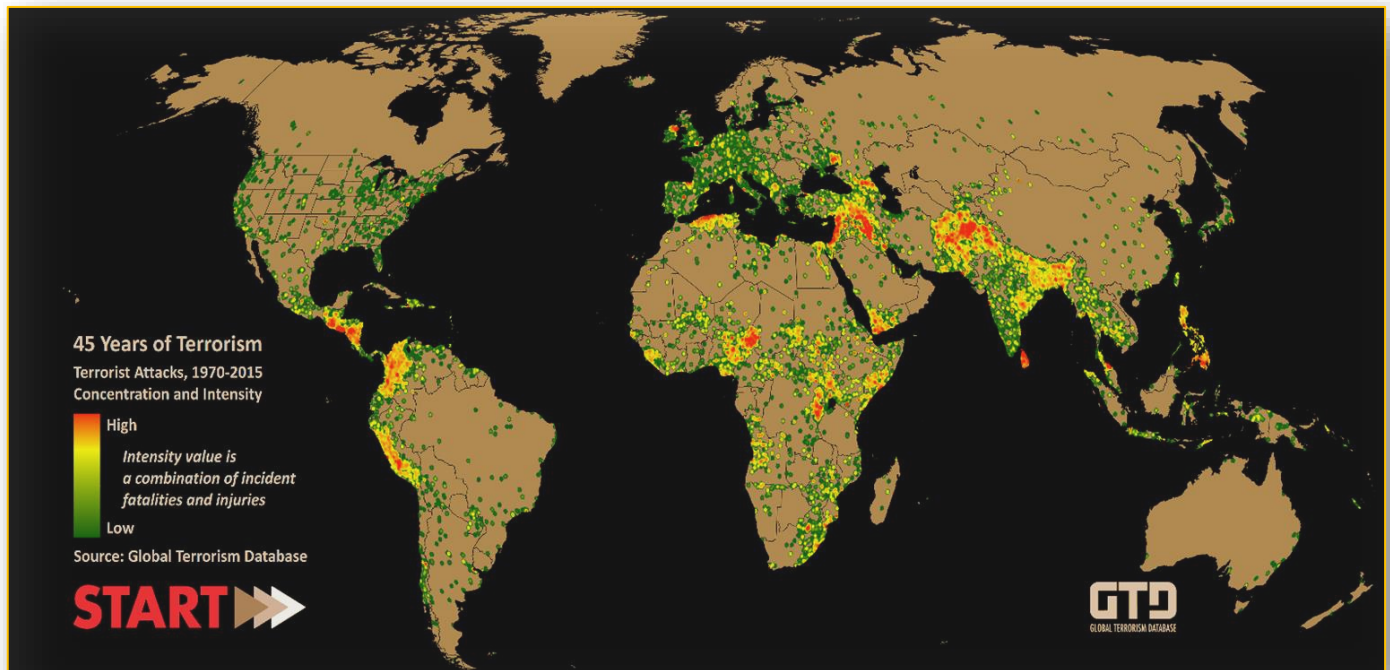


“CAPSTONE PROJECT”

Exploratory Data Analysis

GLOBAL TERRORISM ANALYSIS



BY,

VAIBHAV SANTOSH MANDE

ABSTRACT - The objective of this work is to analyze the region and country of a terrorist attack using exploratory data analysis, in order to acquire valuable information about the predicted attacks and attackers.

The work has been carried out upon the Global Terrorism Database (GTD), which is an open database containing list of terrorist activities from 1970 to 2017.

Keywords: Exploratory data analysis, Terrorist activities, Terrorist groups, Attack region.

1. PROBLEM STATEMENT

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.

2. INTRODUCTION

The GTD defines terrorism as-

"The threatened or actual use of illegal force and violence by a non-state actor to attain political, economic, religious or social goal through fear, coercion, or intimidation."

In this EDA project, we are doing the exploratory data analysis on the global terrorism database to find out various insights of terrorism activity and their effect over region, country and city. Also, we did some visualizations for easy understandings of the analysis.

After analyzing the cleaned data, we divided the project into various types of analysis. The first analysis based on Trend of Terrorism Per Year. the second analysis is analysis based on Most affected

areas of Terrorism. Third analysis based on Terrorist Organization and the last analysis is based on country.

