Capstone Project-1

Global Terrorism Dataset Analysis

Team Members

Potdar Vinayak

Pathak Sameer

Contents

Abstract

Violence and Terrorism have been a part of the society for a while now. The whole world is suffering from this curse. The motive for such negative activities may be political, emotional or even personal. Whatever the motive may be, the society will be suffering with the direct and indirect consequences.

Here, in this analysis we will try to understand the data that has been collected regarding terrorism in the world and try to get some insights. The questions of Who, What and Why can be extracted from this data.

Problem Statement

We are required to find out the key findings from the given dataset.

The key contributors to these attacks, along with the affected regions and people will be our focus. We will be looking at the trends, total share to the overall activities, consequences of attacks and fatality of these terrorist organizations.

Introduction

The world wide terrorist activities have been documented in a global and open source data base called Global Terrorism Database. The dataset consists of data from 1970 to 2017. This database is currently published by WHO and maintained by The National Consortium for the Study of Terrorism and Responses to Terrorism.

The data was maintained by multiple organizations viz.
*Pinkerton Global Intelligence Service (PGIS) 1970-1997
*Center for Terrorism and Intelligence Studies (CETIS) 1998-2008
*Institute for the Study of Violent Groups (ISVG) 2008-2011

Libraries Used

1. Pandas- We used this library for data wrangling and clean presentation.

2. Matplotlib- We used Matplotlib for data visualization.

3. Plotly.express- Features such as zooming in and out, panning, detailed values etc. can be achieved easily and in a more organized manner using this library.

4. Seaborn- It is built on top of matplotlib and helps in achieving more colorful and pleasing visualizations.

Approach

The data is huge, so we had to have a certain way to approach the analysis process. We first approached the world data as a whole and after getting a template to analyze, we used the same on countries level analysis.

We created two utility functions. The first would generate a comprehensive report for the desired country. Based on the first report the second function generates the terrorist organization's comprehensive report.

EDA

1. Library Imports- We imported some essential libraries like numpy, pandas, matplotlib,etc. We added some more libraries and droped some as per the experiment progressed.

2. Data Loading- We started off by loading the data into the google colab notebook from our google drive using pandas.

3. Understanding Data- We crudely analyzed the data using head(),tail(), describe(), info(). We renamed and extracted the features that we could work with and which made some sense in the first glance. We had about 20 features to work with now.

EDA

4. Null Value Analysis

```
# Measuring the amount of null values in the current working dataset.
data.isnull().sum()
```

```
eventid             0
Year                0
Month               0
Day                 0
Country             0
state             421
Region              0
city              434
latitude         4556
longitude        4557
AttackType          0
Killed          10313
Wounded         16311
Target            636
Summary         66129
Group               0
Target_type         0
Weapon_type         0
Motive         131130
success             0
dtype: int64
```

Here, Motive, killed, wounded, summary had most of the missing values.

EDA

5. Null Value Treatment-

-The major chunk of missing motive can be handled by replacing it as 'Unknown Motive'. Dropping this column won't make sense as we can still extract the motive where it is known.

-Summary column can be dropped, as it is not contributing much to the analysis process even where it is present.

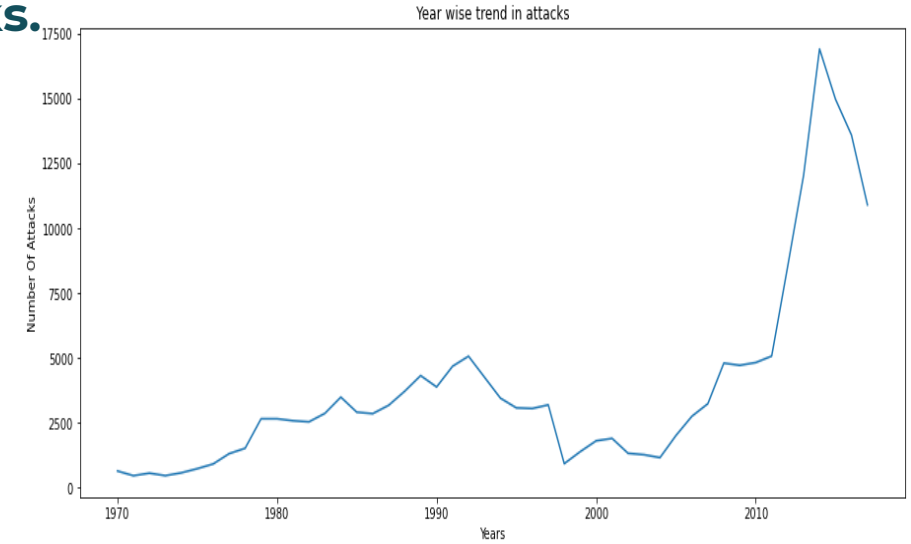
-Killed and Wounded can be left untouched for now. The missing amount is also less than 10% of the column data.

- Missing cities can be replaced by ‘Unknown’ and excluded while analysis.

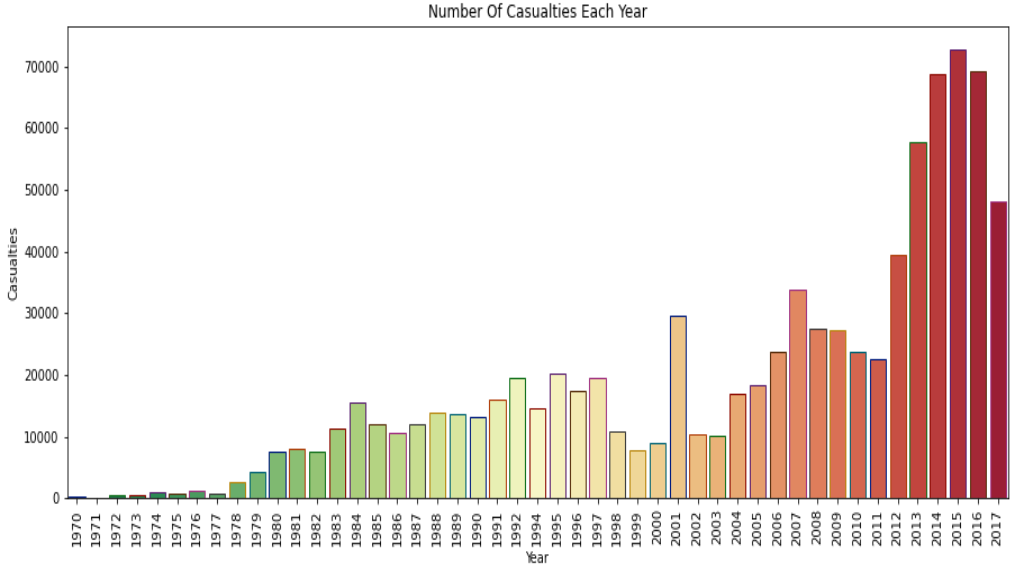
EDA

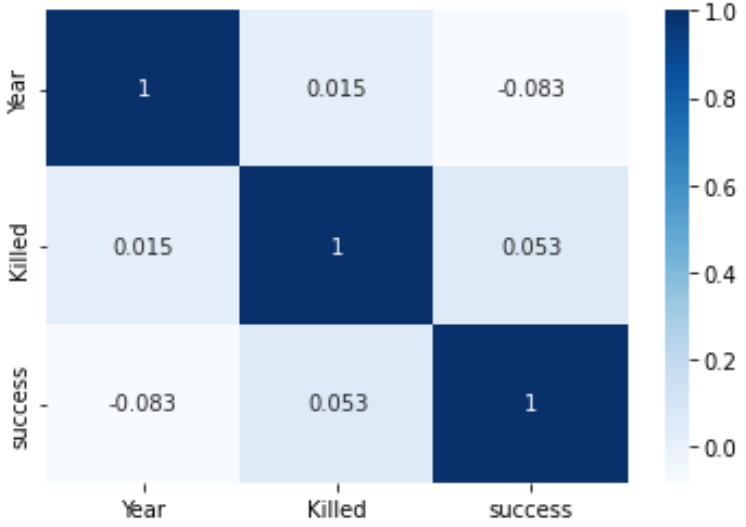
6. Yearly Trend of Attacks over the world- Massive Upward Trend in attacks.

