Comprehensive Report on Global Terrorism Analysis

Introduction

This report presents a comprehensive analysis of global terrorism incidents using the dataset provided. The analysis includes data cleaning, exploratory data analysis, and various visualizations to uncover trends, patterns, and insights.

Data Loading and Performance Comparison

The dataset was loaded using both Pandas and Dask to compare performance:

- **Pandas Load Time:** 6.23 seconds
- **Dask Load Time:** 0.12 seconds

Dask proved to be faster for handling large datasets efficiently.

Data Cleaning

1. **Missing Values**: Several columns had a high percentage of missing values. Columns with more than 50% missing values were dropped. The remaining missing values were imputed using appropriate methods (e.g., mode for categorical columns, median for numeric columns).

Key Statistics

Numeric Columns

- **Mean ,Median and Standard Deviation Values**: Provided for each numeric column.

Categorical Columns

- **Most Frequent Values**: Identified for each categorical column.

Yearly Trends

- **Number of Attacks per Year**: Visualized using a line plot, showing a clear trend over time.
- **Yearly Statistics**: Number of attacks, total killed, and total wounded were calculated for each year. **(High number of attacks was 16903 in 2014).**

Regional Analysis

- **Most Affected Regions**: Middle East & North Africa, South Asia, and Sub-Saharan Africa were the most affected regions.

Country Analysis

- **Most Affected Countries**: Iraq, Pakistan, and Afghanistan were the most affected countries.

Attack Type Analysis

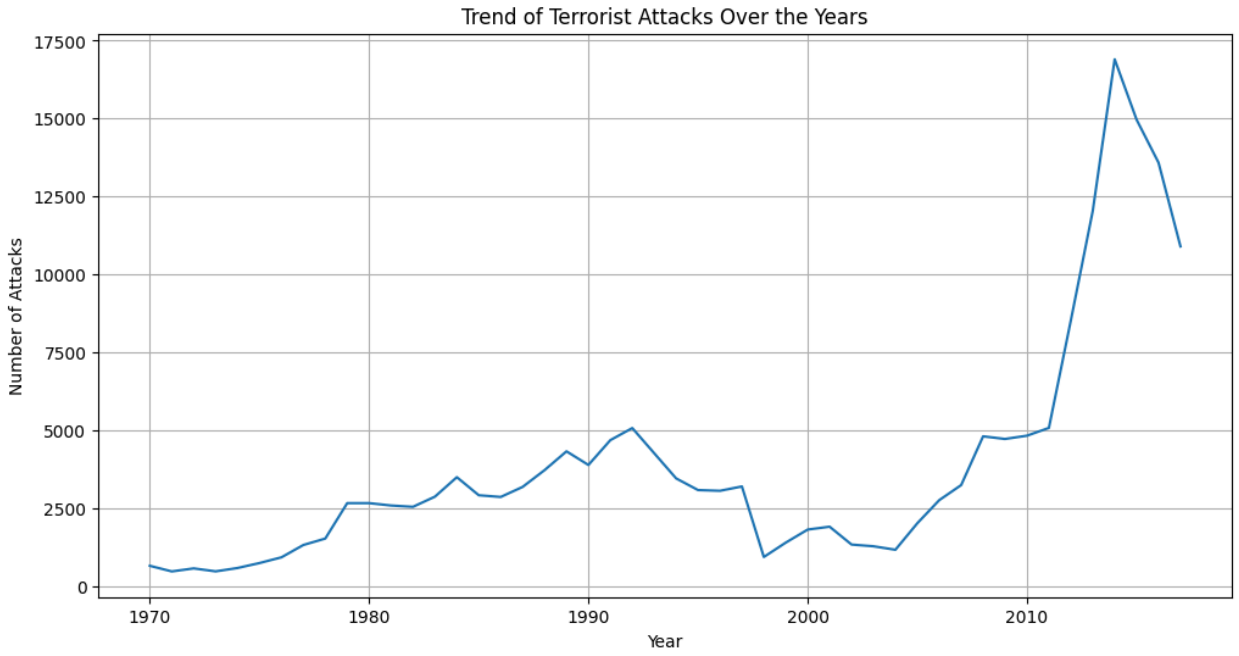
- **Most Common Attack Types**: Bombing/Explosion, Armed Assault, and Assassination were the most common attack types.