In the EDA we have conclude that after 2011 there is a large growth recorded in terrorist activity and 2014 appears to be a witness of huge terrorist attack.

3. GLOBAL TERRORISM DATASET

The Global Terrorism Database Includes information on terrorist attacks around the world from 1970 through 2017.

It 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.

The data set contains the following columns:

- **Year:** - Year of Terrorist Attack
- **Month:-** Months of Terrorist Attack
- **Day:-** On which Day Terrorist Attack is Happened
- **Latitude:-** Y Co-ordinates of Terrorist Attack point
- **Longitude:-** X Co-ordinate of Terrorist Attack point
- **Region:-** Regions in which Terrorist Attack happens
- **Country:-** Countries in which Terrorist Attack happens
- **State:-** States in which Terrorist Attack happens
- **City:-** Cities in which Terrorist Attack happens

- **Attack_Type :-** Type of Terrorist Attack
- **Killed:-** Number of people killed in the Terrorist Attack
- **Gang_Name :-** Name of Terrorist Organization
- **Wounded:-** Number of people Injured in the Terrorist Attack
- **Main_Target :-** Main target of the Terrorist Attack
- **Sub_Target :-** Sub target of Terrorist Attack
- **Weapon_type :-** Weapon used in the Terrorist Attack.
- **Motive :-** Motive Behind the Terrorist Attack
- **Nationality :-** Nationality of Terrorist Organization
- **Casualties :-** It is the Sum of Number of people killed and Number of people Injured in the Terrorist Attack

3.1 PYTHON

Most of the info scientist use python due to the good built-in library functions and therefore the decent community. Python now has 70,000 libraries. Python is simplest programming language to select up compared to other language. That is one among the most reasons to use python. Specifically, for data scientist the foremost popular data inbuilt open-source library is named panda. The Python standard library consists of more than 200 core modules. All these work together to make Python a high-level programming language. We used following libraries,

-NumPy : The name “NumPy” stands for “Numerical Python”. It is the commonly used library. It is a popular machine learning library that supports large matrices and multi-dimensional data

-Pandas: Pandas are an important library for data scientists. It is an open-source machine learning library that provides flexible high-level data structures and a variety of analysis tools. It eases data analysis, data manipulation, and cleaning of data. Pandas support operations like Sorting, Re-indexing, Iteration, Concatenation, Conversion of data, Visualizations, Aggregations, etc.

-Matplotlib: This library is responsible for plotting numerical data. And that’s why it is used in data analysis. It is also an open-source library and plots high-defined figures like pie charts, histograms, scatterplots, graphs, etc.

-Seaborn: is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics. For a brief introduction to the ideas behind the library, we can read the introductory notes or the paper.

3.2 STRUCTURES

- Import Packages
- Dataset Information
- Data Cleaning
- Analysis of various data
- Analysis of attack per year
- Analysis of number of people killed per year
- Analysis based on most affected area of the terrorism
- Analysis based on Terrorist Organization
- Analysis Based on county
- Visualization of the analysis using graphs
- Conclusion of the analysis

3.3 DATA CLEANING AND PREPARATION

Preprocessing is important into transitioning raw data into a more desirable format. Undergoing the preprocessing process can help with completeness and compellability. For instance, you'll see if certain values were recorded or not. Also, you'll see how trustable the info is. It could also help with finding how consistent the values are. We need preprocessing because most real-world data are dirty. Data can be noisy i.e. the data can contain

outliers or simply errors generally. Data can also be incomplete i.e. there can be some missing values.

The available data is raw and unusable for Exploratory data analysis, so before we do anything with the data we will have to explore and clean it to prepare it for data analysis.

- **Step1:** We write a function `info ()`, that will display 5 attributes about all the columns: Data type, Count of non-null values, Count of null values, number of unique values in that column and percentage of null value in that columns in the Global Terrorism dataset.
- **Step2:** We use `describe ()` function to get information about statistical inference from dataset
- **Step3:** we start with the checking the null values in the data so that we get rough overview how much missing values in the data. Then we think what can we do with this null values, there are two option in front of us. 1. Drop the null values and 2. Fill null values with 0 of that column.
- **Step 4:** We can see that the Killed and Wounded column has null values. We replace the null values with 0 of that corresponding column. and fill this value in place of null values using the `fillna()` function.
- **Step 5:** We created one column name casualties which is the sum of killed and wounded column for analysis.

