

Analyzing Global terrorism data

Anoushka Shinde
Shikha Punjabi
Shriyansh Singh

anshinde@iu.edu
shpunjab@iu.edu
singshri@iu.edu

*Department of Computer Science
Indiana University Bloomington
Bloomington, IN 47405, USA*

Abstract

Terrorism endangers world peace, security, and stability. It is essential to comprehend the intricate terrain of international terrorism, encompassing its sources, incentives, strategies, and patterns, for multiple reasons. Firstly, thorough analysis makes it possible to recognize changing trends and new dangers. Policymakers, law enforcement organizations, and security specialists can create preventive initiatives to reduce the likelihood of terrorist occurrences by identifying these trends. Second, examining international terrorism sheds light on the underlying factors and motivations of extremism. Societies can address underlying issues and promote long-term solutions and counter-radicalization initiatives by investigating the causes behind terrorist activities. Third, research on international terrorism promotes information exchange and collaboration between nations. Since terrorism frequently crosses national boundaries, cooperation is crucial. Data analysis on terrorism fosters mutual knowledge and a coordinated response to the worldwide threat. Finally, data-driven analysis provides a means of evaluating the efficacy of counterterrorism policies and initiatives. By assessing previous achievements and shortcomings, countries can enhance their strategies, ultimately mitigating the effects of terrorism. Analyzing global terrorism is not just a scholarly endeavor in a world where the effects of terrorism are felt deeply. However, it is also an essential instrument for preserving life, promoting peace, and working toward a more secure and peaceful future.

Keywords: Terrorism Analysis, Global Security, Data Mining, Knowledge Discovery, Predictive Modelling, Counterterrorism, Text Mining, Clustering Analysis, Geospatial Visualization, Threat Assessment

1. Introduction

Terrorism spreads across borders like a horrible illness, infecting humanity with dread and chaos. It's jarring that there are organizations out there playing a violent version of tag, employing bombs and terror as opposed to childhood rules. It's an awakening that the globe isn't all sunshine and rainbows, from the terrible 9/11 atrocities to lesser but no less significant tragedies. Governments are constantly playing whack-a-mole with extremist organizations, but it's sometimes like fighting shadows. The difficulty is determining how to halt the spread of this harmful ideology while protecting the rights and liberties that define our nations. It's a difficult balancing act that keeps world leaders awake at night. Measuring terrorism on a global scale allows us to better grasp how it affects people's lives and livelihoods. However, assessing terrorist attacks and casualties is fraught with difficulties. People frequently do not concur on what constitutes a terrorist act. Even when they have been established,

these features are difficult to measure and separate from other forms of government aggression and violent crime. While the number of terrorism deaths can change a lot from year to year — in 2014, almost 45,000 people died in terrorist attacks, while in 2010, it was less than 8,000 — the average over the last decade was close to the 2019 number, at around 24,000 per year^[3]

To bring these insights into highlight, we would employ the GTD database, alongside different data mining techniques like clustering and association mining. By putting the dataset in a model and running sentiment analysis through it we get to understand how deep the scars of terrorism truly run.

2. Methodology:

The database contains information on multiple dimensions of each event. Over 100 structured variables characterize each attack's location, tactics and weapons, targets, perpetrators, casualties and consequences, and general information such as definitional criteria and links between coordinated attacks. Unstructured variables include summary descriptions of the attacks and more detailed information on the weapons used, specific motives of the attackers, property damage, and ransom demands (where applicable) ^[1]. Through advanced data mining techniques, we aim to gain valuable knowledge that can inform decision-making, policy formulation, and strategic planning to counteract terrorism worldwide.

2.1 Global Terrorism Dataset

The project aims to utilize data mining techniques to glean insightful information from the Global Terrorism Database, a sizeable open-source dataset that includes incidences of worldwide terrorism. From 1970 through 2020, it has compiled information on over 200,000 terrorist attacks, both domestic and foreign. The database contains systematic data on domestic, transnational, and international terrorist acts that have happened throughout this time, in contrast to many other event databases. Every incident includes details on the time and place of the event, the kind of target and weapons used, the number of casualties, and, if identifiable, the group or person responsible. Only after the reliability of the sources is established is information added to it. The GTD defines terrorist attacks 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."

