**Principles of Business Data Mining**

**Project Report**

**Terrorism In the United States**

A group of men holding guns

Description automatically generated

Sanika Acharekar

**INDEX**

Business background 3

Business problem 4

Objectives 5

Dataset description 6

Data preprocessing 8

Expected Initial results 8

Data cleaning 10

Data visualization and prediction techniques 11

Data modelling 17

Unconventional Insights 29

Managerial Insights 31

Conclusion 32

**Business Background:**

In this project, we will be studying a sample of a dataset that collects information about the known terrorist incidents that have taken place in the USA in the last 50 years, with detailed information and variables for each one. The reason we chose this topic for the project was that, between all the different datasets we found, this was complete data we had, with enough information, cases, and variables to study, and also was interesting enough to extract some conclusions.

In response to this complex security challenge, the United States government has assembled a team of highly skilled Security Analysts. Their mission is to thoroughly analyze this extensive dataset, extracting valuable insights that are crucial for shaping strategic plans and strengthening risk management practices to reduce terrorism. The goal of this endeavor is to reinforce the security of our nation, making it more resistant to threats and better prepared to face adversity.

This initiative is significant as it equips decision-makers with a potent tool to enhance security measures, safeguard assets, and protect lives. Using data mining techniques, we aspire to provide businesses and government agencies with practical knowledge vital for strategic planning and risk mitigation. Ultimately, our project will contribute to creating a safer and more resilient nation, while mitigating the economic and societal impact of acts of terrorism.

**Business Problem:**

Our report will analyze the extent of attack severity across all U.S. states using historical data. Subsequently, we will focus our analysis on areas where the attacks exhibited the highest severity level. Depending on the number of people killed, number of people wounded, target type, etc. We would like to determine the success of the attack and advise the government on which location to safeguard first. We would also like to give insights on what regulatory policies the government should contemplate concerning the utilization of weapons, including biological weapons, chemical weapons, and explosive weapons.

**The objectives of this project include:**

* **Objective 1:** Develop a predictive model to determine the success or failure of attacks, aiming to provide valuable insights that could assist the government. This model takes into consideration factors such as target type, the number of casualties (both killed and wounded), and the method of attack. The insights derived from this analysis can offer valuable assistance to the government in its decision-making processes.

Target Variable: success – 0: Attacks Unsuccessful (18%)

1: Attacks Successful (82%)

Model: Logistic Regression / Classification

Prediction: If a terrorist attack will be successful or unsuccessful depending on the various factors.

* **Objective 2:** What is the relationship between different types of weapons (such as explosives, chemicals, and biological weapons) and the severity of terrorist attacks, measured by the number of people killed or wounded (nkill)? How can this analysis of data on terrorist attack patterns inform government decisions on implementing more stringent regulations for specific weapon types to enhance security measures in high-severity areas?

Target Variable: nkill

Model: Logistic Regression

Prediction: Depending on the level of severity of the weapon type, we can predict the number of people killed and inform the government on implementing more stringent regulations on specific weapons.

**Source of the dataset:** <https://www.kaggle.com/START-UMD/gtd>

**Description Of Dataset:**

The Global Terrorism Database (GTD) is a publicly accessible database that contains information on terrorist incidents occurring worldwide between 1970 and 2017. It covers both domestic and international acts of terrorism, comprising over 180,000 attacks. This database is maintained by researchers at the National Consortium for the Study of Terrorism and Responses to Terrorism (START), based at the University of Maryland.

Each incident recorded in the GTD is detailed with various variables, such as the type of weapon used, the number of casualties, and fatalities. These variables enable us to analyze trends like the most commonly used weapons and which cities have experienced the highest number of lethal incidents.

At first, we had 135 columns with numerous repeated details, and some of them were not even being utilized. Consequently, we refined the data by filtering out unnecessary information and introduced new columns, including one for total casualties and another for the date.

In our data mining project, we've focused on narrowing the database's scope to the United States, eliminating entries from other regions. This decision was made because the dataset includes specific attributes related to U.S. citizen casualties and injuries, which are not as relevant for other countries or regions. For U.S. terrorism data, we have a dataset that includes 2,835 attacks spanning from 1970 to 2017.

|  |  |
| --- | --- |
| Number of Observations: | 2835 |
| Number of variables: | 26 |
| Number of numerical variables: | 12 |
| Number of binary variables: | 5 |
| Number of categorical variables: | 7 |

**List of Variables are below:**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Meaning** | **Type** |
| id | Identifier, unique for each row | Num. |
| date | Date of the attack | Date |
| provstate | State where the terrorist attacks were perpetrated | Cat. |
| city | City where the terrorist attacks were perpetrated | Cat. |
| latitude | Coordinates of the attack | Num. |
| longitude | Coordinates of the attack | Num. |
| doubterr | It indicates whether there is doubt or not about the attack being a terrorist one | Bin. |
| success | It indicates whether the terrorist attack was a success or not | Bin. |
| attacktype1\_txt | This field captures the general method of attack and often reflects the broad class of tactics used. | Cat. |
| targtype1\_txt | It captures the general type of target | Cat. |
| nperps | Total amount of terrorists participating in the incident | Num. |
| nperpcap | Total amount of captured perpetrators | Num. |
| claimed | Whether a group or person claimed responsibility for the attack | Bin. |
| claimedmode\_txt | Mode used by the claimant to claim responsibility | Cat. |
| weaptype1\_txt | Type of the (main) weapon used in the attack | Cat. |
| nkill | Total confirmed fatalities of the attack | Num. |
| nkillus | Total confirmed U.S. citizen fatalities of the attack | Num. |
| nkillter | Total confirmed perpetrator fatalities of the attack | Num. |
| nwound | Total confirmed non-fatal injuries the attack | Num. |
| nwoundus | Total confirmed U.S. citizen non-fatal injuries of the attack | Num. |
| nwoundte | Total confirmed perpetrator non-fatal injuries of the attack | Num. |
| propvalue | U.S. dollar amount of total damages | Num. |
| INT\_IDEO | It indicates whether a perpetrator group attacked a target of a different nationality | Bin. |
| INT\_MISC | It indicates if the attack was miscellaneous international | Bin. |
| president\_party | Governing party of the U.S. at the moment of the attack | Cat. |

**Data Pre-Processing Steps:**

In our data mining project, we initially extracted and focused on relevant data pertaining to United States Terrorism from the original dataset, which consisted of a total of 135 columns. During this process, we observed that certain columns contained extremely rare occurrences. As a result, we made the decision to eliminate these infrequent columns from the dataset. This not only streamlined the dataset but also aligned it more closely with our specific area of interest, which is U.S. Terrorism. Before reading a file in Python, preliminary data cleaning was conducted in Excel, but the majority of the cleaning process was executed using Python. Numerous incomplete rows and variables were removed during this phase.

