**Project Report**

**On**

**Global Terrorism Data Analysis and Visualization**



*Submitted*

*In partial fulfilment*

*For the award of the Degree of*

**PG-Diploma in Big Data Analytics**

**(C-DAC, ACTS (Pune))**

**Guided By: Submitted By:**

Mr. Prakash Sinha Kunal Bitey (240340125018) Anish Sao (240340125009)

Arsalan Khan (240340125013)

Ujjwal Kumar Garg (240340125053)

Shivi Chourey (240340125042)

**Centre for Development of Advanced Computing (C-DAC), ACTS (Pune- 411008)**

## Acknowledgement

This is to acknowledge our indebtedness to our Project Guide, **Ms. Prakash Sinha,** C-DAC ACTS, Pune for his constant guidance and helpful suggestion for preparing this project **Global Terrorism Data Analysis and Visualization.** We express our deep gratitude towards her for inspiration, personal involvement, constructive criticism that she provided us along with technical guidance during the course of this project.

We take this opportunity to thank Head of the department **Mr. Gaur Sunder** for providing us such a great infrastructure and environment for our overall development.

We express sincere thanks to **Mrs. Namrata Ailawar** **(Process Owner)** for their kind cooperation and extendible support towards the completion of our project.

It is our great pleasure in expressing sincere and deep gratitude towards **Mrs. Risha P R (Program Head)** and **Ms. Pratiksha Gacche (Course Coordinator, PG-DBDA)** for their valuable guidance and constant support throughout this work and help to pursue additional studies.

Also, our warm thanks to **C-DAC ACTS Pune**, which provided us this opportunity to carry out, this prestigious Project and enhance our learning in various technical fields.

Kunal Bitey (240340125018) Anish Sao (240340125009)

Arsalan Khan (240340125013)

Ujjwal Kumar Garg (240340125053)

Shivi Chourey (240340125042)

## ABSTRACT

The analysis of global terrorism data is crucial for understanding patterns, trends, and the impact of terrorist activities worldwide. In this project, we conducted a comprehensive data cleaning and exploratory data analysis (EDA) of a global terrorism dataset. The initial phase involved handling missing values and addressing approximate dates within the dataset using custom parsing methods, ensuring the integrity and accuracy of the temporal data.

Subsequently, we explored the dataset's various features, including incident dates, locations, and attack types, to identify significant patterns and correlations. The data was prepared for further statistical analysis and machine learning tasks by selecting relevant features. Additionally, we conducted a comparative analysis of different data preprocessing techniques, focusing on their impact on model performance.

Our project demonstrates the potential of combining advanced data analysis techniques with modern data visualization tools to address the complex challenges of global security. By providing a robust tool for analyzing and understanding terrorism trends, we aim to support policymakers and security agencies in making data-driven decisions and mitigating the risks associated with terrorism. The insights derived from this analysis can contribute to more effective strategies in combating terrorism worldwide.

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **S. No** | **Title** | **Page No.** |
|  | **Front Page Acknowledgement Abstract**  **Table of Contents** | **I II III**  **IV** |
| **1**  **1.1**  **1.2** | **Introduction**  Introduction  Objective and Specifications | **01-04**  **01-02**  **02-04** |
| **2** | **Literature Review** | **05-06** |
| **3** | **Methodology/ Techniques** | **07-13** |
| **3.1** | Approach and Methodology/ Techniques | **07-11** |
| **3.2** | Dataset | **11** |
| **3.3** | Model Description | **11-13** |
| **4** | **Implementation** | **14-16** |
| **5** | **Results** | **17-30** |
| **6** | **Conclusion** | **31-35** |
| **7** | **References** | **36** |
|  |  |  |

**Chapter 1**

**Introduction**

Detection of text regions either from handwritten or printed document images containing various non-textual information is a difficult task, and it can be more challenging to locate the position of the text regions when we deal with a doctor’s prescription.

**1.1 Introduction**

Terrorism remains a significant global threat, affecting societies, economies, and political landscapes worldwide. Understanding the patterns, causes, and impacts of terrorism is crucial for developing effective counter-terrorism strategies and policies. With the advent of big data and advanced analytics, it is now possible to analyze vast amounts of data to gain deeper insights into the nature and evolution of terrorist activities.

This report focuses on the analysis and visualization of global terrorism data using modern data processing and analytical tools. The primary objective is to identify key trends, patterns, and correlations within the data, offering insights that can inform policy-making and enhance counter-terrorism efforts.

To achieve this, the project utilizes Spark for efficient data processing and MongoDB for data storage and retrieval. The data undergoes rigorous cleaning and preprocessing before being subjected to exploratory data analysis (EDA). Visualization techniques are employed to present findings in an intuitive and accessible manner, helping to highlight temporal and spatial trends in terrorist activities. Furthermore, time series analysis and machine learning models are applied to predict future trends and potential hotspots of terrorist incidents.

This report demonstrates the application of data science methodologies to a critical global issue, illustrating how data-driven approaches can enhance our understanding of complex phenomena like terrorism. The insights gained from this analysis have the potential to contribute meaningfully to the development of more effective counter-terrorism strategies and to the broader field of security studies.

**Exploratory Data Analysis** – Exploratory Data Analysis (EDA) is a critical step in the data analysis process, serving as the foundation for uncovering underlying patterns, trends, and relationships within a dataset. In this project, EDA was employed to gain an initial understanding of the global terrorism data. This involved examining the distribution of key variables, identifying missing values, and detecting outliers or anomalies that could impact the analysis. Visualizations such as Maps, bar charts, Heat Maps, Line chart, Pie chart, etc. were used to explore the temporal and geographical distribution of terrorist incidents, the frequency of different attack types, and the prevalence of various terrorist groups. Through these visual and statistical techniques, EDA provided valuable insights into the dataset, guiding subsequent analysis, and helping to form hypotheses.

**Data visualization** – It plays a crucial role in transforming complex datasets into intuitive and insightful representations, making it easier to identify patterns, trends, and outliers. In the context of global terrorism data analysis, visualization techniques such as geospatial mapping, time-series plots, and interactive dashboards are employed to present data in a visually compelling manner. These tools enable stakeholders to grasp the temporal and spatial distribution of terrorist incidents quickly, understand the relationships between different variables, and uncover hidden insights that may not be immediately apparent through raw data alone. By turning data into visual stories, data visualization not only enhances the interpretability of the analysis but also supports more informed decision-making in counter-terrorism efforts.

**1.2 Objective**

The global threat of terrorism remains a significant challenge, impacting nations and communities worldwide. Understanding the underlying patterns, trends, and dynamics of terrorist activities is crucial for developing effective counter-terrorism strategies. To this end, the need for a comprehensive analysis of global terrorism data has become increasingly important.

This project focuses on the systematic examination and analysis of a detailed dataset that records terrorist incidents across the globe. The dataset includes a wide array of information related to these incidents, such as the geographical location, date, type of attack, target category, terrorist group involvement, number of casualties, and more. However, the complexity and vastness of this dataset present several challenges:

**Data Quality and Integrity:**

The dataset contains inconsistencies, missing values, and imprecise information, particularly regarding dates and locations. These issues must be addressed through rigorous data cleaning and preprocessing to ensure that the analysis is based on accurate and reliable data.

**Exploration and Pattern Discovery:**

Given the multidimensional nature of the dataset, a thorough Exploratory Data Analysis (EDA) is necessary to uncover meaningful patterns and correlations. This involves analyzing the temporal and spatial distribution of terrorist activities, understanding the frequency and impact of different types of attacks, and identifying trends that may signal emerging threats.

**Predictive Insights:**

Beyond understanding historical data, there is a pressing need to develop predictive models that can forecast potential future terrorist activities. By leveraging machine learning techniques, the project aims to predict high-risk areas and the likelihood of specific types of attacks, providing valuable insights for proactive risk management.

**Comparative Evaluation of Techniques:**

The project also seeks to compare various data preprocessing methods and machine learning models to determine the most effective approach for analyzing and predicting terrorism-related incidents. This comparative analysis is crucial for identifying best practices and enhancing the accuracy and efficiency of predictive models.

**Effective Communication through Visualization:**

Finally, the ability to communicate the findings through clear and impactful data visualizations is essential. The project aims to generate visual representations that not only highlight key insights but also make the data accessible to a wide audience, including policymakers, security professionals, and researchers.

Overall, this project addresses the challenge of analyzing complex and extensive terrorism data to provide actionable insights and support decision-making processes in the field of global security. By cleaning, analyzing, modeling, and visualizing the data, the project seeks to contribute to the ongoing efforts to understand and mitigate the threat of terrorism worldwide.

**Chapter 2**

**LITERATURE REVIEW**

1. LaFree et al. (2015)

In this study, LaFree and colleagues conducted a comprehensive analysis of the Global Terrorism Database (GTD), one of the most extensive datasets available on terrorist incidents worldwide. Their research focused on identifying trends and patterns in terrorism over several decades, exploring factors such as the frequency, geographical distribution, and types of attacks. The study highlighted the importance of data accuracy and consistency, particularly in the context of historical data where inconsistencies are common. The authors emphasized the need for rigorous data cleaning and preprocessing to ensure reliable analytical outcomes, which is a critical step in any terrorism-related research.

2. Young and Dugan (2014)

Young and Dugan's research delved into the spatial and temporal patterns of terrorist activities using advanced statistical methods. Their work demonstrated how EDA could uncover significant insights, such as identifying hotspots of terrorist activity and understanding how these hotspots shift over time. They also stressed the importance of visualizing data to make complex patterns more accessible and understandable. Their research provides a foundation for the use of geospatial analysis in terrorism studies, which aligns with the exploratory data analysis objectives of this project.

3. Enders et al. (2016)

Enders and colleagues explored the economic and political impacts of terrorism using a multi-dimensional analysis of the GTD. They applied various data modeling techniques to assess how terrorism affects different sectors and regions. Their study underlined the necessity of feature selection in predictive modeling, as including irrelevant or redundant variables can lead to overfitting and reduce model accuracy. This research is particularly relevant to the machine learning aspect of this project, where feature selection plays a critical role in developing reliable predictive models.

