Global Terrorism Database: Analysis for Prevention of Global Terror

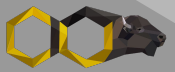
Group 1
Team Members: Liam Keyek

Description

- The Global Terrorism Database (GTD) is a comprehensive resource that compiles data on over 200,000 terrorist incidents globally from 1970 to 2020. The data is meticulously aggregated through open-source intelligence utilizing various methods to ensure the breadth and reliability of the information. Apart from offering detailed incident data, I chose the database because it is one of the largest unclassified databases related to human conflict. The dataset is from a reliable source developed in partnership with the Human-Computer Interaction Lab at the University of Maryland.
- Scope of Data:
    - The GTD database contains 209,706 recorded terrorist incidents, capturing up to 120 attributes for each event. These attributes span eight broad categories, including details like incident date, location, perpetrator details, attack tactics, target specifics, weapon types, incident outcomes, and unique variables for specific types of incidents like kidnappings and hijackings.
- Methodology:
    - The GTD sources its information entirely from publicly available, open-source materials such as electronic news archives, books, journals, and legal documents. While efforts are made to corroborate information across multiple sources, the database only reflects what's reported without further verification. The data collection has undergone various phases, with contributions from Pinkerton Global Intelligence Service, CETIS, ISVG, and START staff at the University of Maryland. Additionally, the database has integrated cases from other terrorism archives, including works by Alex P. Schmid, Yonah Alexander, Christopher Hewitt, and several other sources.

Interesting Questions

- Temporal Analysis:
  - How has the frequency of terrorist incidents changed over the years?
  - Are there specific months or days of the year that see a spike in terrorist activities?
- Spatial Analysis:
  - Which countries or regions have experienced the highest number of terrorist incidents over the past decade?
  - Are there any patterns or clusters in the geographical distribution of these incidents?
- Tactic and Weapon Analysis:
  - What are the most commonly used tactics in terrorist attacks?
  - How has the choice of weapons in terrorist incidents changed over the years?
- Network Analysis:
  - Can I construct a network to visualize the relationships between different terrorist groups?
  - How do these groups collaborate or compete with each other over time?



Prior Work

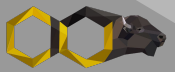
- Predictive Policing in Counter-terrorism:
  - Project: Palantir's Gotham platform has been used by intelligence agencies and the military to integrate, manage, secure, and analyze data. It's been applied in counter-terrorism efforts to predict potential threats.
  - Application: Helps in identifying patterns and networks among various entities, predicting potential terrorist activities, and aiding in decision-making.
- Social Media Monitoring for Radicalization:
  - Project: Programs like the U.S. Defense Advanced Research Projects Agency's (DARPA) Social Media in Strategic Communication (SMISC) aim to detect and counter the propaganda, misinformation, and radicalization efforts on social media.
  - Application: By analyzing social media data, these programs can identify potential threats and understand the narratives driving radicalization.
- Additional Academic Papers:
  - https://eric.ed.gov/?id=ED579753
  - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9246180/

Datasets

Dataset: https://www.start.umd.edu/gtd/using-gtd/

Source: START (National Consortium for the Study of Terrorism and Responses to Terrorism). (2022). Global Terrorism Database 1970 - 2020. https://www.start.umd.edu/gtd

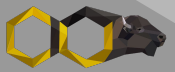
Storage: I have downloaded the full dataset. It is stored on JupyterLab. I am working on uploading to GitHub: https://github.com/liamkeyek/Keyek_Data-Mining-Project_CSPB-4502_F23/tree/main



Data Cleaning

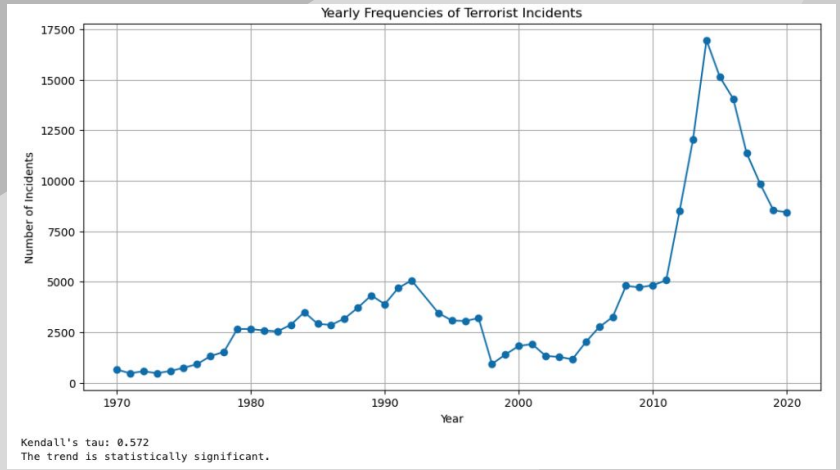
Data Cleaning:

