**GLOBAL TERRORISM**

|  |
| --- |
|  |
| This is a Individual project which is entirely done by me.    Name:D.Ankitha  Mail id : dakithareddy1997@gmail.com |
| **https://github.com/ankitha-dudla/EDA-GLOBAL-TERRORISM.git** |
| Github Link:-https://github.com/ankitha-dudla |
|  |
| **Introduction**:  Terrorism is one of the biggest global challenges faced by humanity, killing and wounding millions of people, and responsible for the long-term destruction of wealth opportunities and the well-being of people. Our project is an attempt to dig deep into data to find out the possible ways to prevent terror attacks, initiatives that can be taken for the short and long term for tackling global terrorism, and changing trends in terror attacks to effectively plan mitigation measures for the future.  **Objectives :**  The main objectives of our project are:   * Which regions were affected the most? * Which countries were most attacked? * Is there an increase or decrease in the number of successful terror operations? * Is there an increase or decrease in recent years? * What are the most common types of attacks globally and regionally? * Who are the most targeted globally and regionally? * Which terrorist organization is responsible for the majority number of attacks? * What are the most used weapon types globally and regionally? |
| **Work Flow:**   1. Understanding problem statement and Raw data 2. Data wrangling: cleaning, manipulation 3. Exploratory Data Analysis 4. Data Visualization 5. Analyzing and sharing the conclusions 6. **Understanding problem statement and Raw data :**   First of all carefully, We understand the problem statement and the questions for which we are doing in this study. what we want to find out or can find out from the given data set. then we collect data, import it into the system/software. Here we are using "Colab" for this study. Import libraries such as NumPy, pandas, matplotlib, seaborn.Then load the raw dataset. We read the data and try to understand the minute details like what information each column giving and how useful it can be for the study. we found out that there are 135 variables/Columns with around 1,81,691 entries/Rows. There were few columns which we felt very important to base our study on i.e. Year, Region , Country, State/provs, City, Target type, Attack type etc.   1. **Data wrangling: cleaning, manipulation:**   For prompt decision-making,Process of cleaning, organizing, and transforming raw data into the desired format for analysts to use . Here we start finding in the data is there any i.e. Null values, missing values, NaNs, redundant values of data. **a. Handling Null Values** Around 106 columns out of 135 columns were having more than 85% null values, by rule of thumb we have dropped those columns since we cant make better insights out of them. We left with 47 columns after dropping. **b. Dropping unnecessary columns** The columns with the completely numeric representation i.e, nominal data of textual data available in other columns representing the same fact/data were dropped. The dropped columns are as follows country, region, attacktype1, targtype1, targsubtype1,weaptype1, weapsubtype1, natlty1. Also the columns deemed unnecessary for our analysis were dropped. Those columns are guncertain1, individual, dbsource, INT\_LOG, INT\_IDEO, INT\_MISC, INT\_ANY, specificity, vicinity, crit1, crit2, crit3, eventid, property, weapsubtype1\_txt, targsubtype1\_txt, doubter, iday, eventid, extended, target1. **c. Data Manipulation: - Creating new variable/columns by combining other two columns for an effective study i.e.** Victims = No. of Killed + No. of Wounded. **3. Exploratory Data Analysis** Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns,to spot anomalies,to test hypothesis and to check assumptions with the help of summary statistics and graphical representations. **Data study****i) UNIVARIATE ANALYSIS:** Univariate analysis is the simplest form of analyzing data i.e study of one variable. Its major purpose is to describe; distribution of single data, and find patterns in the data. **ii) BIVARIATE ANALYSIS:** Bivariate analysis between two variables. One of the variables will be dependent and the other is independent. The study is analyzed between the two variables to understand to what extent the change has occurred. **iii) MULTIVARIATE ANALYSIS** Multivariate data analysis is the study of relationships among the attributes, classify the collected samples into homogeneous groups, and make inferences about the underlying populations from the sample. **4. Data Visualization** Data visualization is the practice of translating information into a visual context, such as a map or graph, to make it easier to understand and gain insights from them. The graphs used here for study are: -  Histogram.  Pie Chart.  Bar Plot.  Line Plot.  Sunburst Chart  Tree map Chart  Word Cloud  Heat map Chart  Scatter Plot.  Geo Mapping. **5. Analyzing and sharing the conclusions**  * Terrorist attacks reached a peak during 2014 and then in 2015 started to drop. * The Drop in the Number of Terror Attacks is not a sign of improvement in security checks in society as the ability to carry out a successful terror attack has not reduced significantly. * The Middle East & North Africa is the most affected region in terms of the number of terror attacks among all the regions followed by South Asia, South America, and Sub-Saharan Africa. * Iraq is the most affected nation of all and it has suffered the highest number of attacks followed by Pakistan, Afghanistan, and India. * Most attacks were Bombing or Explosion followed by firearms. * Most commonly used weapons in terror attacks are Explosives and assault weapons. * Most Attacks are Targeted towards Private Citizens & Property, Military, Police, and Government. * Majority of Terror Attacks are not claimed by any terrorist organization. * Taliban and ISIL are the most active Organizations in the world. |