

**Name:** Muhamed Ahmed Abdelnaby

Data Engineer

## Table Of Content

<b>INTRODUCTION.....</b>	<b>3</b>
<b>DATA PREPROCESSING.....</b>	<b>3</b>
<b>DESCRIPTIVE STATISTICS: .....</b>	<b>4</b>
<b>DATA VISUALIZATION.....</b>	<b>5</b>
<b>PERFORMANCE COMPARISON.....</b>	<b>8</b>

## ■ Introduction

Terrorism remains a significant global concern with far-reaching consequences for societies and nations worldwide. Understanding the patterns, trends, and dynamics of terrorist incidents is crucial for effective counter-terrorism measures and policy-making. In this report, we analyze a comprehensive dataset on global terrorism to derive insights and provide a detailed understanding of the phenomenon.

Analyze a dataset on global terrorism and provide insights into various aspects of terrorist incidents. The dataset used is the "globalterrorismdb\_0718dist.csv" file, which contains information about terrorist attacks worldwide. The analysis includes data preprocessing, descriptive statistics, trend analysis, data visualization, and performance comparison between Pandas and Dask.

## ■ Data Preprocessing

**Renaming columns:** The column names were updated to more descriptive names for better understanding.

**Selecting relevant columns:** Only the important columns related to the analysis were retained in the Data Frame.

**Handling missing values:** Missing values in the 'Killed', 'Wounded', 'state', 'city', and 'Motive' columns were filled with appropriate values.

**Checking for null values:** The Data Frame was checked for null values, and it was found that there were no null values remaining.

## ■ Descriptive Statistics:

The statistics included the number of attacks, killed, and wounded per year. Additionally, the number of attacks, killed, and wounded per region was also calculated. The descriptive statistics provided insights into the overall trends and distribution of terrorist incidents.

Attacks per Year: Year	Killed per Year: Year	Wounded per Year: Year
1970 651	1970 174.0	1970 212.0
1971 471	1971 173.0	1971 82.0
1972 568	1972 566.0	1972 409.0
1973 473	1973 370.0	1973 495.0
1974 581	1974 539.0	1974 865.0
1975 740	1975 617.0	1975 617.0
1976 923	1976 674.0	1976 756.0
1977 1319	1977 456.0	1977 518.0
1978 1526	1978 1459.0	1978 1600.0
1979 2662	1979 2100.0	1979 2506.0
1980 2662	1980 4400.0	1980 3645.0
1981 2586	1981 4851.0	1981 3337.0
1982 2544	1982 5136.0	1982 3342.0
1983 2870	1983 9444.0	1983 4047.0
1984 3495	1984 10450.0	1984 5291.0
1985 2915	1985 7094.0	1985 5130.0
1986 2860	1986 4976.0	1986 5814.0
1987 3183	1987 6482.0	1987 5775.0
1988 3721	1988 7208.0	1988 6960.0
1989 4324	1989 8152.0	1989 5539.0
1990 3887	1990 7148.0	1990 6128.0
1991 4683	1991 8429.0	1991 7591.0
1992 5071	1992 9742.0	1992 9915.0
1994 3456	1994 7690.0	1994 7569.0
1995 3081	1995 6103.0	1995 14292.0
1996 3058	1996 6966.0	1996 10795.0
1997 3197	1997 10924.0	1997 9072.0
1998 934	1998 4688.0	1998 8183.0
1999 1395	1999 3393.0	1999 5341.0
2000 1814	2000 4403.0	2000 5869.0
2001 1906	2001 7729.0	2001 22774.0
2002 1333	2002 4805.0	2002 7162.0
2003 1278	2003 3317.0	2003 7384.0
2004 1166	2004 5743.0	2004 12022.0
2005 2017	2005 6331.0	2005 12784.0
2006 2758	2006 9380.0	2006 15550.0
2007 3242	2007 12824.0	2007 22524.0
2008 4805	2008 9157.0	2008 18998.0
2009 4721	2009 9273.0	2009 19138.0
2010 4826	2010 7827.0	2010 15947.0
2011 5076	2011 8246.0	2011 14659.0
2012 8522	2012 15497.0	2012 25445.0
2013 12036	2013 22273.0	2013 37688.0
2014 16903	2014 44490.0	2014 41128.0
2015 14965	2015 38853.0	2015 44043.0
2016 13587	2016 34871.0	2016 40001.0
2017 10900	2017 26445.0	2017 24927.0

