

GLOBLE TERRORISM EDA CAPSTONE PROJECT(2022)

TECHNICAL
DOCUMENTATION

BY SANKET CHOURIYA AND TEAM

AImaBetter



ABSTRACT

The entire world in today's time is highly undergoing through terrorism. This topic not only includes the incessant obstacles thrusting in the path of global progress but also a deep-down effect on development of every county.

Here we have an impactful dataset drawn from Global Terrorism Database which comprises of data of all the terrorist attacks taken place around the world from 1907 to 2017. It includes more than 180000 attacks and a big range of massacre. It has distressed the political and economic stabilities on worldwide level.

Here our study can help understand what are the basic factors riding behind this growing problem. And some core points are milestones to take a proper action to reduce this problem at some extent.

1. PROBLEM STATEMENT

TERRORISM AROUND THE WORLD

Countries that suffered the greatest number of Terror attacks/Hostile situations

Number of deaths from terrorism at Different Regions around the world

Most Targeted Cities in the world by Terrorism

Correlation Matrix

Top 10 active Terrorist groups as per region and world

Total Casualties caused by Top 15 Terrorist Groups across the world since 1970

TERRORISM ACTIVITY TREND IN INDIA

Number of Attacks since 1970

Total Number of Killed and wounded personals caused by Terrorism in India

Top 10 Terrorist Group activities and there most used Attack

CONVENTIONAL STRATEGIES OF TERRORISTS

Hostage and Kidnapping situation Analysis

Most used Main Weapon and Sub-Weapon Type by Terrorist Groups across the World

Targeted Vandalism caused by Terrorism since 1970

And therefore, we need to present a solid data analysis which help people in predicting which country needs more security, which country is safer to live and which kind of weapons and strategies the terrorists use, so that people can be alert.





2. LOADING LIBRARIES AND DATASET



DATASET FACTS

- Dimension:
181691 x 135

IMPORTING LIBRARIES

- **Pandas**

We have used pandas for data analysis and manipulation and of better data representation.

- **Matplotlib**

We have used matplotlib for data visualization and graphs & Plots.

- **Seaborn**

We have added seaborn to make data visualization more colorful and meaningful.

- **Dataprep**

It has been used for cleaning the country elements into ISO 3166-1 alpha-3 codes for map plotting

- **Plotly**

An interactive amazing library has been used to cater more dynamic representation of our graphs and plots.

LOADING DATA

Using pandas to load the CSV file with ISO-8859-1 encoding for 8-bit character set. The Dataset exist into our drive and then finally loaded it to google Colab notebook by using this command:

```
[ ] # Install_CODE for DataPrep which is needed to get Country Codes
!pip install dataprep
from dataprep.clean import clean_country

# Basic Dependencies
import pandas as pd
import numpy as np
# For basic Plotting graph and charts
import matplotlib.pyplot as plt
%matplotlib inline
# Using Plotly Instance methods to plot your data on interactive maps and charts
import plotly.express as px
import plotly.graph_objects as go
# Convert Country name to ISO-3166 Alpha-3 codes for geographical map plotting
# Install_CODE: !pip install dataprep

# For more plotting options
import seaborn as sns
# To just ignore all the warnings.
import warnings
warnings.filterwarnings("ignore")

[ ] pd.set_option('max_rows', None) # Set max row output limit
pd.set_option('display.max_columns', None) # Set max column output limit

[ ] # Loading file from drive
from google.colab import drive
drive.mount('/content/drive')

Mounted at /content/drive
```




NaN value Management

- Replacing Alphabetical elements with "Unknown".
- Replacing Numerical elements with Mode, as per need.



Null Value count

- Before: 15 out of 25 columns had null values in it.
- After cleaning and Fixing Data Structure: All 25 columns have 0 Null values.