- Data Cleaning Rigor: Utilized Python's Pandas library to import the GTD from a 'globalterrorismdb.csv' file, ensuring all data was correctly encoded with ISO-8859-1 and setting low_memory=False for efficient memory management during the process.
- Missing Value Management: Executed a thorough assessment of missing values using the isnull().mean() function, which led to the removal of attributes with more than 50% missing data to maintain the integrity and reliability of the subsequent analysis.
- Attribute Relevance Evaluation: Conducted a critical review of each attribute within the GTD, removing irrelevant variables, such as 'eventid' and 'vicinity', to streamline the dataset for more focused analysis, and added significant new features like 'day_of_week', 'total_casualties', and 'total_us_casualties' to capture the temporal patterns and human impact of terrorist events.
- GDP Integration for Economic Context: Meticulously merged GDP data from the World Bank with the GTD, aligning economic figures with corresponding terrorist events by careful synchronization of country names and event dates, adding a vital economic dimension to each incident for a multi-layered analytical approach.
- Dataset Optimization for Analysis: Finalized the data preprocessing by exporting the curated and enriched dataset to a new CSV file with utf-8-sig encoding, omitting row indices for a clean dataset, ready for advanced data mining applications and explorations of the interconnections between economic conditions and terrorism trends.



Temporal Analysis

- Rising Global Trend: The data indicates a statistically significant global increase in terrorist incidents over time, with a Kendall's tau coefficient of 0.572, highlighting a trend of growing terrorist activities worldwide.
- Regional Variations: Different regions exhibited contrasting trends, with Southeast Asia and the Middle East & North Africa showing alarming increases in terrorist activities, as indicated by high Kendall's tau values, whereas Central America & the Caribbean demonstrated a notable decrease.
- North America's Decline in Incidents: Post-major terrorism events, North America displayed a negative trend in terrorist incidents, suggesting the effectiveness of counter-terrorism measures in the region, as reflected by a Kendall's tau of -0.313.
- Stability and Changes in Other Regions: South America's stable trend with a low Kendall's tau value hints at a relative calm in terrorist activities, while the significant negative trend in Central Asia suggests changing socio-political landscapes affecting terrorism dynamics.



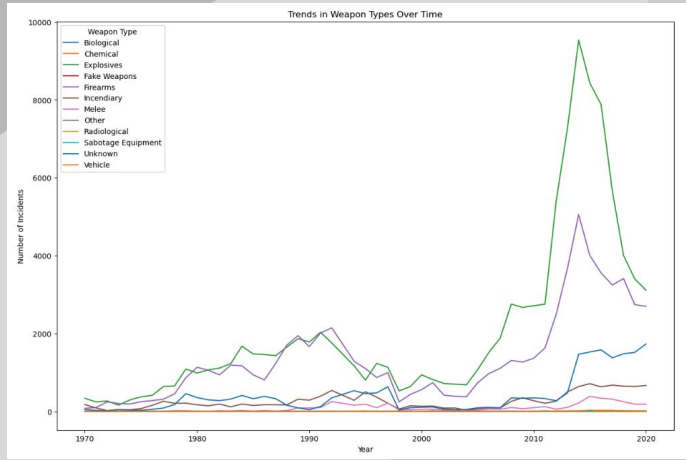
Spatial Analysis

- Regional Distribution and Hotspots: The analysis identified distinct patterns of terrorism incidents, highlighting areas with concentrated activity through scatter plots. This allowed the identification of hotspots, areas of intensified terrorist activity, and regions of relative calm.
- Cluster Analysis for Epicenter Identification: Applying K-means clustering helped categorize incidents based on geographic proximity, leading to the identification of the epicenters of terrorist activities, often correlating with regions known for political instability.
- Economic Correlation with Terrorism Incidents: By merging economic indicators such as GDP with incident data, a significant correlation was found, suggesting that economic challenges in a region might influence the likelihood of terrorist activities. This insight is crucial for informing resource allocation and counter-terrorism strategies.