Financial Crisis of 2007-08 (possible reason)

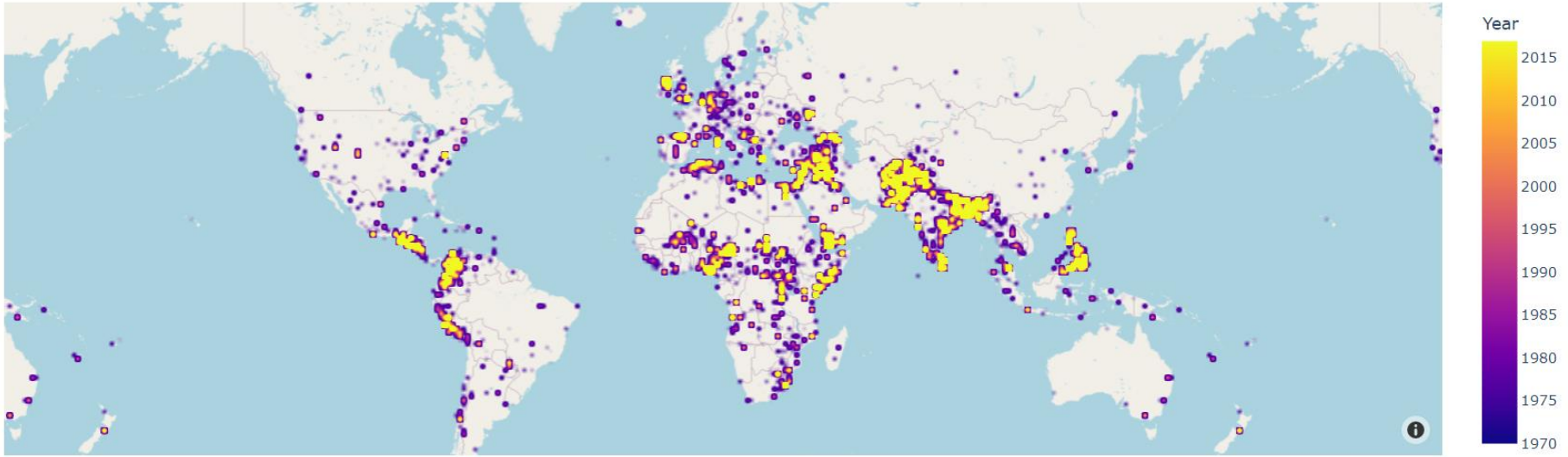


7. Yearly casualties and correlation-

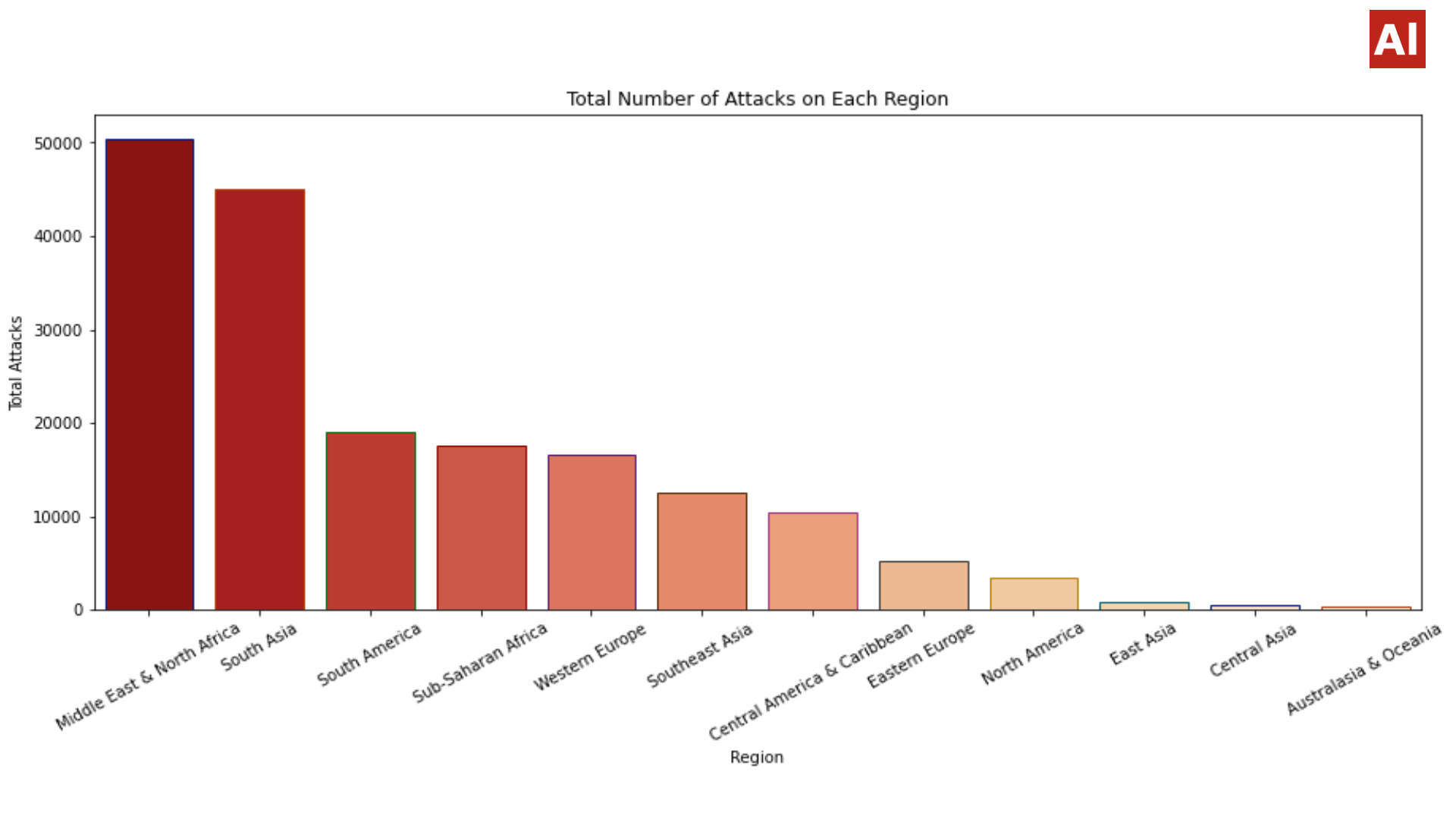