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 180714 entries, 0 to 181690
Data columns (total 27 columns):
#   Column              Non-Null Count  Dtype
---  -
0   Year                180714 non-null  int64
1   Incident_Duration   180714 non-null  object
2   City                180714 non-null  object
3   Country             180714 non-null  object
4   Region              180714 non-null  object
5   Summary             180714 non-null  object
6   Success             180714 non-null  int64
7   Attack_Tactic       180714 non-null  object
8   Attack_type         180714 non-null  object
9   Suicide             180714 non-null  int64
10  Target_type         180714 non-null  object
11  Target_SubCat       180714 non-null  object
12  Terror_Group        180714 non-null  object
13  Terror_Motive       180714 non-null  object
14  Weapon_Type         180714 non-null  object
15  Weapon_Subtype      180714 non-null  object
16  Total_Killed        180714 non-null  int64
17  Total_Wounded       180714 non-null  int64
18  Terrorists_Killed   180714 non-null  int64
19  Property_Damage     180714 non-null  object
20  Hostage_con         180714 non-null  object
21  Ransom              180714 non-null  object
22  Ransom_amt          180714 non-null  object
23  Ransom_amt_paid     180714 non-null  object
24  Hostage_outcome      180714 non-null  object
25  Total_Casualties    180714 non-null  int64
26  Country_code        180714 non-null  object
dtypes: int64(7), object(20)
memory usage: 38.6+ MB
```

3. PREPROCESSING DATA

DATA CLEANING

Data preprocessing is the first step to be done after collecting data. It is a set of operations performed on the START (Study of Terrorism and Response to Terrorism) dataset to modify ambiguous data which can be a bottleneck to analytical results. Data preprocessing methodology helps in converting this raw data into a more meaningful, focused, interpretable and readable format

```
[ ] # Renaming the columns which contains important values for analysis

GT_Data.rename(columns={'year':'Year','extended':'Incident_Duration','city':'City','country_txt':'Country','region_txt':'Region','summary':'Summary','success':'Success','alternative_txt':'Attack_Tactic','attacktype_txt':'Attack_type','suicide':'Suicide','targettype_txt':'Target_type','targetsubtype_txt':'Target_SubCat','name':'Terror_Group','motive':'Terror_Motive','weapon_type_txt':'Weapon_Type','weaponsubtype_txt':'Weapon_Subtype','kill':'Total_Killed','wounds':'Total_Wounded','killert':'Terrorists_Killed','propertent_txt':'Property_Damage','isostkid':'Hostage_con','ransom':'Ransom','ransomamt':'Ransom_amt','ransompaid':'Ransom_amt_paid','hostkidoutcome_txt':'Hostage_outcome'}, inplace=True)

# Now creating a new dataframe variable to keep the desired columns
Modified_GT_data = GT_Data[['Year','Incident_Duration','City','Country','Region','Summary','Success','Attack_Tactic','Attack_type','Suicide','Target_type','Target_SubCat','Terror_Group','Terror_Motive','Weapon_Type','Weapon_Subtype','Total_Killed','Total_Wounded','Terrorists_Killed','Property_Damage','Hostage_con','Ransom','Ransom_amt','Ransom_amt_paid','Hostage_outcome']].reset_index(drop=True)