4.EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis, or EDA, is an important step in any Data Analysis or Data Science project. EDA is the process of investigating the dataset to discover patterns, and anomalies (outliers), and form hypotheses based on our understanding of the dataset. EDA involves generating summary statistics for numerical data in the dataset and creating various graphical representations to understand the data better. In this article, we will understand EDA with the help of an example dataset. We will

use **Python** language (**Pandas** library) for this purpose.

4.1 ANALYSIS BASED ON NUMBER OF ATTACK PER YEAR.

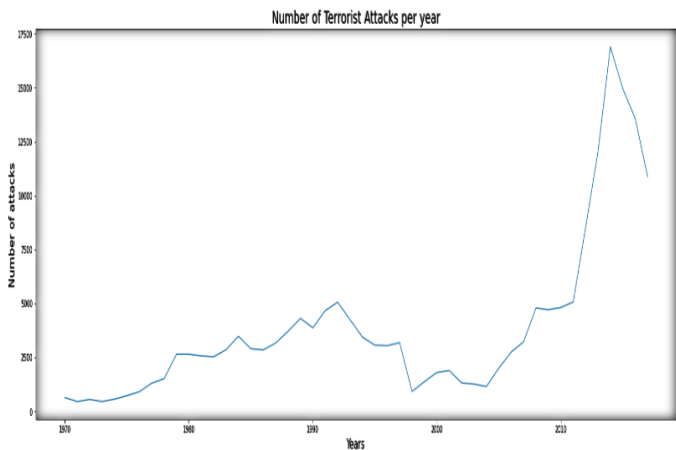


Fig.4.1:- Growth of Terrorism Per Year.

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 shows data from 1970 to 2017 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.

4.2 ANALYSIS BASED ON KILLED PER YEAR

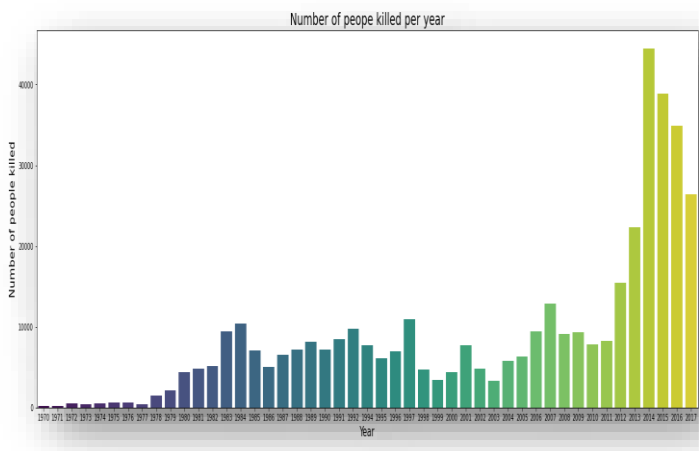


Fig -4.2: Number of People killed per year.

It is clearly seen that due to terrorist attacks number of people killed in the world is increases year by year

Observation:-

- 1.Due to increase in terrorism activity after 2004 ,number of people killed per year is also increases.
- 2.Most Number of People killed in year 2014.
- 3.Minimum Number of People killed in year 1971.

4.3 MOST AFFECTED AREA OF TERRORISM(REGIONS)

from the below scatter plot, we can say that terrorism is spread in almost every region of the world. there are some regions which is highly affected of terrorism while there are some regions in which terrorism is less. but the fact is terrorism is everywhere.

impacted by terrorism during the early 1980s to mid-1990s. Since then, terrorism has been relatively low.

Opposite of this is true for the Middle East and North Africa region. Middle East and North Africa has shown no noticeable rise in terrorism other than the early 2000s. This change in terrorism has been sudden and steep. But otherwise, there is no previous history of such high terrorist activities in the past.

From the scatter plot we get distribution of Terrorist attacks in all regions, but we can't figure out exact number of attacks in particular region.

So, to get clear idea of number of attacks per region we are using bar graph by doing exploratory data analysis on regions.