Country with the most attacks: Iraq

City with the most attacks: Baghdad

Region with the most attacks: Middle East & North Africa

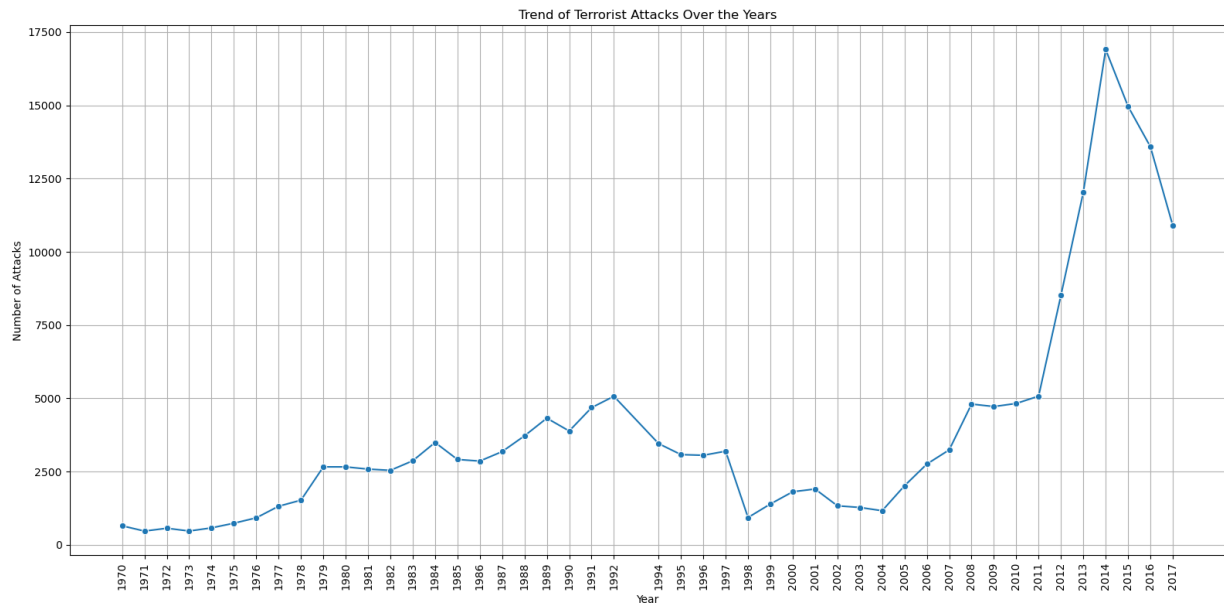
Year with the most attacks: 2014

Month with the most attacks: 5

