Exploratory Data Analysis
Global Terrorism Analysis

By:- Abhishek Kirar,Ajinkya Shingote,
Ganesh Subramanian,Mayur Chulbhare,Syed Aquib
Data science trainees,
AlmaBetter, Bangalore

Abstract:

Terrorism can be described as the use of violence against persons or properties to intimidate or coerce a government or its citizens to some certain political or social objectives. It is a global problem which has led to loss of lives and properties and known to have negative impacts on tourism and the global economy. Terrorism has also been associated with high level of insecurity and most nations of the world are interested in any research efforts that can reduce its affect. The analysis is based on how the terrorist activity happens, what locations that terrorist targeted, which weapons where used most. On that analysis basis we can conclude that how to reduce the activity od terrorist and how can we fight against terrorist.

1.Problem Statement

Data provided by the Global Terrorism Database.csv file is in unformatted manner, uneven data, and duplicate data and also some data columns it is irrelevant, because it's a piled-up data coming from various different countries. For doing the analysis on the data the data needs to be in correct format and well organized formed

The main objective of the analysis is to obtain the meaning full information and facts from the given huge datasets as shown above, by cleaning the datasets, doing a proper analysis and visualization and plotting the useful information into different graph and charts so that the trend and relationship between the various indicators on which the analysis is done can be understand easily.

2. Introduction

World peace was one of the core reasons for forming the United Nations organization. Terrorism is the biggest hurdle to world peace. In this project, we focus on terrorism by analyzing the dataset provided by to explore meaningful patterns and statistics. Terrorism is an unsettled term.

Terrorism is sporadic, widespread and inconsistent with time and nature. Because of these characteristics, International terrorism is difficult to summarize all aspects as a single conclusive solution and make this information available to be easily understood by most people. Exploring this dataset can provide an insight into how different parameters are correlated with each other, which can help identify unknown hidden patterns. This exploration will also assert enough facts to provide justifications for some common misconceptions regarding terrorism. One of misconceptions is that more military can suppress and control terrorism. However, using the instrumental

variable approach, studies show that counter-terrorism solutions like more military spending is not enough to control terrorism and is also dependent on other factors like economy and national politics. Another popular misconception is that terrorism only affects the individuals directly involved in any terrorist event. Terrorism adversely impacts not only the economy of the victim country but also the countries financially associated with international terrorism.

3. Dataset

The dataset contains data of more than 180000 terrorist events(no. of rows) happened since 1970 and has over 135 variables (no. of columns) describing each attack. Some of the key attributes consisting those variables which are taken under consideration for this project are listed below:

- Attacktype1_txt : Attacktype1_txt consists of categories like explosion, armed assault, assassination, kidnapping, unarmed assaults.
- Target1_txt : Target1_txt consists of categorical values like private citizens, military, police, government officials, transportation, education, religious institution, airports, etc
- Success : '1' if attack was a success. '0' if the attack was a failure.
- Multiple : Value for the number of attacks happened in a single terrorist event.
- Natlty1 : Nationality of the attacker

- Weaptype1 : Type of weapon used in the attack. Weaptype1 contains values like firearms, explosives, melee, vehicles etc.
- Nkill : Number of people killed in any event
- Nwonded : Number of people wounded in any event
- Gang_type : Organization that claimed responsibility of an attack
- Region_txt : Name of the region where the attack happened. Region_txt consists of values like East Asia, South Asia, Western Europe, etc.
- Longitude :Longitude of the location
- Latitude : Latitude of the location
- Property : Total property damage happened in any event.
- Motive Known motive of the attacker
- Day, month, year : Calendar details of the event.
-

Note : For the better interpretability we had name converted those attributes in simple manner.

4. Steps involved:

- **Exploratory Data Analysis**
  After loading the dataset we performed this method by comparing our target variable that is Surge_Pricing_Type 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 inorder 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.

- **Feature Selection**
  In these steps we used algorithms like ExtraTree classifier to check the results of each feature i.e which feature is more important compared to our model and which is of less importance.
  Next we used Chi2 for categorical features and ANOVA for numerical features to select the best feature which we will be using further in our model.