During our data preprocessing journey, we engaged in essential procedures such as consolidating and merging data, applying filters, selecting relevant variables, and addressing missing data. These systematic steps played a pivotal role in enhancing and optimizing the dataset, paving the way for effective data mining and subsequent analysis tasks.

**Basic Descriptive Statistics:**

During Basic Descriptive Statistics step, we used python programming language to conduct a thorough exploration of fundamental statistical measures. This involved generating visualizations and descriptive statistics to obtain a comprehensive understanding of the dataset.

**Handled Missing Value:**

By using data description for particular we systematically identify and pinpoint any missing values present within each column. Our approach to addressing these missing values was characterized by a meticulous, case-by-case analysis. For each instance of missing data, we engaged in thoughtful reasoning to determine the most appropriate and contextually relevant treatment strategy. We have imputed data by using mean/median/mode/ specific values.

**Outliers:**

Eliminated outliers, ensuring the dataset was cleansed and values remained within their anticipated ranges.

**Addition Of New Variables:**

Generated new variables using existing variables such as number of causalities and date.

**Data Filtering:**

We applied a data filter to concentrate on the United States region.

**Expected Initial Result:**

1. **Bar Chart: Type of weapons used in the incidents.**

Explosives and Incendiary are the most used weapons.

A graph of blue bars with white text

Description automatically generated

1. **Bar plot: Terrorist attacks across the States.**

California and New York are the most attacked states.

A graph with green bars and white text

Description automatically generated

1. **Pie Chart:**

As we see in the pie chart there are multiple methods of claiming an act with different occurrence rates.

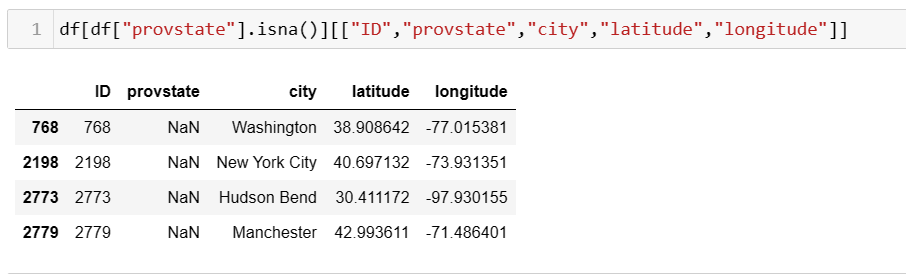
A pie chart with text on it with Crust in the background

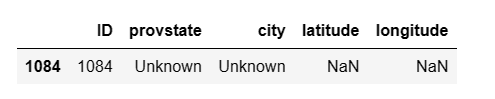
Description automatically generated

**Data cleaning:**

During the data cleaning process, several key transformations were applied to enhance the dataset's integrity and coherence. The 'nperpcap' and 'nperps' columns were converted into numerical formats, and any missing values were consistently replaced with 'NA' to maintain uniformity. The 'claimed' column, initially represented in binary, was transformed into a more interpretable format using 'True' or 'False,' with missing values replaced accordingly. In the case of 'claimmod,' missing values were addressed by replacing them with 'NA' to ensure completeness. Other columns across the dataset had missing values uniformly replaced with 'NA' for consistency. Additionally, the 'int\_ideo,' 'int\_misc,' and 'success' variables were converted into a binary format with 0 and 1. Notably, irrelevant data such as 'gname,' 'nationality,' and 'state' government information was removed from the dataset to focus on the relevant variables for analysis. These adjustments collectively refined the dataset, providing a more standardized and streamlined foundation for subsequent analytical endeavors.  
  
**Provstate:**

During the summarization of the "provstate" data, it came to our attention that there is one unknown value and four NaN (Not a Number) values. Initially, our approach involves adding values for those "provstate" entries that have a corresponding city. However, if the city value is null, making it impossible to determine the state in which the incident occurred, we have decided to remove that particular row from the dataset.



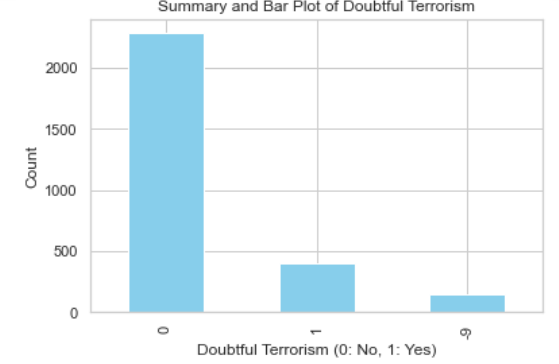


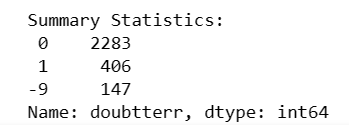
**City:**

Fourteen rows lack city information. Consequently, we populated these rows with external data using the latitude and longitude columns.



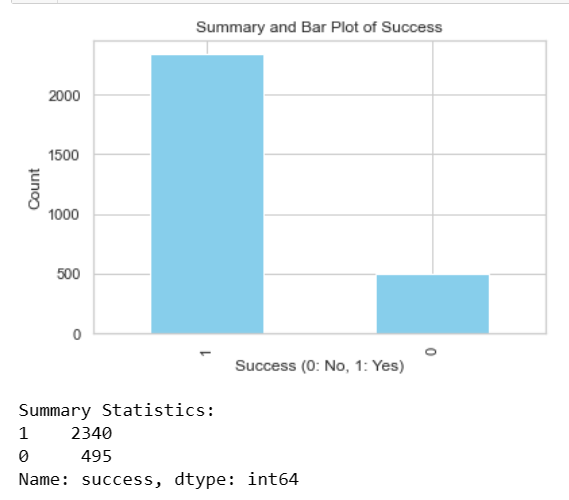
**Doubterr:**  
A total of 147 rows contained the value -9, which was replaced with a NaN (Not a Number) value.