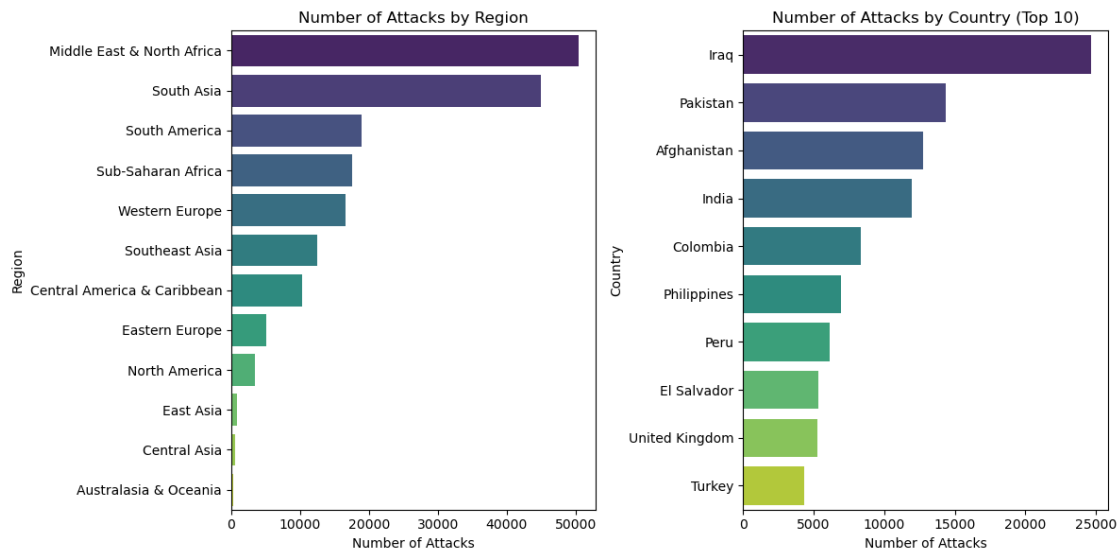
Group with the most attacks: Taliban

Most Attack Types: Bombing/Explosion

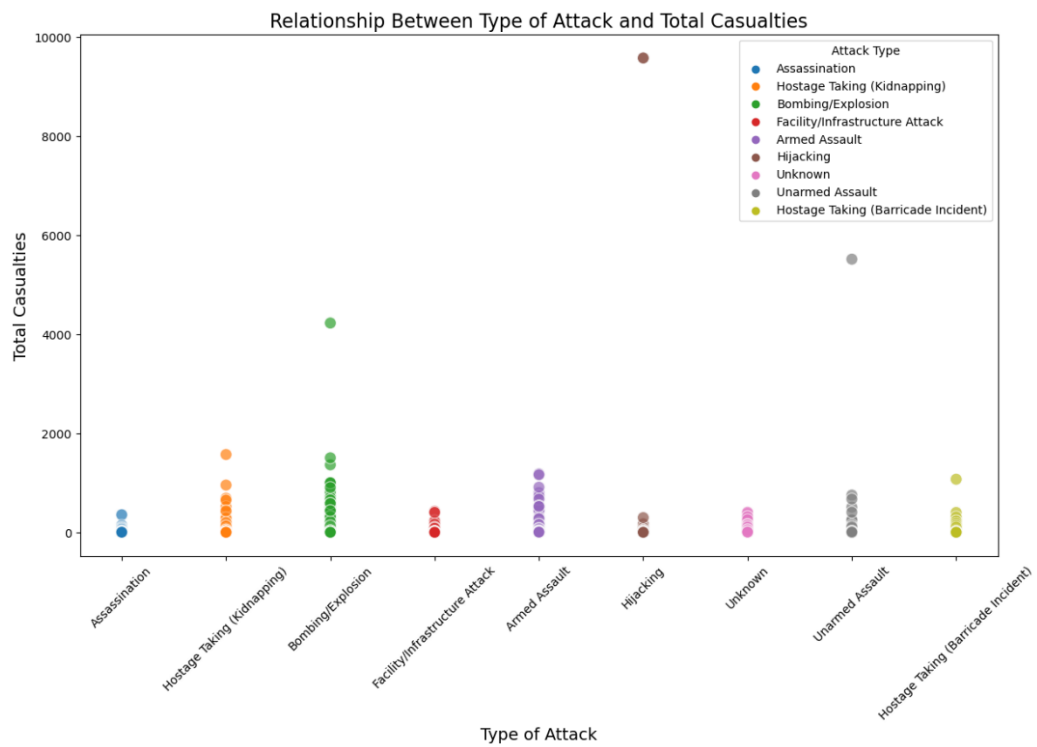
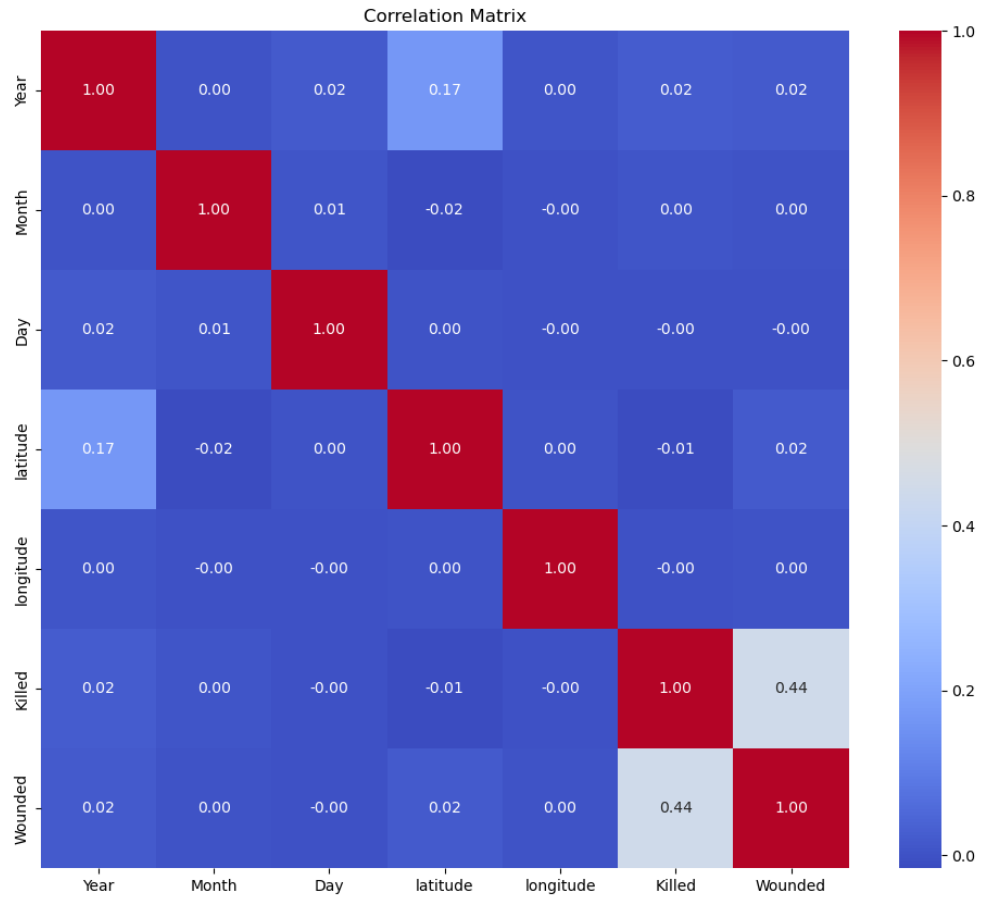
## ■ Data Visualization

