**EDA on Global Terrorism Database**

**Shubham Chandrakar**

**Data science trainee,**

**AlmaBetter, Bangalore**

**Abstract:**

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.

Our goal is to perform EDA on Global Terrorism Database (GTD) to identify trend and pattern of terrorist attack globally all over the world.

***Keywords: EDA, Exploratory data analysis, global terrorism database, data wrangling, python, visualisation.***

**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. This database has features such as year, date, month, approximate dates, country, county name, city, city name, no. of kills, no. of wounded, type of weapon used etc. There is no specific problem statement given. Dependent feature does not exist too.

Aim of project is to understand, explore and analyse the data to discover key findings pertaining to terrorist activities.

**2. Introduction**

The Global Terrorism Database (GTD) is an open-source database including information on terrorist attacks around the world from 1970 through 2017. This database consists of more than 135 features and 18169 rows.

Each feature is independent of other. We need to identify relevant feature and then perform EDA on this database. Some relevant features like **“year”** describes the year of terrorist event, **“country\_name”** points towards name of country, **“nkill”** describes no of people killed in terrorist attacks, **”nwound”** no of people wounded in terrorist attacks, **“weapon\_type”** describe type of weapon, **“no\_of\_affected\_people”** is sum of no of people killed and wounded in terrorist attacks. ”**success**” if value of success is 1 then attack was successful else attack was unsuccessful. **“location”** shows location of event it is combination of latitude and longitude. **”attack\_type”** describe type of attack used. **“city”** name of city where terrorist event happened. **“target\_type”** type of institutional target chosen for attack by terrorists.

### Our goal here is to explore and analyse the database by performing EDA on Global Terrorism Database (GTD) and find various trend in data and identify problems and find solutions

## **3. Types of features**

* Categorical features
* Numerical features

### Categorical features are the features which are discrete in nature.

Categorical features in the dataframe “df” are as follows:

**'year','country\_name','region\_name', 'city', 'attack\_type', 'target\_type', 'targ\_subtype', 'target\_name', 'weapon\_type', 'location'.**

Numerical features in the dataframe “df” are as follows:

**'nkill', 'nwound', 'success', 'no\_of\_affected\_people'**

## **4. Reasons for increase in no of affected people**

The reasons for increase in no of affected people are:

* Increase in bombing activities
* Private citizens and properties are most targeted people
* Weak governments control over terrorists like in Iran, Afghanistan and Pakistan
* Lack of literacy and secularism in top three most affected countries:

Iran, Afghanistan and Pakistan

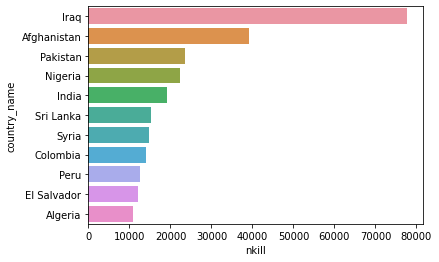
* Illegal arms are widely available for the terrorists.
* Rapidly increasing terrorist activities after 2012 i.e. six times.

# **5. Findings from the data**

## **No of kills are rapidly increasing**

In these database no of identified people who are killed in terrorist activities are rapidly increasing with each year.

After 2012, average no kills is increased six times the average no of kills before 2012, number of kills is maximum for Iran followed by Pakistan



We need to focus in the top countries where terrorist are killing more people

## **Most of the terrorist attacks were successful**

Every year most of the terrorist attacks were successful i.e. the countries all over the world is mostly unable to stop terrorist activities till date.

## **Citizens are most preferred target**

Citizens as well as private property are top priority target for terrorists in most of the countries over the span of 1970 to 2017.

All government must acknowledge this problem and try to protect the citizens

## **Bombing/Explosives followed by Armed assault are most preferred type of attack all over the world**

Bombings/Explosion are most preferred type of attack all over the world. It accounts for 48% of all attacks all over the world.

Armed assault second most preferred mode of attack , it accounts for 24% of attacks all over the world.

Governments all over the world must prevent access of illegal arms and material required to make explosive.

**6. Steps involved:**

* **Exploratory Data Analysis**

After loading the dataset we performed this method by comparing all features i.e. categorical and numerical features. It is found that all features are independent .First step was cleaning the data, then that correlation of independent features is established. Following to that univariate and multivariate analysis is performed. This process helped us figuring out various aspects and relationships among the independent variables. It gave us a better idea of which feature behaves in which manner compared to the each other.

* **Null values Treatment**

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

* **Feature engineering**

The dataset contains too many features most of them were irrelevant from our perspective. Irrelevant features were dropped, some features were combined together. This process is performed to analyse the dataset efficiently.

* **Visualisation of data**

In these step, after analysis of data various type of visualisation is performed with the help of seaborn and matplotlib to plot various plot like distribution plot, count plot, bar plot, pie plot etc.

This process is essential to understand and analyse the data in the holistic way.

* **Univariate analysis**

In this analysis each feature is analysed individually to understand the distribution of data.

It is found that most of the numerical features are distributed year wise and have positive correlation with each other.

Using distribution plots such as histograms, count plots etc. helps us to identify the distribution of each numerical feature and categorical features individually.

**Multivariate analysis**

* In multivariate analysis two or more feature are analysis together to understand relationship between the feature. In this phase of analysis categorical as well as numerical features can be analysed together.

**8. Conclusion:**

The project started with loading the data we performed have EDA , null value treatment, feature engineering, univariate and bivariate analysis.

* There is rising trend in growth of global terrorist activities all over the world.
* The no of people killed over span of 1970 to 2017 is constantly rising, after 2012 average no of people killed per year is increased six times
* The most affected countries are Iran, Pakistan, India etc. 11 countries have share of 57% of all terrorist activities all over the world.
* The most preferred attack type by terrorist is Bombing/Explosion etc.

The threat of global terrorism is rising every year. We need to curb terrorist activities globally as soon as possible. We need to stop bombing activities, trade of illegal armaments and increase security of citizens. Governments of each countries and UNSC must take necessary action to solve this global problem at a high priority level. Citizens must need to be provided with food security, education, justice, health and jobs so that they must protect themselves from terrorism and prevent any citizen to join terrorist organisation.

**References-**

1. Geeks for Geeks
2. Analytics Vidhya