# Printing first three rows
Modified_GT_data.head(3)
```

	Year	Incident_Duration	City	Country	Region	Summary	Success	Attack_Tactic	Attack_type	Suicide	Target_type	Target_SubCat	Terror_Group	Terror_Motive
0	1970	0	Santo Domingo	Dominican Republic	Central America & Caribbean	NaN	1	NaN	Assassination	0	Private Citizens & Property	Named Civilian	MANO-D	NaN
1	1970	0	Mexico city	Mexico	North America	NaN	1	NaN	Hostage Taking (Kidnapping)	0	Government (Diplomatic)	Diplomatic Personnel (outside of embassy, cons...	22nd of September Communist League	NaN
2	1970	0	Unknown	Philippines	Southeast Asia	NaN	1	NaN	Assassination	0	Journalists & Media	Radio Journalist/Staff/Factory	Unknown	NaN

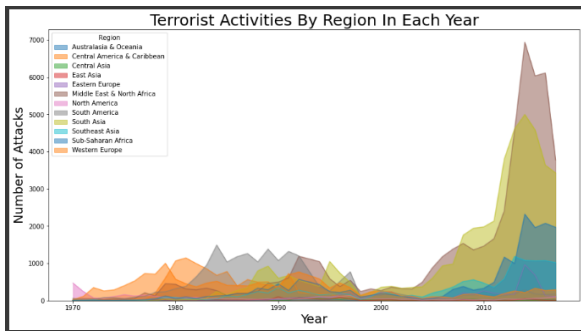
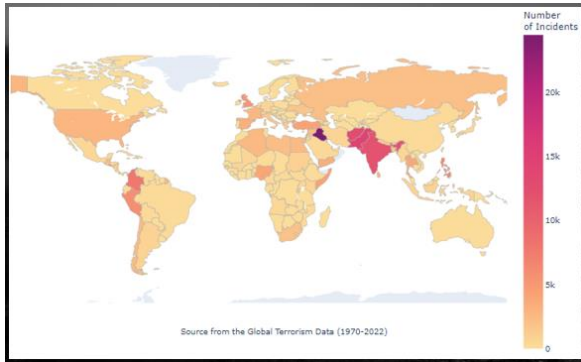
FIXING STRUCTURAL ERRORS

- Most of the column names of the dataset needed to be renamed for the purpose of easy to use and understand. A new column is created with the name casualties containing the sum of the values of total people killed and wounded.
- Above we can see that most of the important column contains a lot of NaN or missing values Thus to make a sense out of all that we will start re-placing those values with 'UNKNOWN' value to make it is when we visualize the data via Plots and Graphs. Also, you can't ignore missing data because many algorithms will not accept missing values.
- For Numerical Data, Replacing NaN values with Mode of the respective column element.
- Minor Data Improvements and re-assigning the proper Data Type.
- Converting Country name to ISO 3166-1 alpha-3 codes for map plotting
- Now we will drop above NaN values aka Non-Existent Countries from Country_code column as those countries doesn't exist anymore.
- Now Dropping useless NaN value throughout the data frame.

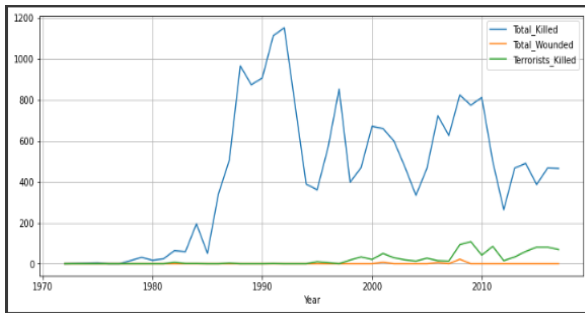
4. DATA ANALYSIS

This section consists of details regarding some of the visual results and insights used in the main project

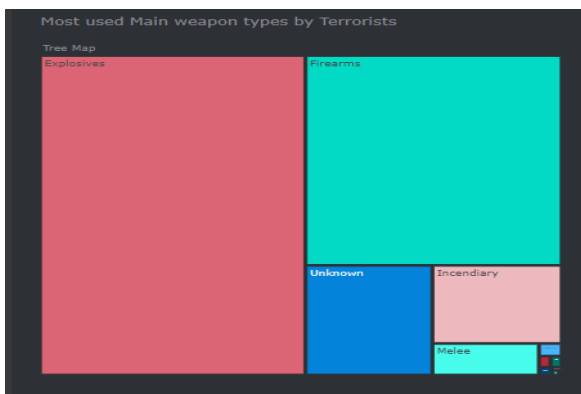
From above map, we see the number of terrorism attacks by country on this map. Iraq – the country has faced staggering 24636 number of such attacks between 1970 to 2017 which accounted for 13.62% of terrorism attacks in the world.



The below graph is plotted using an area type of graph. It shows the total number of attacks in corresponding years and regions. Most attacks were made in 2014 on Australasia & Oceania. In the 1970s we can see that most terrorist attacks occurred in Western Europe, followed by Latin and South America in the 1980s. But then there is an overall decline in the number of attacks between the mid-1990s and 2003.



In India, till now 18,663 people has been confirmed dead, 34 confirmed wounded against 933 terrorists have been encountered. As per calendar year in 1992, India seen the deadliest year of terrorism at which total 1076 personal had been killed. On the other hand, on year 2009 India Encountered the most of number of terrorists which is 108.

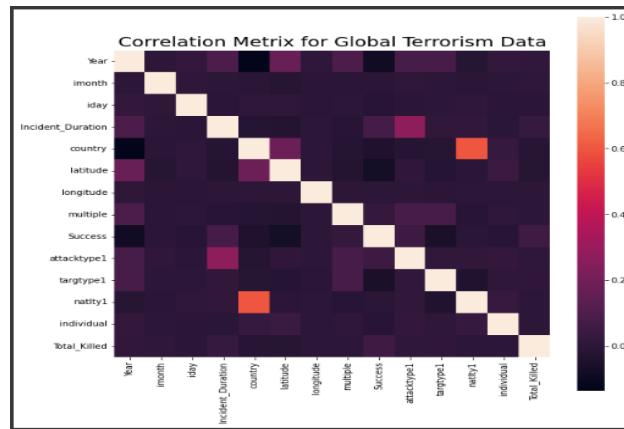


Seems like Terrorist do like 'Explosives' very much like this is the most used Weapon category across the world terror groups. There have been around 90626 incidents related to Explosives and Bombings. Whereas Terrorists also have used confirmed 56397 Firearms related attacks.

5. CORRELATION METRIX

