Capstone Project
EDA on Global Terrorism
Team: Data Titans

Team Members:

SOUMYA RANJAN RATH

ABHINASH NAYAK

ABHISEKH PATRO

SHAILEE SWASTIK

Introduction

- World peace was one of the primary motivations for establishing the United Nations

- Terrorism is the greatest obstacle to global peace. Terrorism is frequently disregarded by those who are not immediately impacted by its dangers

- Terrorism is generally regarded as an unpredictable and tragic tragedy that disproportionately affects certain regions of the world

- On the basis of the location of the incidents, the general public has extremely limited information about similar occurrences in other areas of the world and therefore responds differently

- This research focuses on terrorism by analysing the dataset supplied by an identified source in order to discover significant trends and statistics.

- It's difficult to define terrorism. The United Nations General Assembly is unable to come to an understanding on a single definition of terrorism at this time

- Different governments and organisations define terrorism differently as a result of this issue

- This ambiguity generates many conflicts regarding which events constitute terrorism and which do not. Various organisations define terrorism differently and conduct their business appropriately

- As a result, the contents of terrorism-related datasets collected by independent groups may differ significantly Consequently, the analysis and outcomes of this study may differ from identical studies performed on a different dataset

Global Terrorism Database

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.
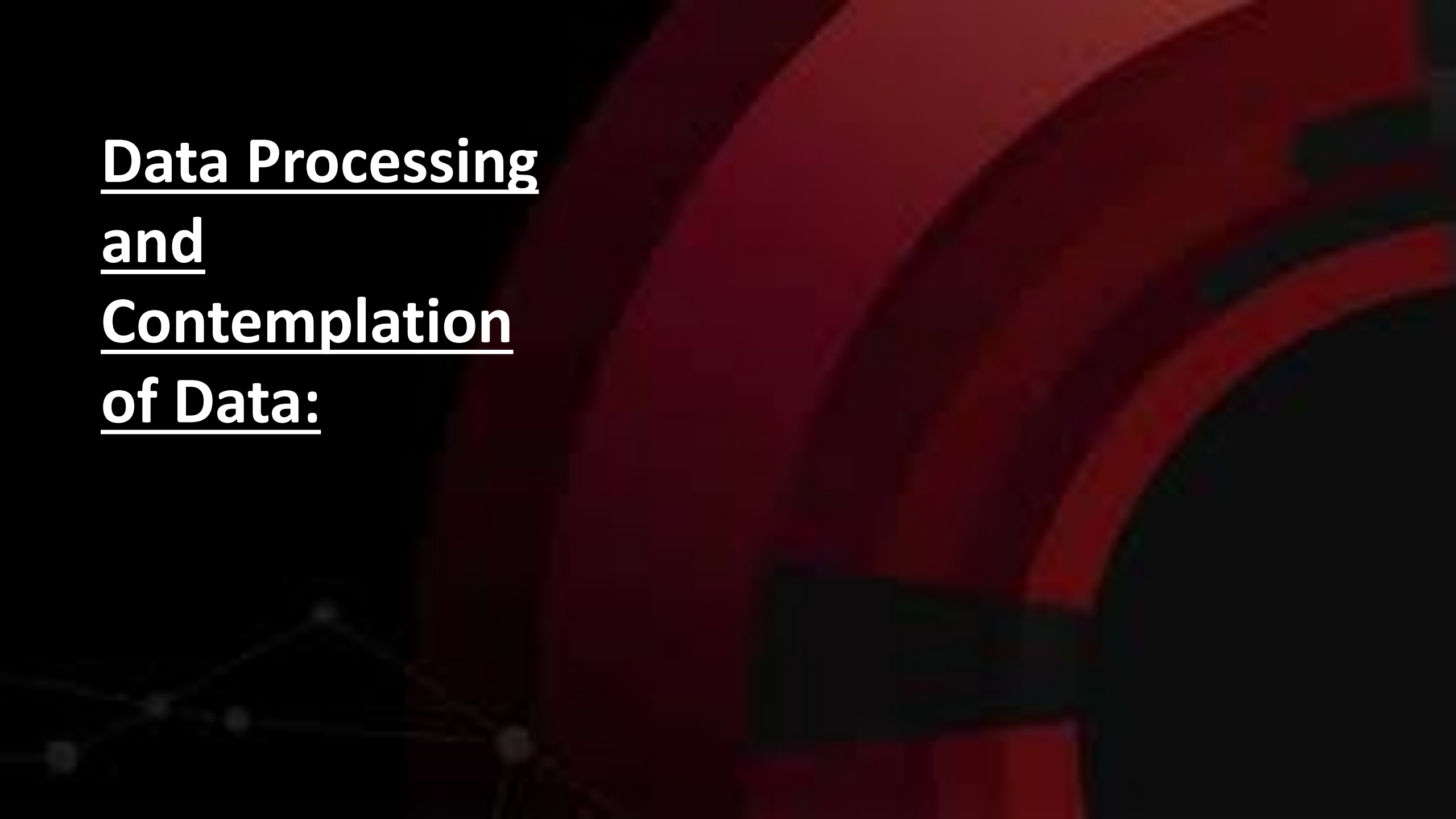
Now We have the Database to work on the Analysis on

Global Terrorism from 1970 to 2017.

Misconceptions About Terrorism

- Terrorism is intermittent, pervasive, and inconsistent with time and nature

- Due to these qualities, it is impossible to synthesise all aspects of international terrorism into a single, convincing solution and make this knowledge accessible to the general public

- Exploring this dataset can reveal how different factors are associated with one another, hence facilitating the identification of unknown hidden patterns

- This investigation will also present sufficient evidence to justify certain popular fallacies about terrorism

- One of the common misunderstandings is that a larger military can repress and control terrorism

Factors affecting the Terrorism

- Identifying terrorism's dependant elements is one of the objectives of this study.

  - Religion

  - Nationality

  - Government

Data Processing and Contemplation of Data:

Dataset Challenges

- This dataset presents a significant problem due to the fact that individual investigations lead to contradictory outcomes.

- Current deficiencies and limits in data gathering methodologies, arguments over definitions, and irregularities in coding and processing give rise to disagreements among researchers, so invalidating their conclusions .

- Critical disagreement over the definitional arguments surrounding major terrorist acts has a negative impact on the growth of this area .

- This issue necessitates exercising the requirement for common grounds that can be acknowledged by the majority of specialists and relevant authorities in order to agree on what may be the standard norms and method to be deemed a valid piece of information on terrorism upon which proper study can be conducted .

DATA PRE-PROCESSING

- After data collection, the first stage is data pre-processing. It is a series of operations done on dataset to change unclear data that can impede analytical conclusions. Raw data is essentially a collection of interconnected information. Oftentimes, raw data is unstructured and contains a great deal of information that is unnecessary to the project's requirements. Methodologies for data preparation facilitate the transformation of these raw data into a more meaningful, focused, interpretable, and readable manner.