This dataset is made available online by the National Consortium for the Study of Terrorism and Responses to Terrorism (START) to comprehend terrorist violence better and facilitate research into and defeat it. The database familiarizes analysts, policymakers, scholars, and journalists with patterns of terrorism by providing details on numerous aspects of each assault. With data on over 88,000 bombings, 19,000 assassinations, and 11,000 kidnappings since 1970, it is presently the world's most extensive declassified database on terrorist attacks.

2.2 Exploratory Data analysis and visualization

- a. Data Preprocessing:
 - * Handling Missing Values: Missing values can introduce bias and compromise the integrity of analysis. Addressing them ensures a more accurate representation of the dataset. Therefore, we conducted a comprehensive review of missing values and employed suitable strategies, such as imputation or removal, to maintain data completeness.
 - * Subset Selection for Relevance: Focusing on a subset of columns enhances the clarity and relevance of subsequent analysis. Unnecessary columns can introduce noise and complexity. To avoid those scenarios, we carefully selected columns based on research objectives,

prioritizing features essential for meaningful exploration.

- * **Data Type Conversion for Consistency:** Uniform data types facilitate seamless analysis and prevent inconsistencies that might arise from mixed data types in the dataset. So, we ensured consistency by converting data types using functions like `pd.to_numeric()` and `pd.to_datetime()` where necessary.
- * **Effective Handling of Categorical Data:** Categorical variables necessitate suitable encoding to translate them into a numerical format conducive to analysis. This involved employing techniques like one-hot encoding and label encoding, enabling the transformation of categorical data for effective statistical exploration.

b. Analysis and data visualization

i. Descriptive Statistics

Trend Analysis in the context of terrorism involves the systematic identification and evaluation of patterns and changes in terrorist activity over time. This approach is crucial for understanding the dynamic nature of terrorism. By examining trends, we can discern shifts in the frequency, methods, locations, and severity of terrorist attacks.

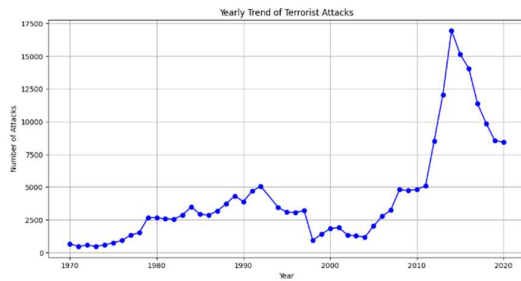


Figure 1: Number of terrorist incidents per year globally

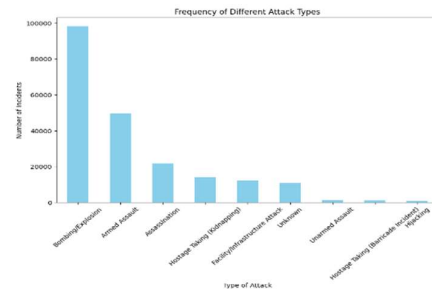


Figure 2: Types of Attacks

ii. Comparative Analysis and Correlation

Comparative analysis is a systematic approach employed in this research to discern patterns, variations, and relationships within global terrorism data.

- To understand the varying impact of terrorism across different regions, we conducted a comparative analysis to find the average lethality of attacks within each region. By assessing the severity of incidents in this manner, we aimed to identify regions experiencing higher or lower levels of violence, contributing to a comprehensive understanding of the global terrorism landscape.

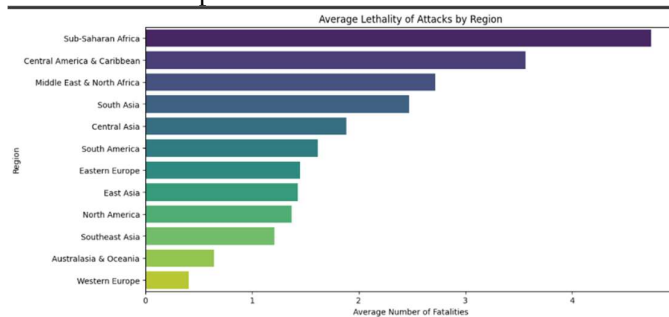


Figure 3: Lethality of attacks

- Our comparative analysis extended to evaluating countries with the maximum number of terrorism incidents. This examination sought to

identify hotspots where terrorism occurrences were most concentrated, shedding light on regions facing heightened security challenges and necessitating focused counter-terrorism efforts. In addition to assessing the sheer frequency of incidents, we delved into the comparative analysis of average fatality per country. This nuanced approach aimed to identify not only the most affected countries in terms of incident occurrence but also those witnessing higher casualties on average, providing a more comprehensive perspective on the human toll of terrorism.