```
# Doing correlation Metrix on untouched data for more details
data_corr = GT_Data[['Year', 'imonth', 'iday', 'Incident_Duration', 'country', 'latitude', 'longitude',
                    'multiple', 'Success', 'attacktype1', 'targtype1', 'natlty1', 'individual', 'Total_Killed']]

# Set up the matplotlib plot configuration
plt.subplots(figsize=(13, 13))
# Correlation between different variables
sns.heatmap(data_corr.corr(), annot = False, square=True)
# add overall title
plt.title('Correlation Metrix for Global Terrorism Data', fontsize =20)
plt.show()
```



Another relation we can see is among 'natlty1' and 'country'. There correlation index is 0.6 Here 'natlty1' defines the nationality of the attacker and 'country' defines the country where the attack took place. This observation shows that most of the attacks are done by the citizen of their own country. Such a relation provides an interesting insight into how to perceive international terrorism as the proportion of international terrorism is significantly less in comparison with domestic terrorism.

There a another mildly strong yet interesting correlation we can observe which is in between 'attacktype1' and 'Incident_Duration', here the index is 0.27 which shows the relation between the different type of attack tactic can be a game changer for more resilience in attack duration.

The block representing 'year' and 'success' has a darker shade which means that both these parameters are inversely related to each other. There correlation index is -0.086 So, over time, the rate of success of any attack has reduced.

6. CONCLUSION

The goal of this project was to build a tool which helps users to understand and interpret the nature of terrorism. Users can perceive the START dataset through visual designs. A visualization which can be used to calculate the total number of attacks, total kill counts and location based on the selected region and year provides interactive interface to explore this dataset. Users can understand various patterns, trends and correlation in terrorism through visual interpretation and its provided explanation.

Users can also explore START dataset and other terrorism related sources for additional research purposes provided in this tool. This work can be used by curious civilians, security related policy-makers, international organizations hosting worldwide events, foreign investors and academic researchers for the purpose of understanding terrorism and its nature.



7. REFERENCE

- Official Global Terrorism Site for insights and dataset: <https://ourworldindata.org/terrorism>
- AlmaBetter Capstone Project Tab: <https://grow.almabetter.com/data-science/projects/Global-Terrorism-Dataset>
- Wikipedia: <https://en.wikipedia.org/wiki/Terrorism>
- My GitHub Repository: https://github.com/Sanket7994/Globle_Terrorism_Capstone_Project