4. Piazza (2013)

Piazza's research focused on the relationship between socio-political factors and the occurrence of terrorism. By employing regression analysis on the GTD, the study identified key predictors of terrorism, such as political instability and economic inequality. Piazza's work demonstrated the value of integrating socio-economic data with terrorism data to enhance the understanding of the root causes of terrorism. This approach highlights the potential benefits of comparative analysis in identifying the most effective data preprocessing and modeling techniques.

5. Krueger and Laitin (2008)

Krueger and Laitin conducted a pioneering study that examined the motivations behind terrorist activities, using the GTD to analyze patterns in attack types and target selection. Their research emphasized the importance of accurate data representation and the challenges associated with missing or ambiguous data. They proposed methodologies for addressing these challenges, which are crucial for ensuring the integrity of the dataset before conducting further analysis or developing predictive models. This study is directly applicable to the data cleaning and preprocessing phase of this project.

6. Chenoweth et al. (2009)

Chenoweth and her team explored the effectiveness of counter-terrorism strategies by analyzing trends in terrorist activities before and after specific interventions. Their work utilized advanced data visualization techniques to illustrate changes in terrorist behavior over time. The study underscored the importance of interactive and dynamic visualizations in communicating complex data trends to policymakers and the public. This research supports the objective of using data visualization tools in this project to enhance the accessibility and impact of the findings.

**Chapter 3**

**Methodology and Techniques**

**Data Cleaning and Preprocessing**

* Data Source: The Global Terrorism Database was obtained and loaded into a suitable data analysis environment (e.g., Python using Pandas).
* Data Cleaning:
  + Handling Missing Values: Identified and handled missing values appropriately, either by imputing them or excluding them from specific analyses.
  + Data Type Conversion: Converted columns to appropriate data types (e.g., dates, categorical, numeric).
  + Feature Engineering: Extracted relevant features such as year, month, and day from date columns and created new categorical or numeric variables as needed.

**Time Series Analysis**

* Annual Trends:
* Objective: Analyze the overall trend of terrorist attacks over time.
* Method:

- Grouped the data by year and plotted the number of annual

terrorist attacks from 1970 to 2020.

- Applied a rolling mean to smooth short-term fluctuations and highlight

long-term trends.

* Regional Analysis:
* Objective: Understand regional differences in the number of terrorist attacks.
* Method:

-Segregated the data by region and plotted the number of attacks over the region.

- Focused on identifying regions with the highest frequency of attacks.

**Detailed Time Series Analysis for Specific Period (1980-1990)**

* Objective: Conduct an in-depth analysis of the 1980-1990 period to understand seasonal patterns and stationarity.
* Method:

- Grouped data by year and month.

- Conducted the Augmented Dickey-Fuller (ADF) test to check for stationarity.

- Performed seasonal decomposition to separate the trend, seasonality, and residual components.

- Generated Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots to identify the lag structure.

**Time Series Prediction (1990-1995)**

* Objective: Forecast terrorist attacks using time series models.
* Method:

- Built ARIMA and SARIMAX models on the 1980-1990 time series data.

- Used these models to predict the number of terrorist attacks for the years 1990-1995.

- Evaluated the models based on accuracy and reliability of predictions.

**Target Analysis**

* Objective: Identify the most frequently attacked target types and analyze trends over time.
* Method:

- Calculated the frequency of attacks on different target types (e.g., Private Citizens, Military, Police).

- Plotted the number of attacks per year for the top 5 most attacked target types.

- Highlighted significant changes in attack patterns, particularly the spike in Military targets during 2010 and 2015.

**Chi-Squared Tests for Categorical Associations**

* Objective: Determine associations between categorical variables related to terrorist incidents.
* Method:

- Conducted chi-squared tests between:

- Terrorist Groups and Weapon Type

- Terrorist Groups and Target Types

- Region and Attack Types

- Terrorist Groups and Attack Types

- Evaluated the significance of associations and interpreted the results.

**Correlation Analysis and Heatmap**

* Objective: Explore relationships between numeric variables.
* Method:

- Calculated correlation coefficients between:

- Number of perpetrators (nperps)

- Number of civilians killed (nkill)

- Number of terrorists killed (nkillter)

- Number of civilians wounded (nwound)

- Number of terrorists wounded (nwoundte)

- Visualized correlations using a heatmap and identified strong and weak correlations.

- Created a pair plot for these numeric columns to examine relationships between pairs of variables.

**Distribution Analysis Objective**

* Objective: Analyze the distribution of attack frequencies and other numeric columns.
* Method:

- Calculated descriptive statistics (mean, median, skewness, kurtosis) for the frequency of attacks per month.

- Assessed normality of numeric columns using the Shapiro-Wilk test and QQ-plots.

- Identified and analyzed the presence of outliers in numeric columns using appropriate detection methods.

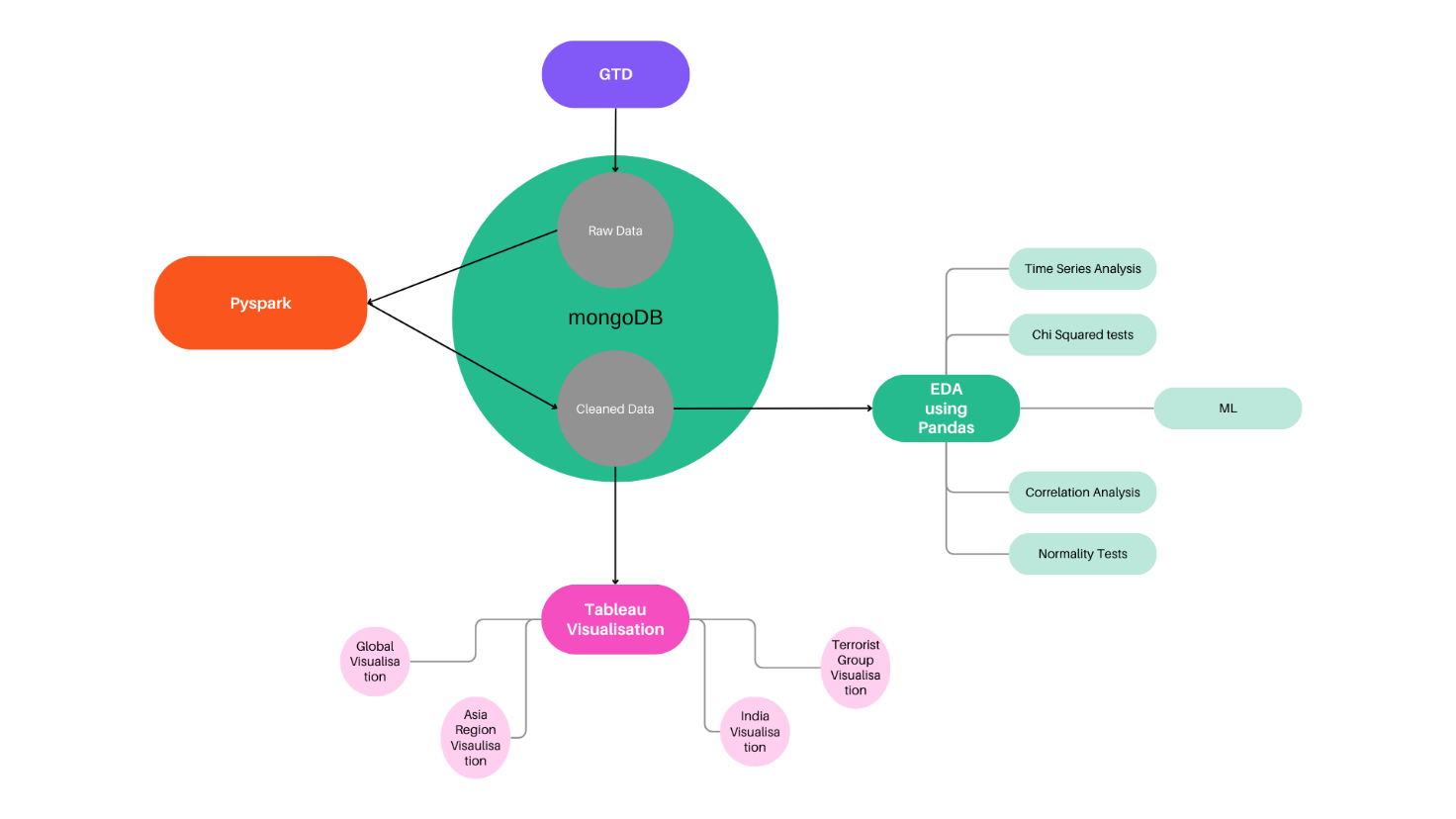
**Categorical Data Visualization**

* Objective: Visualize the distribution of categorical variables.
* Method:

- Generated count plots for each categorical column (e.g., Attack Types, Target Types, Weapon Types, Terrorist Groups).

- Used these visualizations to identify the prevalence of certain categories and observe any notable patterns.

**Technique/Workflow**



**Fig: Workflow Diagram**

**Dataset**

The Global Terrorism Database (GTD) is a comprehensive, open-source database that contains information on terrorist attacks around the world from 1970 to the present. The database is maintained by the National Consortium for the Study of Terrorism and Responses to Terrorism (START) at the University of Maryland. The GTD is one of the most detailed datasets available for understanding global patterns and trends in terrorism.