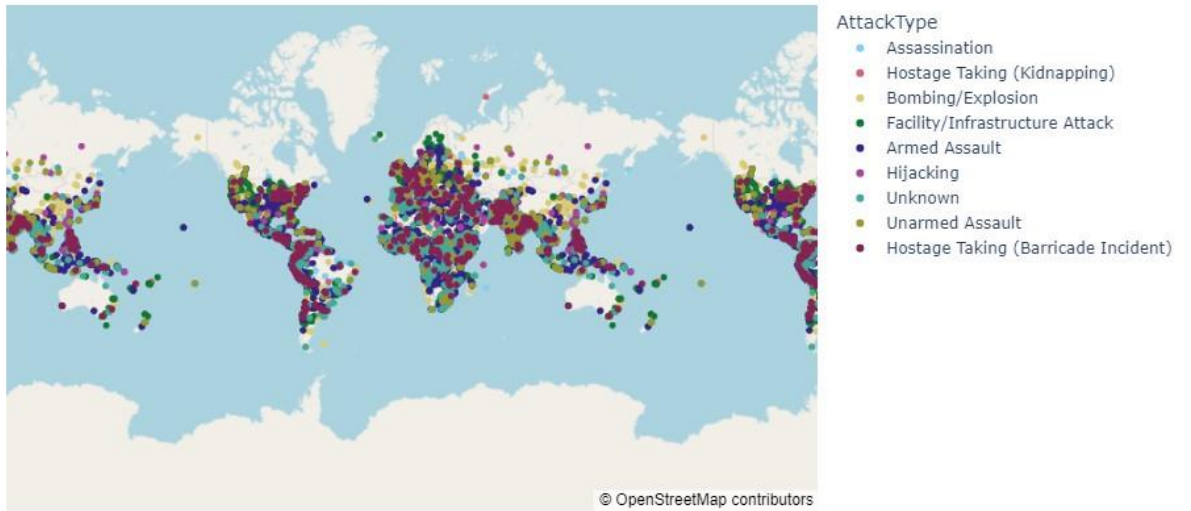


line graph showing the trend of terrorist attacks over the years.



Bar chart comparing the number of attacks by region and country.





*1Interactive map to show the geographic distribution of attacks*

## ■ Performance Comparison

### Pandas vs. Dask

Compared the performance of loading and aggregating data using Pandas and Dask. Below are the observed load times and aggregation times for each library:

#### – Pandas

Load Time: 8.91 seconds

Pandas is a powerful data manipulation library well-suited for small to moderately large datasets. However, the load time can be significantly longer when dealing with large datasets. The aggregation time is very efficient once the data is loaded into memory.

#### – Dask

Load Time: 0.20 seconds

Dask is designed for parallel computing and can handle larger-than-memory datasets efficiently. The load time is substantially lower compared to Pandas due to Dask's ability to distribute the load across multiple processors. Although the aggregation time is slightly higher than Pandas, Dask remains advantageous for large-scale data processing.