- The dataset from the Global Terrorism Database is insufficient, inconsistent, and contains several errors, missing attribute values, outliers, improper tags, and duplicate entries. These disparities can be resolved through data preparation.

- The steps employed in this project's data pre-processing methodology are as follows:

  Data Cleaning: Filling in missing numbers, removing outliers, and handling irregularities in data constitute the data cleaning process. Numerous categories, such as 'motives' and 'responsible organisations,' are absent from the terrorist dataset due to a lack of information or because the field was irrelevant to the incident in question.

DATA PRE-PROCESSING (continued)

- **Data integration:** In this phase, data discrepancies are settled. To avoid misunderstanding and redundancy, different representations of the same data are combined, such as multiple heads killed and wounded to Casualty

- **Data transformation:** Here data aggregation, generalisation, and normalisation are conducted. The dataset contains several subtypes of target and victim. All of these subtypes were aggregated to reflect a single value by adding all subtypes with similar characteristics. This method minimises the overall number of attributes in the dataset, hence minimising the data's variability

- The dataset features a significant degree of data sparsity, which raises its overall dimensionality.

- This strategy diminishes the efficacy of density-related procedures such as clustering and outlier detection. Multiple fields include a greater number of missing or null values than legitimate ones

Technologies Used:

- **Python** is an advanced programming language that supports multiple platforms, including Windows, Linux, Mac, and Raspberry Pi, among others. Python can be used to build online applications, database systems, manage large amounts of data, and conduct complicated mathematical calculations. Python is object-oriented, functional, and procedural. These are some of the Python packages utilised by this project

- **Matplotlib** is a 2D plotting tool that includes the necessary modules and functions. A developer can adjust font properties, styles, axis properties, etc

- **Pandas** is used for data manipulation and analysis. Pandas can transform data structures and dataset formats to data frames on which operations like as loading data, renaming attributes, mapping, crosstab, sub-data frames, etc. can be carried out

- **NumPy** offers structures for multidimensional array objects as well as tools for associated operations. Typically, NumPy is employed for high-performance scientific computations

- **Seaborn:** It is a library that uses Matplotlib underneath to plot graphs. It will be used to visualize random distributions

Technologies Used: (continued)

- **Correlation** refers to a process for establishing the relationships between two variables.

  - Pearson Correlation coefficient used for linear dependency between the data sets. The value of the coefficient lies between -1 to +1. When the coefficient comes down to zero, then the data is considered as not related. While, if we get the value of +1, then the data are positively correlated, and -1 has a negative correlation

- **Statistics- Arithmetic Mean and Std Deviation:**

  - The arithmetic mean is the simple average, or sum of a series of numbers divided by the count of that series of numbers

  - The *Standard deviation* describes how dispersed a set of data is

- **Google Collab**: Collaboratory, or "Collab" for short, is a product from Google Research. Collab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education

- **Plotly's** Python graphing library makes interactive, publication-quality graphs

Exploratory Data Analysis(EDA):

Key Findings

89%
Attacks- Success Rate (%)

Maximum **1570** **Killed**

Most Affected Nation **IRAQ**

Deadliest Terrorist Group **TALIBAN**

Countries affected by Terrorism **190+**

1.Number of Attacks in each Year(1970-2017)

Most Number of Attacks in year 2014:16903

Timeline of Attacks

Least Number of attacks in year 1971:471
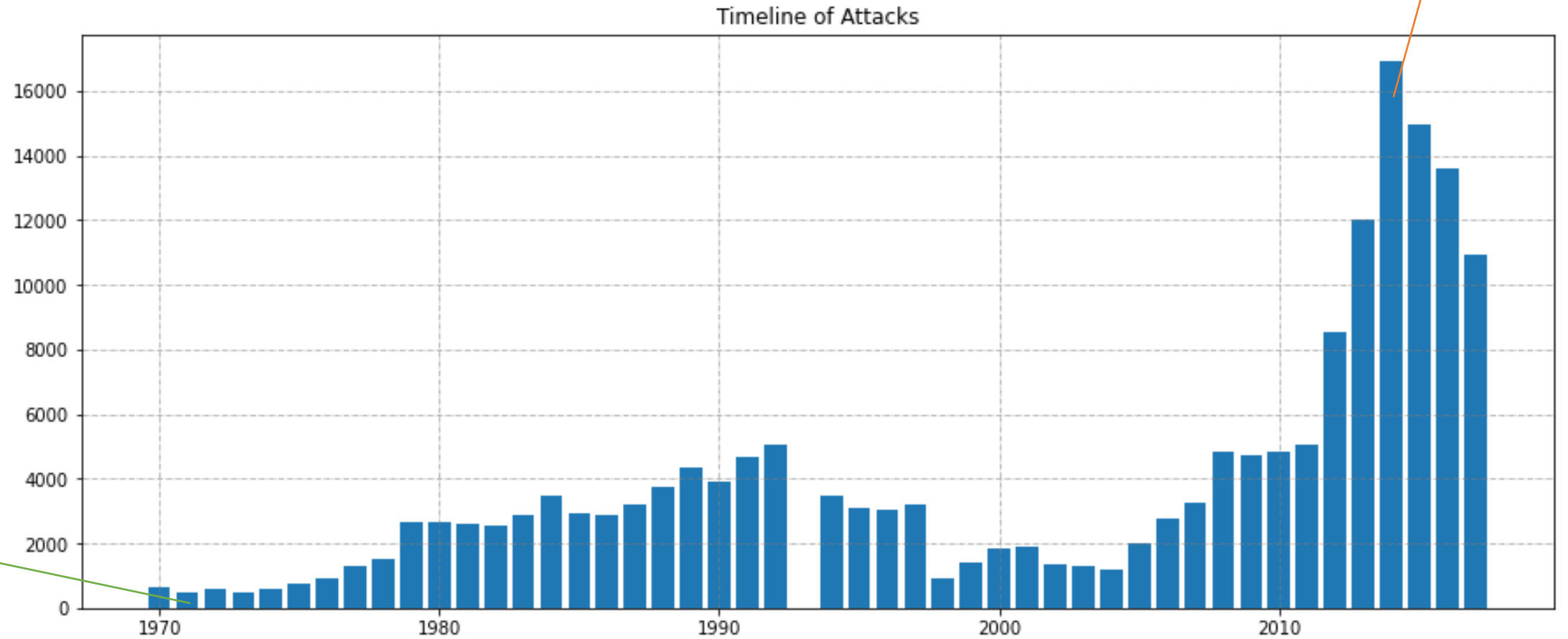
Explanation:

- The objective of this analysis is to better understand the terrorist activities through finding patterns and trends from the data

- From this analysis, all activities that occurred in a given year in each available geographical location in the world are represented. This study allows us to determine whether terrorist activities are increasing or decreasing per year

