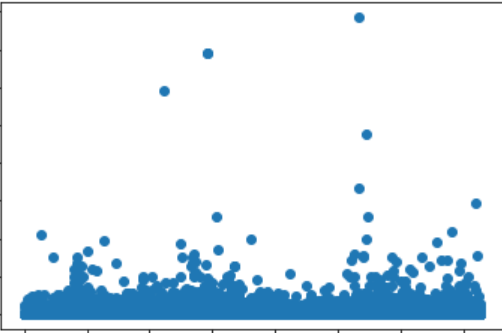
**Kde plot**

****

**Countplot**

****

Barplot

****

**scatterplot**

1. **Dataset accuracy:**

That's it! We reached the end of our exercise.

Starting with loading the data so far we started with EDA , null values treatment, Data visualization.

Our Dataset Readability has been increased as we perform some operations, data manipulation that helps to reduce null values from 56% to 0.30%.

Later, Visualizations helps us to get more insights present inside dataset.

**References-**

1. Github
2. GeeksforGeeks
3. Analytics Vidhya