Key Features of the Global Terrorism Database (GTD)

1. **Coverage and Scope:**
   * **Time Period:** The dataset includes information on terrorist events from 1970 to the present, although data for 1993 is missing due to a loss of records.
   * **Global Reach:** The GTD covers incidents from all over the world, making it one of the most comprehensive resources for analyzing global terrorism.
2. **Data Sources:**
   * The GTD compiles data from publicly available sources such as news articles, books, journals, legal documents, and other open-source intelligence.
   * Each incident in the database is corroborated by multiple sources, ensuring reliability and accuracy.
3. **Definition of Terrorism:**
   * The GTD defines a terrorist attack as "the threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation."
4. **Key Variables:**
   * **Date and Location:** Detailed information about the date and location of each incident, including the country, region, city, and specific coordinates (latitude and longitude).
   * **Attack Type:** The method of attack, such as bombings, armed assaults, hijackings, and assassinations.
   * **Weapons Used:** The type of weapons used in the attack, including firearms, explosives, incendiaries, and unconventional weapons.
   * **Target Type:** The type of target attacked, such as businesses, governments, police, military, or civilians.
   * **Perpetrators:** Information about the individuals or groups responsible for the attack, including whether they claimed responsibility.
   * **Casualties:** Data on the number of people killed, injured, or kidnapped during the attack.
   * **Property Damage:** Information about the extent of property damage caused by the attack.
   * **Motivation:** The ideological, religious, or political motivations behind the attack, when available.
5. **Data Quality and Reliability:**
   * The GTD is regularly updated and reviewed to ensure accuracy and completeness. The data undergoes rigorous validation processes to verify the information from multiple sources.
6. **Applications:**
   * The GTD is widely used by researchers, policymakers, security agencies, and educators to study trends in terrorism, assess the impact of counter-terrorism measures, and develop strategies to prevent future attacks.
   * It serves as a foundational dataset for various academic studies, government reports, and international security analyses.

**Chapter 4**

**Implementation**

1. Use of Python Platform for writing the code with **Spark, Pandas, Matplotlib, Seaborn, SciPy, statsmodels.**
2. Hardware and Software Configuration:

Hardware Configuration:

* + CPU: 8 GB RAM, Quad core processor
  + GPU: 16GB RAM **Nvidia's GTX 1080Ti**

Software Required:

* + **Anaconda**: It is a package management software with free and open-source distribution of the Python and R programming language for scientific computations (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify deployment.
  + **Jupyter Notebook**:

Jupyter is a web-based interactive development environment for Jupyter notebooks, code, and data.

Jupyter is flexible: configure and arrange the user interface to support a wide range of workflows in data science, scientific computing, and machine learning.

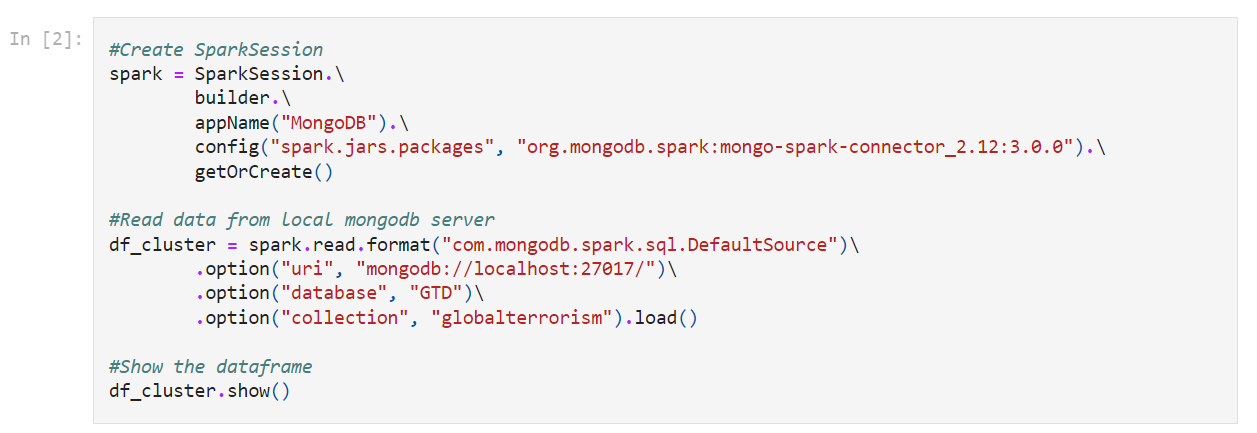
Jupyter is extensible and modular: write plugins that add new components and integrate with existing ones.

* + **Spyder**: Spyder, the Scientific Python Development Environment, is a free integrated development environment (IDE) and open source scientific environment that is included with Anaconda written in Python, for Python, and designed by and for scientists, engineers and data analysts.

It includes editing, interactive testing, debugging, and introspection features with the data exploration, interactive execution, deep inspection, and beautiful visualization capabilities of a scientific package.

* **Tableau:** Tableau is a powerful data visualization tool that enables users to create interactive and shareable dashboards, making it easier to understand and analyze data. Known for its user-friendly interface, Tableau allows individuals, from beginners to advanced users, to connect to a variety of data sources, including spreadsheets, databases, and cloud services. With Tableau, users can drag and drop data to create a wide range of visualizations, such as charts, graphs, and maps, that reveal patterns, trends, and insights hidden within the data. Tableau’s ability to handle large datasets efficiently and its advanced features like real-time data analytics, forecasting, and AI-driven insights make it a popular choice for business intelligence and decision-making across industries. Additionally, Tableau’s collaborative features enable teams to share their insights and findings easily, fostering a data-driven culture within organizations.