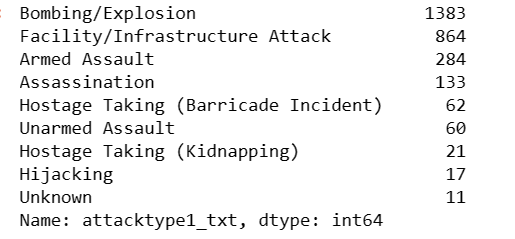
**Success:**

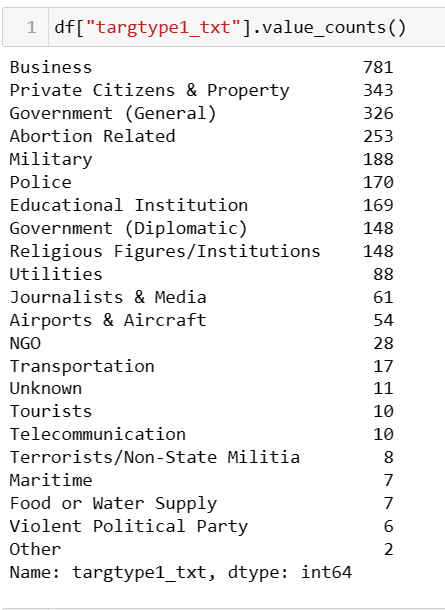
No missing value handling was necessary for the "success" variable, as the data was already clean.



**Attacktype\_txt: and Target\_type**

We imputed a total of 11 unknown values with the mode value**.**





**Nationality:**



Few names are simplified and converted into categories. Following are the names which are simplified:

'Multinational': 'International',

'Virgin Islands (U.S.)': 'United States',

'West Germany (FRG)': 'Germany',

'Branch': 'United States'

**Gname, Weapontype:**

Replaced unknown value with "Other”.

**Killed:**

Null values of killed is replaced by median.

**Total\_casualities:**  
Added this column by adding nkill and nwound column to find number of people affected by the attacks.

**Data Visualization and Prediction Techniques:**

We will be using Tableau for Data Visualization and mentioned below are a few prediction techniques:

* **Attacks across the United States:**

The below chart shows the geographic distribution of terrorist attacks across a region of the United States. On the map, individual attacks are represented by colored dots, where each dot's color indicates the success of the attack: green dots represent successful attacks (True), and red dots signify unsuccessful ones (False). This visualization helps in identifying geographical patterns in the success rates of attacks. For instance, one might observe clusters of green dots, indicating areas with higher rates of successful attacks, which could suggest vulnerabilities or specific tactics that are effective in those regions. Conversely, red dots would highlight areas where attacks tend to fail, which correlate with stronger security measures or other protective factors.

By analyzing this data, the government or security agencies could identify hotspots that require more attention or could study the characteristics of successful attacks to develop better counter-terrorism strategies. Furthermore, this visualization could be used for resource allocation, emergency response planning, and security policy development.

A map of the united states

Description automatically generated

* **A Dual-axis line chart with two different variables over time: The number of people killed (Nkill) and the number of people wounded (Nwound) in various incidents:**

The chart below is a dual-axis line chart that represents two different variables over time: the number of people killed (Nkill) and the number of people wounded (Nwound) in various incidents.

There are two lines, each representing one of the variables: The orange line represents the number of people killed over the years. The blue line represents the number of people wounded over the years.

The peaks in the lines represent years with particularly high numbers of killed or wounded individuals. These correspond to years with more severe incidents or a greater number of incidents resulting in casualties.

A graph with blue and orange lines

Description automatically generated

* **The bar chart displays "Total Casualties," a combination of confirmed fatalities and non-fatal injuries from terrorist attacks, categorized by attack type:**

The categories of attack type include Bombing, Unarmed Assault, Armed Assault, Hijacking, Facility/Infrastructure Attack, Hostage Taking, Assassination, and other forms of Hostage Taking. From the graph, we can conclude that Bombing and Armed Assault have the highest number of total casualties, followed by Unarmed Assault, and then Hijacking. The other categories show significantly fewer casualties in comparison.

**In terms of government decisions on security measures:**

The data suggests that a focus on controlling and monitoring explosives and firearms could be beneficial, as they are associated with the highest number of casualties. For explosives, the government might consider regulations on the sale and distribution of materials that could be used to make bombs.

For firearms, measures might include stricter gun control laws, enhanced background checks, and limitations on the types of firearms available to the public. This analysis helps us prioritize which types of attacks to prepare for, such as improving emergency response protocols for bombings and armed assaults since they are associated with higher casualties.

A graph of different colored bars

Description automatically generated

* **Success rates associated with different attack types:**

The chart shown below is a horizontal stacked bar chart, which visualizes the success and failure of terrorist attacks across states. Each bar corresponds to a location and is divided into two segments:

1. Green Segment: Represents the number of successful events in the location. The length of the green portion is proportional to the count of successes.

2. Red Segment: Indicates the number of unsuccessful events. Similarly, the length of the red portion corresponds to the count of failures.

The total length of each stacked bar represents the total number of events (both successful and unsuccessful) in each location. The locations are listed on the y-axis (labeled as 'Provstate Set'), and the counts of success or failure are on the x-axis. This visualization allows us to quickly assess which locations have the highest number of events and to compare the success rate versus the failure rate within each location. For instance, we can see that California has a high number of successful events (477) and a certain number of unsuccessful ones (102). In contrast, Puerto Rico has fewer events in total, with 166 successes and 60 failures.

A graph with green and red squares

Description automatically generated

* **Outcome by Target Type:**

The graph below shows a stacked bar chart titled "Outcome by Target Type ". This chart displays the success or failure of various incidents, which are terrorist attacks, grouped by the type of target they were aimed at. Each bar in the chart represents a different type of target, such as "Private Citizens & Property," "Military," "Police," etc. Within each bar, there are two colors:

Green: Indicates the number of successful incidents related to the target type.

Red: Shows the number of unsuccessful incidents related to the target type.

The height of each color within the bar corresponds to the count of each outcome, and the entire height of the bar represents the total number of incidents both successful and unsuccessful for that target type. The chart is organized such that the target types are arranged along the x-axis, and the y-axis likely represents the count of incidents. From this chart, one can identify which target types are more frequently attacked and which ones tend to have higher success rates for the attackers. For example, if one target type has a particularly tall green portion, it indicates that incidents aimed at that target type are often successful. Conversely, a tall red portion would indicate many unsuccessful attempts. This visualization helps to quickly identify the effectiveness of security measures across different target types. Such insights are crucial for resource allocation, risk assessment, and strategizing security enhancements.