Target Type Analysis

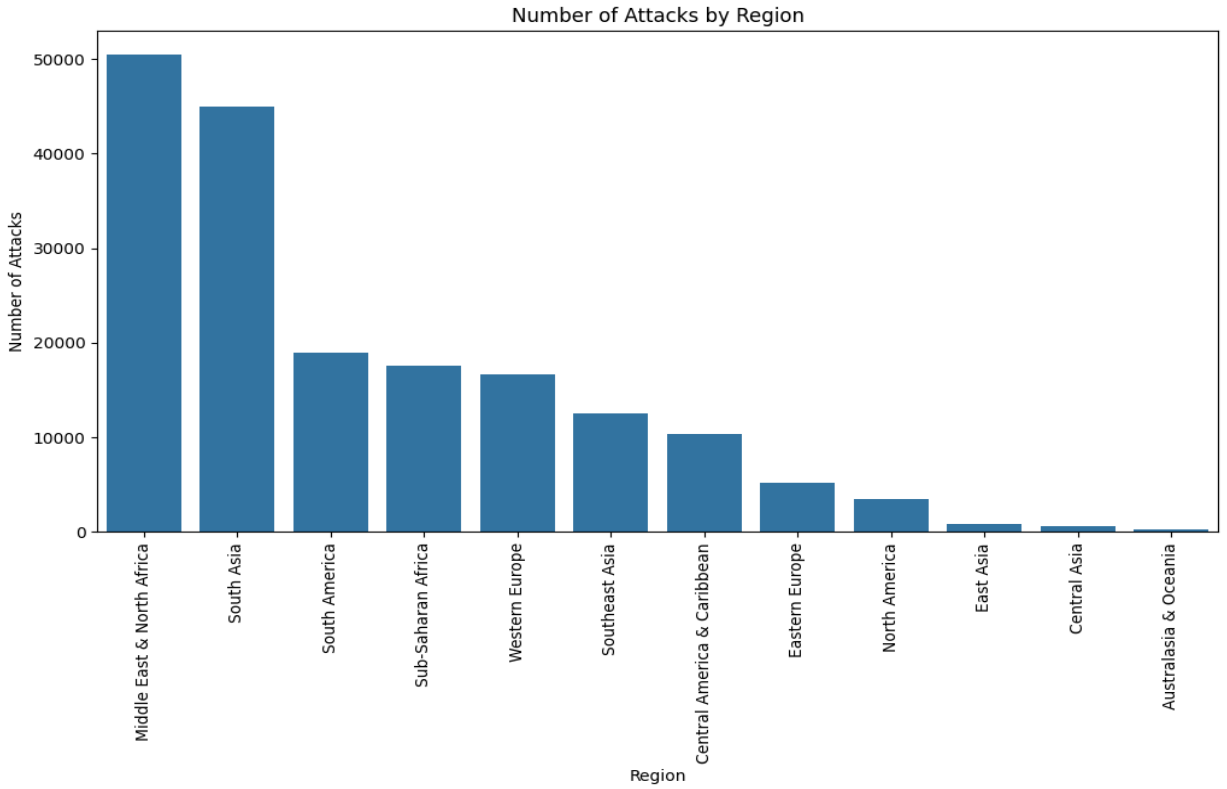
- **Most Common Targets**: Private Citizens & Property, Military, and Police were the most common targets.