**Reading Data from MongoDB –**

****

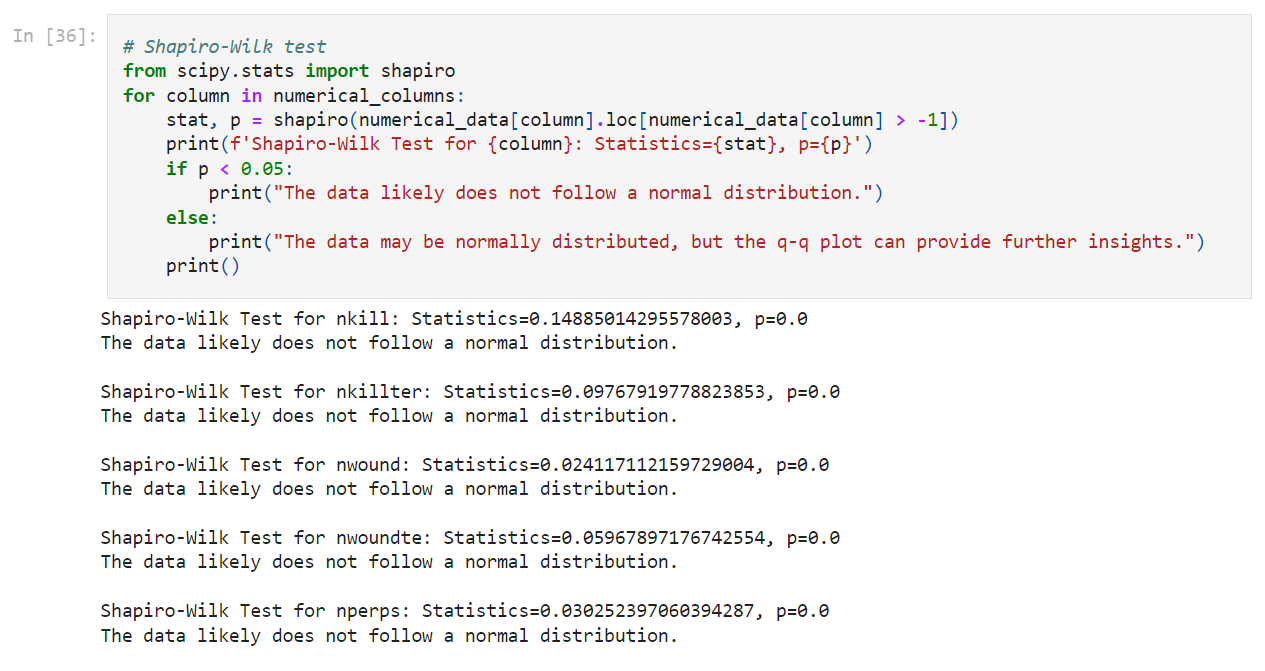
**Writing Data in MongoDB –**

****

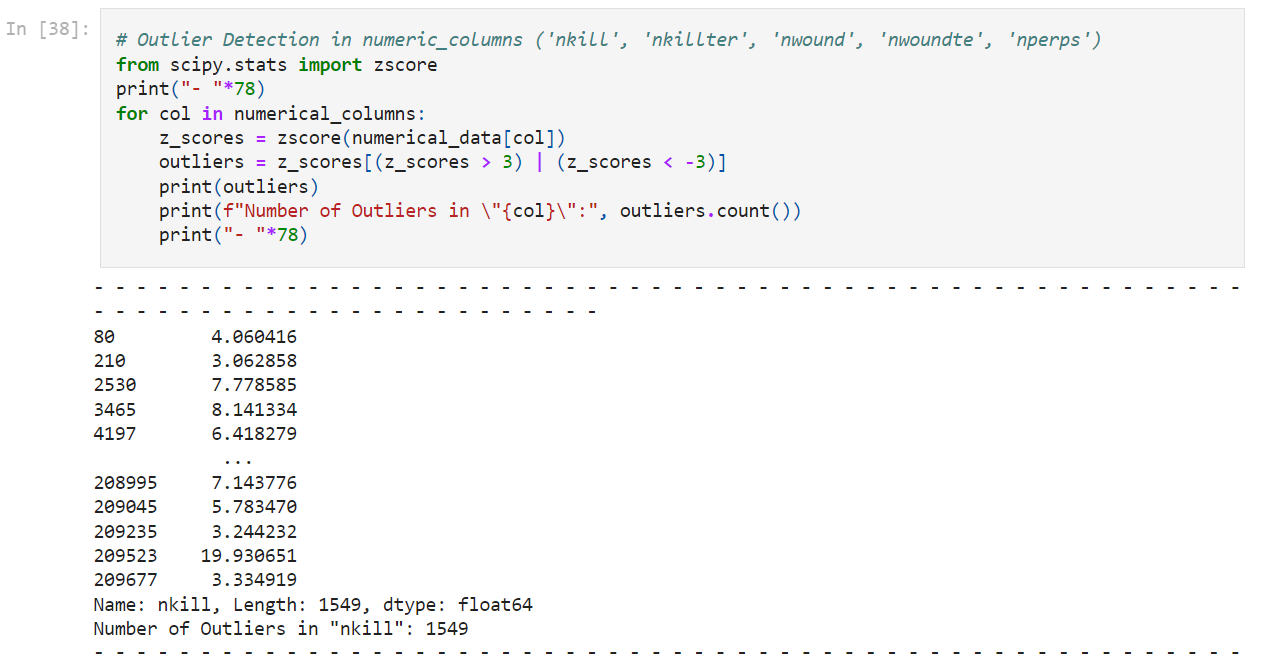
**Stationarity Check –**

****

**Shapiro-Wilk Test –**

****

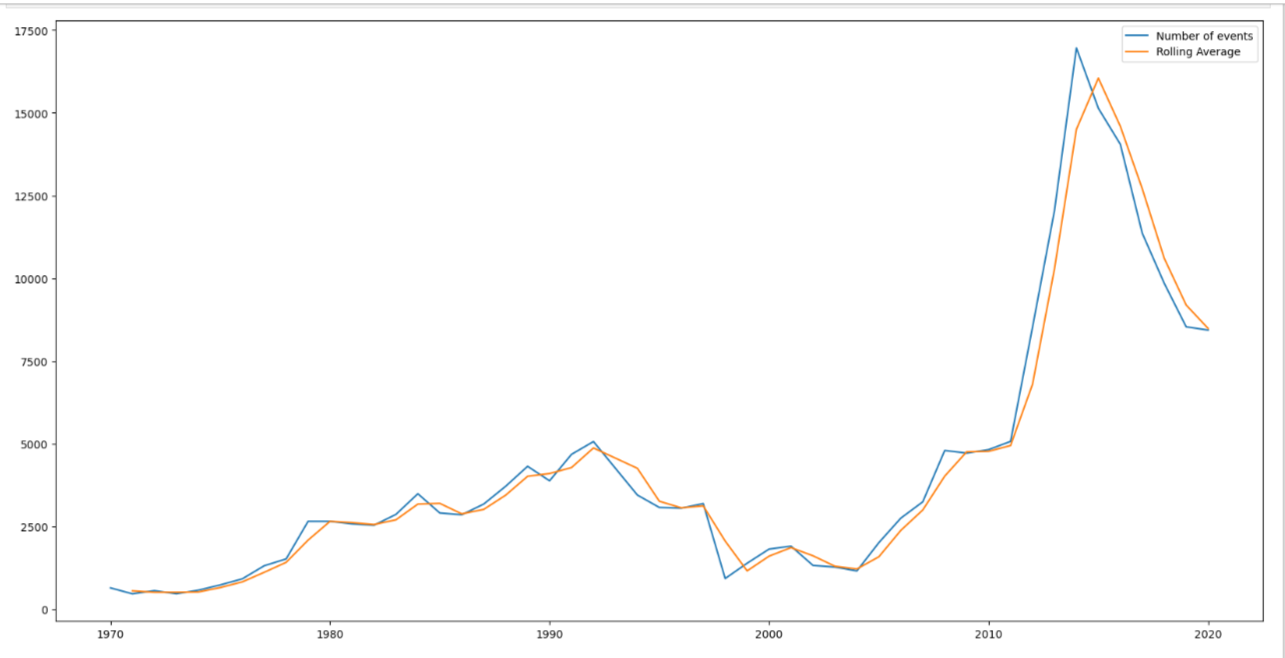
**Outlier Detection**

****

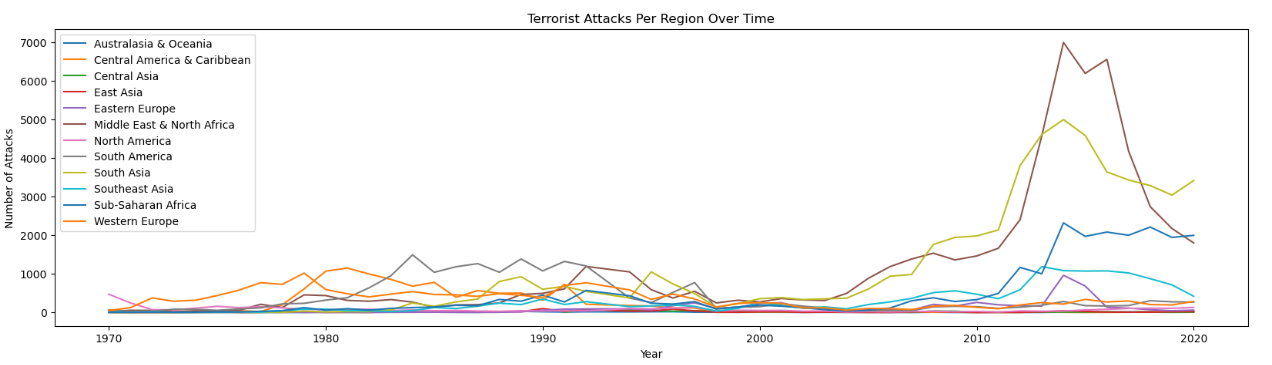
**Chapter 5**

**Results**

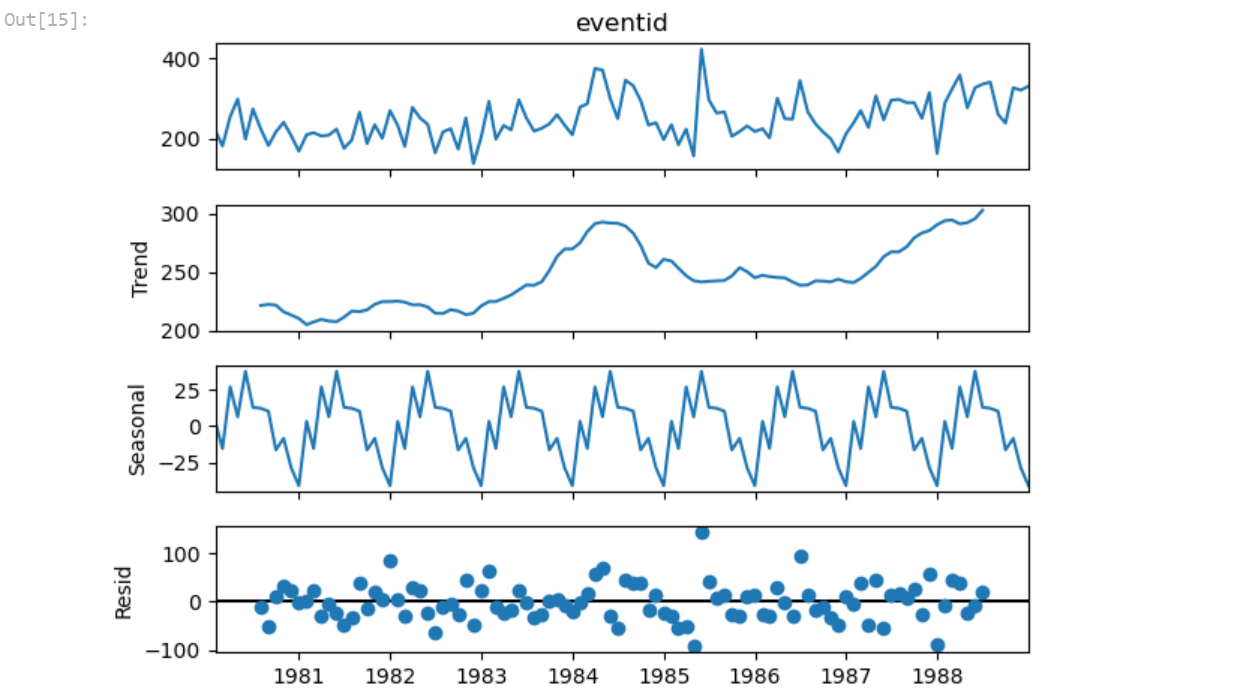
**Terrorist events per year**

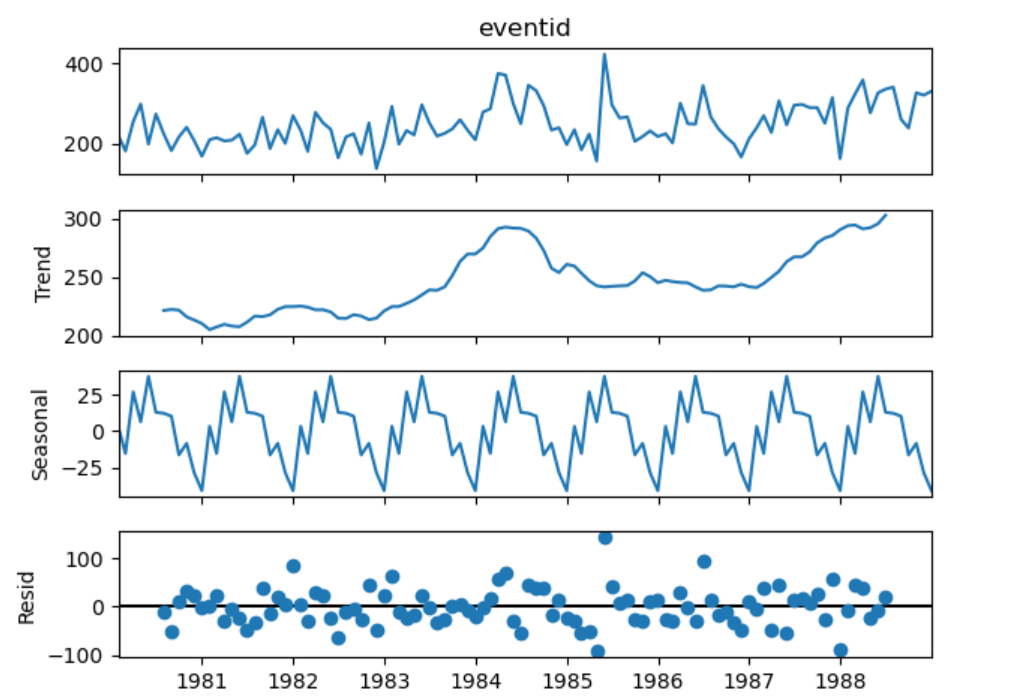


**Terrorist attacks per region**

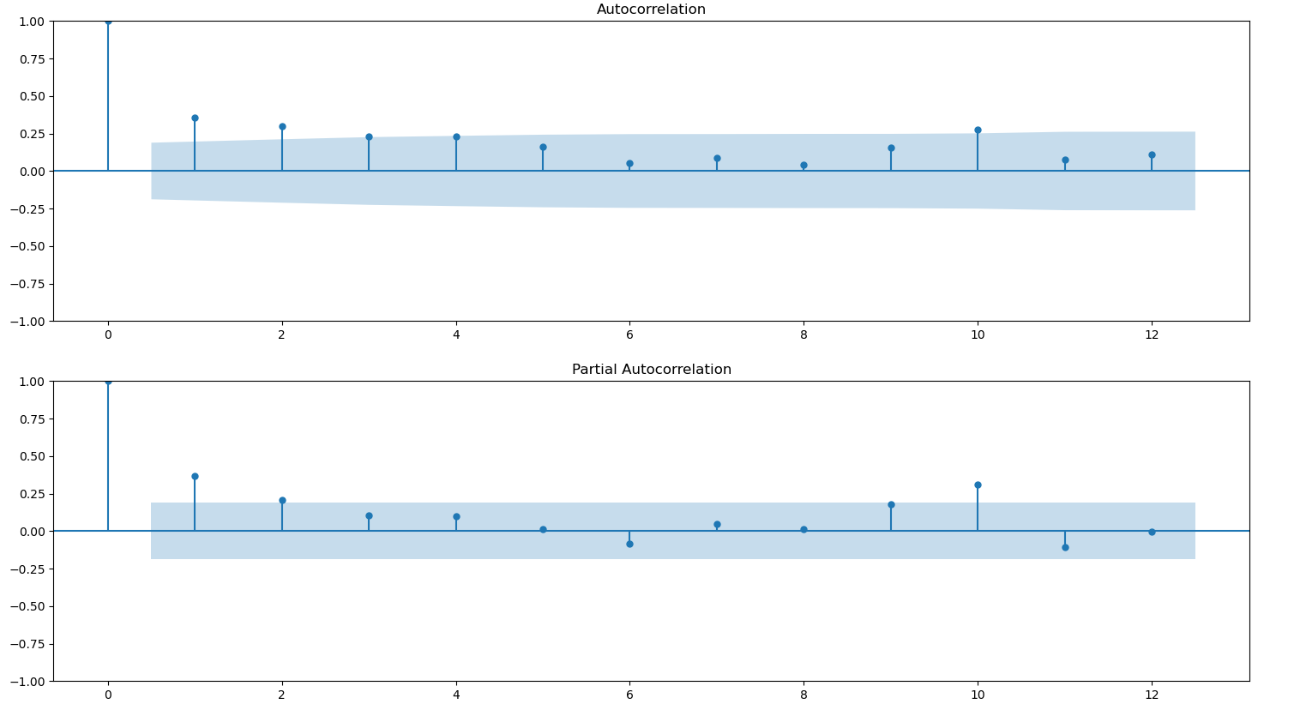
****

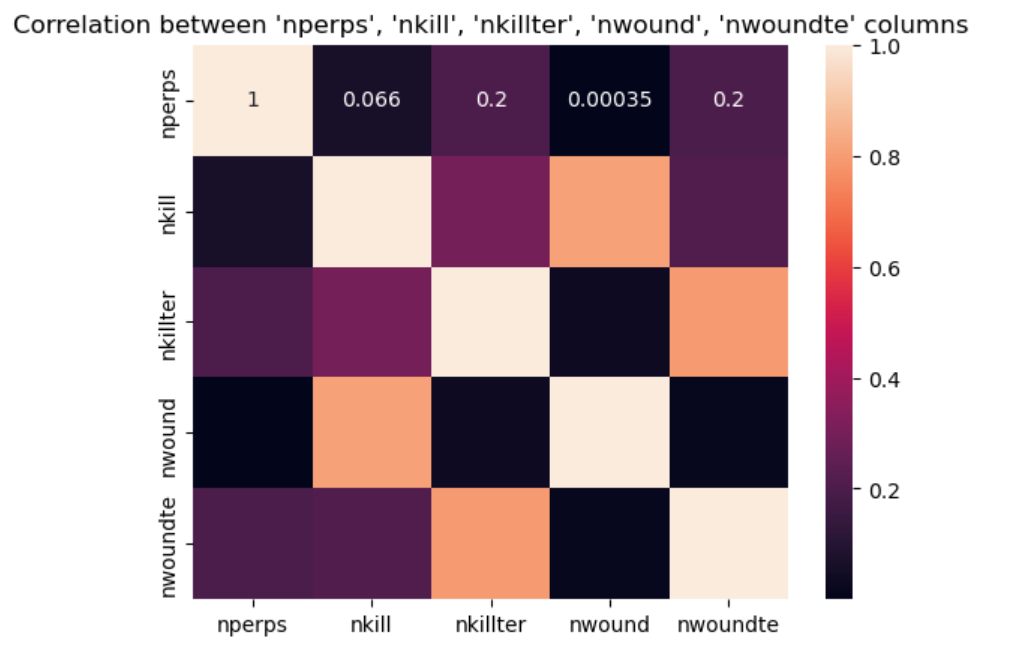
**Trend and Seasonality**

****

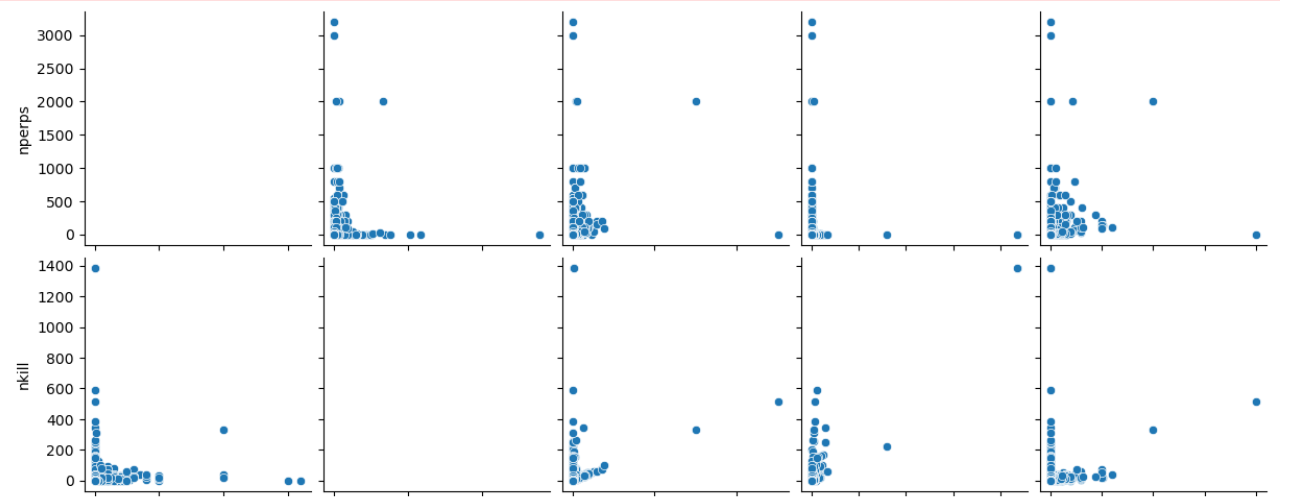
****

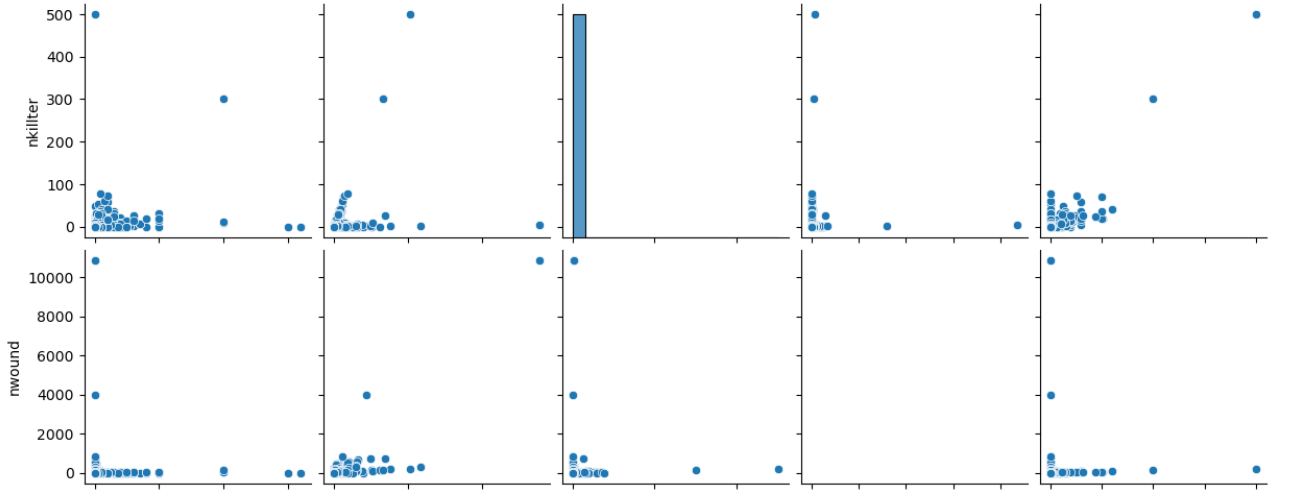
**Correlation**

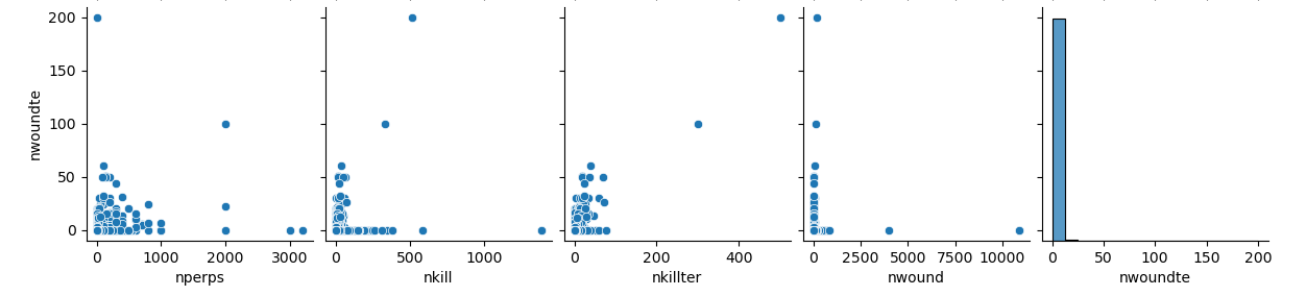
****

****

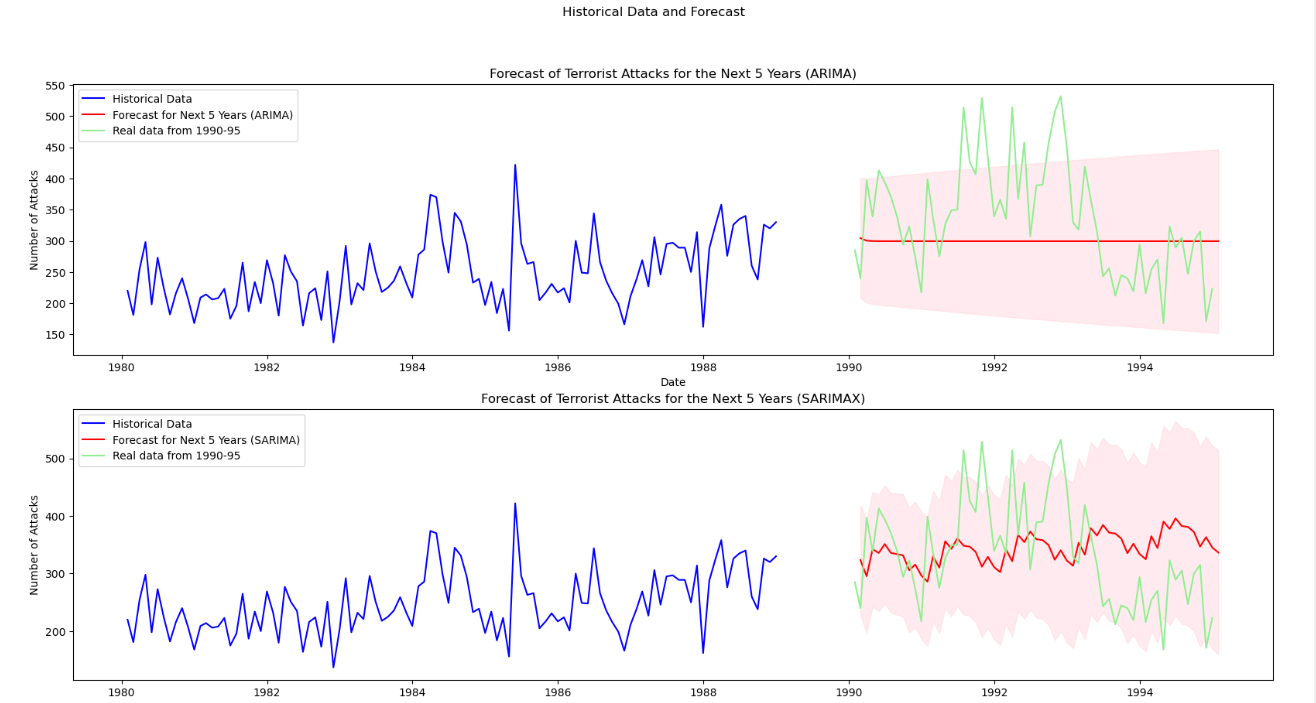
**Pair Plot**

****

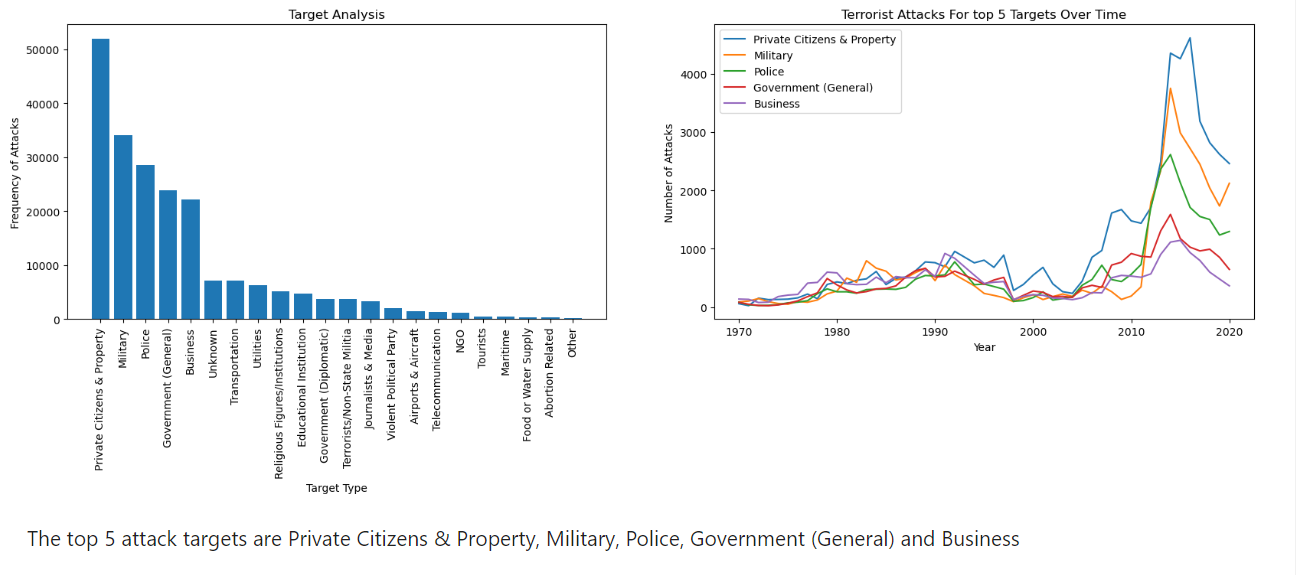
****

****

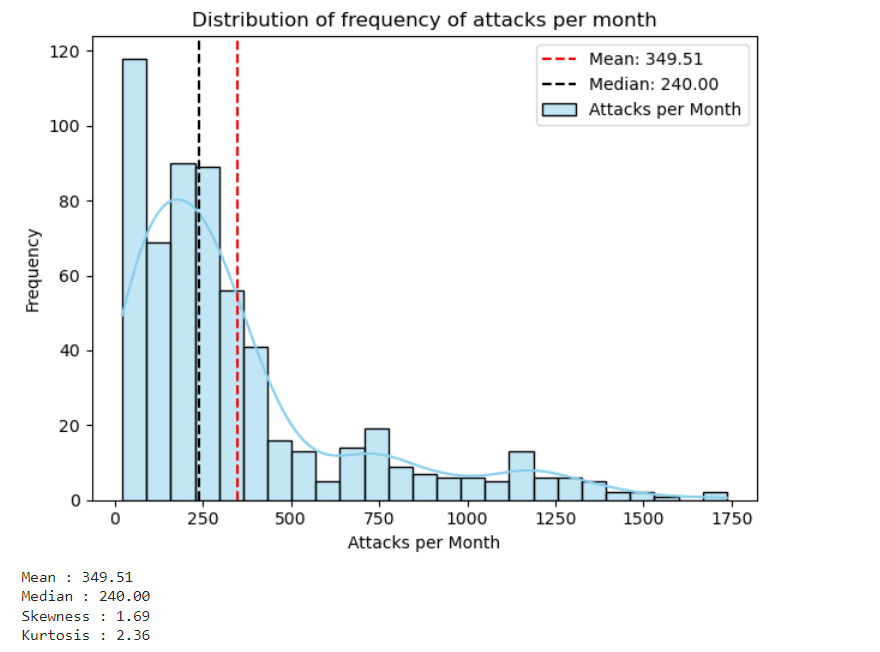
**Some Predictions/Forecasting –**

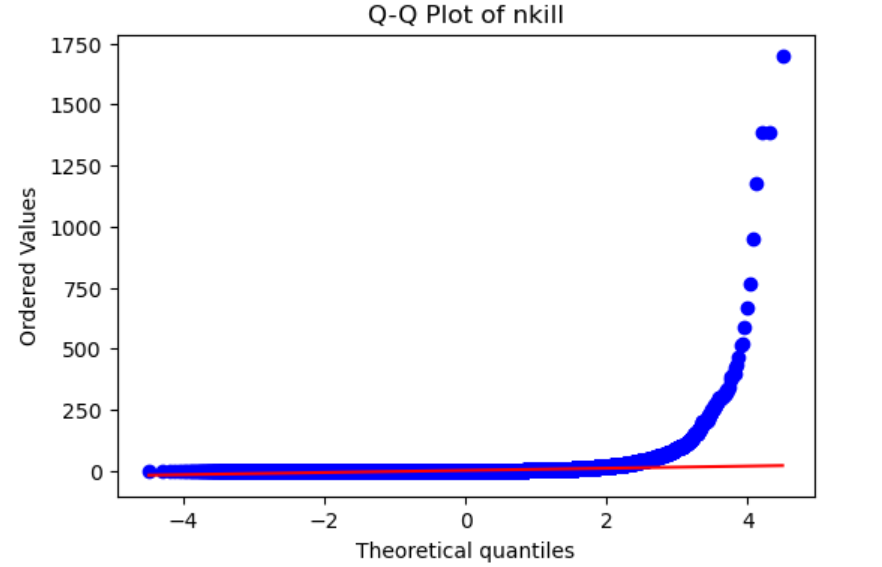
****

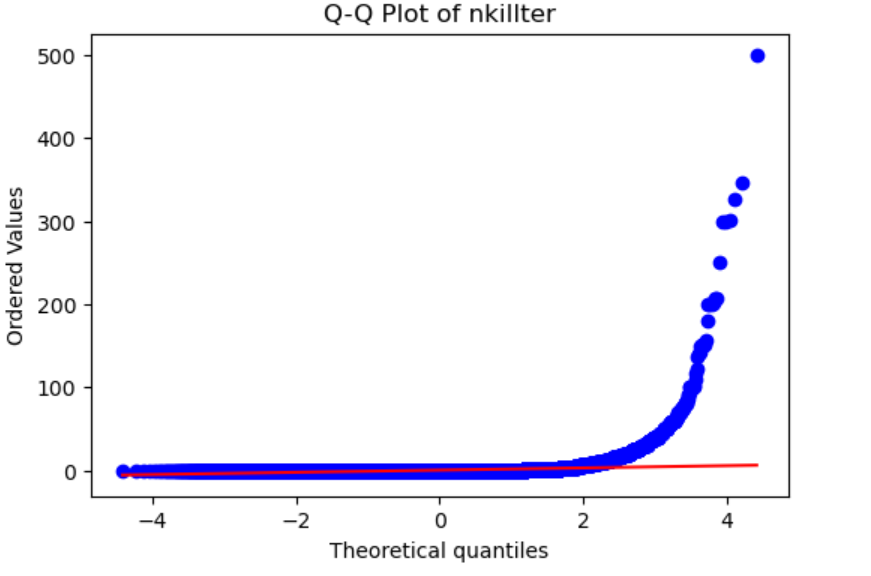
**Target Analysis –**

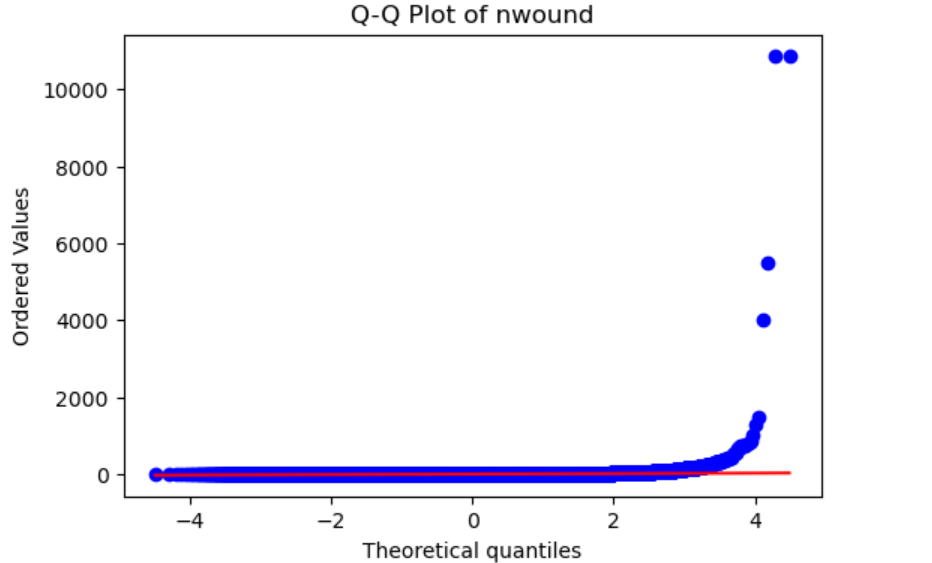
****

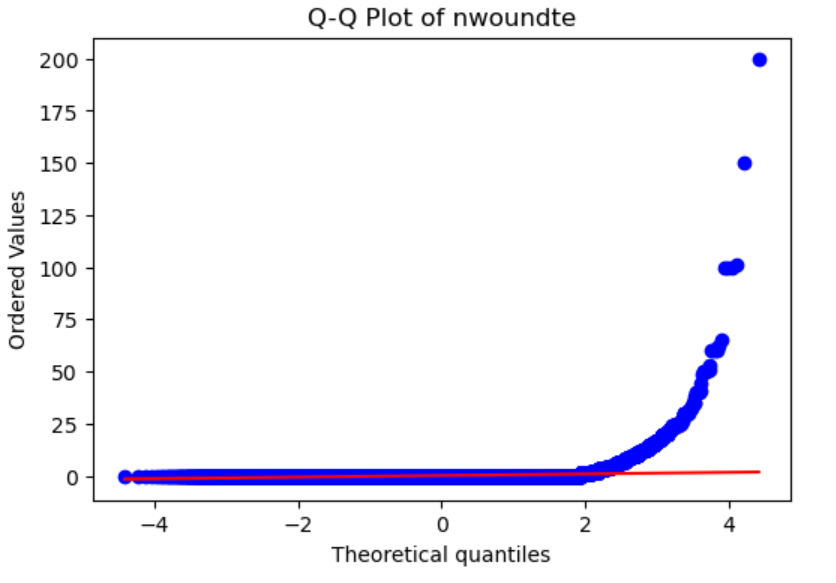
**Checking Normal Distribution**

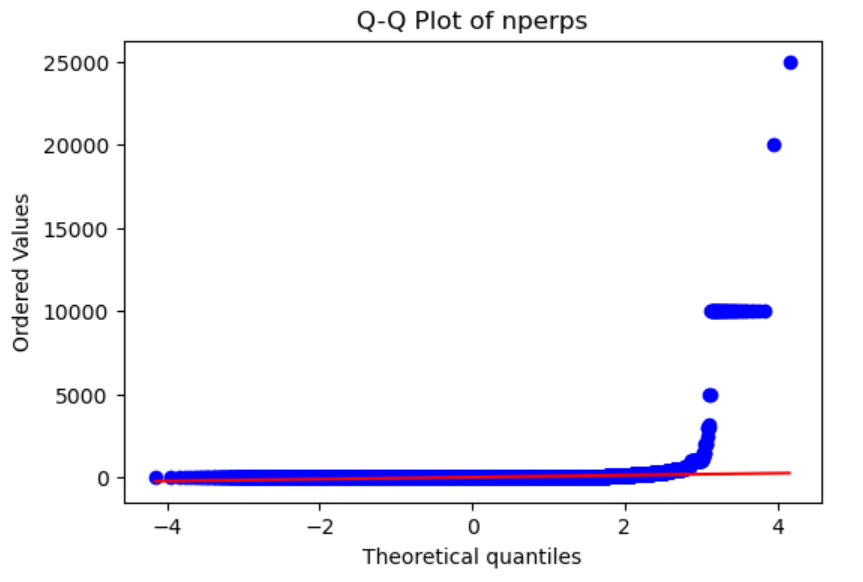
****

****

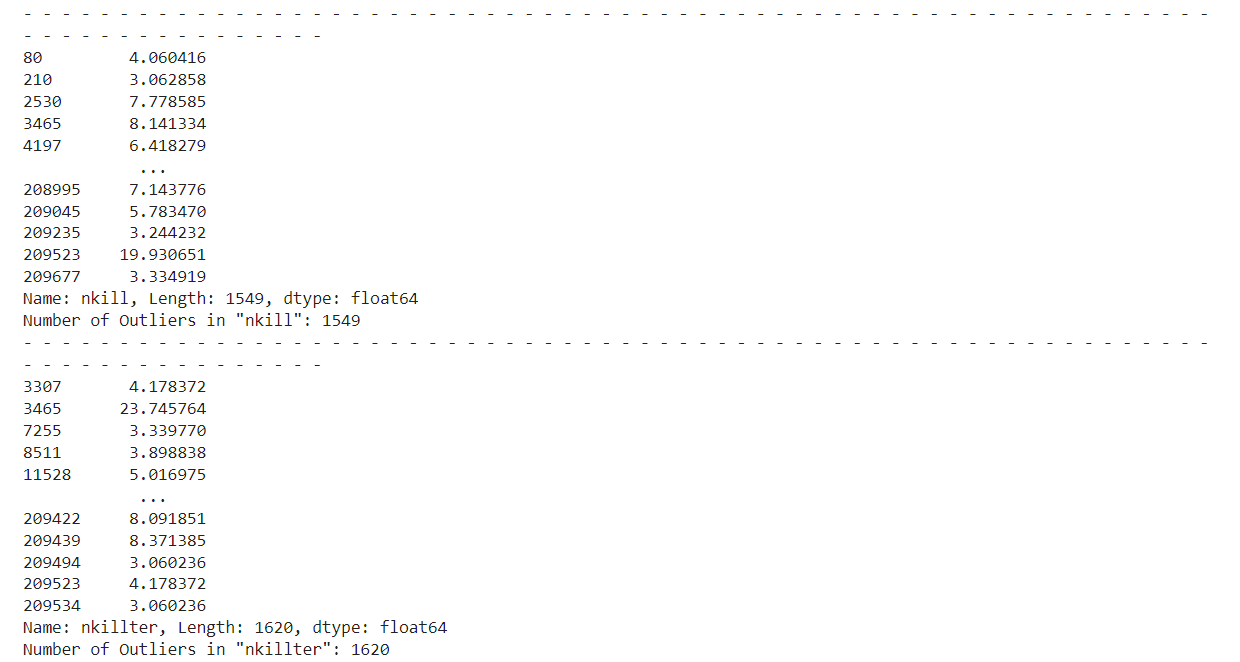
****

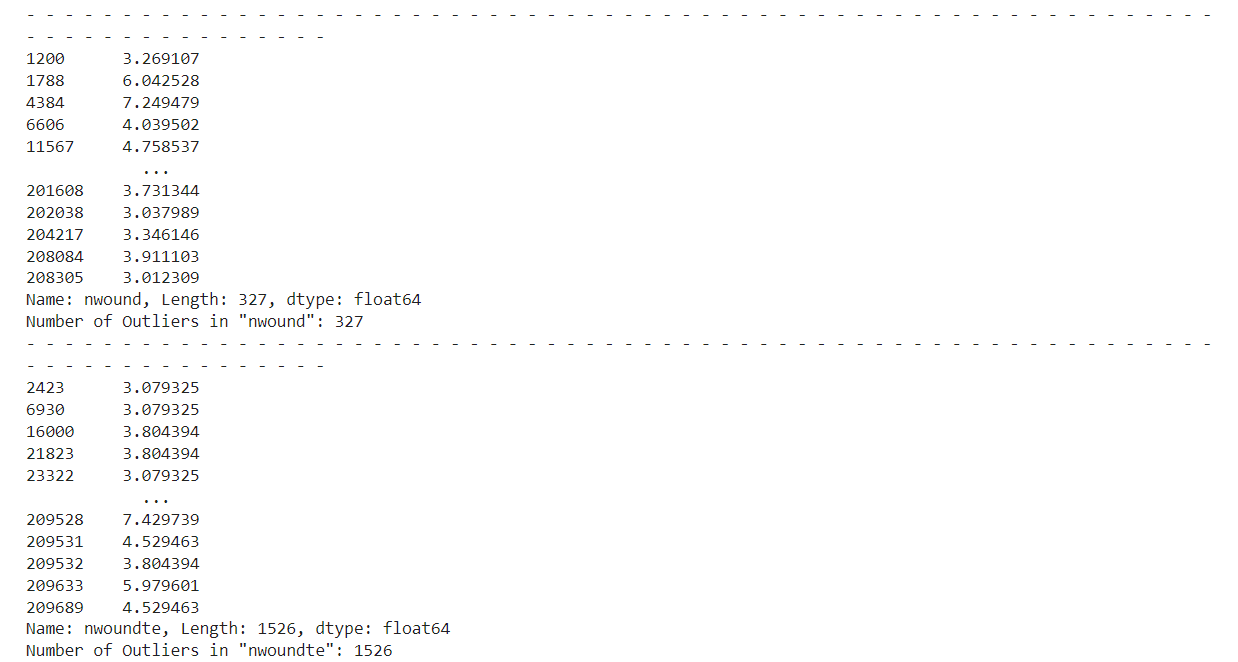
****

****

****

**Outliers Detection**

****

****

****

**ML**

Logistic regression

* ROC curve

**A graph with a red line

Description automatically generated**

* **Kfold Confusion Matrix**

**A screenshot of a graph

Description automatically generated**

**KNN**

**A graph of a number and a number

Description automatically generated with medium confidence**

**Decision Tree**

**A diagram of a mathematical equation

Description automatically generated with medium confidence**

**A chart of a graph

Description automatically generated with medium confidence**