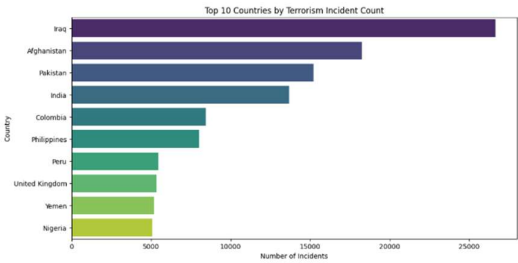


Figure 4: Countries with max terrorist attacks

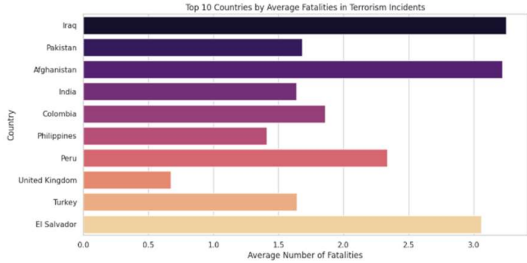


Figure 5: Average fatality per country

c) Correlational Analysis :
To understand the impact of terrorism we correlated the damage done by these attacks and the years when they took place.

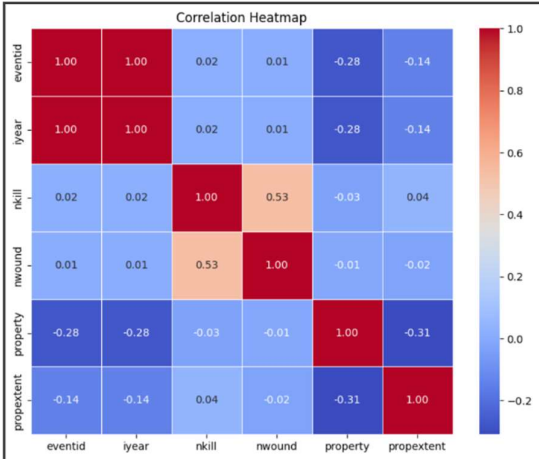


Figure 6: Impact of terrorism

To deepen our understanding of the economic ramifications of terrorism, we conducted a comparative analysis by merging another dataset containing Gross Domestic Product (GDP) information. The execution of the same was done by introducing a new dataset [6]. This allowed us to explore the correlation between a country's economic strength and the incidence of terrorism, providing valuable insights into the complex interplay between economic factors and security challenges.

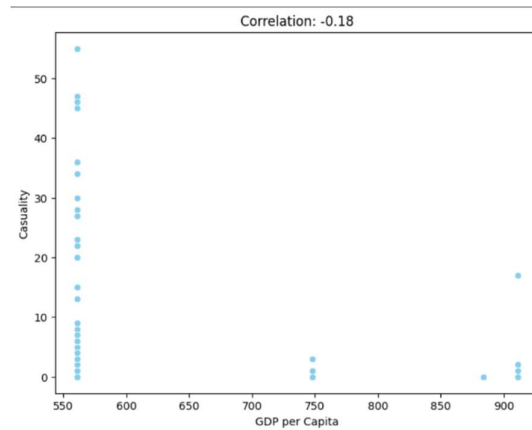


Figure 7: Effect of GDP and deaths

iii. Thematic Analysis

Thematic analysis is a qualitative research method used to identify and interpret recurring themes or patterns within a dataset. In the context of analysing trends in terrorist attacks over time, this involves scrutinizing textual or categorical data to extract overarching themes related to the nature, motives, or characteristics of the incidents. Graphs illustrating these trends can provide visual representations of the frequency and variations in terrorist activities. Figure 8 depicts trends in terrorist attacks over time may showcase patterns such as fluctuations, spikes, or declines in specific types of attacks or regions. Meanwhile, figure 9 illustrates the frequency of terrorist attacks over time offers a comprehensive view of the overall occurrence of such incidents, potentially revealing periods of heightened or subdued activity. These visualizations are crucial for discerning temporal patterns and informing strategic counter-terrorism efforts and policies.