- We can see that in the year 2014 we had the most amount attacks and least amount attack in the year 1971. We can see that the number of attacks occurring increasing as the year progresses from 1970-2017 , so to understand the cause of   drop in the number of attacks we made analysis further analysis on region wise attacks

- So that we could see which region is worst affected and which is best at tackling terrorism
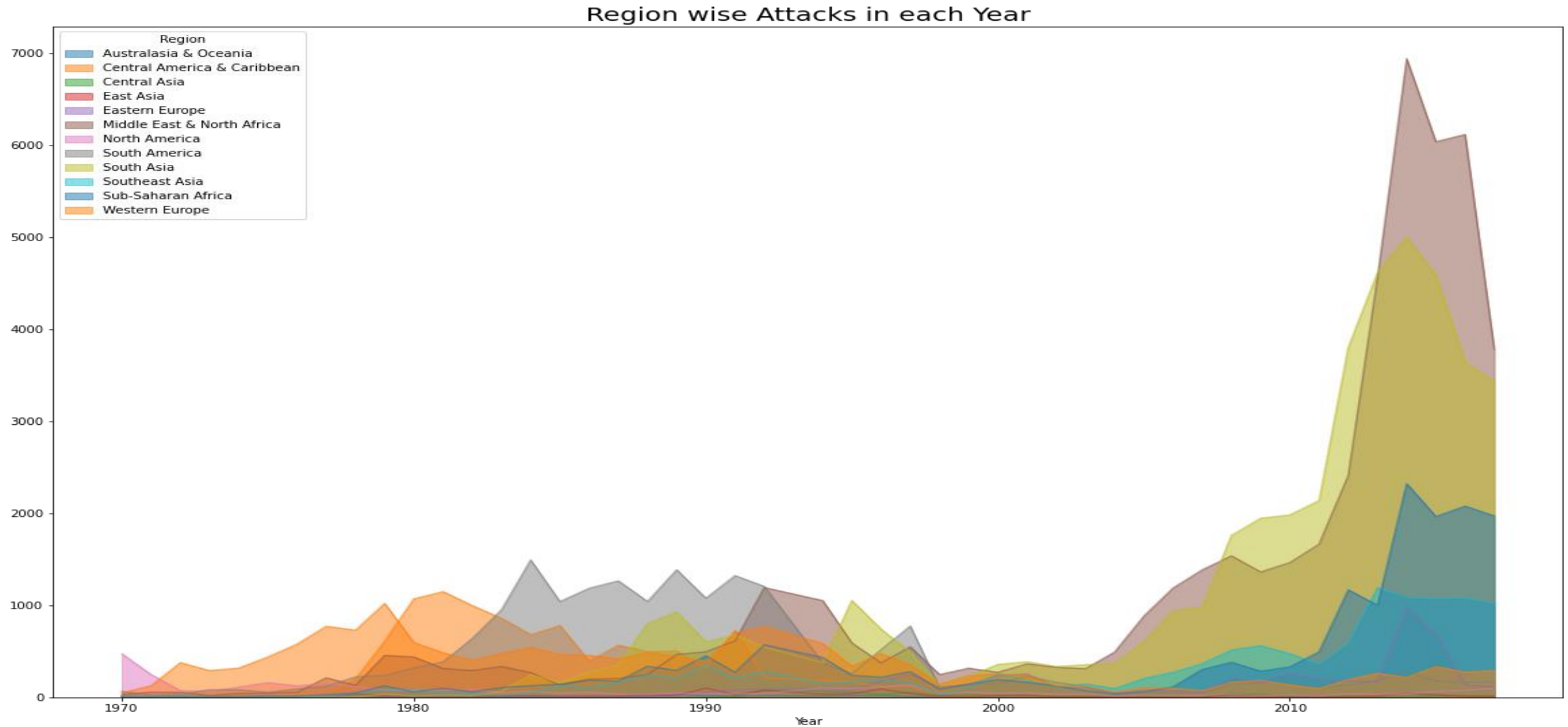
Measuring the Impact of Terrorism:
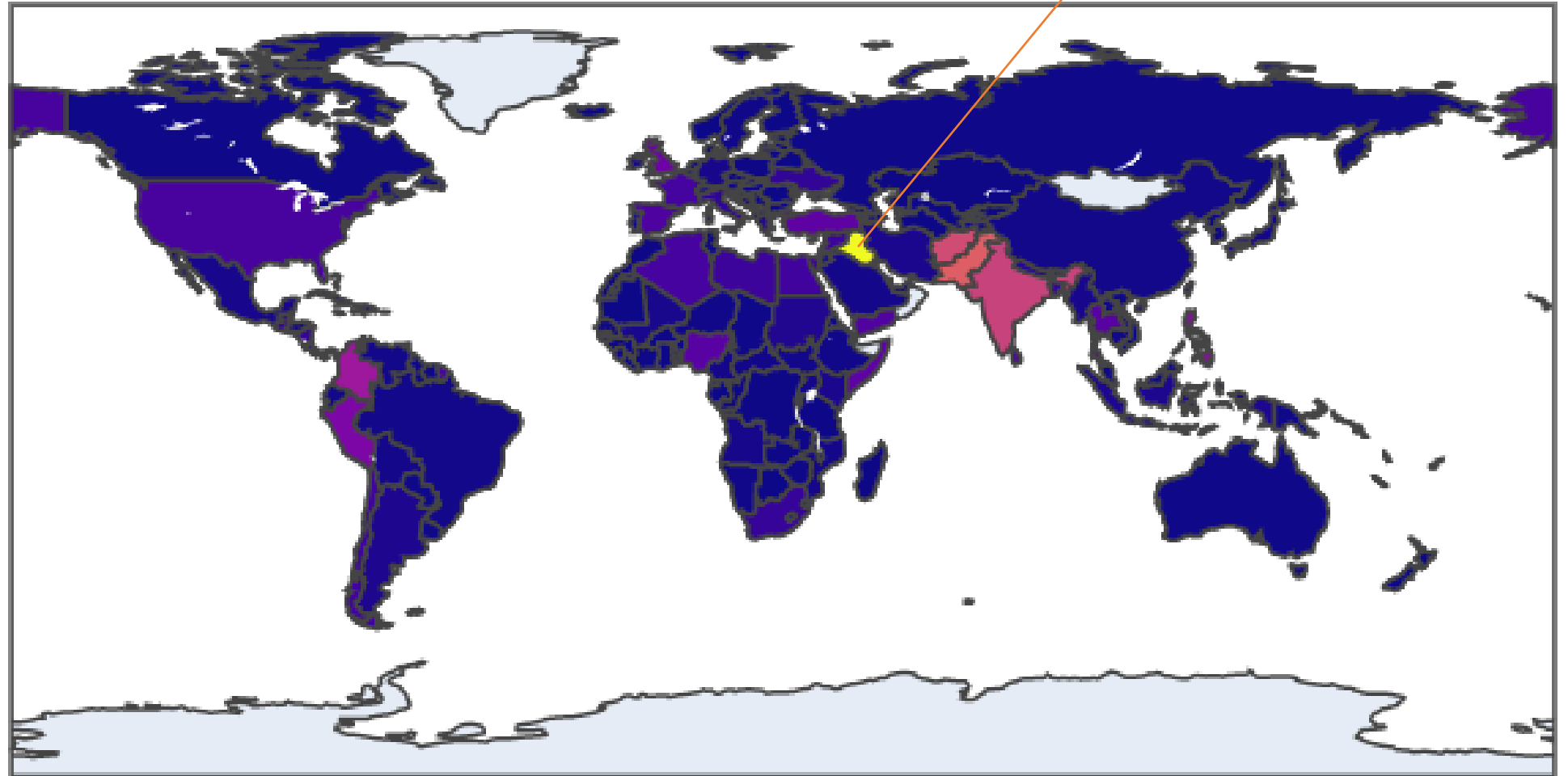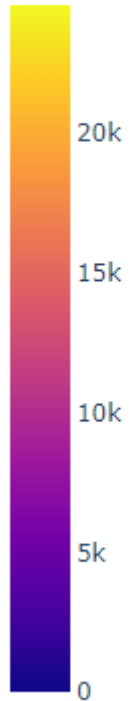
IRAQ-Most Impact Nation