- **Standardization of features**
  Our main motive through this step was to scale our data into a uniform format that would allow us to utilize the data in a better way while performing fitting and applying different algorithms to it.
 The basic goal was to enforce a level of consistency or uniformity to certain practices or operations within the selected environment.

5. Analysis

Making analysis and plotting graphs for the same.
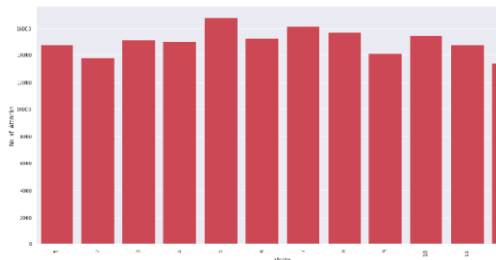
- **Yearwise Attack Analysis**

Summarizing all the terrorist attacks over the years can provide us an idea about how terrorism has evolved and what rate has it impacted the world each year. Figure 6 shows data from 1970 to 2016 for the total number of attacks that happen each year.
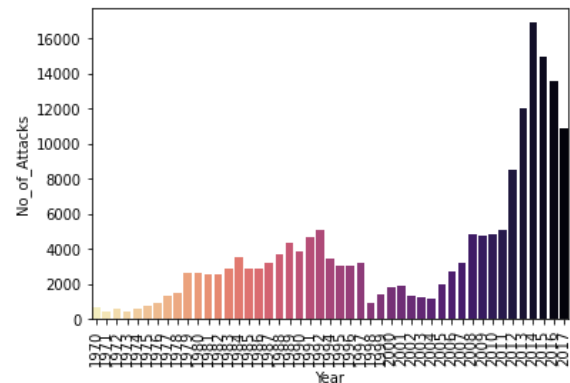
Terrorist attacks were quite low in numbers in the decade of 1970. Terrorism then had a fairly rise in the 1980s and early 1990s and was considerably low in the next decade but then terrorism rose from early the 2000s topping the charts like never before in the history. Hostile environment and global tension have increased because of the number of attacks in recent years. This observation can help investigate factors that
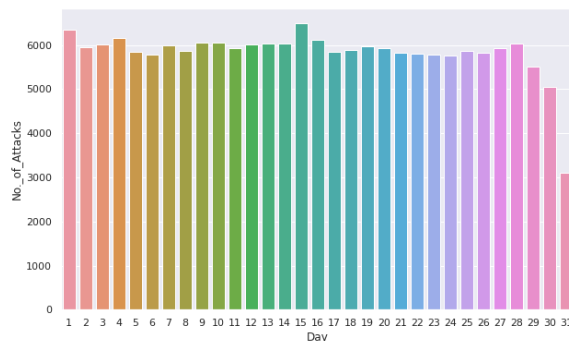
adversely impacted the sudden rise in the number of attacks.

- **Monthwise Attack Analysis**

By the month wise attacks analysis almost every month has the same no.of attacks.

- **Daywise Attack Analysis**

By looking at Day wise attack analysis it doesn't give much inference but we can conclude that on 1st and 15th day of month attacks were high and at last date of month attacks were low.
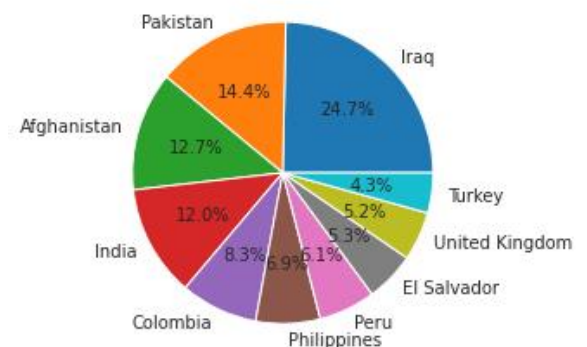
- **Top 10 Affected Regions in the World**

The Middle East and North Africa are the most affected regions.