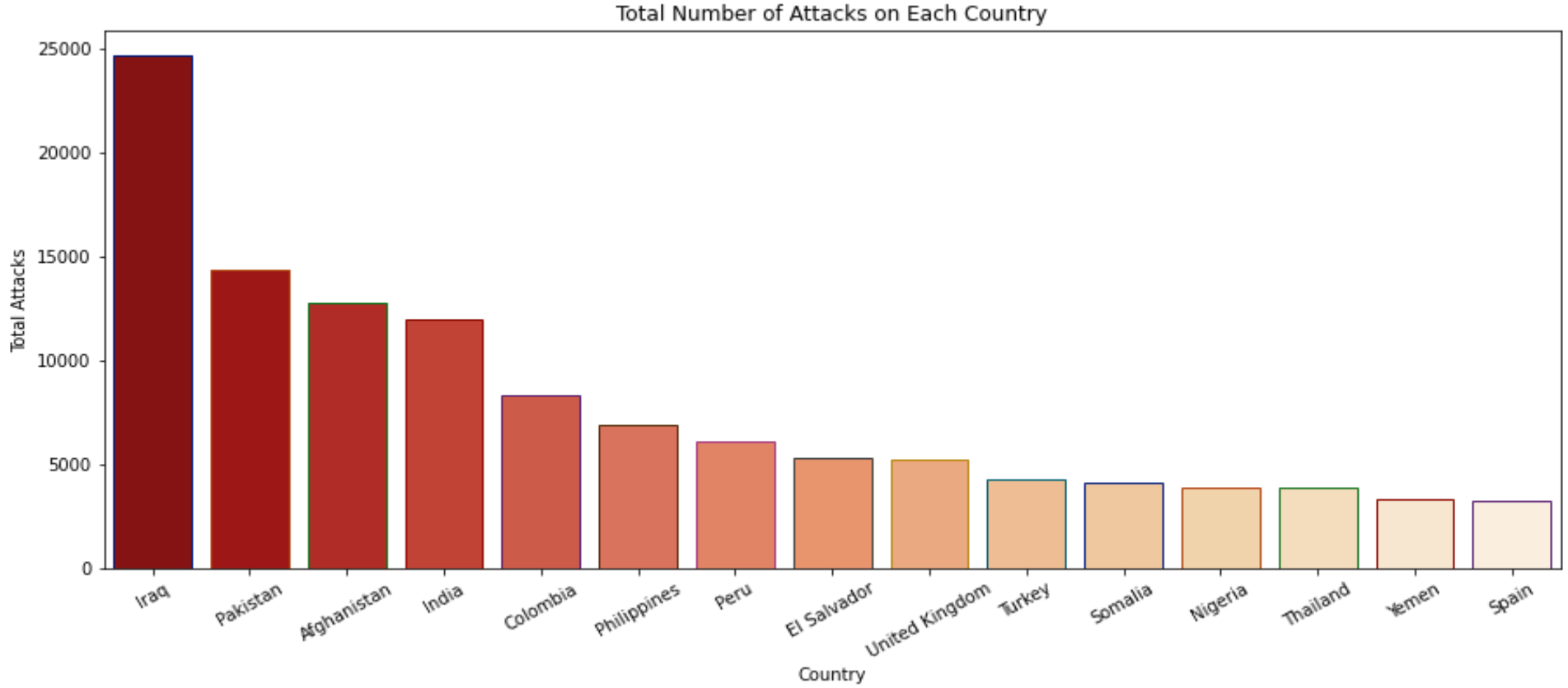
8. World Wide Attacks Heat Map-



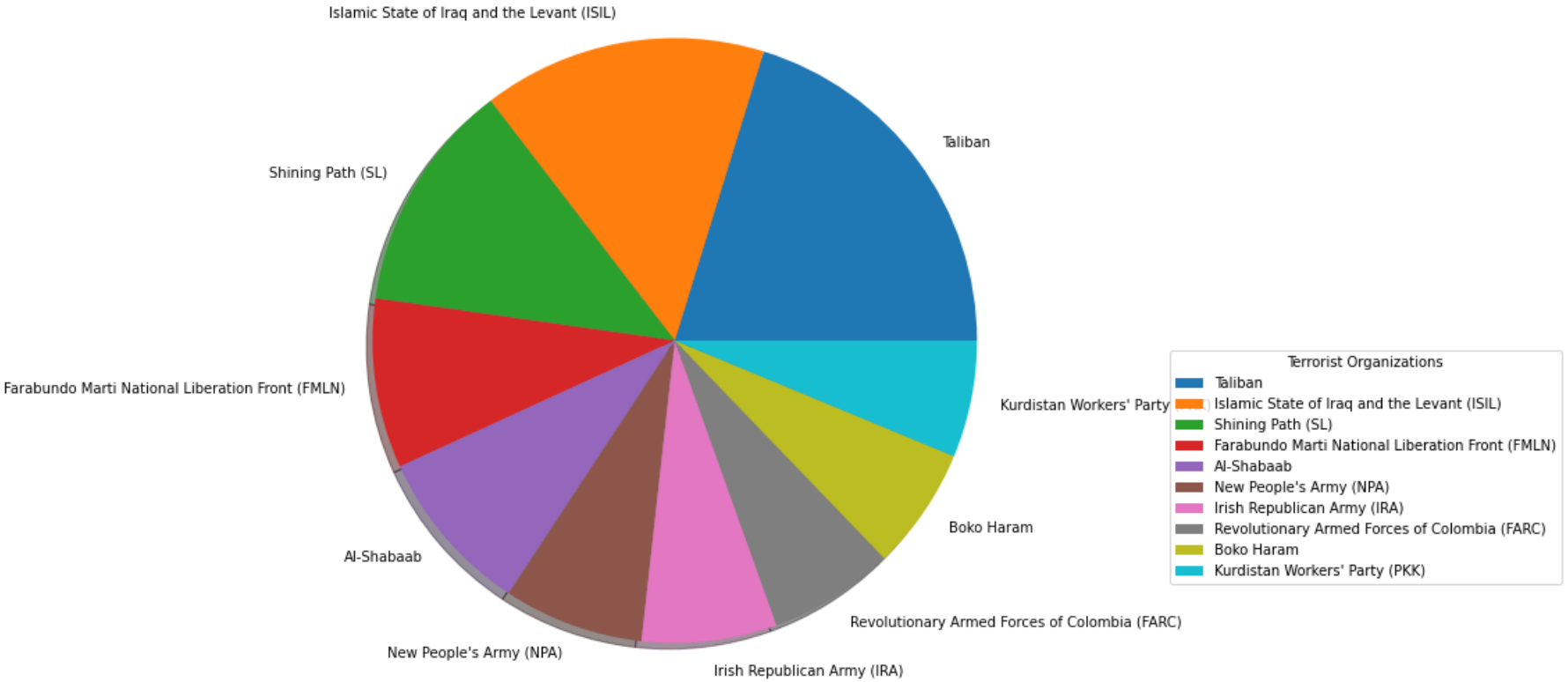
Concentration in the Middle East and South Asian Region.