# 2.Region wise Impact:

Region wise Attacks in each Year

Explanation:

- In the region wise attacks we focused on year and the region features to find how the distribution of attacks in the region across the years

- This distribution will helps us to better understand that which are the region have to face most of the attacks through out the years

- We can see that recently middle east and north Africa are the worst affected regions followed by South Asia

3.Success rates of Attacks by Terrorist group in each Regions

Most Impacted Region

Success Rate of Attacks by Terrorist groups in different Regions
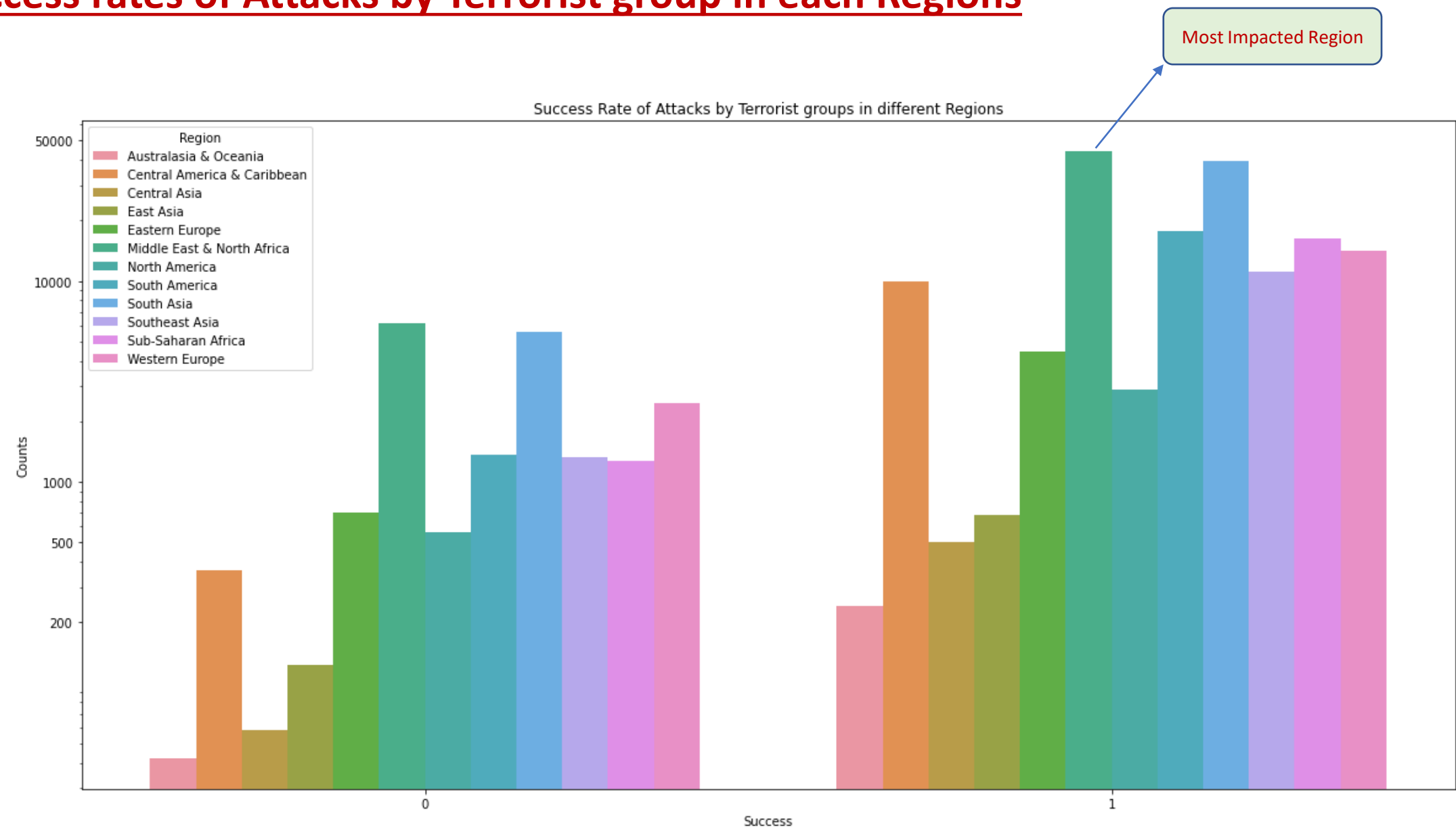
Explanation:

- To understand how well the regions have been in controlling the situation and preventing the attacks from being successfully executed

- We can see that maximum attacks successfully executed and failed by terrorists in Region Middle East & North Africa and minimum attacks successfully executed was in region Australasia and Oceania

4.Total Number of Attack in Each Contry:

