

TERRORISM DATASET ANALYSIS REPORT

Introduction

This project involved analyzing a dataset of global terrorist attacks spanning from 1970 to 2017. The objective was to extract key insights, transform the data to address inconsistencies, and visualize significant patterns in terrorist attacks worldwide. The analysis focused on identifying trends in attacks over time, regions most affected, common types of attacks, victims, and weapon types used by terrorists.

Data Preparation and Cleaning

1. Initial Review of the Dataset

The dataset consisted of over 180,000 rows and multiple columns providing details of each terrorist attack. However, several columns were inconsistent or contained missing values. The 'approxdate' column, which provided an approximate date for each attack, was found to have significant inconsistencies and ambiguity, rendering it unsuitable for analysis. As a result, I excluded this column from further processing.

2. Handling Missing Values

Several columns, including latitude and longitude, contained missing values, which were essential for geographical analysis. Rows with missing data in these critical columns were removed to ensure accuracy in the analysis of regional trends.

3. Handling Duplicate and Incorrectly Formatted Data

The 'eventid' column, which was designed to uniquely identify each record, contained duplicate values and incorrectly formatted entries. Given its irrelevance to the core analysis, I removed this column and replaced it with an automatically generated unique identifier to maintain data integrity.

Data Transformation

1. Date Transformation

The dataset had separate columns for iyear, imonth, and iday, representing the year, month, and day of each attack. The month column (imonth) contained values ranging from 1 to 12, but some entries were recorded as "0," representing unknown months. To improve readability and analysis, I created a new column 'monthName' using DAX, where the numbers were converted into text values from "January" to "December," with unknown months labeled as "Unknown."

2. Unique Identifier Column

Since the original eventid column contained duplicates, I created a new index column in Power BI to assign a unique value to each record, which was used for counting the number of attacks in subsequent visualizations.

Verification of Data Accuracy

During the analysis, I verified the dataset by searching for known large-scale terrorist attacks, such as the September 11, 2001 attacks (9/11), the Oklahoma City bombing (1995), and the Beslan school siege (2004). Upon review, it was found that the highest recorded number of deaths in a single attack was 47. This is significantly lower than expected for events like 9/11, where nearly 3,000 people died, indicating potential issues with data completeness or accuracy in recording high-casualty events.

Visualizations and Insights

1. Yearly Distribution of Terrorist Attacks

I created a line chart to visualize the number of terrorist attacks from 1970 to 2017. The x-axis displayed the months, while the y-axis represented the count of attacks (based on the unique identifier). A slicer was added to filter the data by year, enabling a detailed view of attacks in specific years. The trend showed a significant increase in attacks over time, especially in the 2000s.

2. Country-wise Distribution of Attacks

A bar chart was used to show the countries with the highest number of terrorist attacks. Iraq, Pakistan, Afghanistan, India, and Colombia emerged as the most affected countries, with the first three also being located in the Middle East.

3. Regional Distribution of Attacks

Using a map visualization, I displayed the regions with the most terrorist attacks, highlighting the concentration of attacks in the Middle East, South Asia, and Latin America. These regions accounted for the majority of attacks globally.

4. Most Common Attack Types

A bar chart was used to visualize the most common types of attacks. Bombings/explosions were found to be the most frequent, followed by armed assaults, assassinations, and hostage-taking incidents.

5. Common Victims of Terrorist Attacks

Initially, a pie chart was considered to show the common victims of terrorist attacks. However, due to the large number of categories, a tree map was chosen for better visual representation. The analysis revealed that private citizens and property, police, military, and government were the most frequent victims.

6. Weapons Used by Terrorists

A pie chart was used to visualize the top 5 types of weapons used in terrorist attacks. Vehicle-based attacks, suicide bombs, sticky bombs, and time fuses were among the most commonly used weapons.

Key Insights

- **Geographical Concentration:** Terrorist attacks are highly concentrated in specific regions, with the Middle East and South Asia experiencing the most incidents.
- **Victim Profiles:** Private citizens and law enforcement were the most frequent targets of terrorist activities, indicating a high degree of civilian impact.
- **Attack Methods:** Bombings and explosions are the most commonly employed methods, followed by armed assaults.

Conclusion

The analysis of the global terrorism dataset from 1970 to 2017 reveals crucial insights into the patterns of terrorist attacks. While the dataset provides a comprehensive overview of the global distribution of attacks, it suffers from notable gaps in recording extreme cases of high-casualty events. Additionally, missing or inconsistent data, especially in columns like 'approxdate' and 'eventid', affected the completeness of the analysis. Nevertheless, the visualizations offer valuable insights into the frequency, geographic distribution, and nature of terrorist attacks over time.

To improve the accuracy of future analyses, it is recommended to address the inconsistencies in high-casualty event reporting and ensure the inclusion of major terrorist incidents in the dataset.