Global Terrorism Insights

Comprehensive Analysis of Global Terror Data

Capstone Project Presentation

Team Members:

Mohd. Shehzad Khan | Akansha | Vedanshi Chaudhary | Aseem Gulati | Anand Mehto

Guided by: Dr. Debashish Roy

Great Learning, Gurugram

Introduction

> Terrorism continues to be a **complex global challenge** requiring advanced analytical solutions

> This project leverages the **Global Terrorism Database (1970-2021)** with over 214,000 incidents

> The database contains **135 attributes** for each incident, providing rich data for analysis

> Our goal: Use **machine learning** to predict the terrorist group responsible for attacks

> Data-driven methods show potential to **enhance counter-terrorism strategies**

Problem Statement

Challenge

⚠ Many terrorist attacks remain **unattributed** due to absence of claims

⚠ This delays effective countermeasures and response planning

Our Approach

💡 Build **machine learning models** to predict likely perpetrator groups based on attack characteristics

💡 Address challenges: Class imbalance, noisy data, and complex patterns

Benefits

✔ Support **faster investigations**, better resource allocation, and more effective preventive strategies

Dataset Overview

Global Terrorism Database (GTD)

- **Open-source database** covering terrorist events worldwide

- Time Period: **1970 through 2021**

- Scale: Over **200,000** documented incidents

Key Information

- Coverage: Both **domestic and transnational** terrorist incidents

- Data includes: Date, location, weapons, target type, casualties, and responsible groups

- Contains **135 attributes** for each incident

Source: START (National Consortium for the Study of Terrorism and Responses to Terrorism)

Data Preprocessing

Duplicate Detection

Checked for duplicate values in the dataset; **no duplicate records** were found

Missing Values

Identified null values and retained columns with at least 95% data present

Categorical Variables

Applied **One-Hot Encoding** for low cardinality (<30) and **Label Encoding** for high cardinality (>30)

Feature Engineering

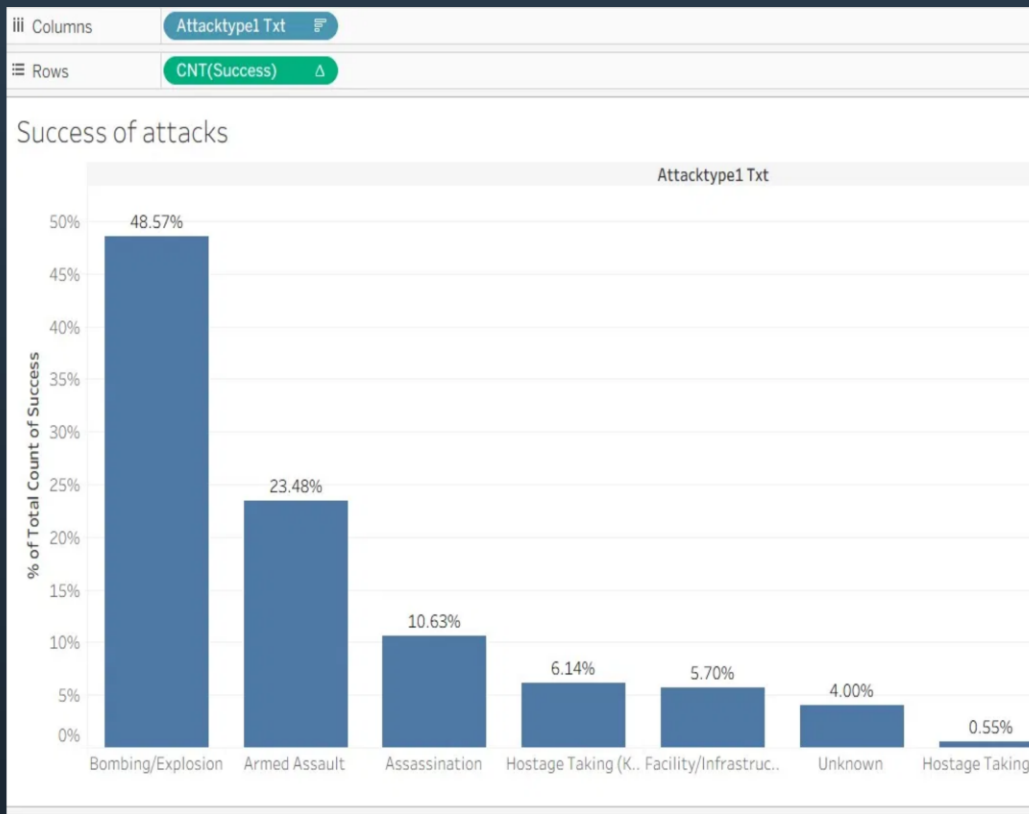Extracted relevant features after removing columns with excessive null values