4.4 TOP 10 MOST AFFECTED COUNTRIES OF TERRORISM.

4.3.Analysis based on Most affected area of Terrorism

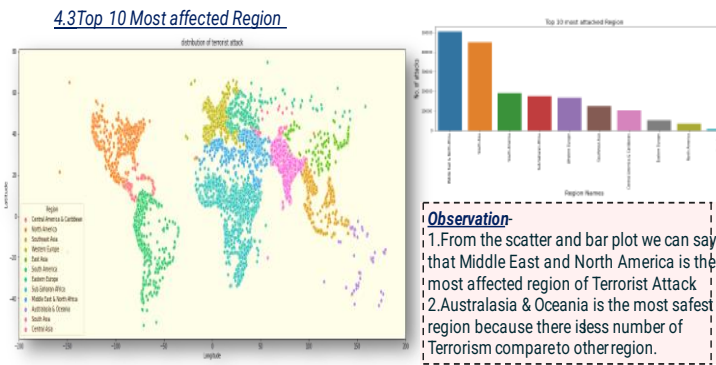


Fig -4.3 Most affected region of terrorism

Based on the geographic location of countries, they have been subcategorized into twelve regions to compare the rate of terrorism in each one of them as shown. Middle east and north Africa have the highest number of attacks followed by South Asia and South America. Terrorism here does not show an equal distribution among all regions. As a result, based on the number of attacks, different levels of attention are required for each individual region.

The South America has been the third-highest affected region in the world after North Africa & the Middle East and South Asia. But from graph, see South America has no significant contribution to the current trending terrorism. South America was

4.4 Top 10 Most affected Country

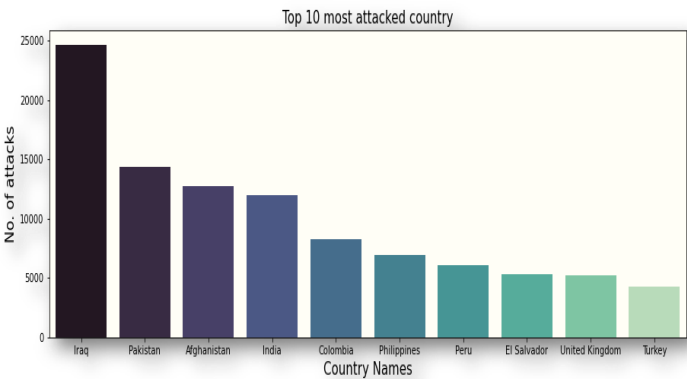


Fig.4.4:- Most affected countries of terrorism

There are some countries which mostly targeted by the terrorist organization.

- 1.Iraq is the most affected country of terrorism which records highest number of terrorism activity in any country.
- 2.India ranked number 4 in terms of most affected country of the terrorist activity

4.5 ANALYSIS BASED ON TERRORIST ORGANIZATION.

4.5.1 TOP 10 ACTIVE TERRORIST ORGANIZATION

Taliban and ISIL are the major contributors in the recent rise in attacks, especially in North Africa and the Middle East regions. Today there are more active terrorist groups spreading violence.

4.5Top 10 Active Terrorist Organizations who attack the most

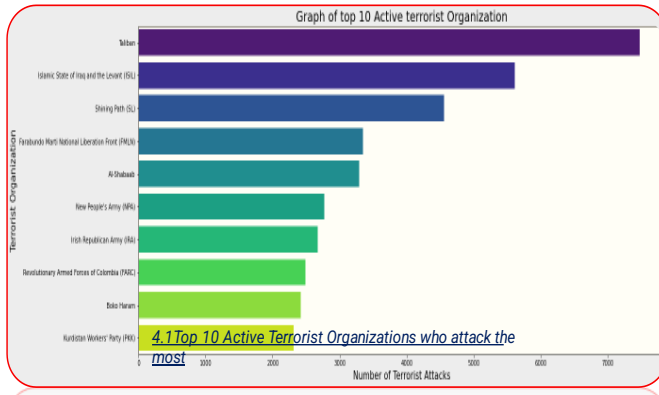


Fig-4.5.1:-Active terrorist organizations

Almost all groups show multiple crests and troughs during their existence.