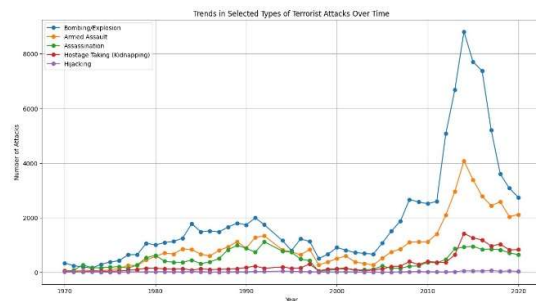


Figure 8: Terrorist attack types over time

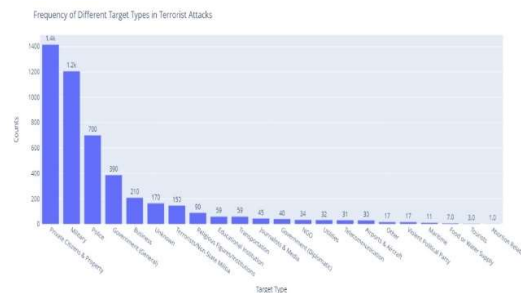


Figure 9: Frequency of terrorist attacks over time

c. Data Mining Techniques

i. Clustering

Utilizing unsupervised learning algorithms like K-means clustering, we aim to group similar terrorism incidents based on various attributes. This approach will allow us to identify hotspots and patterns in terrorist activities across different regions.

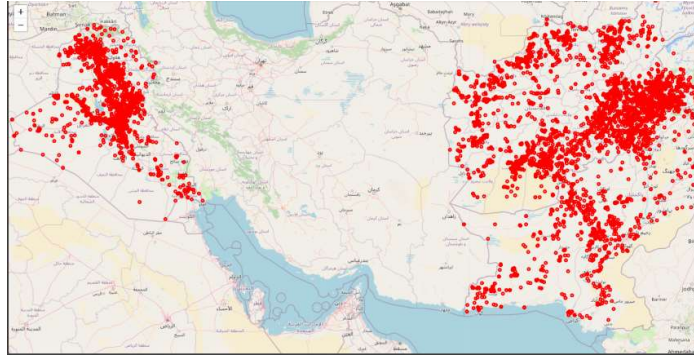


Figure 10: Bombing Incidents from 2010 to 2020

ii. Association Mining

Our project utilizes advanced association rule mining techniques, such as the Apriori algorithm, to uncover concealed relationships and connections among various factors, including attack methods, target types, and terrorist groups. Our objective is to identify the relationships between different variables within a terrorism dataset. Specifically, we focus on discovering frequent itemsets and generating rules that illuminate patterns between countries/regions and types of terrorist attacks.

To begin our analysis, we select and transform relevant columns, such as country and attack type, using one-hot encoding to prepare the data for analysis. Next, we employ the Apriori algorithm to identify frequent itemsets that occur together frequently in the dataset.

Using confidence measures, we generate association rules from these itemsets that indicate the likelihood of an itemset Y being present when itemset X is present. These rules are further assessed using the 'lift' metric, which compares the observed frequency of A and B appearing together to the frequency expected if A and B were independent.

Our analysis yields valuable insights into the patterns of terrorist activities, highlighting which types of attacks are more prevalent in specific countries or regions. These findings are crucial for comprehending the dynamics of terrorism worldwide and can be used to inform prevention and response strategies.

In summary, our project successfully applies association rule mining to Global Terrorism Dataset, employing the Apriori algorithm and metrics such as support, confidence, and lift to extract meaningful patterns and relationships without any mention of AI-powered assistance.

3. Results

The data presented in Figure 1 reveals compelling patterns in the frequency and characteristics of terrorist attacks on a global scale. The number of attacks varies from year to year, with distinct spikes and drops in certain periods. The nature of these attacks has also shifted over time, indicating the adaptability of terrorist organizations. Geographical trends demonstrate which areas have been hit the hardest by terrorism. The severity of casualties emphasizes the urgency to implement effective

counterterrorism strategies. This analysis offers valuable intelligence on the complexities of global terrorism, enabling more informed and targeted approaches to combat this persistent threat.

As depicted in Figure 2, the bar chart analysis reveals that certain types of attacks are more prevalent than others. Such insights can provide valuable information on the preferred methods of terrorist organizations, which could be influenced by factors like available resources and organizational expertise. Utilizing this knowledge can aid in devising effective counterterrorism measures and prevention strategies.

From Figure 3 we can derive that the Sub-Saharan region has the most deaths world-wide. This can be linked to the Boko Haram, an Islamist extremist group, which has been active in Nigeria and the surrounding countries, contributing to a high incidence of terrorist attacks, particularly in the northeastern part of Nigeria.