Visualizations

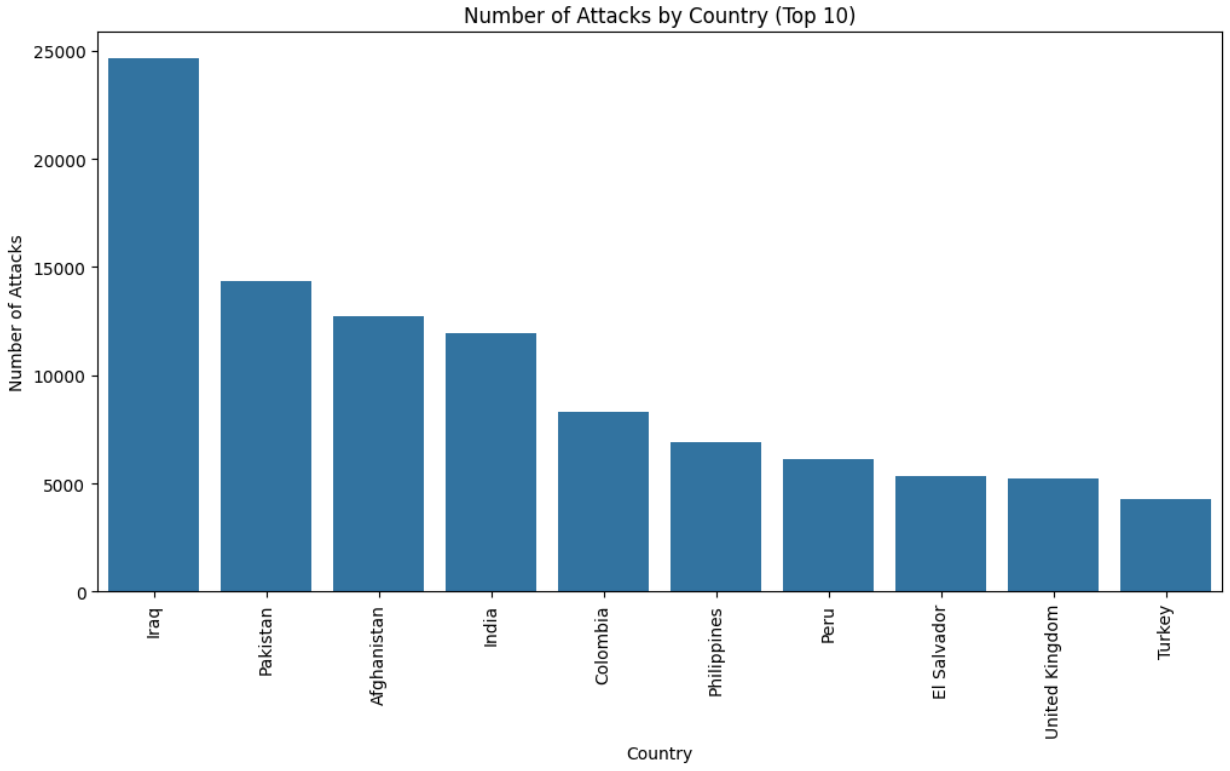
1. **Trend of Terrorist Attacks Over the Years**: Line plot showing the number of attacks per year.