Multicollinearity Treatment

Removed highly correlated columns to prevent redundancy

Train-Test Split

Created separate datasets for model training and evaluation

Attack Types Analysis

Success of attacks

Key Finding

Bombing/Explosion dominates with **48.57%** of successful attacks, followed by Armed Assault at **23.48%**

Strategic Implications

Terrorists favor bombing tactics for their **high impact-to-effort ratio**, anonymity, and psychological effect
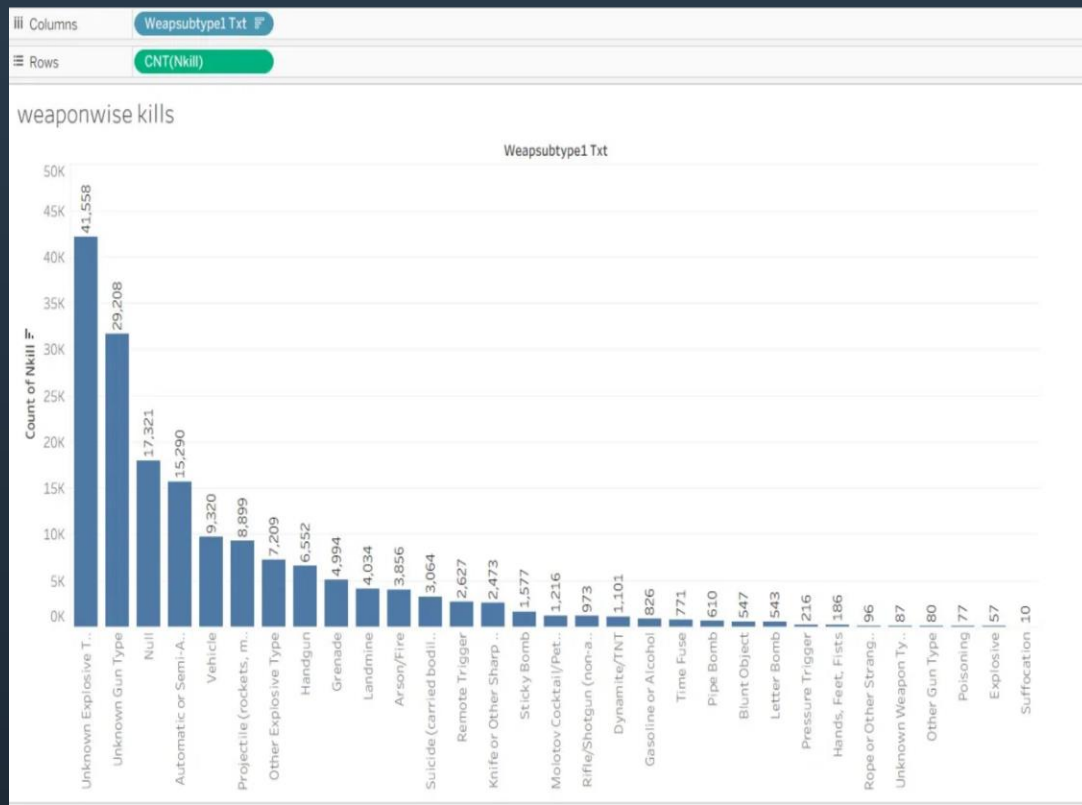
Security Recommendations

Enhance **explosive detection** in public spaces and increase monitoring of **precursor materials**

Trend Analysis

Low percentage of sophisticated attacks (e.g., hijacking at 0.36%) indicates preference for **high-impact, low-complexity methods**

Weapons Analysis

weaponwise kills

Key Finding

Unknown Explosives (41,558) and **Unknown Gun Types (29,208)** cause the most fatalities, highlighting significant intelligence gaps

