**Understanding Gun Violence in the United States: A Comprehensive Data Analysis**

, Dinasarapu Raja Shekar Reddy,   
 Kasarla Chandraprakash Reddy, Mamde Srinivasreddy Venkatreddy, Manam Lokesh

Saint Peter’s University, New Jersey, United States

**Author Note**

Our analysis is based on data sourced from reputable sources, including government agencies, research institutions, and non-profit organizations specializing in gun violence research. We prioritized data integrity and reliability, ensuring that our analysis is grounded in the most up-to-date and comprehensive information available.

**Abstract**

Gun violence in the United States is a multifaceted issue with far-reaching implications. Through comprehensive data analysis, this study delves into the complexities of gun-related incidents, examining key factors such as incident demographics, perpetrator profiles, geographic patterns, and legislative influences. Drawing from reliable sources and employing rigorous statistical methodologies, our analysis unveils nuanced insights into the dynamics of gun violence across diverse communities. We highlight disparities in victimization rates, identify high-risk populations, and explore the impact of socioeconomic factors on incidence rates. Additionally, our study assesses the effectiveness of existing gun control measures and evaluates their potential for mitigating violence. Ethical considerations underpin our research, guiding our efforts to present findings in a manner that respects the sensitivity of the issue and fosters informed dialogue. Ultimately, this analysis serves as a critical resource for policymakers, advocates, and stakeholders seeking evidence-based strategies to address gun violence and promote public safety.

**Keywords**: Gun violence, United States, Comprehensive data analysis, Incident demographics, legislation, public safety, Socioeconomic factors, Gun control measures, Violence mitigation, Informed dialogue, Evidence-based strategies

**Table of Contents**

[List of Figures 3](#_Toc158029168)

[Introduction 4](#_Toc158029169)

[Literature Review 4](#_Toc158029170)

[Data Collection 5](#_Toc158029171)

[Methodology 8](#_Toc158029172)

[Data Processing 8](#_Toc158029173)

[Exploratory Data Analysis 9](#_Toc158029174)

[Forecasting modals 14](#_Toc158029175)

[Random Forest model 14](#_Toc158029176)

[K-Nearest Neighbors model 15](#_Toc158029177)

[Linear Regression model 16](#_Toc158029178)

[Time-Series Analysis 16](#_Toc158029179)

[References 17](#_Toc158029180)

# List of Figures

Table 1: List of variables

Table 2: List of variables after removing duplicates

Figure 1: How many cases have been registered in each state.

Figure 2: Compare gun violence incidents between urban (city) and rural (county) regions.

Figure 3: Determine whether there is a link among the existence of particular representative and either the severity or frequency of occurrences.

Figure 4: Determine whether particular demographics are particularly susceptible or prone to participating in such situations.

Figure 5: Determine whether cases using stolen guns differ from those involving legally owned guns.

Figure 6: To identify participant relation type in violent events.

Figure 7: Identify what particular ages are particularly engaged in these types of situations.

Figure 8: Find the overall amount of people murdered in each state.

Figure 9: Identify which age group plays the biggest part in this violence.

Figure 10: In which year did the most cases occur.

# Introduction

The Gun Violence dataset is a comprehensive collection of data on incidents involving weapons in a variety of circumstances. This dataset is an invaluable resource for studying the scope, trends, and dynamics of gun-related occurrences, including their locations, impact, conditions, and other relevant elements. It is useful in the fields of public safety, criminal justice, policy creation, and lobbying, and it provides critical insights into the complicated subject of gun violence.

A data set usually includes details such as the time and date of the occurrence, the location (city, state, and coordinates), the total number of casualties (fatalities and injuries), the type of gun violence event (e.g., mass shooting, suicide, robbery), law enforcement involvement, and various contextual details. The dataset could include information about the victims, suspects, and firearms utilized. Gun violence remains a critical issue in the United States, with profound implications for public health and safety.

In conducting this analysis of gun violence in the United States, our aim was to provide an objective examination of a complex and pressing issue. We recognize the sensitivity surrounding this topic and approach our findings with a commitment to accuracy, impartiality, and empathy for those affected by gun violence.

The dataset can be used to enhance public awareness, educate communities, and promote educated public debate about gun violence and its consequences. Understanding the complexities of gun violence is vital for creating a safer society and encouraging evidence-based solutions to this critical issue. The Gun Violence dataset is critical to attaining these goals and furthering research and policy efforts targeted at decreasing the devastation caused by gun violence.

# Literature Review

The existing literature on USA Gun Violence Data Analysis provides valuable insights into the multifaceted nature of this pervasive issue. Numerous studies have examined incident demographics, highlighting the disproportionate impact on specific populations based on factors such as age, gender, and socioeconomic status. Perpetrator profiles have been scrutinized, revealing patterns and motivations that contribute to gun violence dynamics. Geographic analyses underscore regional disparities, emphasizing the need for targeted interventions. Legislative influences, particularly the effectiveness of gun control measures, have been subject to rigorous evaluation.