Terrorist Attack per Country
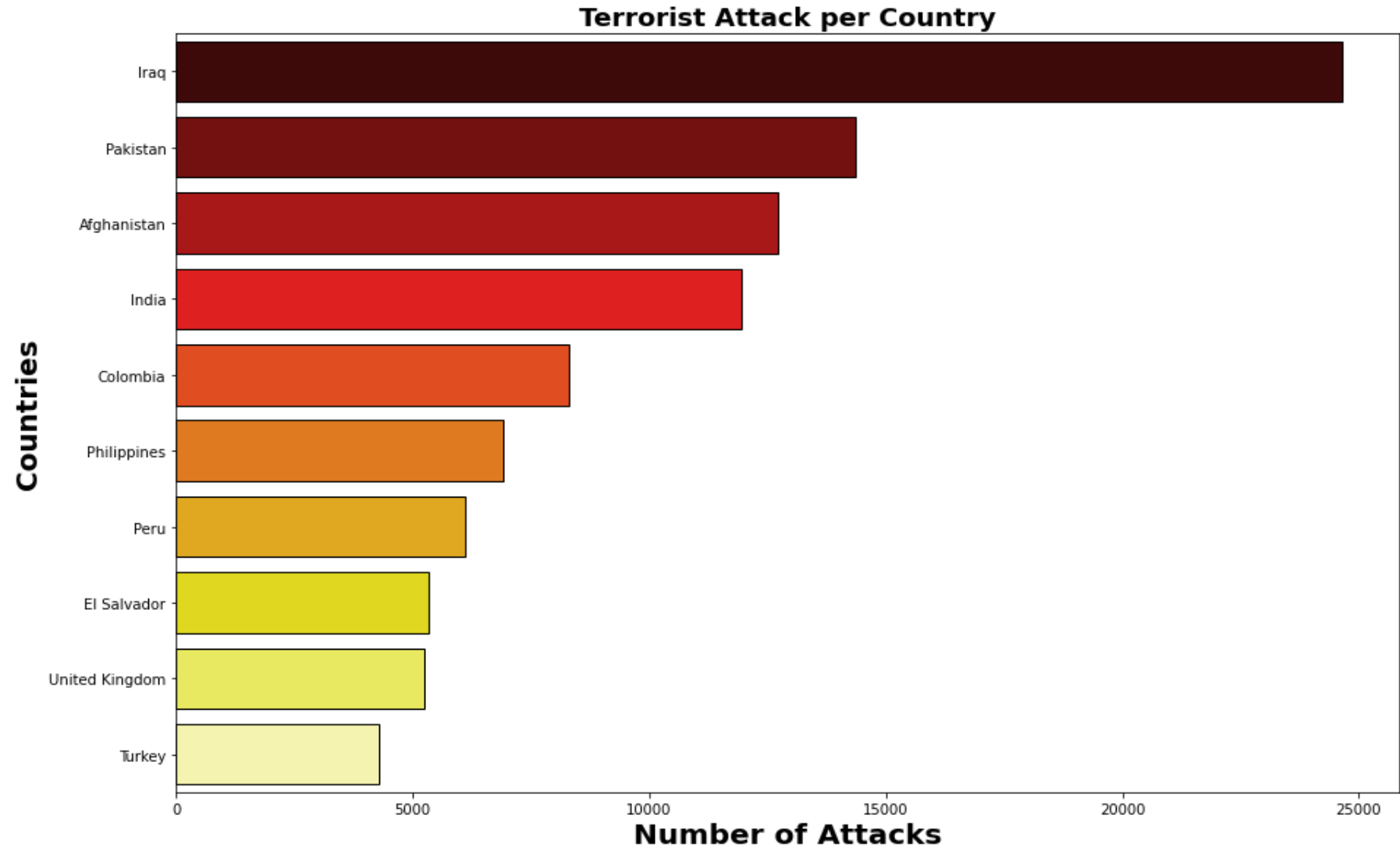
Explanation:

- This analysis was carried out to see how each country in our dataset is affected by these attacks

- We have calculated the number of times the country has been there in the different terrorist attacks and through that we could find the total attack on each country

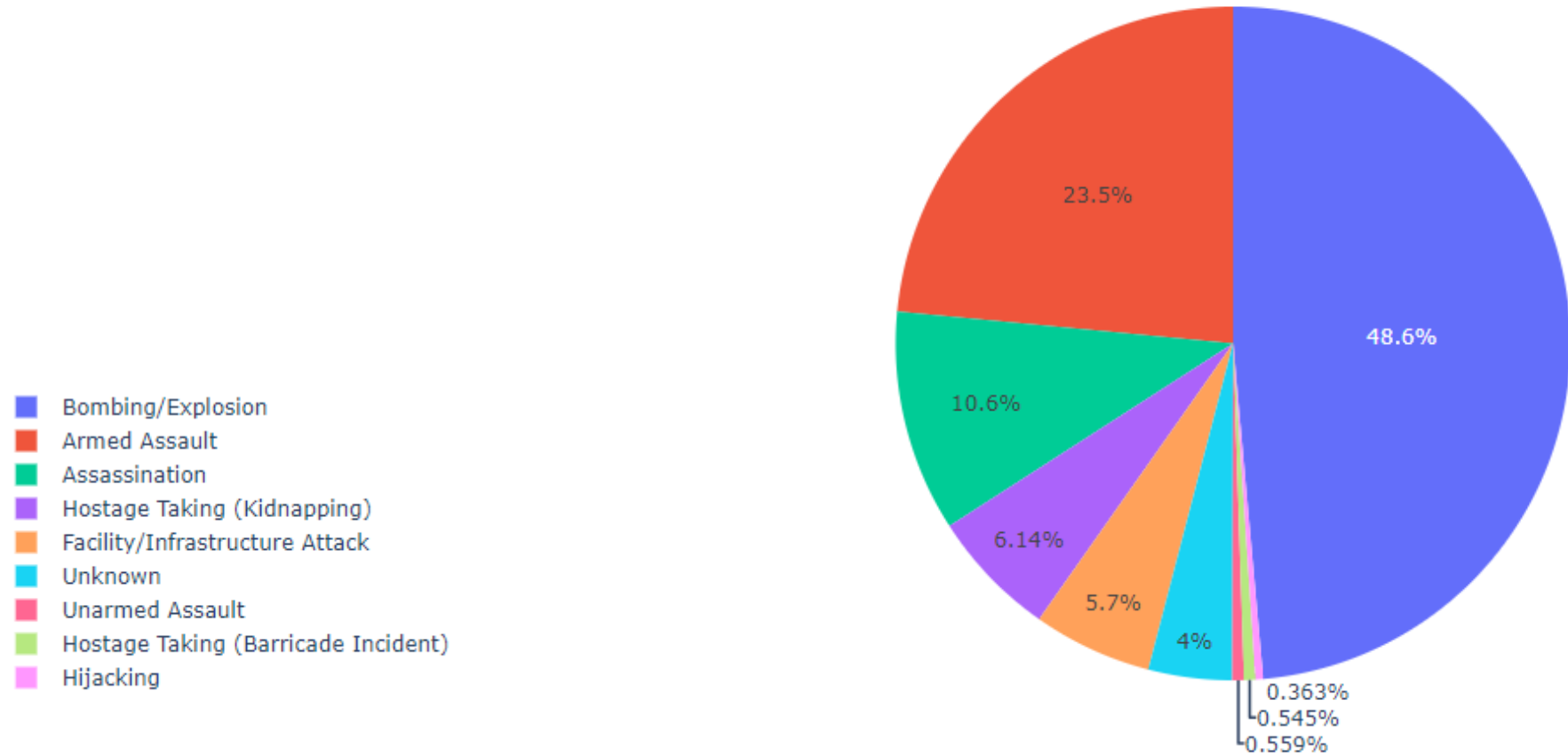- We have plotted the top 10 countries with the highest attacks attempts