A graph of a growing graph

Description automatically generated with medium confidence

* **Total Casualties (sum of the number of people killed or wounded) resulting from terrorist attacks in a set of states, broken down by the type of weapon used in the attacks.**

A graph of different colored squares

Description automatically generated

The height of the bars indicates the severity of attacks, measured by the number of total casualties. Each bar is segmented into colors representing different types of weapons, such as biological, explosives, vehicles, firearms, and incendiary. This segmentation allows us to see which types of weapons have been associated with the most casualties in these states. By examining the data at the state level, we identify which states have experienced higher casualties and which types of weapons have been most lethal in those areas. For example, in Oregon, the chart shows a very high number of casualties associated with biological weapons, which is significantly different from the other states shown. Weapons that are most associated with high casualty numbers can help inform government policy. For example, if explosives are consistently leading to high casualties across multiple states, this might prompt a review of regulations surrounding the sale and storage of explosive materials. Similarly, if a particular state like Oregon shows a high number of casualties from biological weapons, it may lead to specific biosecurity measures in that state.

**Public Awareness:** Beyond government policy, such a chart can also be used in public safety campaigns to raise awareness about the most common threats and encourage preparedness in high-severity areas.

**Data Modelling:**

1. **Objective:** Develop a predictive model to assess the likelihood of success or failure of attacks. The goal is to provide valuable insights that can assist the government in making informed decisions related to security and counterterrorism efforts. The predictive model considers various factors that contribute to the outcome of an attack. These factors include number of killed, number of wounded, types of attacks, targets, weapon, etc. The methods used for prediction are Logistic Regression and Classification.

A screenshot of a computer

Description automatically generated

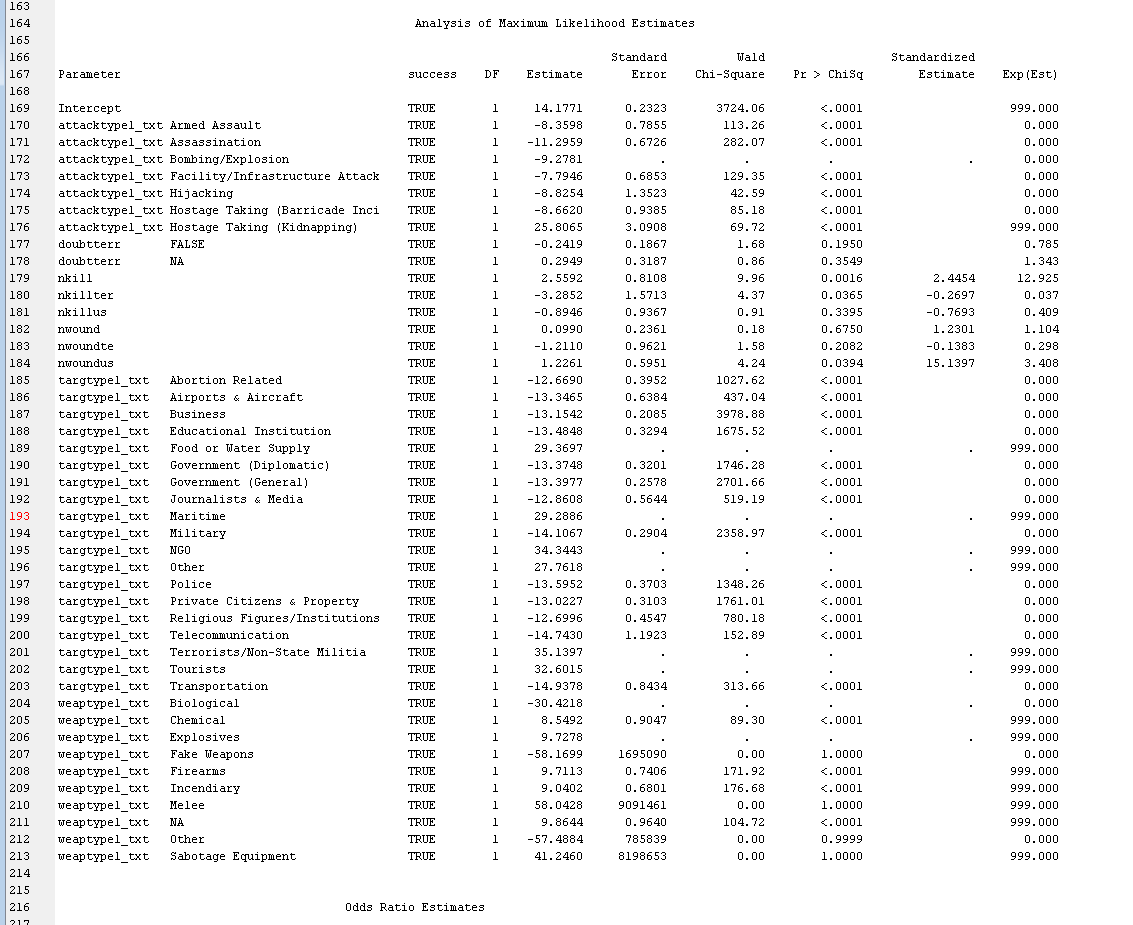
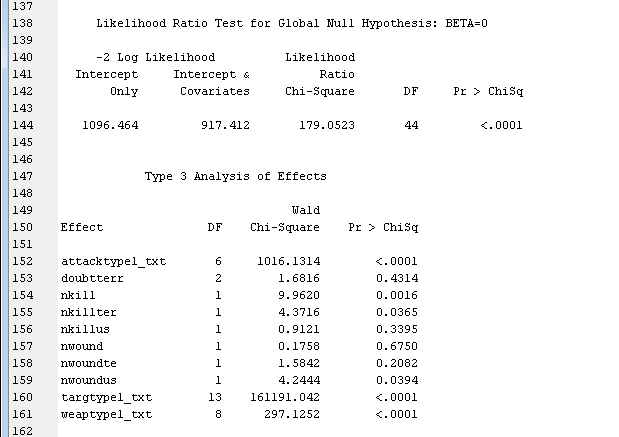
* **Approach 1 - Logistic regression:**