Total Number of Attacks on Each Region
Total Attacks
50000
40000
30000
20000
10000
0
Middle East & North Africa
South Asia
South America
Sub-Saharan Africa
Western Europe
Southeast Asia
Central America & Caribbean
Eastern Europe
North America
East Asia
Central Asia
Australasia & Oceania
Region

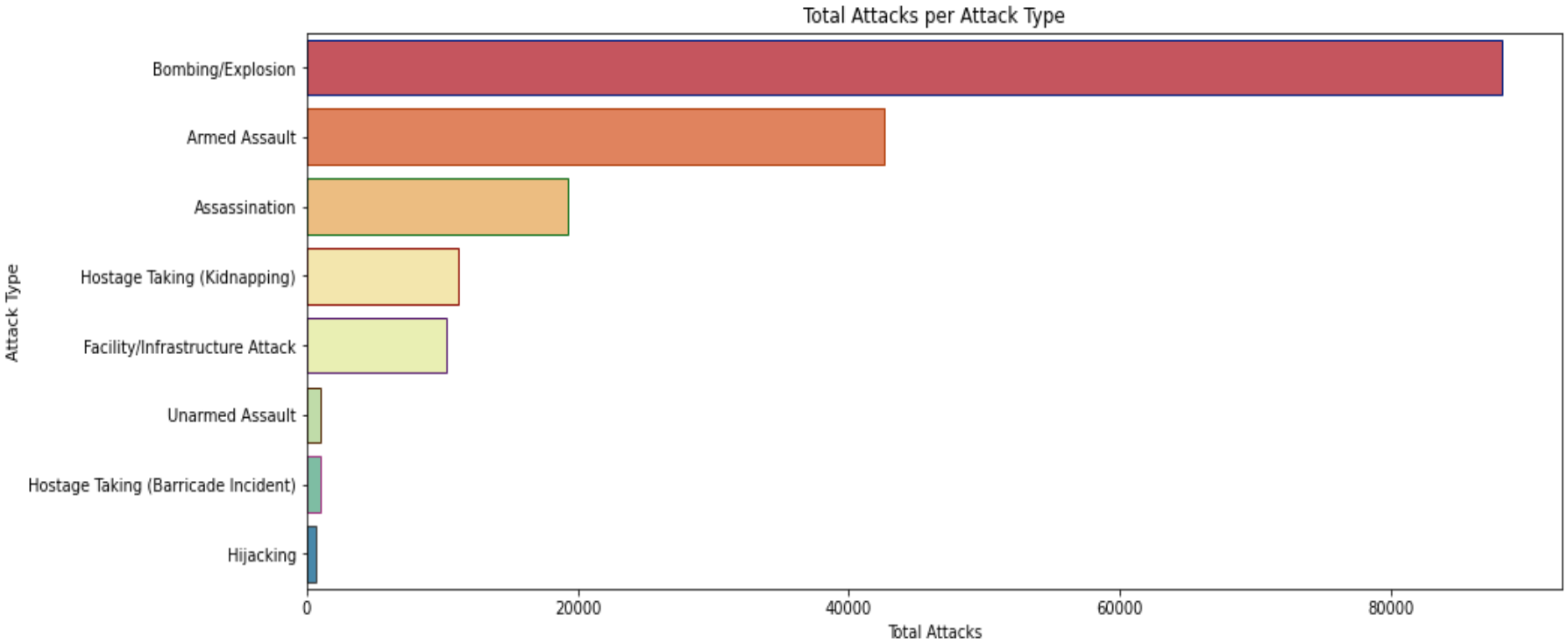
9. Most Affected Countries-



10. Lead Terrorist Organizations-



11. Preferred Mode of Attack-



Utility Functions

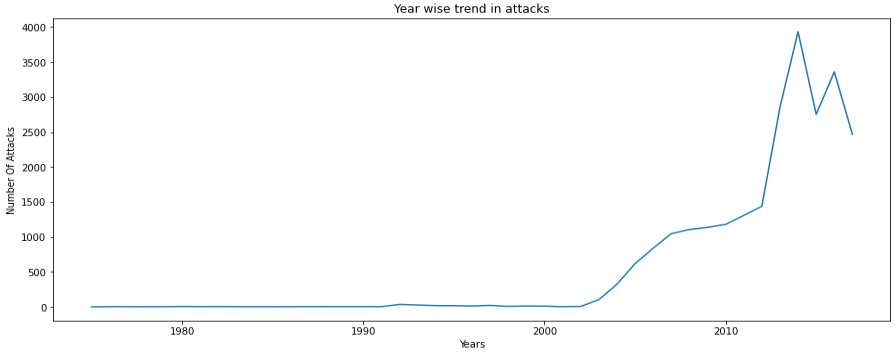
We can now do the similar analysis on each country and its terrorist groups. We achieved this by writing two utility functions which take country and terrorist organization as the arguments and generate a complete report.

1. get-country-report(“country name”)

2. get-terrorist-report(“terrorist organization name”)

Here, we analyzed two of the most attacked countries i.e. Iraq and Pakistan.

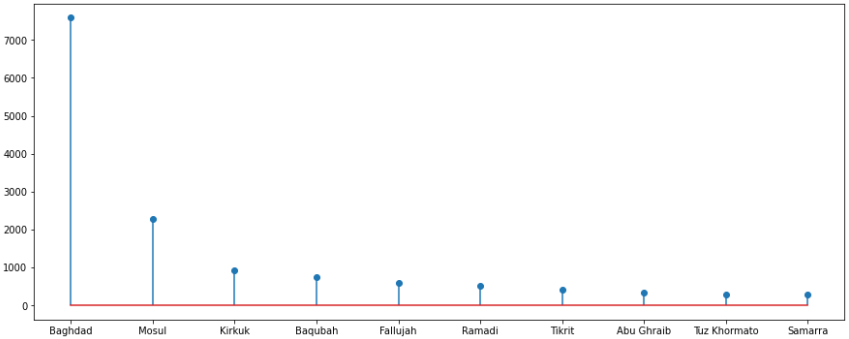
IRAQ -Yearly Attack Trend



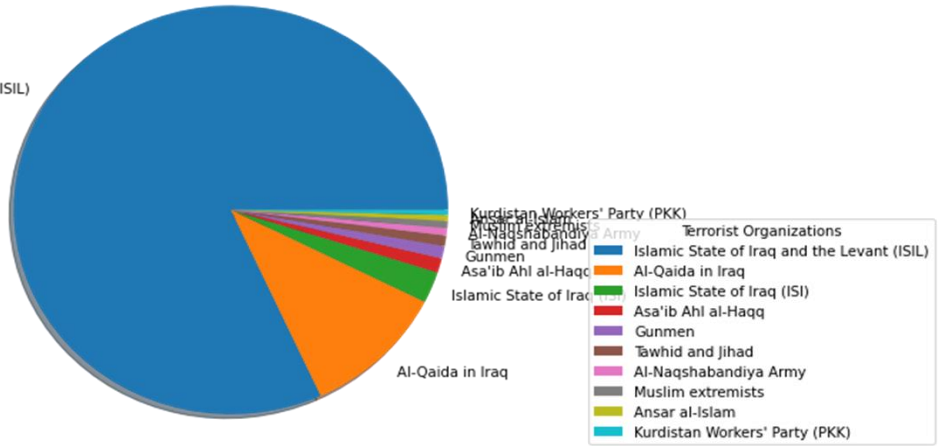
Increase in attacks post 2000.
Sudden spike post 2010.

Attacks concentrating near the middle part of the country.