Figure 4 and 5 shows that while Pakistan has maximum terrorist incidents, it doesn't have a high average fatality. Instead, the third position is taken by El Salvador.

The maximum correlation of 0.53 was found between the 'nwound' and 'nkill' columns which provides insights into the relationship between the severity of attacks in terms of fatalities and injuries can be derived from Figure 6.

While finding relation between the GDP of a country (socioeconomic relation with deaths) and deaths of a country we discovered that there isn't any. This can be backed up by Figure 7.

Figure 8 and 9 exhibit bombing to be the most famous terrorist attack around 2014-2016 with private citizens and property being the main targets.

The analysis of bombing incidents from 2010 to 2020, as seen in Figure 10, indicates a clear pattern in the frequency and distribution of bombing attacks over the decade. Although the number of attacks varied each year, there was a consistent trend of bombings being a preferred method of attack for terrorist groups. This pattern highlights the persistent nature of bombing as a tactic in global terrorism, emphasizing the need for sustained vigilance and adaptable counterterrorism strategies to address this specific threat. It can be concluded that Iraq, Afghanistan, and Syria had the highest number of bombing incidents.

4. Discussions and future work

Analysis of the Global Terrorism Database (GTD) has uncovered significant insights into the global patterns and impacts of terrorism. Key findings include variations in the frequency and nature of terrorist attacks, geographical trends, and the relationship between terrorism and socioeconomic factors. Notably, the data showed shifts in the modes of attacks and identified regions and countries disproportionately affected by terrorism. The study also delved into the economic implications of terrorism, although no direct correlation was found between a country's GDP and the incidence of terrorist attacks.

Despite these findings, limitations were encountered. The GTD, while extensive, does not encompass all aspects of terrorism, such as the socioeconomic conditions and policy reformations. Additionally, defining terrorism poses challenges to comprehensive analysis due to its complexity and subjective nature.

Future research should expand the scope of data analysis to include socioeconomic factors and policy impacts, providing a more holistic view of terrorism. Additionally, employing more advanced data mining techniques and incorporating artificial intelligence could enhance the accuracy and depth of analysis. Collaborative efforts with international organizations and governments would also be beneficial to gain a broader perspective and access to more diverse data sources.

Investigating the root causes of terrorism, such as political instability, economic disparity, and social grievances, can offer insights into preventive measures. Additionally, exploring the effectiveness of different counterterrorism strategies across various contexts would be valuable. Such research could significantly contribute to developing more effective, context-specific counterterrorism policies and initiatives globally.

5. Author Contribution Statement:

Each team member played an equal role in contributing to the project. Shikha Punjabi proposed the project, Shriyansh Singh created the slides, and Anoushka Shinde authored the final report. Shriyansh Singh conducted Descriptive Statistics and association mining, Anoushka Shinde performed comparative analysis and correlation, while Shikha Punjabi led thematic analysis and clustering. It is worth noting that, despite the task allocation, every group member offered assistance and collaboration as needed, reflecting our collective commitment to the project's success.

References:

1. Global Terrorism Database (GTD) | START.umd.edu.
<https://www.start.umd.edu/research-projects/global-terrorism-database-gtd>
2. <https://www.ojp.gov/ncjrs/virtual-library/abstracts/building-and-analyzing-comprehensive-open-source-data-base-global>
3. Bastian Herre, Veronika Samborska, Hannah Ritchie, Joe Hasell, Edouard Mathieu and Max Roser (2023) - "Terrorism" Published online at OurWorldInData.org. Retrieved from: '<https://ourworldindata.org/terrorism>' [Online Resource]
4. Global Trends in Terrorism, 1970-2011 | 6 | Peace and Conflict 2014 | <https://www.taylorfrancis.com/chapters/edit/10.4324/9781315284378-6/global-trends-terrorism-1970%E2%80%932011-gary-lafree-laura-dugan?context=ubx&refId=ed26512a-36c0-4502-a9c5-92b334a561fa>
5. Country Reports on Terrorism - Statistical Annex | START.umd.edu.
<https://www.start.umd.edu/research-projects/country-reports-terrorism-statistical-annex>
6. Maddison Project Database, version 2020. Bolt, Jutta and Jan Luiten van Zanden (2020), "Maddison style estimates of the evolution of the world economy. A new 2020 update".