The logistic regression analysis reveals important numerical insights into the dataset's predictors and their impact on the likelihood of success. The predictors `attacktype1\_txt`, `targtype1\_txt`, and `weaptype1\_txt` are statistically significant. Coefficients with large absolute values indicate a strong effect on the probability of 'success'. Certain types of targets and attack types have large coefficients, indicating they are strong predictors of 'success'. For Example, the coefficients for 'Armed Assault' (-8.3598), and 'Assassination' (-11.2959) suggest a strong negative association with ‘success’ is True. The coefficient for 'Hostage Taking (Kidnapping)' (25.8065) provides a strong positive indication of their associations with the log odds of success. Transforming these coefficients into odds ratios further quantify these associations, with the odds ratio for 'Armed Assault' being a very small number (exp (-8.3598)), indicating lower odds of success for such attacks. In contrast, 'Hostage Taking (Kidnapping)' boasts a considerably higher odds ratio (exp (25.8065)), emphasizing a much greater likelihood of success for this specific attack type. The equation for the logistic regression model:

**ln(success) = 14.17 - 28.14 attacktype + 25.78 targetype + 0.1 weapontype**

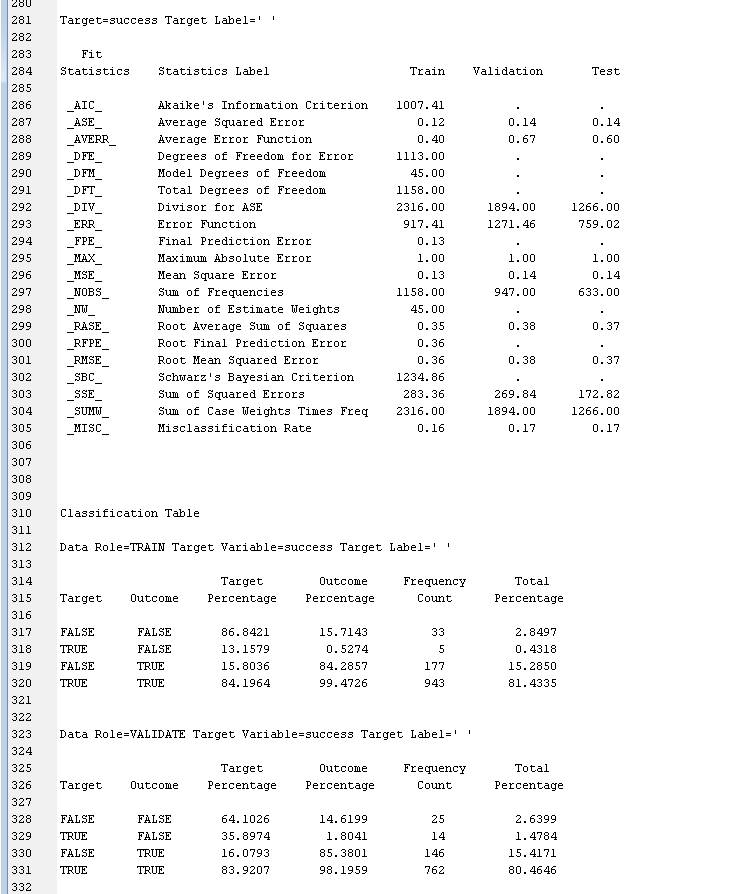
The model's overall fit is supported by a significant chi-square statistic, affirming the collective significance of predictors. Several predictors, including attack and target types, exhibit statistical significance, reinforcing their numerical relevance in predicting success.

The misclassification rate of 0.16 in the training dataset and 0.17 in the validation dataset quantifies the model's accuracy, predicting success correctly approximately 83-84% of the time.

****

In the above image, some of the significant coefficients are highlighted.

In the image below, the output statistics are provided which shows that the model is a good fit.

****

* **Approach 2 - Classification:**

The decision tree analysis provides valuable insights into the factors influencing the prediction of 'success' in the dataset. Notably, `attacktype1\_txt` and `weaptype1\_txt` emerge as the most crucial variables, contributing significantly to the model's performance. Although other variables like `nkillus`, `nwoundus`, `targtype1\_txt`, `nkill`, `nkillter`, and `nwound` have lesser importance, they still play a role in the tree's decision-making process. Examining specific nodes, Node 12 stands out as a deeper node with a high percentage of 'TRUE' outcomes, while Node 3, close to the root, indicates the significance of the initial split.

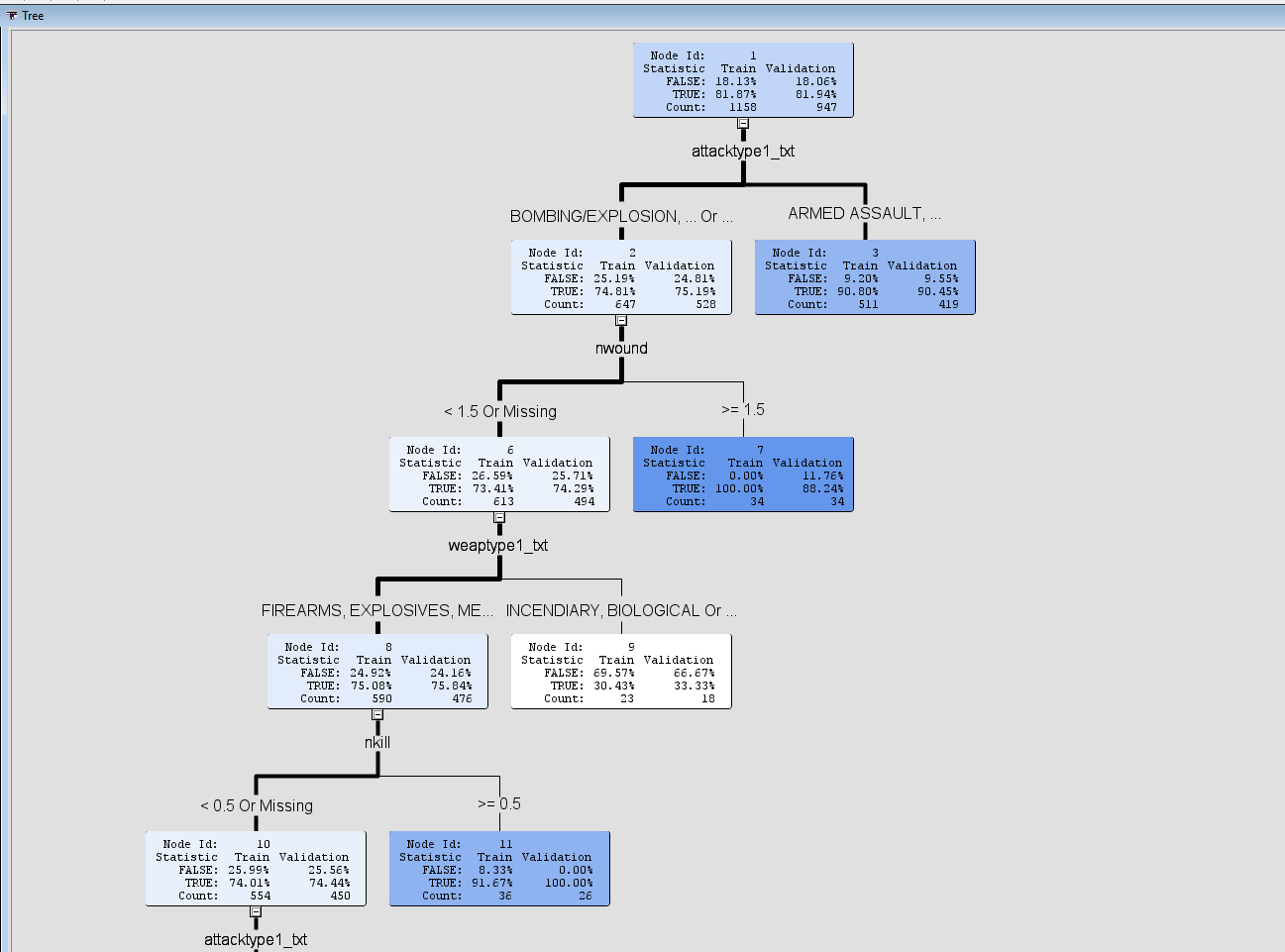
The model demonstrates good predictive performance, with a misclassification rate of 0.17 for the training set and 0.16 for the validation set, highlighting relatively low error rates. The average squared error (ASE) and root average squared error (RASE) consistency across training and validation suggest a stable and not overfitted model. The classification table details correct and incorrect predictions, with a slightly better performance on the validation set. Assessment score rankings and distribution reveal the model's confidence, especially in predicting 'TRUE' outcomes. In conclusion, the decision tree effectively predicts 'success,' particularly highlighting the importance of certain attack types and weapons, while maintaining consistency and stability across training and validation sets.