IRAQ-Most Attacked Cities

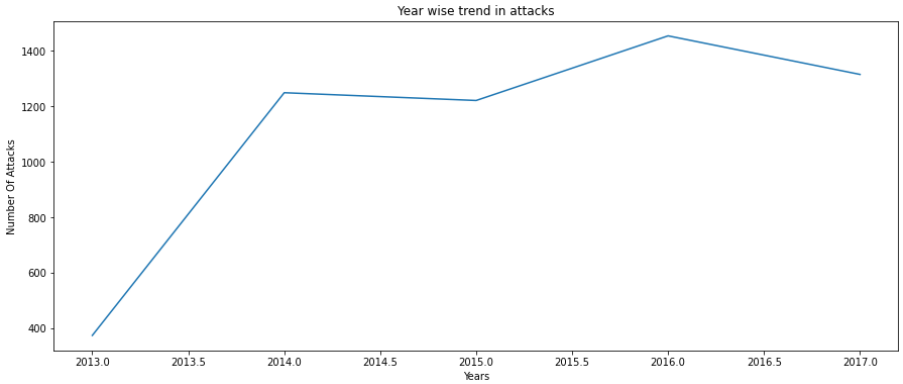


Baghdad, Mosul and Kirkuk. Cities with Financial and Political Importance.



Islamic State of Iraq and the Levant (ISIL) contributing to more than 75% of attacks.

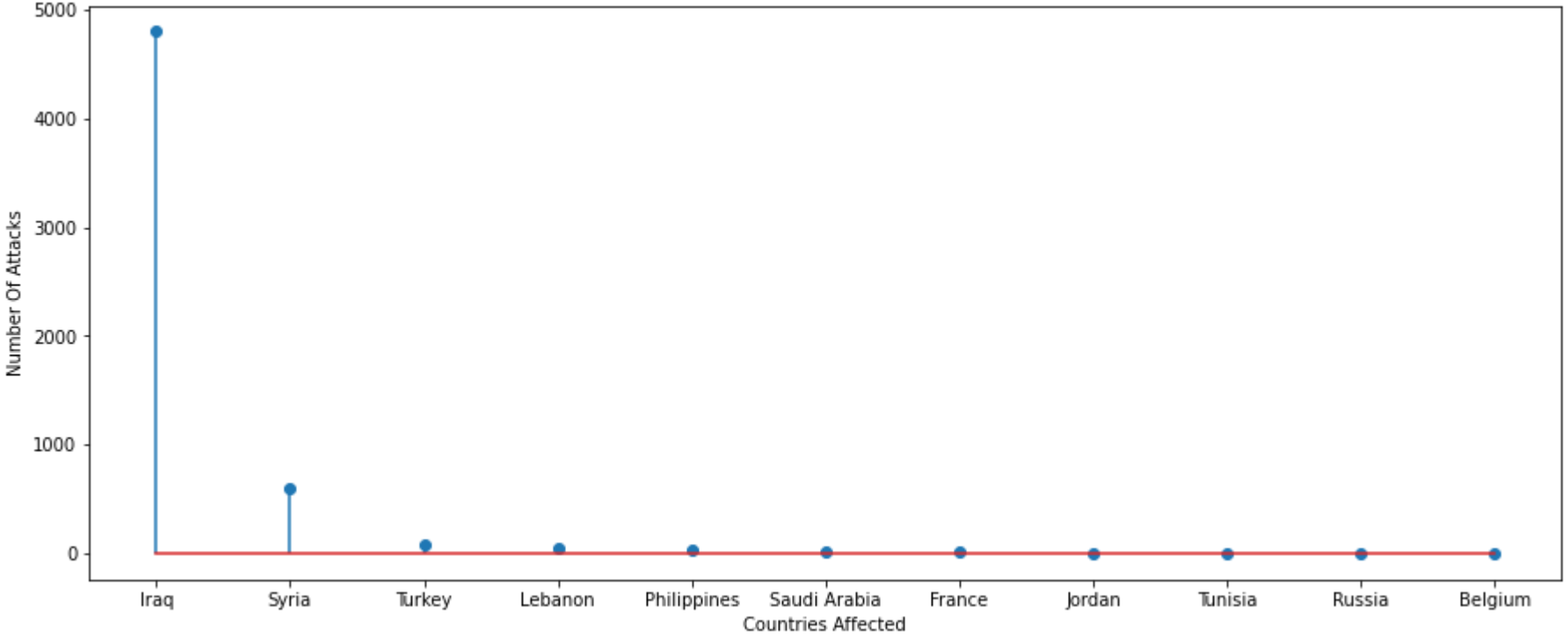
IRAQ- ISIL Attack Trend

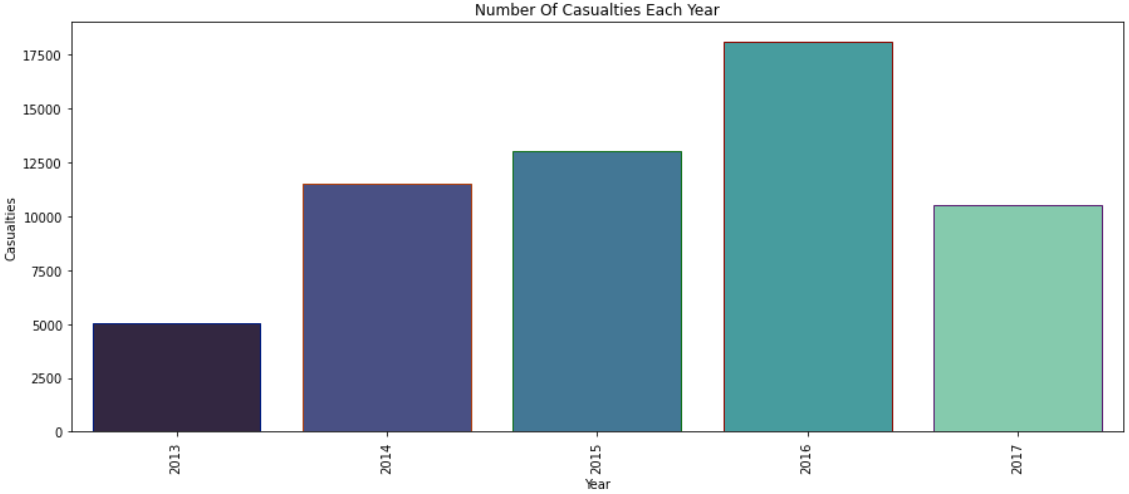


Started in 2013 and picked up pace very fast.

Early attacks can be seen near the borders of the country.