Weapon Preference Analysis

After unknown categories, **automatic weapons (17,321)** and **vehicle bombs (15,290)** are most lethal, indicating preference for high-casualty methods
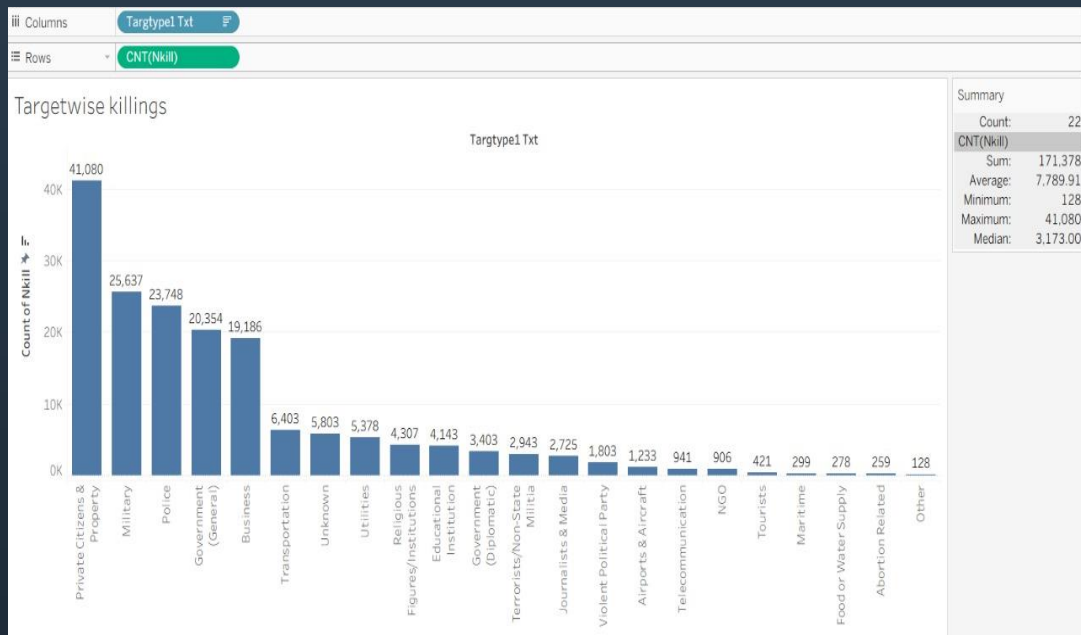
Intelligence Implications

The high "unknown" categories suggest **forensic limitations** in post-attack investigations and need for improved weapon identification techniques

Counter-Terrorism Strategy

Focus on **controlling access** to explosive materials and automatic weapons while

Target Analysis

Targetwise killings

Targtype1 Txt

Summary	
Count:	22
CNT(Nkill)	
Sum:	171,378
Average:	7,789.91
Minimum:	128
Maximum:	41,080
Median:	3,173.00

Bar chart values (Count of Nkill by Targtype1 Txt):
41,080 — Private Citizens & Property
25,637 — Military
23,748 — Police
20,354 — Government (General)
19,186 — Business
6,403 — Transportation
5,803 — Unknown
5,378 — Utilities
4,307 — Religious Figures/Institutions
4,143 — Educational Institution
3,403 — Government (Diplomatic)
2,943 — Terrorists/Non-State Militia
2,725 — Journalists & Media
1,803 — Violent Political Party
1,233 — Airports & Aircraft
941 — Telecommunication
906 — NGO
421 — Tourists
299 — Maritime
278 — Food or Water Supply
259 — Abortion Related
128 — Other

Caption

This graph shows number of killings according to Target Type. From this we can infer that when target was privatecitizens & property the killings were higher. We see that least killings were in ohter target types that are not mentioned. Abortion realted killings are the killings of doctors and medical staff who are involved in abortion related activites, by the protestors having ideology against abortion . Utilities killings indicates that attack on amenites like power house, dams, petrol pumps etc.

Key Finding

Private Citizens & Property are overwhelmingly targeted with **41,080 fatalities**