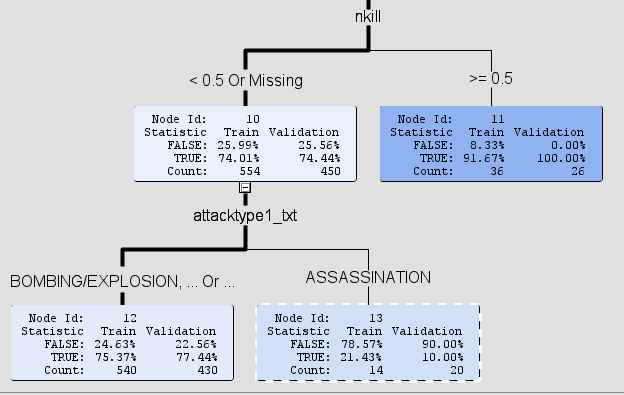
The image below shows the statistical output for the decision tree created.

**A screenshot of a document

Description automatically generated**

The image below is decision tree for success target variable





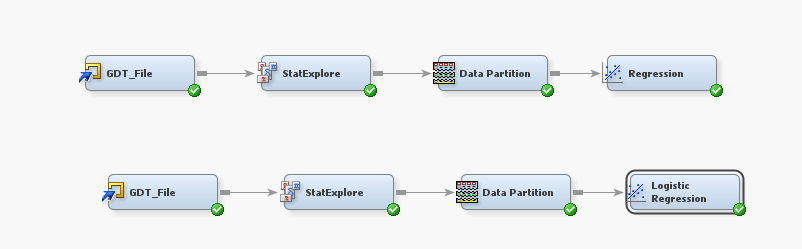
**Decision Tree for the model using SAS Enterprise Miner**

* **Model Selection:**

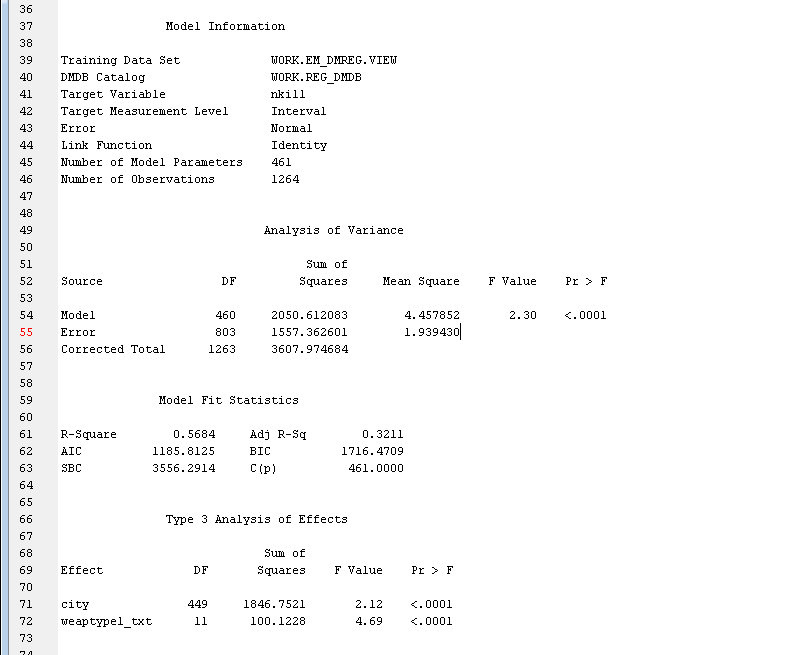
The decision tree is effective at predicting 'success', with certain types of attacks and weapons being strong predictors. The consistency of performance metrics across the training and validation sets suggests that the model generalizes well. However, the model appears to be more confident in predicting 'TRUE' outcomes than 'FALSE' ones, which could be due to class imbalance or other factors in the dataset.

The logistic regression model appears to be a good fit for predicting the target variable 'success'. It is significant, with several significant predictors indicating their importance in the model. The model performs well in terms of classification accuracy, as indicated by the misclassification rate and classification tables for both training and validation datasets. The odds ratios suggest that certain attack types, target types, and weapon types are particularly influential in predicting a successful attack. Hence, we have selected the Logistic Regression model for this objective.

1. **Objective:** Develop a predictive model to analyze the relationship between different types of weapons (explosives, chemicals, and biological weapons) and the severity of terrorist attacks, measured by the number of people killed (nkill). The aim is to inform government decisions on implementing more stringent regulations for specific weapon types and enhancing security measures in high-severity areas.