**Random Forest**

**A graph of a number and a number

Description automatically generated with medium confidence**

**Feature Importance**

**A graph with red bars

Description automatically generated**

**Chapter 6**

**Conclusion**

**6.1 Conclusion**

Exploratory data analysis (EDA) of the Global Terrorism Database provided valuable insights into the patterns and characteristics of terrorist activities from 1970 to 2020. The key findings are summarized as follows:

1. Trends in Terrorist Attacks Over Time:

- The analysis revealed significant spikes in the number of terrorist attacks during the periods of 1985-1995 and 2010-2020. These decades experienced heightened global instability and conflict, which likely contributed to the increase in terrorist activities.

2. Regional Distribution of Terrorist Attacks:

- When analyzing terrorist attacks by region, the Middle East and North Africa (MENA) emerged as the most affected region, followed closely by South Asia. This aligns with historical geopolitical tensions and conflicts within these areas.

3. Time Series Analysis (1980-1990):

- A focused time series analysis on the decade from 1980-1990 showed that the data was stationary based on the Augmented Dickey-Fuller (ADF) test. The seasonal decomposition, ACF, and PACF analyses provided further insights into the underlying patterns and seasonal effects during this period.

4. Time Series Prediction (1990-1995):

- Using ARIMA and SARIMAX models, a prediction was made for the number of terrorist attacks from 1990-1995. These models leveraged the stationary time series data to forecast future attack trends during this period.

5. Target Analysis:

- The most frequently attacked targets were identified as Private Citizens and Property, Military, Police, Government (General), and Business, in that order. Notably, Military targets saw a significant spike in attacks during 2010 and 2015, surpassing Police, Government, and Business targets to become the second most attacked target type.