5.Top 10 Terrorist Groups who has Successfully executed their Attack Mission

Attempt VS Success

Failed attempts

Successful attempts
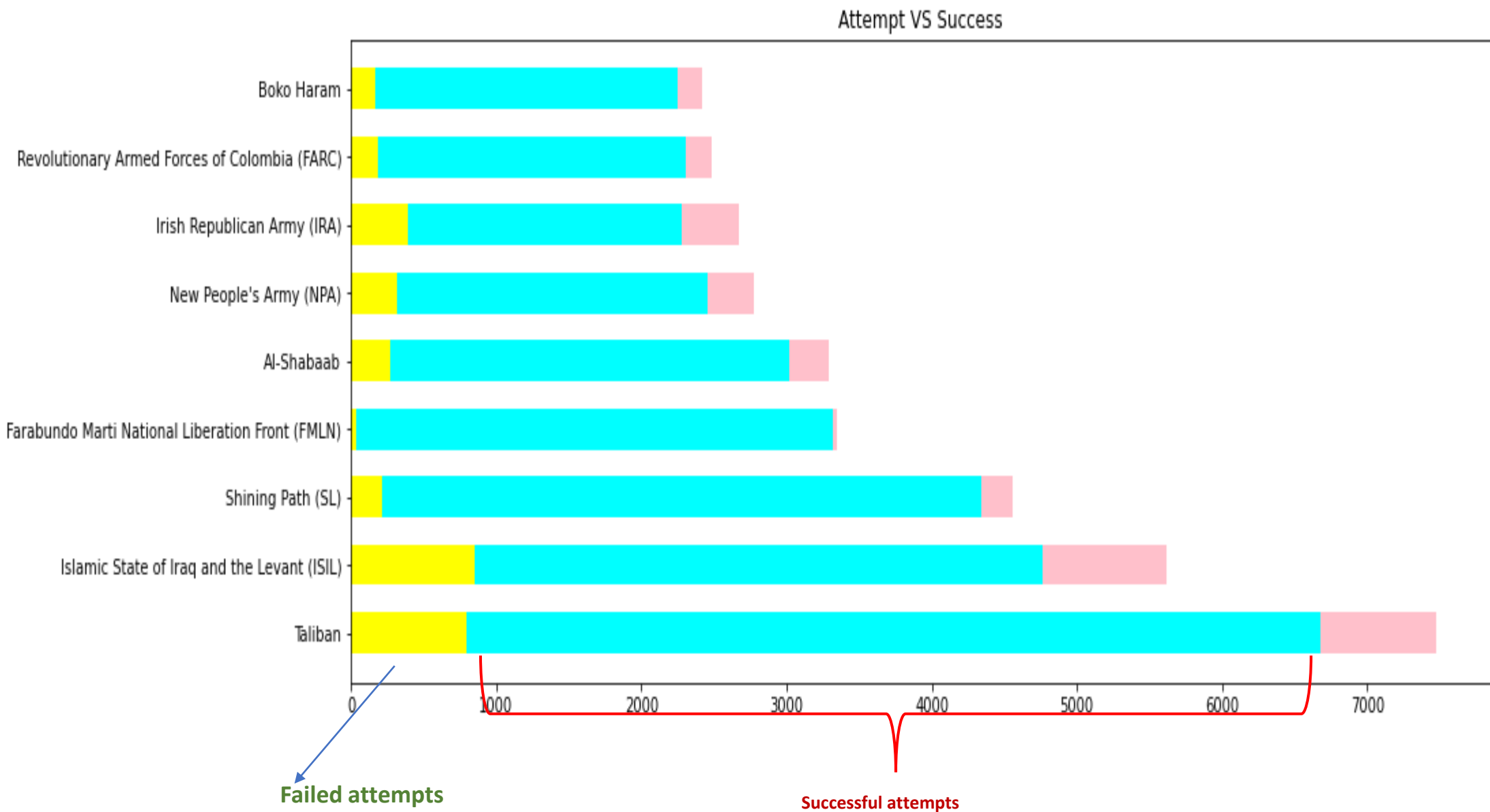
Explanation:

- This analysis will tell us which are the most dangerous terrorist organisations in world. Here we are using features like terrorist group, success, number of attacks as our features to find the ranking of the terror organisation with respect to successfully executed missions

- Terrorist Group : Farabundo Marti National Liberation Front (FMLN) has the highest success rate of their Attack mission with 3351 total attempts and Success rate 98.98%

- Taliban is the Top terrorist group with the attack mission with 7478. It has the success rate of 89.32%

6.Most Attacked Methods used by the Terrorist groups

Legend:
- Bombing/Explosion
- Armed Assault
- Assassination
- Hostage Taking (Kidnapping)
- Facility/Infrastructure Attack
- Unknown
- Unarmed Assault
- Hostage Taking (Barricade Incident)
- Hijacking

48.6%
23.5%
10.6%
6.14%
5.7%
4%
0.363%
0.545%
0.559%

Explanation:

- Here from this analysis we are trying to understand which attack methods are favoured by the terrorist groups for their attack missions

- We select Attack_Type and Terrorist_Group as features to do this analysis

- We see that most of the Deadliest terrorist groups have preferred Bombing and Explosion for their Attack mission about around 50%(actual-48.6%). This means they have the target on masses and explosive sources available for their Deadliest missions