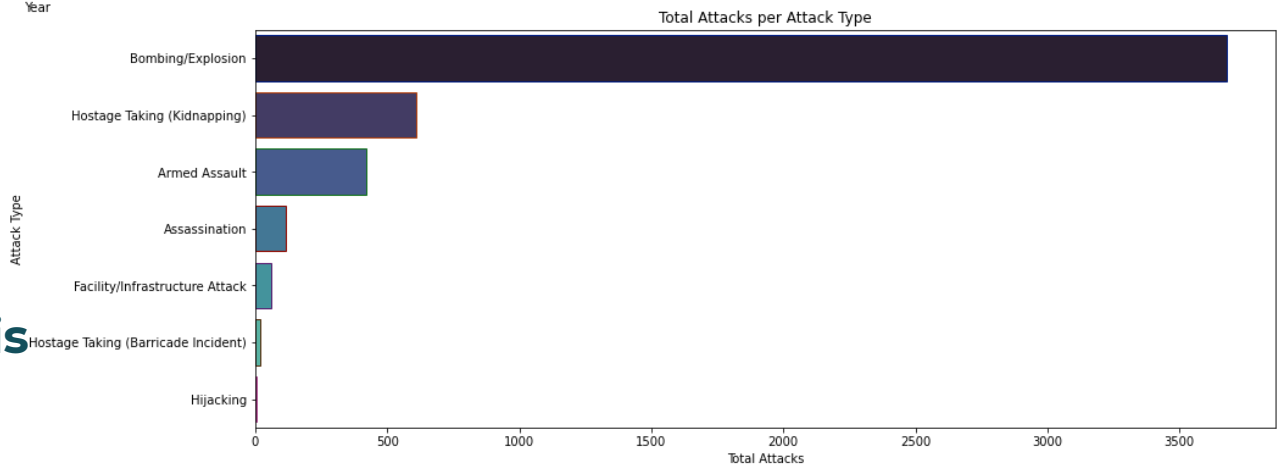
ISIL- Countries Affected by ISIL



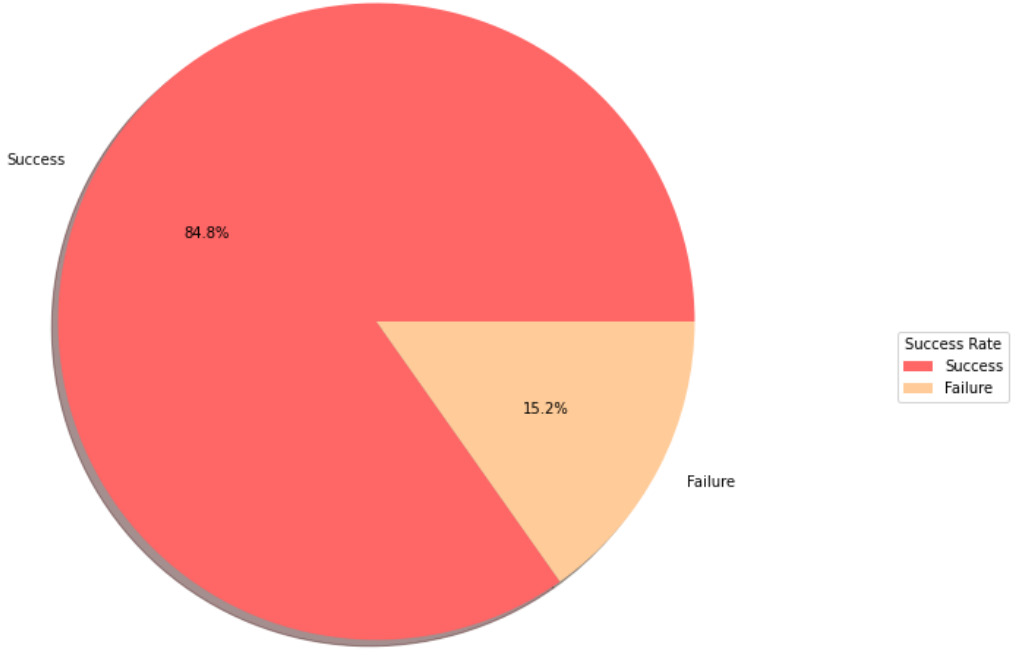


- -Casualties Per Year
- -Decreasing Trend
- -Highest in 2016

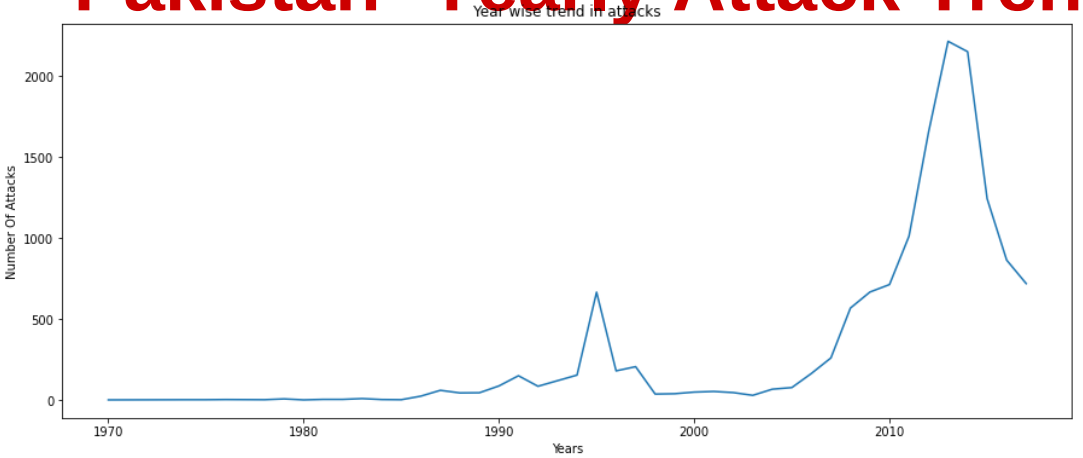
Preferred mode of attack
-Bombing
-Hostage Taking
unlike world analysis

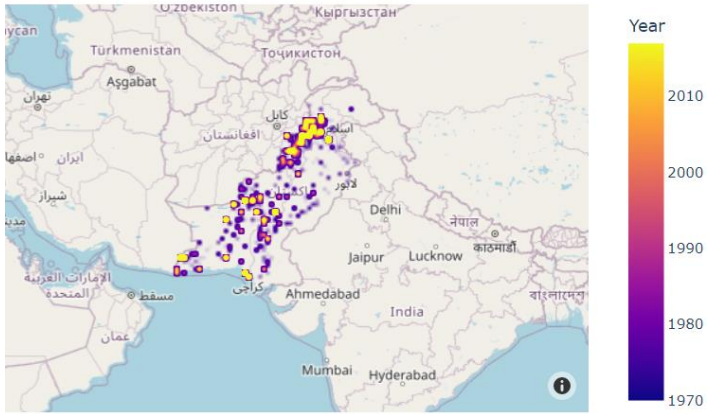


The success rate is 84.8%
-Fairly High
-15.2% attacks stopped or controlled early.



Pakistan- Yearly Attack Trend

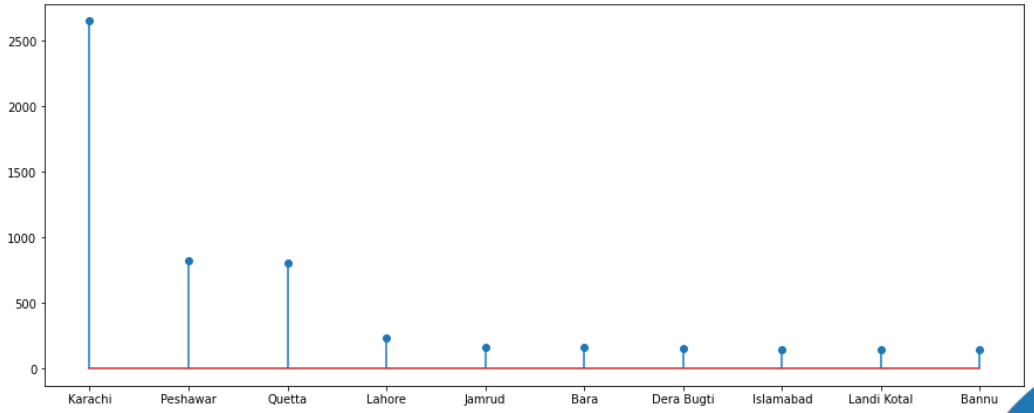




Two spikes in the attacks were observed, one between 1990-2000 and the other post 2010.

