**GLOBAL TERRORISM ANALYSIS**

**SOURAV DUTTA**

**Data Science Trainee**

**Cohort Montreal, ALMABETTER, BENGALURU**

**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. Explore and analyze the data to discover key findings pertaining to terrorist activities.

# Library Packages:

# We will import NumPy, pandas, matplotlib, seaborn, plotly.express in Google Colab then mount the drive. After that we will read csv file from working directory path and assigned it into the data frame called “Terror \_Data”.

# NumPy - NumPy is a Python library used for working with arrays.

# It also has functions for working in domain of linear algebra, Fourier transform, and matrices. NumPy was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely.

# NumPy stands for Numerical Python. C:\Users\Sourav Dutta\Downloads\Desktop\index.png

# Pandas - Pandas is an open source Python package that is most widely used for data science/data analysis and machine learning tasks. It is built on top of another package named NumPy, which provides support for multi-dimensional arrays.

# C:\Users\Sourav Dutta\Downloads\Desktop\images.png

# Matplotlib - Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible. Create publication quality plots. Make interactive figures that can zoom, pan, update.

# C:\Users\Sourav Dutta\Downloads\Desktop\1 I22BY6JUw6XGPN3O5MRgHg.png

# Seaborn - Seaborn is a library for making statistical graphics in Python. It builds on top of matplotlib and integrates closely with pandas data structures. Seaborn helps you explore and understand your data.

# C:\Users\Sourav Dutta\Downloads\Desktop\index.png

# Plotly.express - The plotly.express module (usually imported as px) contains functions that can create entire figures at once, and is referred to as Plotly Express or PX. Plotly Express is a built-in part of the plotly library, and is the recommended starting point for creating most common figures.

# 

# Summary of Data:

# Terror\_Data – It contains the information of raw data. After data cleansing, the new data assigned into it.

# Attack\_Year – It contains the data of total number of attack counts per year.

# df- It contains the data of killed and wounded in correlation analysis.

# Attack\_Attack\_Type - It contains the data of total number of attack counts per attack type.

# Attack\_Target\_Type – It contains the data of total number of attack counts per target type.

# Attack\_Country – It contains the data of attack counts per top 20 countries.

# Attack\_Region – It contains the data of attack counts in each region.

# Attack\_City – It contains the data of attack counts in top 15 cities.

# Attack\_Terrorist\_Groups – It contains the data of total number of attacks done by each terrorist group. (Top 15)

# Attack\_Region\_Year – It contains the data of total number attacks per region by year.

# Region\_Killed – It contains the data of killed people in each region. (Top 10)

# Region\_Wounded – It contains the data of wounded people in each region. (Top 10)

# Region\_Casualties – It contains the data of casualties in each region. (Top 10)

# Country\_Killed – It contains the data of killed people in each country. (Top 10)

# Country\_Wounded – It contains the data of wounded people in each country. (Top 10)

# Country\_Casualties – It contains the data of casualties in each country. (Top 10)

# City\_Killed – It contains the data of killed people in each city. (Top 10)

# City\_Wounded – It contains the data of wounded people in each city. (Top 10)

# City\_Casualties – It contains the data of casualties in each city. (Top 10)

# Data Cleaning:

# Data Cleaning or data scrubbing is the process of fixing incorrect, incomplete, duplicate or otherwise erroneous data in a data set. It involves identifying data errors

# and then changing, updating or removing data to correct them. Data cleansing improves data quality and helps provide more accurate, consistent and reliable information for decision-making in an organization.

# After data cleaning, the data is stored in the data frame named “Terror\_Data”. The information consists of , ‘Year’, ‘Month’, ‘Day’, ‘City’, ‘State’, ‘Country’, ‘Region’, ‘Latitude’, ‘Longitude’, ‘Attack Type’, ‘Target’, ‘Target Type’, ‘Weapon Type’, ‘Killed’, ‘Wounded’, ‘Casualties’, ‘Terrorist Group’, ‘Motive’, ‘Summary’.