6. Chi-Squared Tests:

- Chi-squared tests were conducted to examine the associations between various categorical variables. Significant associations were found between:

- Terrorist Groups and Weapon Types

- Terrorist Groups and Target Types

- Region and Attack Types

- Terrorist Groups and Attack Types

7. Correlation Analysis:

- A heat map and pairplot of numeric variables revealed strong positive correlations between:

- Number of civilians killed (nkill) and number of civilians wounded (nwound)

- Number of terrorists killed (nkillter) and number of terrorists wounded (nwoundte)

- Weak correlations were observed between the number of perpetrators and other outcomes (e.g., nkillter, nwoundte), indicating that the number of attackers does not strongly influence the number of casualties or wounds.

8. Distribution of Attack Frequency Per Month:

- The frequency distribution of attacks per month showed a mean of 349.51 and a median of 240.00, with a skewness of 1.69 and kurtosis of 2.36, indicating a positively skewed distribution with some heavy tails.

9. Normality Tests:

- The Shapiro-Wilk test and QQ-plots indicated that none of the numeric columns followed a normal distribution. This non-normality suggests the presence of outliers and skewness in the data.

10. Outlier Detection:

- Outliers were detected in all numeric columns, with the following counts:

- nkill: 1549 outliers

- nkillter: 1620 outliers

- nwound: 327 outliers

- nwoundte: 1526 outliers

- nperps: 187 outliers

- The presence of these outliers could significantly influence the mean and skewness of the data, reflecting the extreme nature of some terrorist events.

11. Categorical Data Visualization:

- Countplots for each categorical column provided a clear visual representation of the distribution of different attack types, target types, weapons used, and terrorist groups, highlighting the prevalence of certain categories over others.