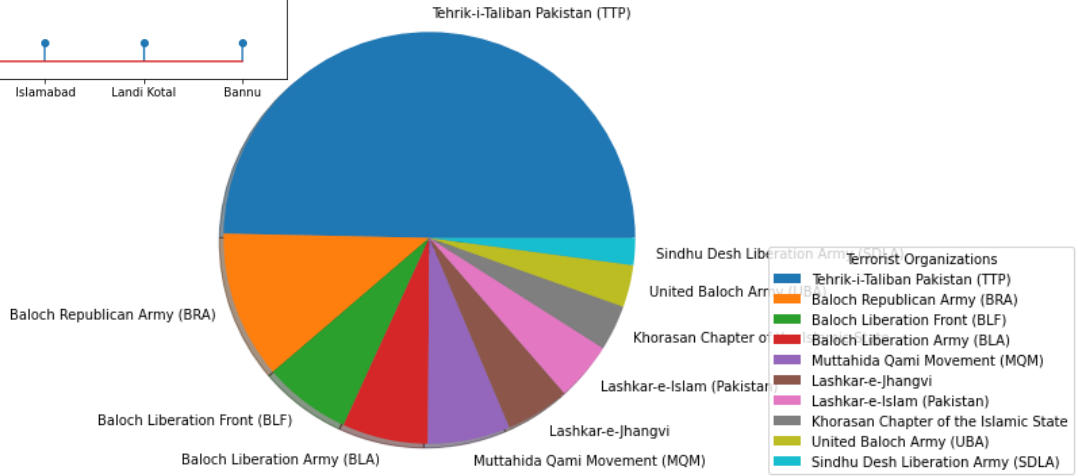
The attacks are concentrated towards the major cities of the country.

Pakistan- Most Attacked Cities

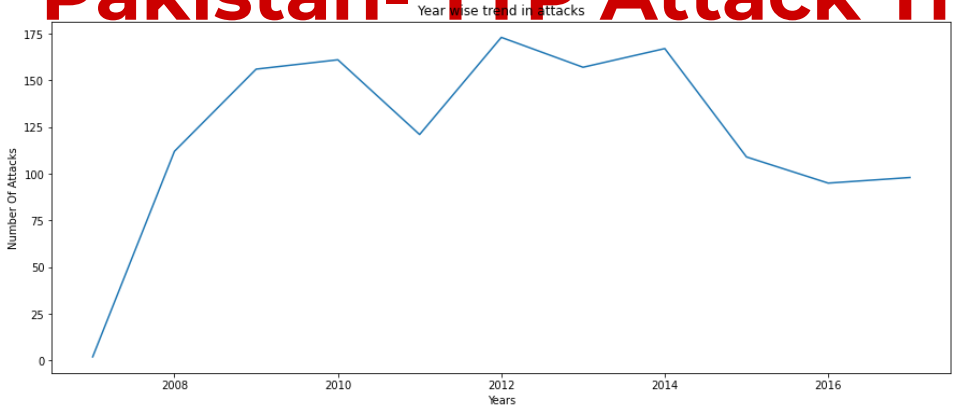


Karachi, Peshawar, Quetta are the most attacked cities, holding industrial importance.

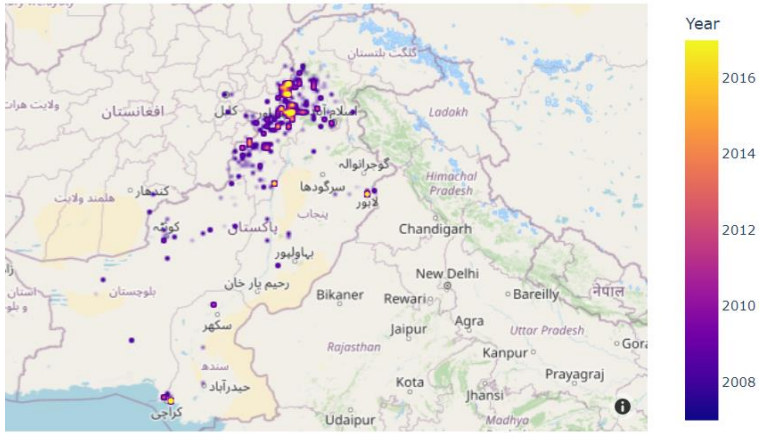
Tehrik-i-Taliban Pakistan is the major terrorist organization responsible for about 50% attacks.