# 7.Targets of the Deadliest Terrorist Groups
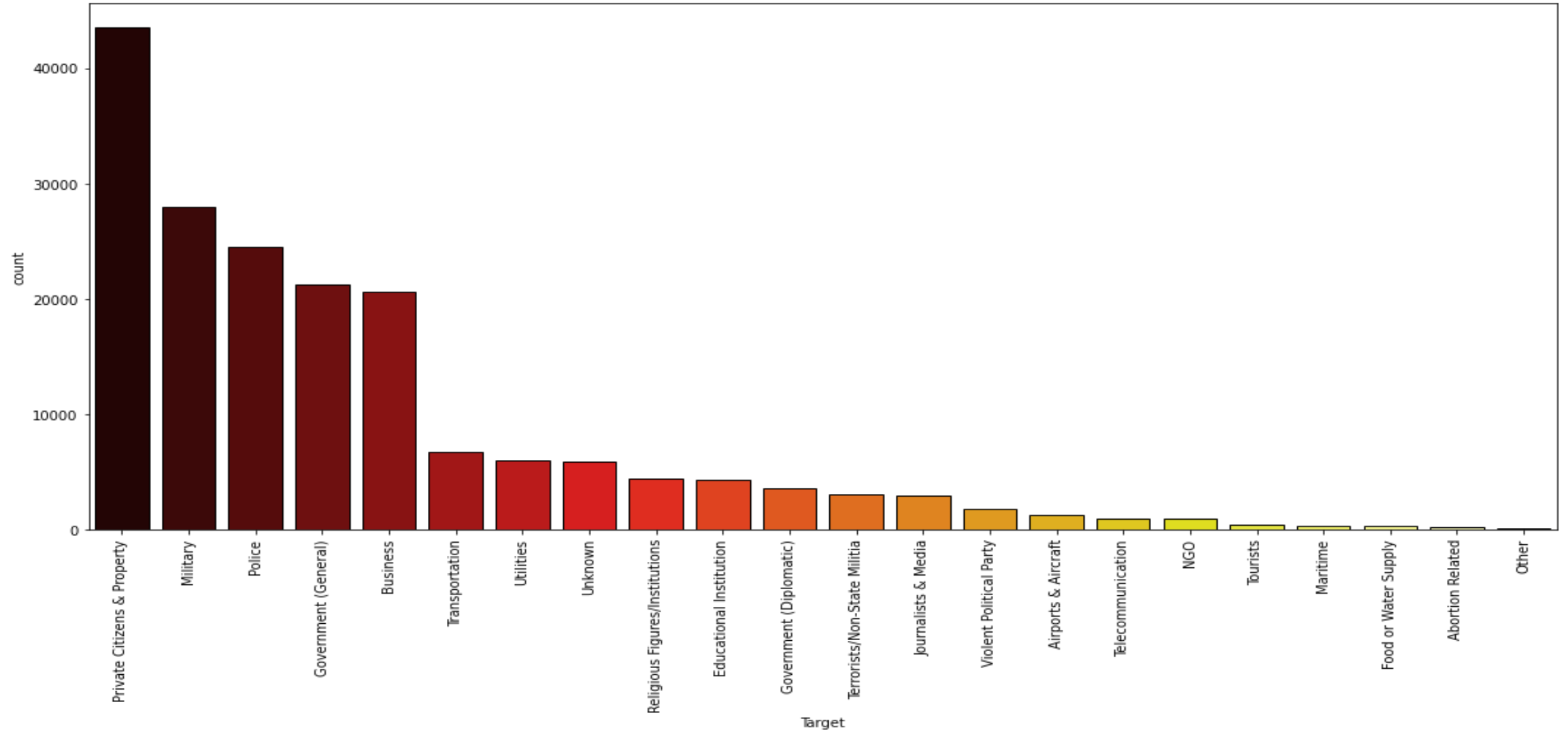
Explanation:

- In this analysis we have considered features like Target, country, and terrorist group

- This was done to understand the which are the targets more prone to the terrorism organisations attacks

- Terrorist Groups has targeted the Private citizens and Property with 43.511k attacks followed by Military with 27.984k attacks