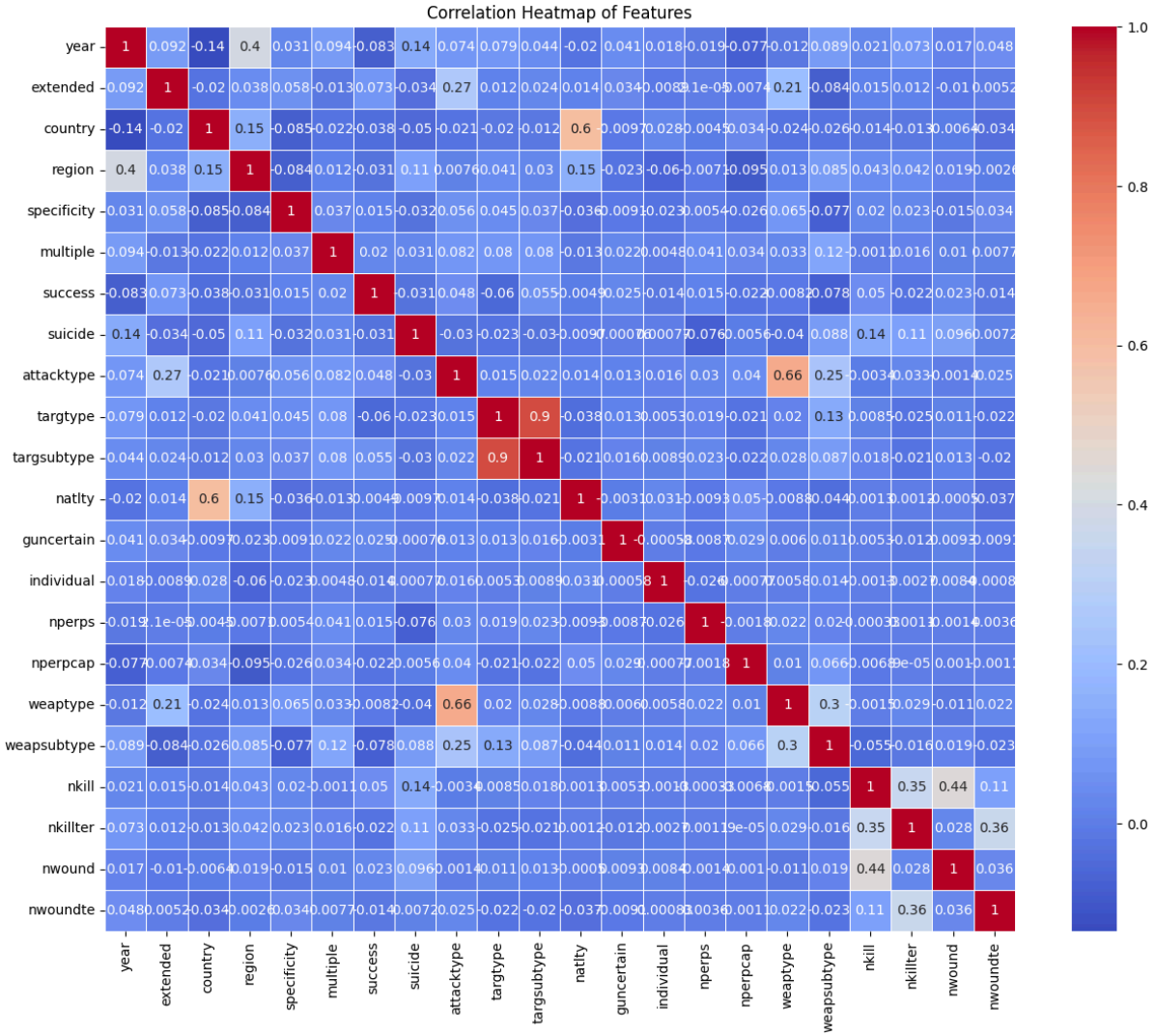
2. **Number of Attacks by Region**: Bar plot showing the number of attacks in each region.