4.5.2 :- TOP 10 DANGEROUS TERRORIST ORGANIZATION

To find out the most Dangerous Terrorist Organization in the world we need number of people killed per Terrorist Organization. If the number of people killed by the terrorist organization is more then we can say that it is most Dangerous organization. So, to get exact figure of number of people killed by the organization we use bar plot.

4.5.2 Top 10 Dangerous Terrorist Organization who killed the most number of people

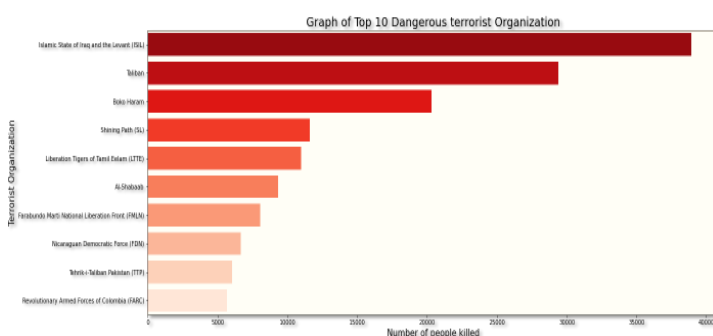


Fig-4.5.2 Dangerous terrorist organization

Observation: -

1.Islamic State of Iraq and the Levant (ISIL) and Taliban is the most dangerous Terrorist Organization in the world which records highest kills 38923.0 and 29410.0 respectively

2.Out of all the top 10 Organizations, around 46% people killed by ISIL and Taliban organization.

4.5.3 TOP 10 MOST ATTACKED TARGETS BY TERRORIST

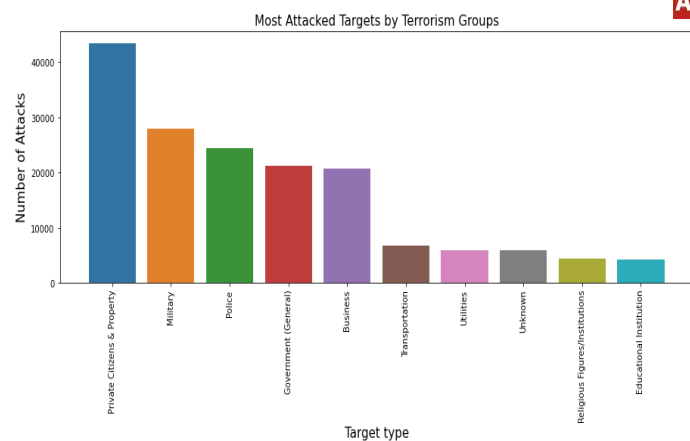


Fig-4.5.3 Top 10 most attacked targets

According to Figure, private citizens are only the modal target. Many of these attacks on private citizens are events in public spaces, such as marketplaces and sidewalk cafés.

The most common target after this is the Military, Police and government.

4.5.4 TOP 5 MOST USE WEAPON BY TERRORIST ORGANIZATION

To get the Percentage use of Weapon by Terrorist Organization we use Pie Chart.

Following pie chart shows exact percentage of most common used weapon by the terrorist organization during the attack.

There are various weapon used by organization but we plot top 5 weapon.

4.5.4 Top 5 Most use weapon by Terrorist Organization

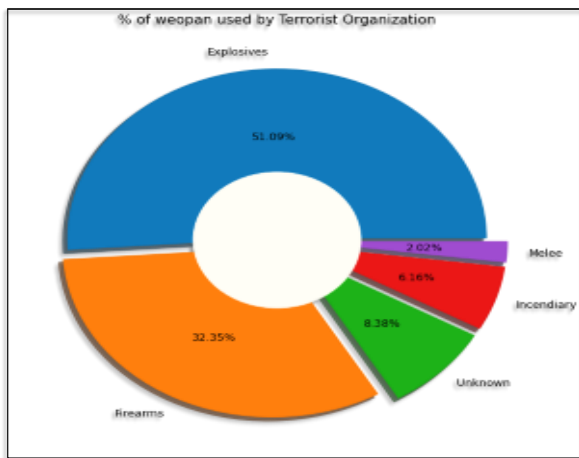


Fig-4.5.4 Most used weapon by terrorist

It seen that explosives were used in around 51.09 % of the attacks, followed by Firearms accounted for 32.35% of the attacks.