8.Correlation between Killed, Wounded and Total Casualty

*Heat Map(for visualizing the Intensity of the Correlation)
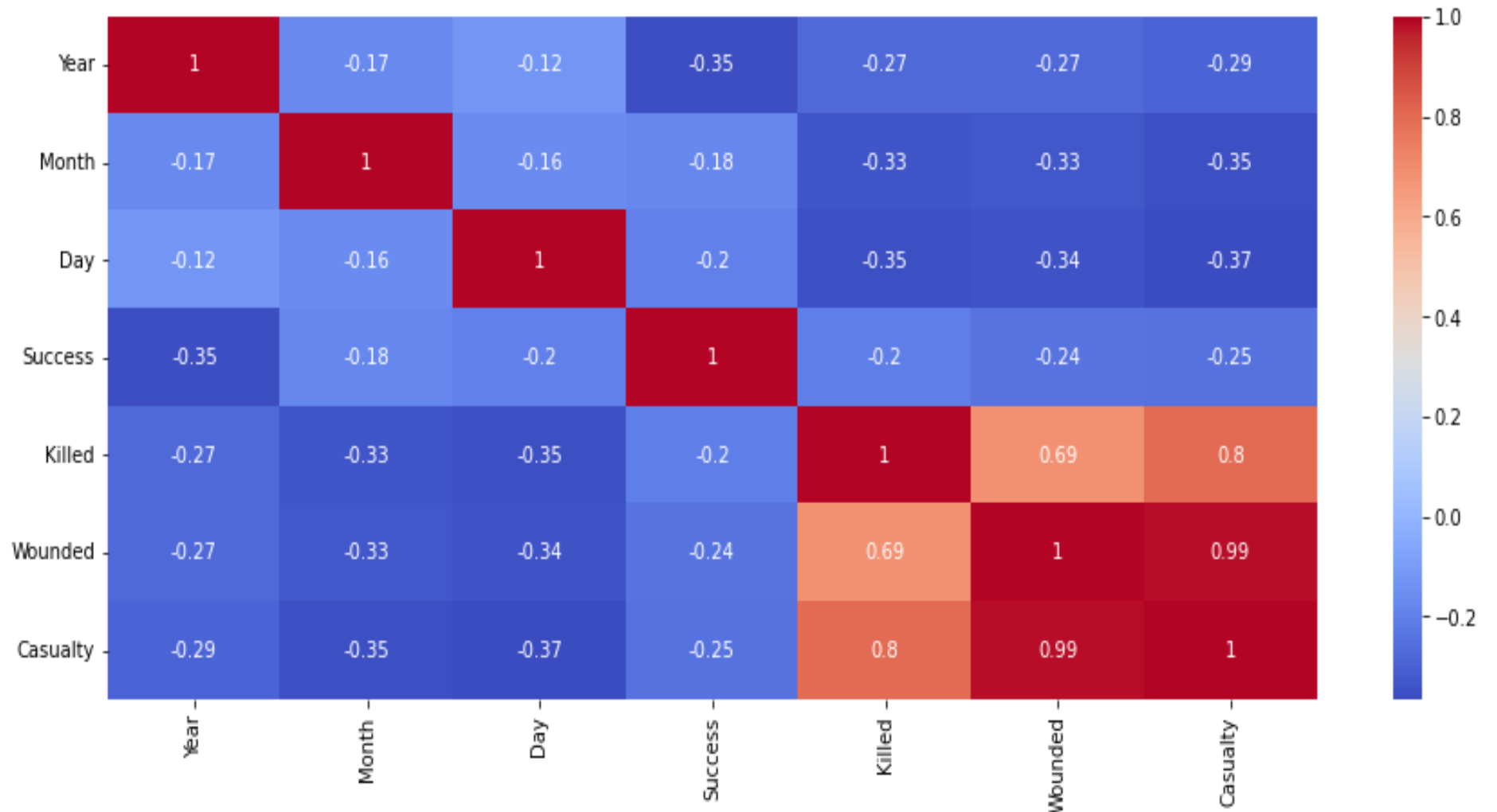
using the scatterplot
to visualize the linear
 relationship
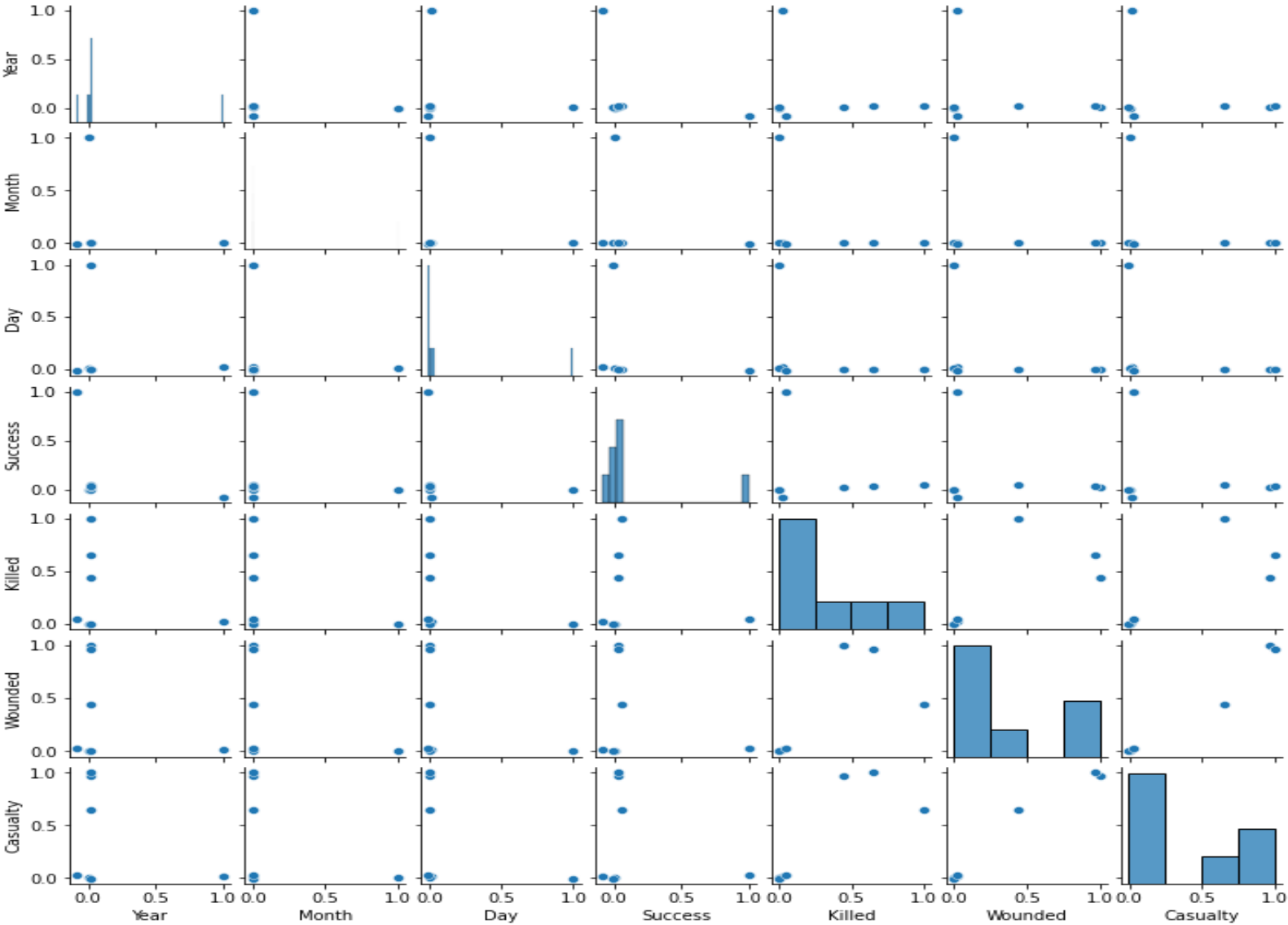
Explanation:

- In this analysis we have attempted to find the linear relationship using correlation method between killed, wounded and total casualty

- We have imported all the numerical features in our analysis data set to find the correlation. The correlation between wounded and Casualty is 0.99 means 99% of the variance in Casualty, is accounted for by the wounded. The correlation between Killed and Casualty is 0.80 means 80% of the variance in Casualty, is accounted for by the Killed

- The correlation between wounded and Killed is 0.69 means 69% of the variance in Killed, is accounted for by the wounded

Conclusions:

- Starting with loading the data so far we have checked the data shape, data head, data info.
- Then since the data contained lot of column labels so we filtered it to our need first then again imported the data to the Collab.
- After that we looked for null values in the data set and started doing data cleaning till we successfully take care of null values.
- Then Data visualisation was done for the analysis to find different trends according to our need for better understanding.
- From the data we can see that middle east and Africa region were prone to more attacks
- Iraq is the country with the highest number terrorist attacks
- Terrorist Group Farabundo Marti National Liberation Front (FMLN) has the highest success rate of their Attack mission with 3351 total attempts and Success rate 98.98%.

- Taliban is the Top terrorist group with the attack mission with 7478. It has the success rate of 89.32%.
- Most of the Deadliest terrorist groups have preferred Bombing and Explosion for their Attack mission about around 50%(actual-48.6%).
- This means they have the target on masses and explosive sources available for their Deadliest missions.
- Terrorist Groups has targeted the Private citizens and Property with 43.511k attacks followed by Military with 27.984k attacks.
- The Most common weapon used for the Deadliest Missions are explosives mainly for bombing and explosions. Followed by Firearms for the Armed Assault.

- The correlation between wounded and Casualty is 0.97 means 97% of the variance in Casualty, is accounted for by the wounded
- The correlation between Killed and Casualty is 0.80 means 80% of the variance in Casualty, is accounted for by the Killed
- The correlation between wounded and Killed is 0.69 means 69% of the variance in Killed, is accounted for by the wounded

References

[1] United Nations, "Chapter-1 Purposes and Principles," [Online]. Available: https://www.un.org/en/sections/un-charter/chapter-i/index.html [Accessed: May 2019]

[2] A. Z. Borda, "Why we react differently to terror attacks depending on where they happen," [Online]. Available: http://theconversation.com/why-we-reactdifferently-to-terror-attacks-depending-on-where-they-happen-57389 [Accessed: May 2019]

[3] START organization, "Global Terrorism Database," [Online]. Available: https://www.start.umd.edu/gtd/about/ [Accessed: May 2019]

[4] ALMABETTER [Online],https://learn.almabetter.com/courses/take/team-capstone-projects/texts/19443222-world-bank-global-education-analysis