In addition to the extensive exploratory data analysis (EDA) conducted on the Global Terrorism Database, machine learning models were applied to predict and classify terrorist activities. The results of these models further enhance our understanding of the data and its predictive potential.

Key Findings from Machine Learning:

1. Logistic Regression:

- Logistic regression was utilized as a baseline model to predict various outcomes related to terrorist incidents. While it provided decent performance, the linear nature of logistic regression may have limited its ability to capture the complex relationships within the data.

2. Decision Tree:

- The decision tree model demonstrated improved predictive power over logistic regression. Its ability to handle non-linear relationships and interact with categorical variables made it a more effective model for this dataset. The decision tree's interpretability also allowed for insights into the most significant features influencing predictions.

3. Random Forest:

- The random forest model outperformed both logistic regression and the decision tree. By aggregating the predictions of multiple decision trees, the random forest model achieved higher accuracy and robustness. This ensemble method effectively captured the complex patterns in the data, making it the best predictor among the models tested.

Summary:

EDA has illuminated key patterns and associations in the Global Terrorism Database, offering a comprehensive view of terrorist activities across time, regions, and target types. The findings emphasize the need for targeted counterterrorism strategies in the most affected regions and against the most vulnerable targets. Additionally, the insights gained from the time series analysis and predictions can inform future security policies and preparedness measures. The presence of significant outliers and non-normal distributions in the data also suggests the importance of robust statistical techniques to account for these characteristics in further analyses.