Top two regions account more than 50% of the total terrorist attacks in the world.
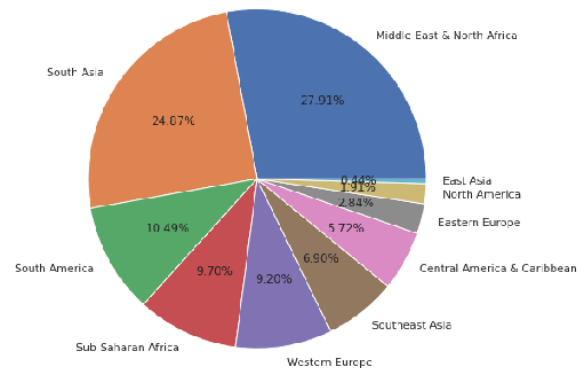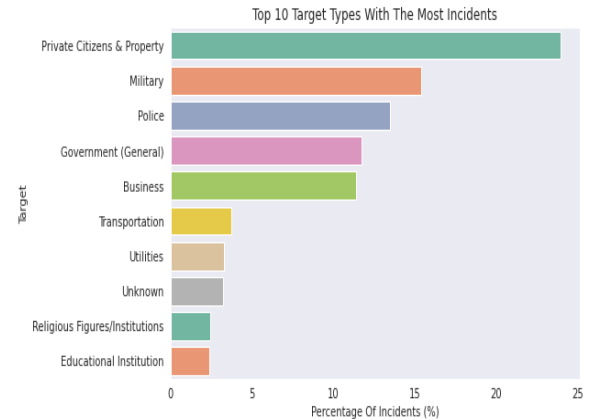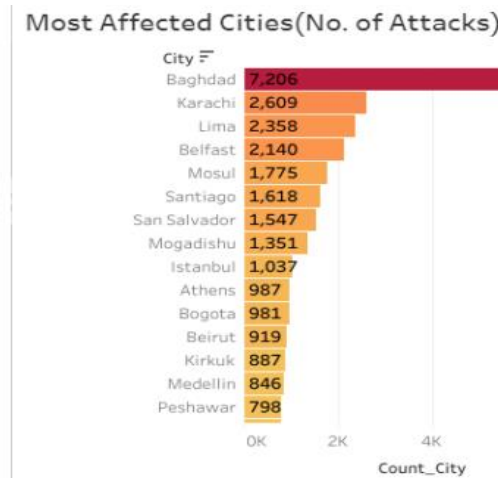

- **Top 10 Affected Countries in the World**

The above graph represents highest attacks in the country.It concluded that the highest attacks in countries such as Iraq,Pakistan,Afghanistan and India follwed by others.

- **Top 10 Affected Cities in world**

Amongst the world's cities Baghdad was the most affected by terrorist attacks.

Most Affected Cities(No. of Attacks)

City	Count_City
Baghdad	7,206
Karachi	2,609
Lima	2,358
Belfast	2,140
Mosul	1,775
Santiago	1,618
San Salvador	1,547
Mogadishu	1,351
Istanbul	1,037
Athens	987
Bogota	981
Beirut	919
Kirkuk	887
Medellin	846
Peshawar	798

Top 10 Target Types With The Most Incidents

● **Top 10 Terrorist group names**

Gang_name	No_of_Attacks
Taliban	7478
Islamic State of Iraq and the Levant (ISIL)	5613
Shining Path (SL)	4555
Farabundo Marti National Liberation Front (FMLN)	3351
Al-Shabaab	3288
New People's Army (NPA)	2772
Irish Republican Army (IRA)	2671
Revolutionary Armed Forces of Colombia (FARC)	2487
Boko Haram	2418
Kurdistan Workers' Party (PKK)	2310

Analyzing the type of target will help understand their objective and most likely their motives. Terrorism is driven by an ideology that tries to make a change or impose an ideology. Looking into the most commonly targeted attributes will signify the attacker's objectives and terrorism in general. There are more than 100 distinct target types. These target types are generalized into 22 categories. Here the graph shows that citizens, military, government, and police are the most common targets. This graph explains that terrorist groups or individuals have a dislike towards the authority of the state or the nation. Their main focus is either to make a change at the political level or force their ideology on government forces by retaliating against their authority.

