Dichotomy of Gun Violence Incedents:

Data Cleansing:

* Other – date, incident\_id, state, city\_or\_county,
* n\_killed
* N\_injured
* Gun type
* N\_guns involved
* Participant ages
* Participant genders
* Suspect relationship
* Participant relationship
* Participant status
* Participant type

A screenshot of a computer screen

Description automatically generatedA screenshot of a computer screen

Description automatically generated$A screenshot of a computer

Description automatically generated

Section 1 – Abstract

Section 1 – Introduction [ 0.5 Pages]

Summary of project

The goals

Dataset selected.

Section 2 – Data Description [1-1.5 Pages]

Feature Engineering

Distribution of different variables

Relationship between variables

Section 3- Model Selection & Methodology [1- 1.5 Pages]

Section 4- Results [1-2 Pages]

Section 5 – Classifications/predictions and conclusions [1-2 Pages]

Section 6 – Appendices [max 10 pages]

Section 7 - Code

**Section 1 – Introduction**

Gun violence is a critical societal issue, necessitating a comprehensive understanding of incidents to inform preventive strategies. This study employs a dataset of gun-related incidents, focusing on cases involving more than one gun, to extract valuable insights. Through advanced statistical techniques and visualizations, this analysis aims to unravel patterns, characteristics, and relationships within the dataset.

* 1. Summary of Project

This research leverages a dataset containing information on gun violence incidents with multiple guns involved. By applying data preprocessing and clustering techniques, we aim to identify inherent structures and patterns within the data. Our goal is to provide a nuanced understanding of the diverse factors associated with these incidents and shed light on potential avenues for targeted interventions.

* 1. The Goals

Cluster Analysis: Apply PCA and k-means clustering to identify inherent structures and groupings in the dataset.

Descriptive Statistics: Generate summary statistics for each cluster, offering insights into the characteristics of incidents within each group.

Visualizations: Utilize visualizations, such as scatter plots, box plots, and heatmaps, to effectively communicate patterns and relationships in the data.

Dataset Selected

The dataset used in this analysis comprises gun violence incidents, specifically focusing on those involving more than one gun. This subset aims to provide a nuanced examination of incidents with heightened severity, offering a targeted perspective on multifaceted occurrences.

**Section 2 – Data Description [1-1.5 Pages]**

2.1. Feature Engineering

The analysis involves preprocessing steps, including the calculation of counts for various aspects such as stolen guns, the number of subjects, victims, and total individuals involved, as well as categorization based on age groups and genders. These engineered features aim to capture essential nuances within the dataset.

2.2. Distribution of Different Variables

An exploration of the distribution of key variables, such as the number of injured and killed, stolen guns, and participant demographics, offers a preliminary understanding of the dataset's characteristics.

2.3. Relationship Between Variables

Investigating relationships between variables, such as the correlation between the number of victims and the type of guns involved, provides insights into potential causal links and dependencies.

**Section 3 – Model Selection & Methodology**

3.1 Data Preprocessing: Prior to modeling, we conducted comprehensive data preprocessing. This involved standardizing the dataset, performing PCA, and preparing the data for subsequent clustering analysis.

3.2. Clustering Analysis: We utilized the k-means algorithm to identify inherent structures in the dataset. The optimal number of clusters was determined through careful consideration of the data and iterative testing.

3.3. Visualization: To enhance interpretability, we complemented our analysis with various visualizations, including scatter plots for PCA representation, box plots for cluster comparisons, and a heatmap to illustrate variable relationships.

Descriptive Statistics: We employed summary statistics and stargazer tables to provide an overview of each cluster's characteristics. This step facilitated a comprehensive understanding of the distinct features associated with different incident groupings.

**Section 4 – Results**

4.1. Clustering Results

The application of PCA and k-means clustering revealed distinct groupings within the gun violence dataset. Clusters were identified based on shared characteristics, allowing for a more granular understanding of the incidents.

4.2. Visualization Insights

Visualizations played a pivotal role in communicating complex patterns. Scatter plots showcased the spatial distribution of incidents, while box plots facilitated the comparison of variables across clusters. The heatmap provided an intuitive representation of variable relationships.

4.3. Descriptive Analysis

Stargazer tables were generated to present summary statistics for each cluster. This included mean values and other relevant metrics for key variables, offering a detailed profile of incidents within each grouping.

4.4. Implications and Recommendations

The results carry implications for policymakers, law enforcement, and community stakeholders. By identifying patterns and characteristics associated with different incident clusters, targeted interventions can be devised to address specific challenges and mitigate risks.

Section 5 – Classifications/Predictions and Conclusions [1-2 Pages]

Section 6 – Appendices [Max 10 Pages]

Section 7 - Code