* **Approach 1: Linear Regression**



The regression model had 461 parameters and was applied to 1264 observations. The model explained 56.84% of the variance in the target variable (R-Square = 0.5684), with an adjusted R-square of 0.3211, indicating a moderate fit of the model to the data. The city and weapon type were significant predictors of the number of killings (nkill) in the dataset. The F-values for city and weaptype1\_txt was 2.12 and 4.69 respectively, both with p-values less than 0.0001, indicating strong statistical significance.

**nkill = β0 + β1 weapontype + β2 cities**

|  |  |  |
| --- | --- | --- |
| Parameter | Coefficient | Significance (Pr > |t|) |
| City Littleton | 13.81 | <.0001 |
| City Oak creek | 5.81 | <.0001 |
| City Orlando | 24.38 | <.0001 |
| City Roseburg | 8.81 | <.0001 |
| City San Bernardino | 15.78 | <.0001 |
| Weapon Firearms | 0.94 | <.0001 |

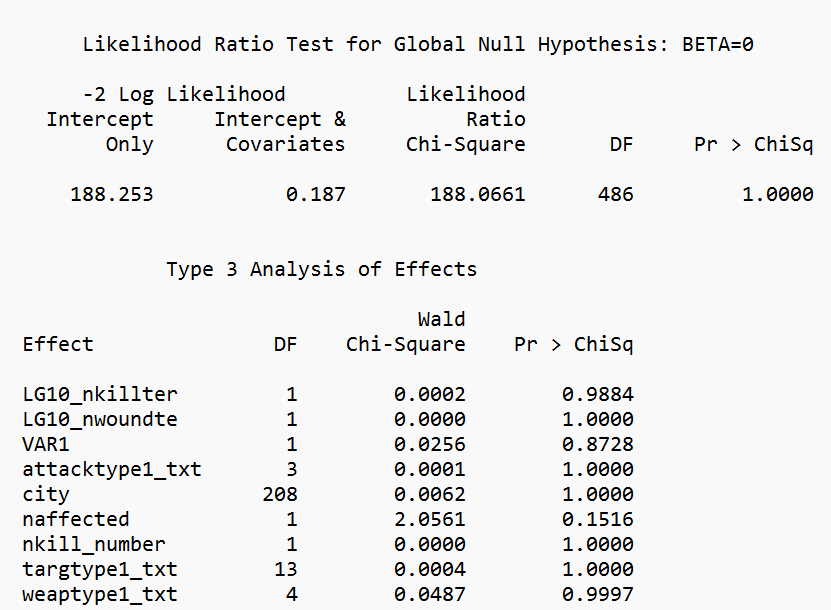
**Table with significant parameters**

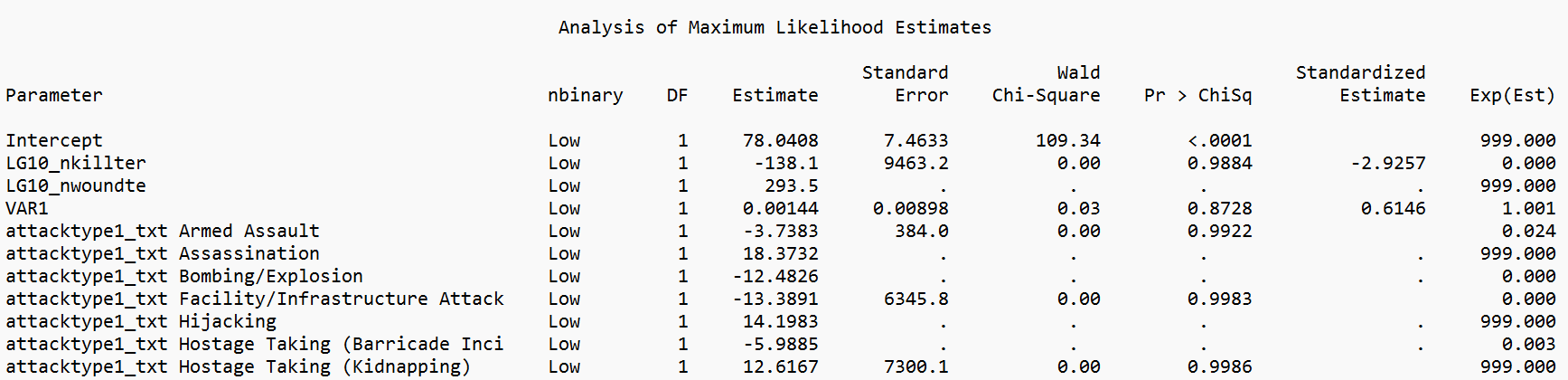
Firearms had a significant positive effect on the target variable (Estimate = 0.9378, p < 0.0001), suggesting that incidents involving firearms tend to have higher numbers of killings. Other weapon types like Biological, Chemical, Explosives, and Incendiary showed no significant effect on the target variable as their p-values were relatively high, indicating that these weapon types do not significantly affect the number of killings in the model.

In conclusion, our regression analysis indicates that the location of an incident (city) and the type of weapon used are significant predictors of the number of killings in terrorist attacks. Specifically, incidents involving firearms are associated with a higher number of killings, underscoring the lethal potential of this weapon type. The model's moderate fit suggests that while these factors are important, other unaccounted variables might also influence the outcome. This analysis highlights the importance of focusing on firearm-related incidents and considering geographical factors in predictive and preventive measures against terrorism-related fatalities.

* **Approach 2: Logistic Regression**

The model’s target is nkill. The variable nkill is binned into Low and High. The order is descending, indicating that "Low" is considered the reference category against which "High" is compared. The model predicts the probability of nbinary being "Low" or "High", and it calculates residuals (the difference between the observed and predicted values) for both levels. The logistic regression model is designed to predict the probability of an event being classified as "Low" or "High" based on various inputs. The training dataset has 1245 observations. The model has 490 parameters, which is quite a large number and could indicate a complex model that runs the risk of overfitting. The model’s output shows that variable and the model is not significant. The Likelihood Ratio Test for the global null hypothesis (Beta=0) showed a chi-square of 188.0661 with 486 degrees of freedom, indicating no significant overall effect. The Type 3 Analysis of Effects indicated that most variables (like 'LG10\_nkillter', 'city', 'attacktype1\_txt') did not have significant individual effects. the intercept for 'nbinary' being 'Low' has an estimate of 78.0408 with a p-value <.0001, indicating a significant effect. However, many city variables show no significant effect with p-values of 1.0000. Overall, the model seems to have a large number of parameters relative to the sample size, with many parameters not showing significant effects. The optimization process appears comprehensive, but the lack of significant effects for many variables may suggest overfitting or the need for model refinement.