The combination of EDA and machine learning analysis on the Global Terrorism Database provided comprehensive insights into the patterns, relationships, and predictive potential of the data. The decision tree and random forest models proved to be effective tools for predicting terrorist activities, offering valuable support for counterterrorism efforts and policy-making. The findings underscore the importance of using advanced machine learning techniques to better understand and anticipate terrorist behavior, contributing to more informed decision-making and enhanced security measures.

**Chapter 7**

**References**

[1] LaFree, G., Dugan, L., & Miller, E. (2015). *Putting terrorism in context: Lessons from the global terrorism database*. Routledge.

[2] Young, J. K., & Dugan, L. (2014). *Surveillance and terrorism: How monitoring terrorists may backfire*. Journal of Peace Research, 51(3), 287-299.

[3] Enders, W., Sandler, T., & Gaibulloev, K. (2016). *The impact of terrorism on public and private investment: Panel data evidence*. Journal of Peace Research, 53(1), 17-30.

[4] Piazza, J. A. (2013). *The determinants of domestic right-wing terrorism in the USA: Economic grievances and racial ideology*. Civil Wars, 15(1), 26-48.

[5] Krueger, A. B., & Laitin, D. D. (2008). *Kto Kogo? A cross-country study of the origins and targets of terrorism*. In P. Keefer & N. Loayza (Eds.), *Terrorism, economic development, and political openness* (pp. 148-173). Cambridge University Press.

[6] Chenoweth, E., Englund, S., & Moore, J. (2009). *What makes terrorism work? Terrorist tactics and their effectiveness*. Security Studies, 18(3), 510-514.

[7] https://github.com/azayz/EDA-ML-Global-Terrorism-Database

[8] <https://www.kaggle.com/code/razamh/exploratory-data-analysis-terrorism>

[9] https://www.kaggle.com/code/ash316/terrorism-around-the-world/notebook