**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.

**Correlation Analysis:**

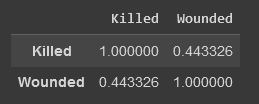
Correlation Analysis in research is a

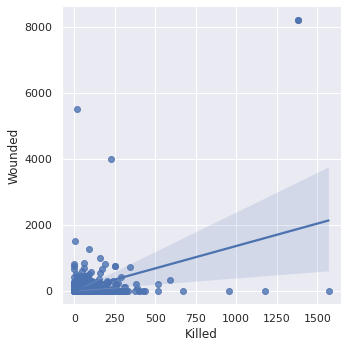
statistical method used to measure the

strength of the linear relationship between two variables and compute their association.

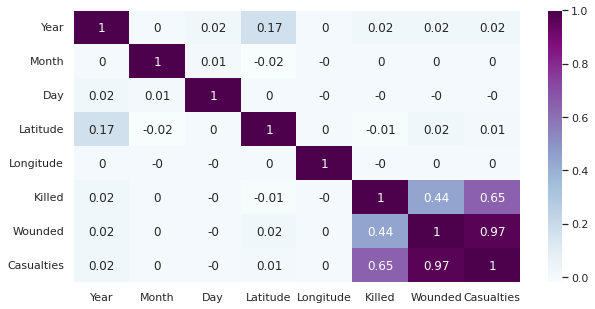
df- It contains the data of killed and wounded in correlation analysis.

df.corr ()





**Correlation Heat map:**

****

From the above analysis, I concluded that both variables, i.e. 'Killed' and 'Wounded' have Low Positive Correlation.

**Data Visualization:**

The process of finding trends and correlations in our data by representing it pictorially is called Data Visualization. To perform data visualization in python, we can use various python data visualization modules such as Matplotlib, Seaborn, Plotly, etc. Library packages of Data visualization and Correlation Analysis is already discussed above.

Bar chart and line chart are mainly used in the Global Terrorism Analysis.

**Bar Chart:**

Matplotlib - A bar plot or bar chart is a graph that represents the category of data with rectangular bars with lengths and heights that is proportional to the values which they represent. It can be created using the **bar()** method.

Seaborn - Bar Plot in Seaborn can be created using the barplot() method.

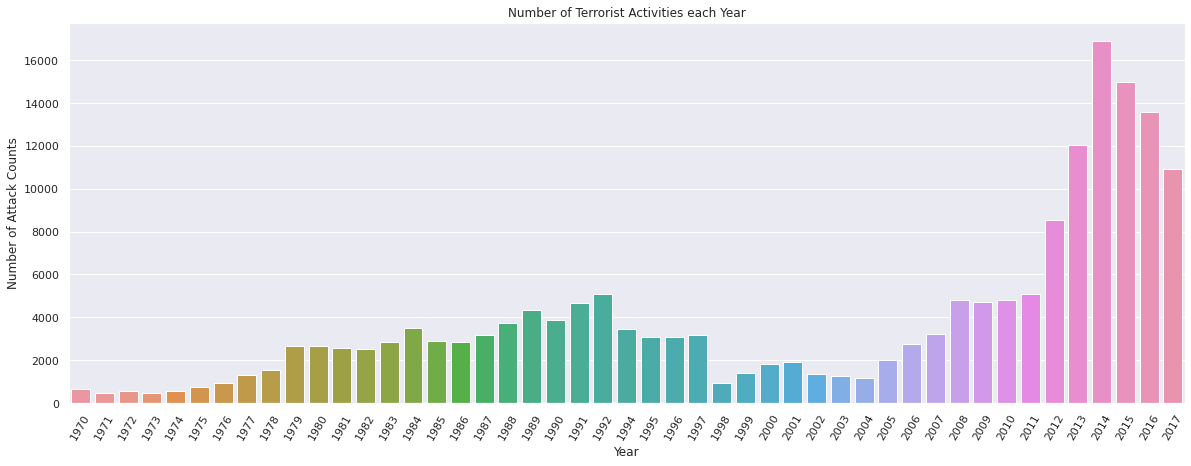
Plotly - Bar Chart in Plotly can be created using the bar() method of plotly.express class.