* **Model Selection:**

The objective is to develop a predictive model that assesses the relationship between different types of weapons and the severity of terrorist attacks, with the severity measured by the number of people killed (nkill). The model should help inform government decisions on weapon regulations and security measures. Logistic regression categorizes the outcome into "Low" and "High", which is less directly aligned with the objective since it does not quantify the severity. The model has fewer parameters (41), suggesting a simpler model which might be less prone to overfitting. The imbalance in the target categories (1227 "Low" and 18 "High") can bias the model towards predicting the majority class.

Linear Regression predicts the continuous variable nkill directly, which aligns with the objective's requirement to measure severity. The model has an R-squared of 0.5684, suggesting that approximately 56.84% of the variance in nkill is explained by the model's inputs. It includes both nominal and interval inputs and has a significant F value (<.0001), indicating the model is statistically significant. The linear regression model is more aligned with the stated objective because it quantitatively predicts the number of people killed, which is the direct measure of severity needed for policy decision-making. The logistic regression model's binary categorization may oversimplify the severity of attacks, which is a nuanced and continuous outcome.

**Unconventional Insights:**

A pie chart with text on it with Crust in the background

Description automatically generated

The initial visual analysis reveals that ‘Bombing’ has the highest impact, closely followed by ‘Unarmed assault.’ These insights provide a snapshot of the severity of different attack types, prompting further investigation into the predictors of attack success through data modeling. Upon conducting a logistic regression analysis, a notable revelation emerges – the coefficient for 'Hostage Taking (Kidnapping)' is remarkably high at 25.8065. This substantial coefficient signifies a strong positive association with the likelihood of success for this specific attack type. In contrast, 'Armed Assault' exhibits a low odds ratio (exp (-8.3598)), indicating lower odds of success. This newfound understanding prompts a reconsideration of 'Hostage Taking (Kidnapping)' as a critical area for intervention and regulatory focus. Armed with insights from the data modeling, governments can now implement more targeted and effective regulations to mitigate the success of 'Hostage Taking (Kidnapping)’ incidents. This might involve strengthening security measures around potential targets, enhancing intelligence efforts, and collaborating with international partners to counteract this specific threat.

A graph with blue squares

Description automatically generated with medium confidence

The data modeling process has revealed a significant positive effect of firearms on the target variable, with an estimated coefficient of 0.9378 and a p-value less than 0.0001. This statistical finding implies that incidents involving firearms tend to result in higher numbers of killings. However, this conclusion stands in apparent contradiction to initial visualizations, raising the need for a more varied interpretation.

Contrary to firearms, other weapon types, including Biological, Chemical, Explosives, and Incendiary, demonstrated no significant effect on the target variable in the statistical model. The elevated p-values associated with these weapon types suggest that they do not significantly influence the number of killings in the analyzed dataset. Yet, the initial visualization seems to suggest a different narrative, with explosives notably standing out as having a substantial impact on fatalities.

This variance prompts a critical analysis of the factors contributing to the misalignment between visual patterns and statistical models. While visualizations provide an intuitive snapshot of impact, statistical models assess overall significance, accounting for potential confounding variables. In this case, the model's conclusion that firearms have a more consistent impact on fatalities across incidents challenges the initial visual assumption.

The discrepancy between visualization and modeling findings underscores the complexity of understanding the true impact of different weapon types in terrorism. It calls for a cautious approach in drawing conclusions solely from visualizations and emphasizes the importance of statistical rigor in the analysis. For counterterrorism strategies, acknowledging the statistical significance of firearms in causing fatalities suggests a need for targeted interventions and regulations surrounding the acquisition and use of firearms.

**Managerial implications:**

Terrorism poses a significant threat to densely populated metropolitan areas, as evidenced by the alarming frequency of attacks in cities like New York and California. The data underscores the need for robust counterterrorism measures to safeguard the lives and well-being of citizens. This highlights the urgent need for proactive security measures to safeguard the population and mitigate potential threats. The predominant use of explosives and firearms emphasizes the accessibility of these weapons, necessitating stricter regulations and monitoring.

Firstly, based on the data, the high incidence of terrorist attacks in New York and California necessitates urgent attention from state authorities. To address this concern, we recommend a strategic increase in security patrolling in these metropolitan regions. By increasing law enforcement presence, particularly in vulnerable areas, authorities can enhance their capacity to prevent and respond to potential threats effectively. Additionally, a visible and proactive security presence can serve as a deterrent, discouraging potential perpetrators from planning and executing attacks.

Furthermore, the success rate of terrorist attacks in these regions is troubling. To curtail this trend, there is a critical need for stricter weapons laws. The data indicates that perpetrators often resort to firearms, emphasizing the urgency of comprehensive gun control measures. The state government should advocate for and implement non-liberal weapons laws, ensuring that the acquisition and possession of firearms and explosives are subject to stringent regulations. This includes thorough background checks, periodic checks on existing weapon owners, and a reduction in the duration of their licenses. Shortening the license duration would necessitate regular renewals, allowing the government to gather updated information about weapon owners and their activities, thereby enhancing overall security.

Additionally, the data indicates that firearms have emerged as a prevalent choice for terrorist attacks, pointing to a need for increased regulation and monitoring. The government must impose stricter restrictions on the purchase and acquisition of firearms, making it more challenging for individuals with malicious intent to access these dangerous materials. Additionally, individuals already in possession of firearms must undergo rigorous periodic checks, including background assessments and assessments of the purpose and frequency of usage. These measures will not only limit access but also provide authorities with valuable data to assess and manage potential security risks effectively.

**Conclusion:**

In conclusion, the safety of citizens in highly populated metropolitan areas demands a proactive and comprehensive approach to counterterrorism. By bolstering security patrolling, implementing stringent weapons laws, and closely regulating explosives, governments can significantly reduce the likelihood and impact of terrorist attacks. The proposed measures aim not only to deter potential perpetrators but also to gather crucial information for informed decision-making. It is imperative that these steps be taken collaboratively, involving law enforcement agencies, policymakers, and the community to create a secure and resilient urban environment.