Latitude	Longitude	Country
29.14609	73.73963	India
35.10133	40.03155	Syria
14.08485	-83.25842	Nicaragua
45.71395	2.909222	France
1.88501	39.59138	Kenya
9.093567	116.4975	Philippines
9.506583	9.710708	Nigeria
-19.92851	-71.82041	Chile

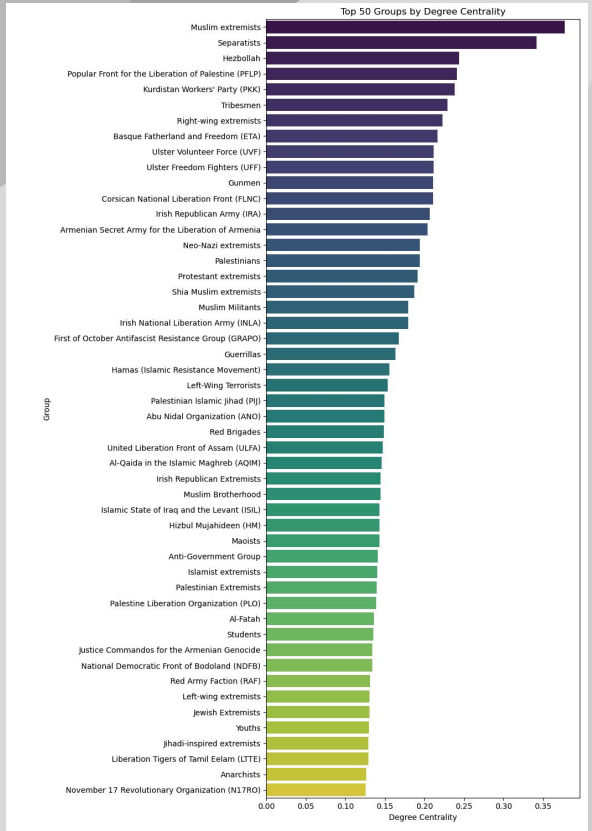
Tactic and Weapon Analysis

- Temporal Shifts in Weapon Use: By examining changes in weapon use over time, I found that certain types, particularly explosives, became more prevalent during periods of global unrest. The multi-line plot from the pivot table offered a visual representation of these shifts and highlighted the adaptability of terrorist groups in their weapon choices.
- Lethality Versus Frequency: I calculated the average fatalities associated with each weapon type to assess their lethality. This analysis provided a deeper understanding of the impact of different weapons, revealing that some less lethal weapons were used more frequently, thus representing a significant threat due to their widespread use.
- Impact of Regional Factors: The data indicated that regional factors, such as local conflicts and the availability of arms, significantly influenced the choice of weapons by terrorists. This regional analysis is crucial for developing targeted counterterrorism strategies that account for local conditions and cultural norms.



Network Analysis

- Unveiling Influential Groups: My analysis leveraged the depth of the GTD to uncover the intricate web of connections between terrorist groups. By focusing on the 'gname' attribute and constructing a detailed network graph, I was able to identify key players within these networks, whose high Degree Centrality suggests they wield considerable influence.
- Identifying Key Relationships: My approach to creating connections within the network was based on the logical premise that groups operating in the same region and year are likely connected. This method allowed me to weave a complex map of relationships, providing a comprehensive view of how terrorist groups may interact or align based on shared objectives or ideologies.
- Strategic Thresholding for Actionable Insights: By applying a thresholding technique to the Degree Centrality values, I was able to focus on the most influential nodes. This strategic choice was crucial for highlighting the groups that could be pivotal in the flow of information, resources, or tactics, offering potential targets for counter-terrorism measures.



Looking to the Future

- Informing Policy and Strategy: Insights from the temporal and spatial analyses can guide policymakers in crafting tailored counter-terrorism strategies that address the unique socio-political dynamics and economic conditions of regions identified as hotspots.
- Resource Allocation: The identification of trends and hotspots through the analysis enables more effective allocation of counter-terrorism resources, ensuring that efforts are concentrated where they are most needed and likely to have the greatest impact.
- Preventive Measures: Understanding the evolution in tactics and weapon usage, such as the rise in vehicular attacks, allows for preemptive security measures and public awareness campaigns, aiming to mitigate the threat before incidents occur.
- Disrupting Terror Networks: By mapping the centrality of nodes within terrorist networks, intelligence and security agencies can target key groups to disrupt the flow of resources and communication, potentially destabilizing larger networks.

Thank you!