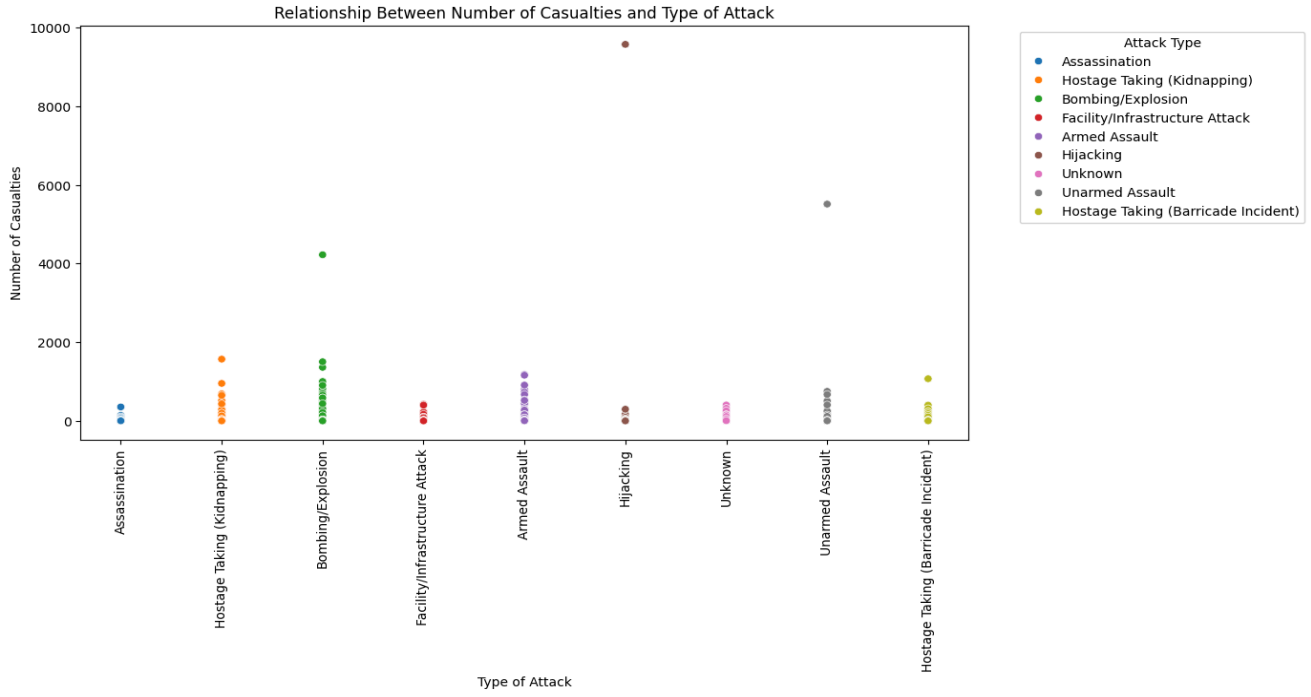
3. **Number of Attacks by Country (Top 10)**: Bar plot showing the number of attacks in the top 10 most affected countries.



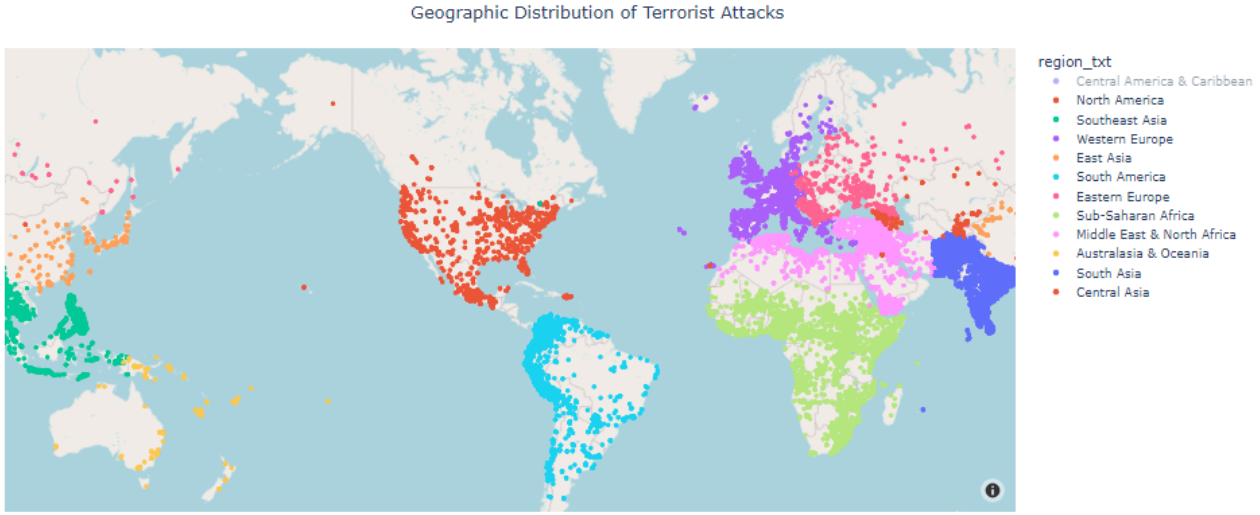
4. **Correlation Heatmap**: Visualizing the correlation between different numeric features.