We have explored the intricate interplay of socioeconomic factors on gun violence, shedding light on how poverty, education, and employment contribute to the prevalence of incidents. Disparities in victimization rates have been a focal point, revealing systemic inequities that demand attention. The literature also delves into ethical considerations surrounding gun violence research, emphasizing the need for sensitivity and responsible dissemination of findings.

Collectively, this body of work not only informs our understanding of the complexities of gun violence but also serves as a foundation for evidence-based strategies. Policymakers, advocates, and stakeholders can draw upon this literature to develop informed interventions that address the root causes of gun violence and promote public safety.

# Data Collection

Our methodology section outlines the statistical techniques and analytical approaches employed to analyze the data and derive meaningful insights. This dataset contains 239678 records and 29 columns. The dataset allows for the research and investigation of a variety of aspects of gun violence, including the number and distribution of incidents, the relationship with sociodemographic variables, its effect on groups, and the success of various prevention and intervention activities. It serves as the basis for policy development based on evidence, supporting legislators, academics, and activists in devising successful methods to reduce the negative societal effects of gun violence.

We obtained a dataset containing relevant information on incidents, gender, state, urban, rural areas, events, age types, density, range, Year.

We considered a range of factors that contribute to gun violence, including incident data, victim demographics, perpetrator information, geographic patterns, legislative and policy factors, and broader social and economic contexts. By considering these factors holistically, we aimed to provide a nuanced understanding of the issue.

We will work with data from 260k firearms occurrences in the United States. The data ranges from January 2013 to March 2018.

The information came from gunviolencearchive.org. This database was compiled using public records and news sources. (<https://www.gunviolencearchive.org/>)

These are the detailed information of dataset columns incident\_id: Unique identifier for each incident. Date: Date of the incident. state: State where the incident occurred. city\_or\_county: City or county where the incident occurred. address: Address of the incident. n\_killed: Number of people killed in the incident. n\_injured: Number of people injured in the incident. incident\_url: URL related to the incident. source\_url: URL of the information source. congressional\_district: Congressional district where the incident occurred. gun\_stolen: Information about whether the gun used was stolen. gun\_type: Type of gun(s) used in the incident. incident\_characteristics: Characteristics describing the incident. latitude: Latitude coordinate of the incident location. location\_description: Description of the incident location. longitude: Longitude coordinate of the incident location. n\_guns\_involved: Number of guns involved in the incident. notes: Additional notes or details about the incident. participant\_age: Age of participants. participant\_age\_group: Age group of participants. participant\_gender: Gender of participants. participant\_name: Names of participants. participant\_relationship: Relationship between participants. participant\_status: Status of participants (e.g., victim, suspect). participant\_type: Type of participant (e.g., victim, suspect). sources: Sources providing information about the incident. state\_house\_district: State house district where the incident occurred. state\_senate\_district: State senate district where the incident occurred.

A screenshot of a computer

Description automatically generated

Table 1: List of Variables

A screenshot of a computer

Description automatically generated

Table 2: List of variables after removing duplicates

# Methodology

## Data Processing

Data preprocessing involved handling null values, removing repeated values, and normalizing numerical features. Categorical variables were encoded, and outliers were addressed to ensure the quality of the dataset for analysis.

Our methodology section outlines the statistical techniques and analytical approaches employed to analyze the data and derive meaningful insights.

Analysis utilized a combination of statistical techniques, data visualization tools, and qualitative insights to identify trends, correlations, and potential causal factors.

The methodology section outlines the statistical techniques and analytical approaches employed to analyze the data and derive meaningful insights.

A screenshot of a computer program

Description automatically generated

## Exploratory Data Analysis

Exploratory Data Analysis was performed to gain initial insights into the dataset. Descriptive statistics, visualizations, and correlation analyses were utilized to understand the distribution of variables and relationships within the data.

By using a combination of visualization and transformation methods, we were able to gain a more comprehensive understanding of cases. The use of bar charts, line graphs, and scatter plots helped to identify patterns and trends in the data that could be used to make informed decisions. The use of Pandas and Python allowed for efficient data transformation and removal of irrelevant data, ensuring the accuracy and relevance of the analysis.

Below is the result where we can see total cases have been registered in each state.

A blue bars on a white background

Description automatically generated

Figure 1: How many cases have been registered in each state.

Below is the result where we compare and contrast gun violence incidents between urban (city) and rural (county) regions.

A graph with blue and green bars

Description automatically generated

*Figure 2: Compare gun violence incidents between urban (city) and rural (county) regions.*

Below result determine whether there is a link among the existence of particular representative and either the severity or frequency of occurrences.

A graph of a number of people

Description automatically generated with medium confidence

*Figure 3: Determine whether there is a link among the existence of particular representative and either the severity or frequency of occurrences.*

Below result determine whether particular demographics are particularly susceptible or prone to participating in such situations.