**Line Chart:**

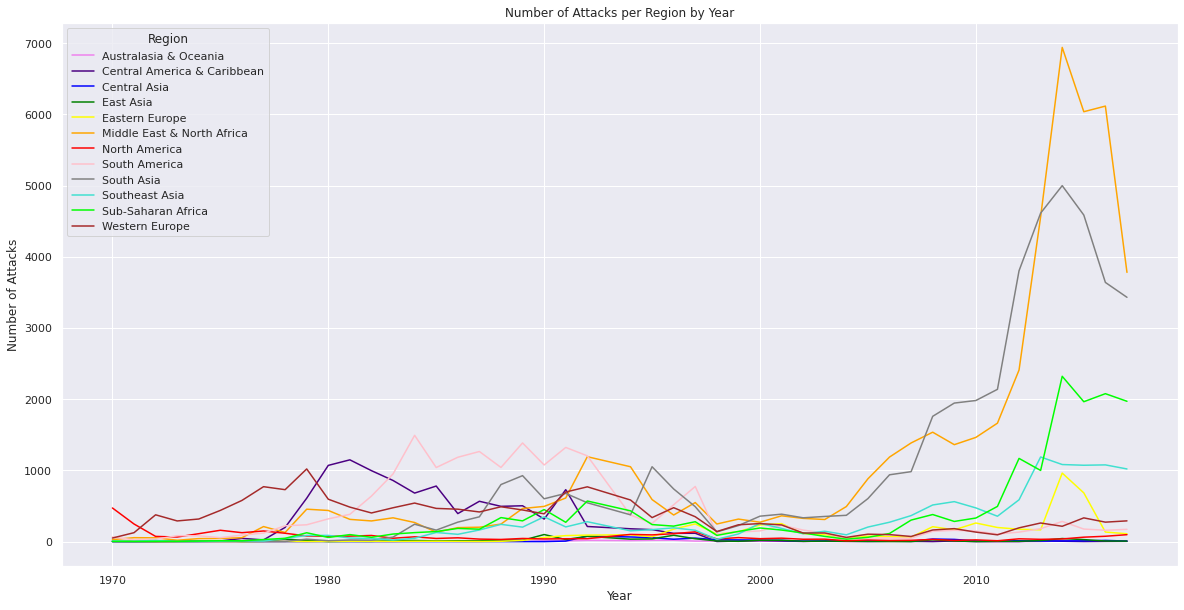
Matplotlib - [Line Chart](https://www.geeksforgeeks.org/line-chart-in-matplotlib-python/) is used to represent a relationship between two data X and Y on a different axis. It is plotted using the **plot()** function.

Seaborn - [Line Plot](https://www.geeksforgeeks.org/data-visualization-with-seaborn-line-plot/) in Seaborn plotted using the [**lineplot()**](https://www.geeksforgeeks.org/seaborn-lineplot-method-in-python/) method.  In this, we can pass only the data argument also.

# Plotly - Line plot in Plotly is much accessible and illustrious annexation to plotly which manage a variety of types of data and assemble easy-to-style statistic. With **px.line**each data position is represented as a vertex.



Above figure is of Bar Chart.



Above figure is of Line Chart.

**Global Terrorism Analysis and its results:**

**Year Wise Attack:**

Attack\_Year – It contains the data of total number of attack counts per year.

Attack\_Year.head() – It contains first five year of the data frame.

Attack\_Year.describe() – It describe the entire data frame including count, mean, standard deviation, maximum, minimum etc.

From the analysis, we concluded that the highest number of Terrorist activities took place in the year 2014 with 16903 attacks. The least number of attacks took place in the year 1971 with 471 attacks.

**Types of Attacks:**

The types of attack done by terrorists are:-

* Assassination
* Hostage Taking (Kidnapping)
* Bombing/Explosion
* Facility/Infrastructure Attack
* Armed Assault
* Hijacking
* Unarmed Assault
* Hostage Taking (Barricade Incident)

The highest Number of Attack Type is Bombing/Explosion - 88255.

The Least Number of Attack Type is Hijacking – 651.

**Types of Targets:**

Different types of targets of terrorists are:-

* Abortion Related
* Airports & Aircraft
* Business
* Educational Institution
* Food and Water Supply
* Government (Diplomatic)
* Government (General)
* Journalists & Media
* Maritime
* Military
* Ngo
* Police
* Private Citizens & Property
* Religious Figures/Institutions
* Telecommunication
* Terrorists/Non State Militia
* Tourists
* Transportation
* Utilities
* Violent Political Party
* Others
* Unknown

The most Target Type by Terrorist is Private Citizens & Property. The second most Target Type is Military and then Police.