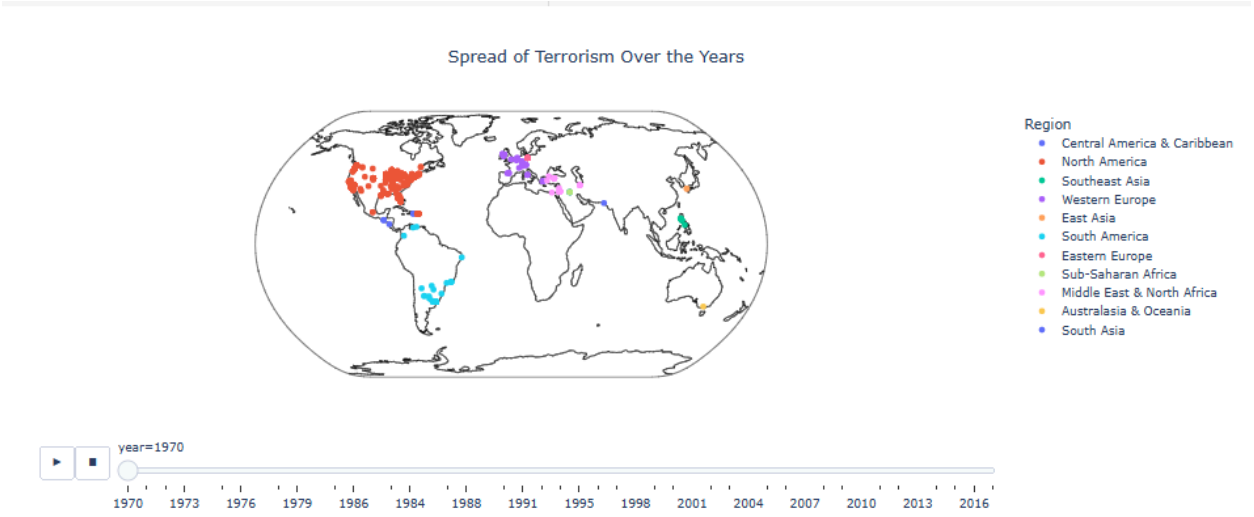
5. **Scatter Plot**: Showing the relationship between the number of casualties and the type of attack.



6. **Geographic Distribution Map**: Interactive map showing the geographic distribution of attacks.



7. **Time Series Animation**: Showing the spread of terrorism over the years.



Performance Comparison with Dask

- **Memory usage before Dask operations:** 1852.375 MB
- **Memory usage after Dask operations:** 2090.29296875 MB
- **Pandas operations took** 36.056161403656006 seconds
- **Dask operations took** 11.875378370285034 seconds

Insights Derived from the Data

1. **Temporal Trends**: The number of terrorist attacks has varied significantly over the years, with noticeable peaks during certain periods.
2. **Geographic Concentration**: Terrorism is highly concentrated in specific regions and countries, with the Middle East & North Africa, South Asia, and Sub-Saharan Africa being the most affected.
3. **Attack Methods**: Bombing/Explosion and Armed Assault are the most frequently used methods by terrorists.
4. **Targets**: Private Citizens & Property, Military, and Police are the most common targets of terrorist attacks.
5. **Casualties**: The number of casualties (both killed and wounded) varies significantly across different types of attacks and regions.

Limitations and Challenges

1. **Missing Data**: The dataset contained a significant amount of missing data, particularly in some columns. Imputation methods were used, but this may introduce some bias.
2. **Data Quality**: The presence of inconsistent data types required significant cleaning efforts.
3. **Temporal Coverage**: The dataset covers a wide time span, and some older records may be less accurate or complete.
4. **Geospatial Accuracy**: Coordinates for some incidents were missing or imprecise, affecting the accuracy of geographic visualizations.

Conclusion

The analysis provided valuable insights into global terrorism patterns, highlighting the temporal and geographic distribution of attacks, common methods, and targets. Despite the challenges and limitations, the findings can inform policy-making, security measures, and further research in counter-terrorism efforts.