A colorful pie chart with numbers and text

Description automatically generated

*Figure 4: Determine whether particular demographics are particularly susceptible or prone to participating in such situations.*

Below output shows whether cases using stolen guns differ from those involving legally owned guns*A screenshot of a computer screen

Description automatically generated*

*Figure 5: Determine whether cases using stolen guns differ from those involving legally owned guns.*

By seeing the below plot robbery is the main catergory where we are seeing more cases are registered and aslo we identify particpant relation type in violent events.

A blue and purple squares

Description automatically generated

*Figure 6: To identify participant relation type in violent events.*

Below images show that Adult 18+ are committing more crimes and also, we can see the other ages are particularly engaged in these types of violations.

**A graph with blue squares

Description automatically generated with medium confidence**

*Figure 7: Identify what particular ages are particularly engaged in these types of situations.*

By going through the below result we get to know that Tennessee State has the highest people got murdered and also, we get to know overall amount of people murdered in each state.

A colorful graph of lines

Description automatically generated with medium confidence

*Figure 8: Find the overall amount of people murdered in each state.*

After applying the logic and we get to know more female age group plays the biggest part in this violence.

A graph of a person and person

Description automatically generated

*Figure 9: Identify which age group plays the biggest part in this violence.*

By seeing the below result we get to know which year did the most cases occur it seems like 2017 has more cases.

A graph with different colored bars

Description automatically generated

*Figure 10: In which year did the most cases occur.*

# Forecasting modals

## Random Forest model

A screenshot of a computer

Description automatically generated

## K-Nearest Neighbors model

A screenshot of a computer

Description automatically generated

## Linear Regression model

A screenshot of a computer

Description automatically generated

## Time-Series Analysis

A graph with blue lines

Description automatically generated

A graph showing a graph

Description automatically generated with medium confidence

# References

1. A Literature Review and Research Agenda by J D Wright; P H Rossi; K Daly; E Weber-Burdin: Weapons, Crime, and Violence in America
2. Planty, Michael G., and Jennifer L. Truman. 2013 ,1993–2011: Firearm Violence.
3. Dewey Cornell - Gun Violence: Prediction, Prevention, and Policy
4. Alpers, P., & Wilson, M. (2013, August 14): Global impact of gun violence: Firearms, public health, and safety by
5. National Institute of Justice (NIJ): The research arm of the U.S. Department of Justice, NIJ funds studies and research initiatives related to gun violence and crime prevention.
6. American Academy of Pediatrics, Council on Injury, Violence, and Poison Prevention Executive Committee. (2012): Firearm-related injuries affecting the pediatric population.
7. FBI Uniform Crime Reporting (UCR) Program: The UCR Program collects and publishes data on crime, including firearm-related offenses, reported by law enforcement agencies across the country.
8. D. Webster & J. Vernick (Eds.): Gun control and homicide reduction.
9. Bandeira, A. R. (2013): Reducing gun violence in America: Informing policy with evidence and analysis (pp. 213–223).
10. The RAND Corporation: RAND publishes research on gun violence, including studies on the effectiveness of various interventions and policies aimed at reducing gun-related harm.
11. Borum, R., Cornell, D., Modzeleski, W., & Jimerson, S. R. (2010): What can be done about school shootings? A review of the evidence. Educational Researcher, 39, 27–37. doi:10.3102/0013189X09357620.
12. Guilford Press by Borum, R., & Verhaagen, D. (2006): Assessing and managing violence risk in juveniles. New York, NY.
13. Centers for Disease Control and Prevention (CDC): The CDC provides comprehensive data on gun-related injuries and fatalities through its National Violent Death Reporting System (NVDRS) and other surveillance programs.
14. The Pew Research Center: Pew conducts surveys and studies on public attitudes toward gun violence, gun ownership, and gun control policies.
15. Bostwick, J. M., & Pankratz, V. S: Affective disorders and suicide risk: A reexamination. American Journal of Psychiatry, 157(12), 1925–1932. doi: 10.1176/appi.ajp.157.12.1925.
16. National Crime Victimization Survey: Criminal victimization in the United States, 2006 statistical tables (NCJ 223436) by Bureau of Justice Statistics.
17. Bureau of Justice Statistics (BJS) - The BJS collects and publishes data on various aspects of crime, including firearm-related crimes, through surveys and reports.
18. Centers for Disease Control and Prevention: Source of firearms used by students in school-associated violent deaths — United States, 1992–1999. MMWR Morbidity and Mortality Weekly Report, 52(9), 169–172.
19. Gun Violence Archive (GVA): The GVA is an independent nonprofit organization that tracks gun-related incidents in the United States through media reports, law enforcement sources, and other sources.
20. Chapman, S., & Alpers, P. (2013): Gun-related deaths: How Australia stepped off “The American path.” Annals of Internal Medicine, 158(10), 770–771. doi:10.7326/0003-4819-158-10-201305210-00624.