**Country Wise Attacks:**

Top 20 countries are taken during analysis.

idxmax() - The idxmax() method returns a Series with the index of the maximum value for each column.

max() - The max() method returns a Series with the maximum value of each column.

idxmin() - The idxmin() method returns a Series with the index of the maximum value for each column.

min() - The min() method returns a Series with the maximum value of each column.

The most affected country is Iraq - 24636 attacks. The least affected country is Egypt - 2479 attacks.

**Region Wise Attacks:**

Terrorist’s attacks in different region are:-

* Australasia & Oceania
* Central America & Caribbean
* Central Asia
* East Asia
* Central Europe
* Middle East & North Africa
* North America
* South America
* South Asia
* Southeast Asia
* Sub-Saharan Africa
* Western Europe

The most affected region is Middle East and North Africa - 50474 attacks.

**City Wise Attacks:**

Top 15 Cities are taken during analysis.

* Athens
* Baghdad
* Beirut
* Belfast
* Bogota
* Istanbul
* Karachi
* Kirkuk
* Lima
* Medellin
* Mogadishu
* Mosul
* San Salvador
* Santiago

Among top15 cities, the most affected city is Baghdad – 9775 attacks and the least affected city is Medellin - 848 attacks.

Terrorist Organization:

Top 15 Terrorist Organizations are taken during analysis.

* Al-Shabaab
* Basque Fatherland and Freedom(ETA)
* Boko Haram
* Communist Party of India – Maoist (CPI - Maoist)
* Farabundo Marti National Liberation Front (FMLN)
* Irish Republican Army (IRA)
* Islamic State of Iraq and the Levant (ISIL)
* Kurdistan Workers Party
* Liberation Tigers of Tamil Elam (LTTE)
* Maoists
* New People’s Army (NPA)
* Revolutionary Armed Forces of Colombia (FARC)
* Shining Path
* Taliban

Taliban is the most active Terrorist Organization.

**Trend Analysis of Attacks per Region by Year:**

In the year 1970 to 2000, the terrorist activities were vey frequent. After 2010, we can see the rise of terrorist activity in Middle East and North Africa, South Asia and Sub Saharan Desert.

**Killed, Wounded and Casualties in top 10 Region**

Total number of people killed in each region (TOP 10):-

* Maximum - Middle East and North Africa - 137642 people killed.
* Minimum - East Asia - 1152 people killed.

Total number of people wounded in each region (TOP 10) :-

* Maximum - Middle East and North Africa - 214308 people wounded.
* Minimum - Central America and Caribbean - 8991 people wounded.

Total number of people casualties in each region (TOP 10):-

* Maximum - Middle East and North Africa - 351950 people casualties.
* Minimum - East Asia - 10365 people casualties.

**Killed, Wounded and Casualties in top 10 Countries:-**

Total number of people killed in each country (TOP 10):-

* Maximum - Iraq - 78589 people killed.
* Minimum - El Salvador - 12053 people killed.

Total number of people wounded in each country (TOP 10):-

* Maximum - Iraq - 134690 people wounded.
* Minimum - Colombia - 10328 people wounded.

Total number of people casualties in each country (TOP 10):-

* Maximum - Iraq - 213279 people casualties.
* Minimum - Philippines - 22926 people casualties.

**Killed, Wounded and Casualties in top 10 Cities:-**

Total number of people killed in each city (TOP 10):-

* Maximum - Baghdad - 21151 people killed.
* Minimum - Aleppo - 2125 people killed.

Total number of people wounded in each city (TOP 10):-

* Maximum - Baghdad - 56725 people wounded.
* Minimum - Mogadishu - 4955 people wounded.

Total number of people casualties in each city (TOP 10):-

* Maximum - Baghdad - 77876 people casualties.
* Minimum - Aleppo - 5748 people casualties.

**Conclusion:-**

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

Starting with loading the data so far we have done EDA, null values treatment, encoding of categorical columns and Data Visualization. We got lots of insights from the data sets of Global Terrorism.

**References-**

* GeeksforGeeks
* Analytics Vidhya
* Kaggle