Pakistan- TTP Attack Trend

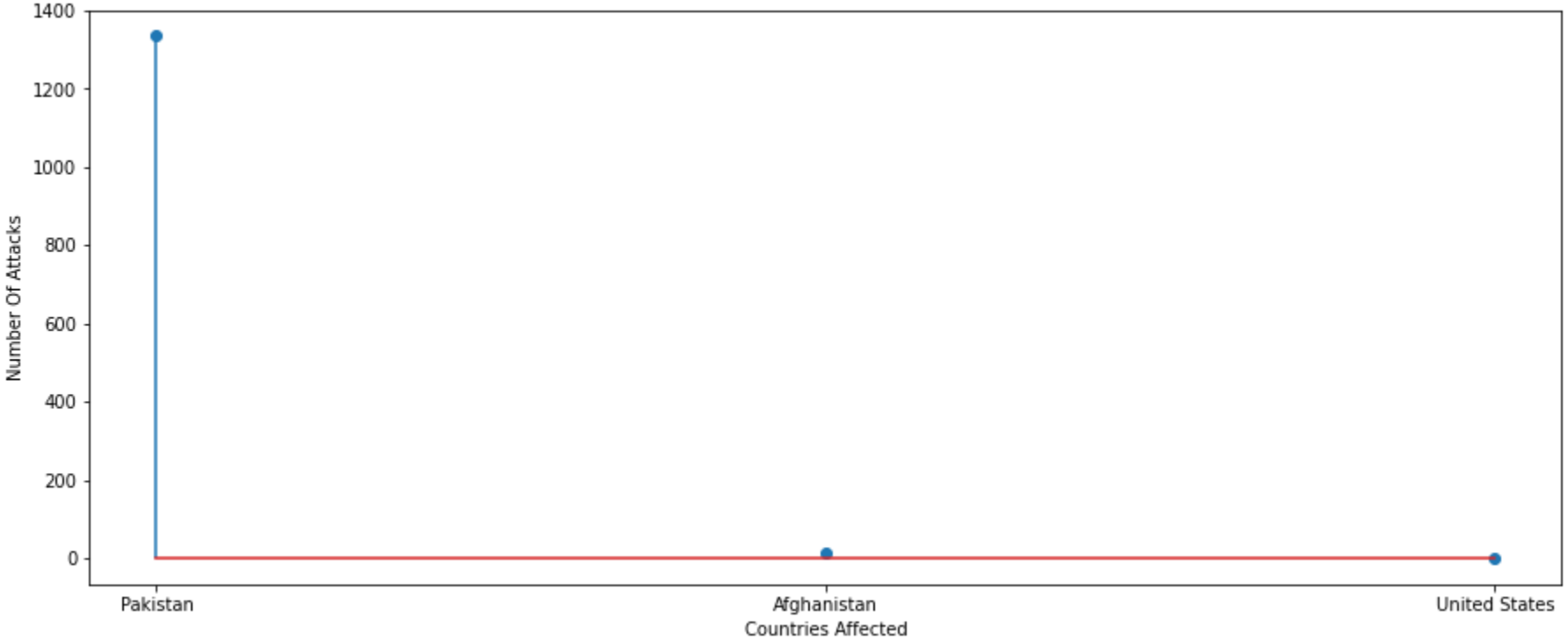


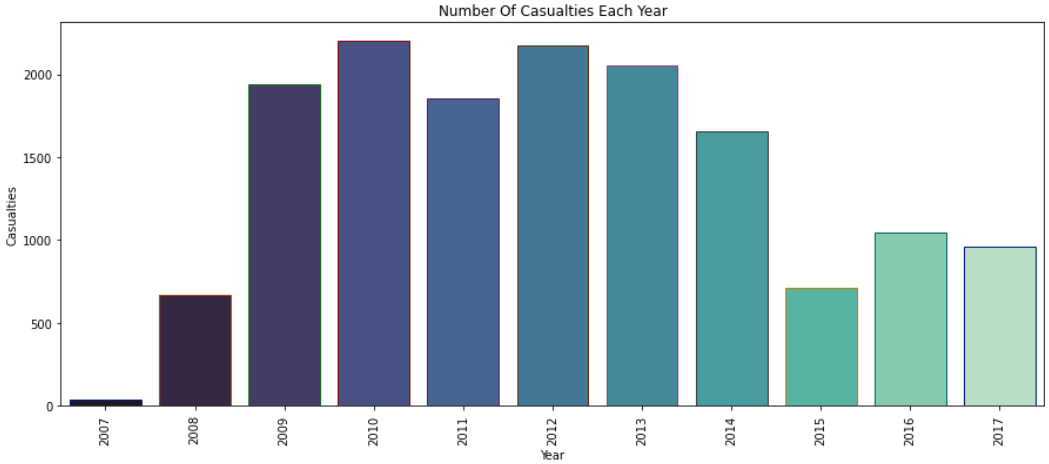
First activity in the year 2007.
Rose to 100+ attacks in a year.



Attacked mostly the Northern region of the country with some occasional attacks on the major port in the South.

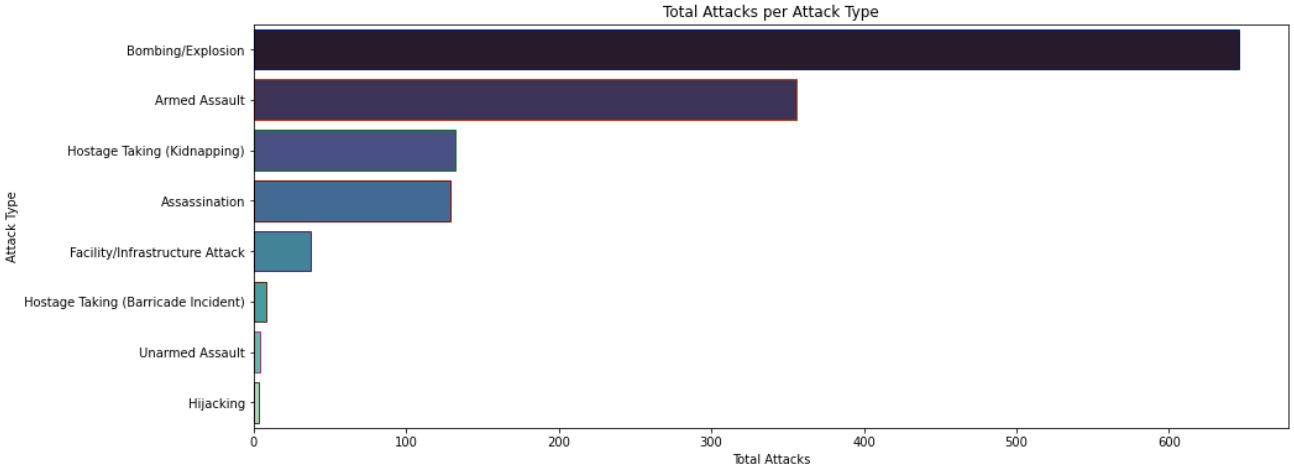
TTP- Countries Affected by TTP



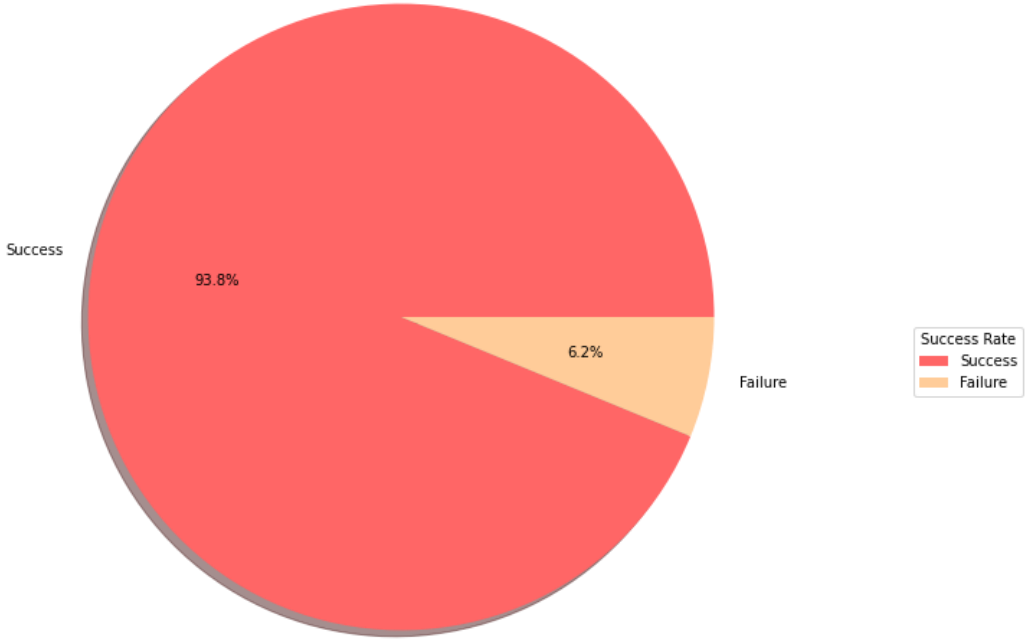


- -Casualties Per Year
- -Decreasing Trend
- -Highest in 2010

Preferred mode of attack
-Bombing
-Armed Assaults like World Analysis



The success rate is 93.8
-Very High
-6.2% attacks stopped or controlled early.



Conclusion and Key Findings

- High growth in attacks in the world after 2007 observed.
- The current scenario shows a decline in attacks.
- Most attacked region is Middle East & North Africa
- Most attacked Countries are Iraq and Pakistan.
- Taliban and ISIL are the most active and damaging terrorist groups.
- Bombing and Armed assaults are the preferred modes of attacks world wide.
- Attacks in Iraq rose after 2007 and peaked in the mid 2010-2017.
- Most attacked cities are Baghdad, Mosul and Kirkuk.
- ISIL is the most active terrorist group here, responsible for more than 75% attacks in the country.
- Attacks in Pakistan had two peaks, one in 1990-2000 and other in 2010-2017.
- Most attacked cities are Karachi, Peshawar, Quetta.
- TTP is the major terrorist group here.

Conclusion and Key Findings

- ISIL in IRAQ.
- Started in 2013 and did more than 2200 attacks in a year.
- Active in countries like Iraq, Syria, Turkey, Russia etc.
- Maximum casualties in 2016 of more than 17500.
- Bombing and Explosion is the most preferred mode of attack followed by Hostage Taking.
- Success rate of 84.8%

- TTP in PAKISTAN.
- Emerged in 2007 and managed to do more than 100 attacks in a year.
- Active in Pakistan, Afghanistan and The United States.
- Maximum casualties in 2010 of about 2500.
- Bombing and Explosion is the most preferred mode of attack followed by Armed Assault.
- Success rate of 93.8%