Target Hierarchy

Top five targets (Private Citizens, Military, Police, Government, Business) account for **76% of all fatalities**
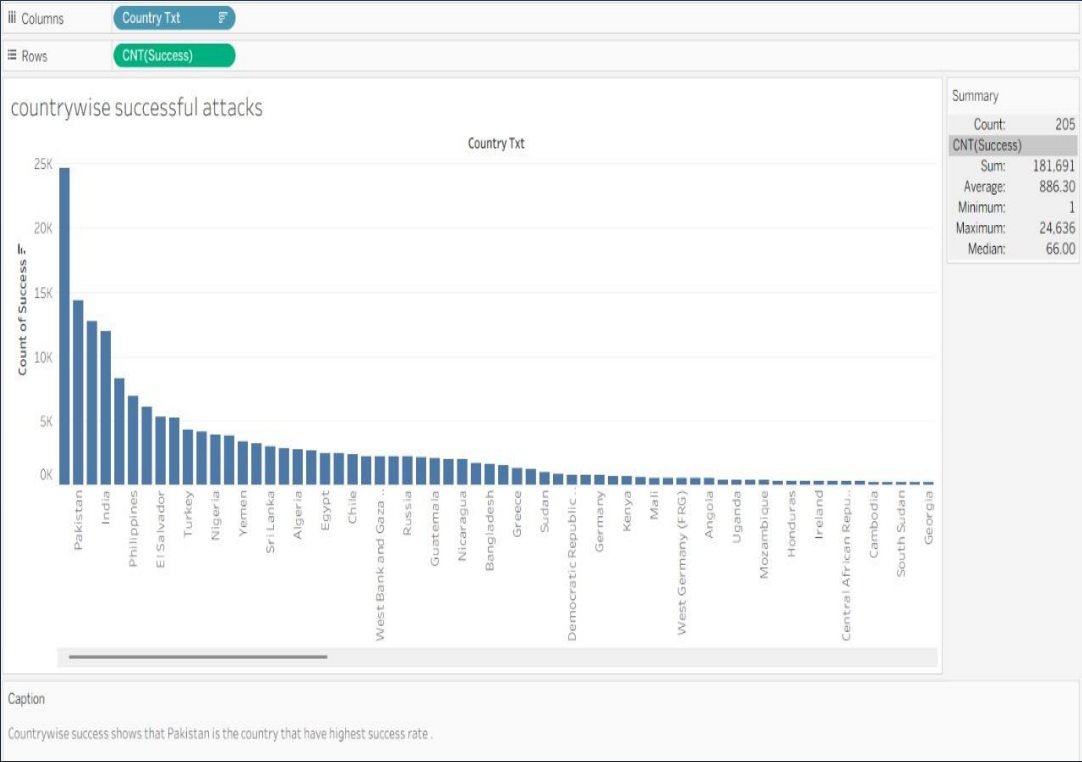
Strategic Pattern

Targeting civilians creates **maximum psychological impact** while attacking authority symbols serves ideological goals

Security Implications

Public spaces require enhanced **soft-target protection** strategies

Geographic Analysis

countrywise successful attacks

Country Txt

Caption

Countrywise success shows that Pakistan is the country that have highest success rate .

Key Finding

Pakistan leads with the highest number of successful attacks, followed by **Iraq**, **India**, and **Afghanistan**

Regional Concentration

South Asia and Middle East regions account for the **majority of successful attacks**, indicating geopolitical hotspots South

Success Rate Analysis

Afghanistan shows a particularly high success rate with **12,731 successful attacks**, suggesting operational advantages for terrorist groups in this region

Security Implications

International counter-terrorism efforts should prioritize **regional cooperation** in these high-risk areas

Model Evaluation

Legend: Accuracy (blue), Precision (green)

Chart showing Accuracy and Precision for: Random Forest, ADA Boost, Decision Tree, Logistic Regression, XGBoost. Y-axis from 0 to 1.0.

Best Performer

Random Forest achieved highest accuracy (0.8167) and precision (0.8080), demonstrating superior classification ability.

Ensemble Advantage

Ensemble methods (Random Forest, ADA Boost) consistently **outperformed** single-model approaches, highlighting the benefit of combining multiple decision trees

Performance Analysis

The 81.67% accuracy of Random Forest indicates strong but not perfect prediction capability, suggesting room for further model optimization

Application Implications

Models can serve as **investigative tools** to narrow down potential perpetrators, but should be used alongside traditional intelligence methods

Conclusion

Key Findings & Implications

Bombing/Explosion (48.57%) is the dominant attack method, requiring specialized prevention strategies

Pakistan, Iraq, and Afghanistan emerge as critical regions requiring focused counter-terrorism efforts

Random Forest model achieved 81.67% accuracy in predicting terrorist group attribution

Private citizens are the most targeted group (41,080 fatalities), highlighting the need for public space security

High numbers of **unknown weapon categories** indicate critical intelligence gaps in forensic capabilities

Data science demonstrates significant potential for **improving counter-terrorism intelligence** and resource allocation

Future Scope

⚙ Advanced Model Optimization

Apply **hyperparameter tuning** and explore advanced ensemble techniques like xgboost to improve prediction accuracy

🗄 Feature Enrichment

Integrate additional datasets such as **political stability indices**, economic indicators, and conflict databases

🖥 Real-Time Prediction Systems

Deploy **cloud-based dashboards** that can take live incident data and predict likely perpetrators instantly

🧠 Deep Learning Applications

Experiment with **sequence models (LSTMs)** for temporal prediction and graph neural networks for analyzing terrorist organization connections