# EDA Report: Global Terrorism Database

## Introduction

This report presents an in-depth Exploratory Data Analysis (EDA) conducted on a Global Terrorism dataset. The dataset comprises variables on location, tactics, perpetrators, targets, and outcomes information. The primary goal of this analysis is to gain insights into patterns, trends, and characteristics of terrorism.

**Data Overview**

**Dataset:** Global Terrorism Database

**Source:** Kaggle

**Rows:** 181,691

**Columns:** 135

## Dataset Description

The Global Terrorism Database (GTD) comprises records of terrorist incidents from 1970 to 2017, except 1993. It includes attributes such as date, location, attack type, target type, etc. With over 180,000 records, this dataset provides a rich source of information for understanding the dynamics of terrorism.

**Dataset Attributes Summary**

**iyear:** Year of the incident

**imonth:** Month of the incident

**iday:** Day of the incident

**country_txt:** Country where the incident occurred

**region_txt:** Region of the world

**attacktype1_txt:** Type of attack

**targtype1_txt:** Type of target

**gname:** Perpetrator group

**nkill:** Number of killed

**nwound:** Number of wounded

*etc...*

# Initial Plan for Data Exploration

The initial plan for data exploration involved:

1. Data loading and initial overview.
2. Data cleaning, which includes handling missing values and outliers.
3. Basic statistical summaries.
4. Visualizations to understand the distribution of various attributes.
5. Identifying trends, patterns, and relationships within the data.

# Data Cleaning and Feature Engineering

In the data cleaning process, I addressed missing values, duplicated records, and outliers. I also performed feature engineering to create additional meaningful features where necessary, such as joining killed and wounded to create a 'casualties' feature.

# Key Findings and Insights

1. Distribution of Terrorist Incidents Over Time

- The dataset spans several decades, and there has been a noticeable increase in terrorist incidents in recent years.

2. Types of Attacks

- Bombing/Explosion and Armed Assault are the most common types of attacks.
- Kidnapping and Hijacking are the least common types.

3. Most Targeted Countries

- Iraq, Pakistan, and Afghanistan are the top three countries with the most terrorist incidents.

4. Most Active Terrorist Groups

- The Taliban, Islamic State, and Shining Path are among the most active terrorist groups.

5. Correlation Analysis

- There is a negative correlation between the year an incident occurred and the amount of property damaged. i.e., As the years have progressed, there is a tendency for less property damage in terrorist incidents.

## Hypotheses Formulation

- **Hypothesis 1:** The number of terrorist incidents has increased over the years.
- **Hypothesis 2:** The type of attack is associated with the region where the incident occurs.
- **Hypothesis 3:** The success rate of terrorist attacks is influenced by the type of attack.

## Significance Test for Hypothesis

I conducted a formal significance test to assess **Hypothesis 3**. I performed a Chi-Square test for independence to determine if there is a statistically significant association between the success of attacks and the type of attack. The significance level (alpha) used in the test was set at 0.05

Chi-Square Test Results:

- Chi-Square Statistic: 5971.82
- P-value: 0.0

Based on the results of the Chi-Square test, we can draw the following conclusion:

The success rate of terrorist attacks is influenced by the type of attack (Reject Null Hypothesis).

This means there is a statistically significant association between the success of attacks and the type of attack. Further analysis is required to explore the nature of this association, and this finding may have implications for counter-terrorism strategies and security measures.

## Suggestions for Next Steps

1. Further analysis of the temporal trends and seasonality in terrorist incidents.
2. In-depth analysis of the most active terrorist groups and their characteristics.
3. Geospatial analysis to identify regions with the highest and lowest incident rates.

## Data Quality Summary and Request for Additional Data

The quality of the GTD dataset is generally good. However, there are some missing values that need to be handled during data cleaning.

Additional data related to geopolitical factors, socio-economic indicators, and counter-terrorism measures would enhance the analysis and provide a more holistic understanding of the factors contributing to terrorism.

In conclusion, the Global Terrorism Database offers a valuable resource for analyzing and understanding the dynamics of terrorist incidents worldwide. The insights gained from this EDA can inform strategic decisions and policy recommendations related to counter-terrorism efforts.

Karen Ngala

03-11-23