● **Analysis Based on Targeted Areas:**

● **Analysis Based on Types of attacks**
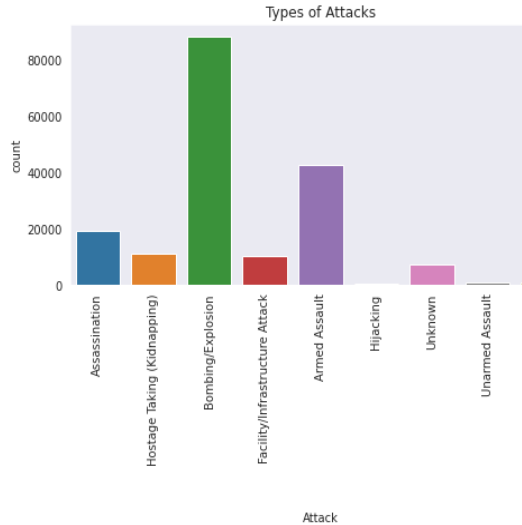
Types of Attacks

The most of attacks Bombing/explosion for the attack as it deals with the higher group damage.

● **Analysis Based on Most weapons Used for Attacks**

Weapon Used for Attack
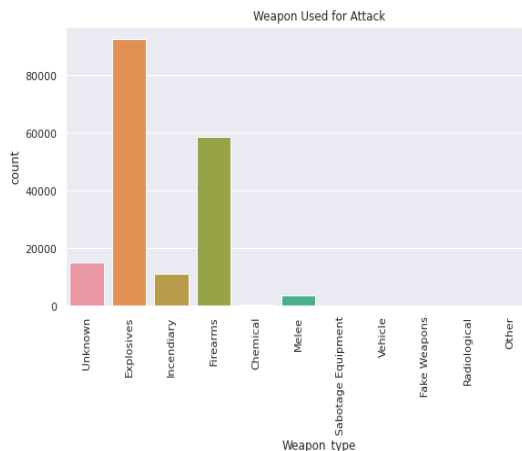
As we can see from the previous analysis of types of attacks there were most attacks done by using explosives and firearms which was the most type of weapon used.

● **Analysis Based on number of Person killed and Injured**
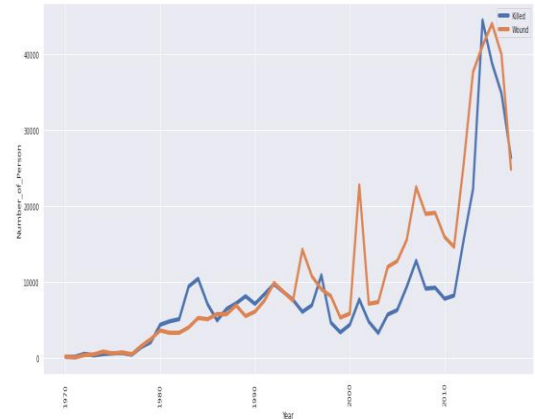
The total number of people killed during the attacks = 171378.
From the graph we can see the most deaths were in 2015.

6. **Conclusions :**

- The year in which maximum attacks took place was during the year 2014.
- The attacks were seen gradually increasing from 2005 to 2014
- The country which was attacked the most are Iraq,Pakistan,Afghanistan and India
- Different kind of attacks were done .The major was through Bombing/Explosion as it gives higher damage and Armed Asult.
- The Middle East and North Africa was the most affected region
- The most affected city in the world is Baghdad and Indian City is Srinagar
- We observed that attack mostly takes place on weekdays(Mon,Tue,Wed,Thur) compared to weekends(Fri,Sat,Sun)

- when we analyzed the attack monthwise we observed that every month has equal contribution and date wise most the attack took place on 1st and 15th of the month and minimum attack on the end of the month
- Overall 171378 people were killed during the attack and Iraq has the highest count of people killed.
- The motive behind the attacks was some unknown means only reason behind attack is to spread terrorism worldwide and some of them known such as "Protest the Vietnam War and the R.O.T.C." & "Protest the War in Vietnam and sabotage the draft".
- The most common Targets are Private Citizens & Property followed by Military sectors and police
- Taliban is the most notorious group

7. Reference:

- **Global Terrorism Database** https://www.start.umd.edu/gtd/
- **Python Pandas Documentation** https://pandas.pydata.org/pandas-docs/stable
- **Python MatPlotLib Documentation** https://matplotlib.org/stable/index.html
- **Our EDA Project Documentation**