4.6. COUNTRY WISE ANALYSIS

4.6.1 COUNTRY IRAQ.

Country wise analysis

Country Iraq

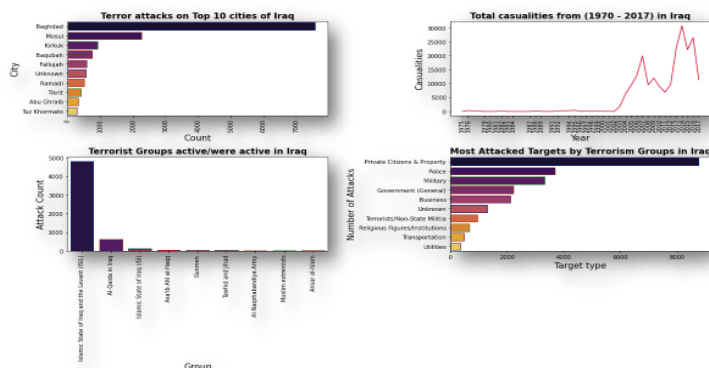


Fig.4.6.1 Analysis of country Iraq

Level of terrorism in Iraq were very low prior to the 2003 invasion, for the last 13 years it has consistently been the country most impacted by terrorism. Most of the attacks are near the Capital of Iraq. This tells us how much these groups hate Government Organizations.

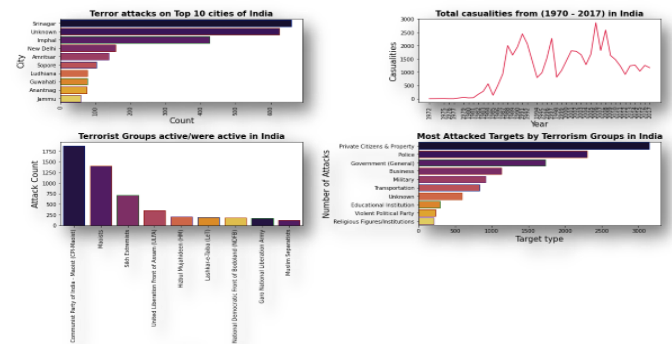
There have been two peak periods of terrorism in Iraq; the first occurred in 2007 and second commenced in 2011.

AI

ISIL is the deadliest terrorist group in Iraq's history and is responsible for over 18,000 deaths between 2013 and 2016.

4.6.1 COUNTRY INDIA.

Country India



It seems like terror Attacks have been increased in India since 1985 and the main group responsible for these is Communist Party of India- Maoist.

India has the lowest rate of deaths per attack among the ten countries most affected by terrorism.

There is many terrorist groups, but many are seeking political recognition and so their attacks are not aimed at killing people.

Conclusion :-

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

After undergoing these algorithms and process, we concluded some important points Through exploratory data analysis we have observed some trends and have made some assumptions.

To deal with those data was fun with many different analyses. Data cleaning and dealing with duplicate value was also most important to deal with correct methodology. Understand the dataset and predicting the solution of the problem was also great big task

After analyzing the Dataset, we got answers to some of the serious and interesting questions. which any of the Government securities would love to know.

Key conclusions: -

1. It was clear that numbers of terrorist attacks were increases from 2002-2004. Most of the attacks were done on year 2014 and Iraq is the most affected country from terrorism because most of the peoples

AI

killed in Iraq. In Iraq maximum of 1570 peoples killed in single attack. As expected, Baghdad is most affected city (and yes this is also called province).

2. Among all the regions "Middle East & North Africa" has the most number of killed people approx. 1.4 Lakhs followed by "South-Asia" & "Sub-Saharan Africa"

3. "Taliban" is the most powerful, dangerous and the most active gang among all of the gangs, followed by "Islamic state of Iraq and the levant (ISIL)" and "shining path(SL)".

4. The most targeted attacks are on "Private Citizens & Property" which is approximately 40% and 10-20% is the target on "Military", "Police", "Government", "Business".

5. "It seen that explosives were used in around 51.09% of the attacks, followed by Firearms accounted for 32.35% of the attacks".



Thank You