VISVESVARAYA TECHNOLOGICAL UNIVERSITY

“JnanaSangama”, Belgaum -590014, Karnataka.

AAT PROJECT REPORT
on

Socio-Police Dynamics

Submitted by

TARUN M M (1BM22CS306)
TISSA MARIA (1BM22CS309)
SRIKAR R OLETY(1BM22CS289)

Under the Guidance of
Dr.K.Panimozhi
Associate Professor, BMSCE

in partial fulfillment for the award of the degree of
BACHELOR OF ENGINEERING
in
COMPUTER SCIENCE AND ENGINEERING

B. M. S. COLLEGE OF ENGINEERING
(Autonomous Institution under VTU)
BENGALURU-560019
Sep-2024 to Feb-2025

B. M. S. College of Engineering,

Bull Temple Road, Bangalore 560019

(Affiliated To Visvesvaraya Technological University, Belgaum)

Department of Computer Science and Engineering

<u>CERTIFICATE</u>

This is to certify that the project work entitled "**Socio-Police Dynamics**" carried out by **Tissa Maria (1BM22CS309), Srikar R Olety (1BM22CS289) AND Tarun M M (1BM22CS306)** who are bonafide students of **B. M. S. College of Engineering.** It is in partial fulfillment for the award of **Bachelor of Engineering in Computer Science and Engineering** of the Visvesvaraya Technological University, Belgaum during the year 2024-25. The AAT report has been approved as it satisfies the academic requirements in respect of **Data Exploration and Visualization (23CS5PCDEV)** work prescribed for the said degree.

Signature of the Guide
Dr.K.Panimozhi
Associate Professor
BMSCE, Bengaluru

Table of Contents

1.Introduction

This project delves into the critical issue of police-involved fatal shootings in the United States, leveraging data from *The Washington Post* database and U.S. Census statistics on poverty, education, income, and race. By examining social and economic trends, the project aims to uncover patterns and disparities that may help explain the occurrence and distribution of these incidents across different regions.

Since January 1, 2015, *The Washington Post* has meticulously compiled a database of every fatal shooting by police officers in the line of duty, encompassing over 2,500 records per file. The study integrates this data with U.S. Census information to analyze key factors such as poverty rates, high school graduation rates, median household income, and racial demographics.

The primary objectives include:

- Visualizing the relationship between poverty rates and high school graduation rates.
- Creating a choropleth map to illustrate police killings by state.
- Analyzing trends in police killings over time to identify potential patterns.
- Exploring correlations between median income, poverty rates, education, and the frequency of fatal police shootings.

This project transcends numerical analysis and visualization. It seeks to provide a deeper understanding of the systemic and social dynamics influencing fatal use-of-force incidents, shedding light on issues that profoundly affect communities across the nation.

2.Data Collection and Inspection

2.1. Sources of Data

Fatal Police Shootings Dataset: Collected from The Washington Post, detailing police-involved fatal shootings in the U.S. since 2015. This dataset contains 2,535 rows and includes information such as the victim's name, age, race, manner of death, whether they were armed, and additional contextual details.

U.S. Census Data: Sourced from official Census Bureau records, providing information on poverty rates, median income, and high school graduation rates across U.S. cities. These datasets contain 29,322 rows each.

2.2. Tools Used

Python Libraries:

pandas: For data manipulation and cleaning.

numpy: For numerical operations

matplotlib and seaborn: For visualization of trends and patterns.

plotly.express: For creating interactive maps and visualizations.

2.3. Data Cleaning Steps

Conversion to DataFrames: All datasets were imported from CSV files and converted into Pandas DataFrames for analysis.

Handling Missing Values: Missing data was addressed using dropna() to remove rows with null values, ensuring clean and usable datasets.

Removing Duplicates: Duplicated rows were identified and removed using drop_duplicates().

Formatting Date Fields: The date column in the police shootings dataset was formatted to datetime format for temporal analysis.

Standardizing State Names: Verified that state abbreviations (e.g., AL for Alabama) were consistent across all datasets to ensure seamless merging during analysis.

2.4. Key Findings from Preliminary Analysis

Top Cities with the Most Incidents

- The cities with the highest number of fatal police shooting incidents are:
    1. Los Angeles: 35 incidents
    2. Phoenix: 28 incidents
    3. Houston: 24 incidents
    4. Chicago: 22 incidents
    5. Austin: 18 incidents
    6. Las Vegas: 17 incidents
    7. Columbus: 16 incidents

8. Miami: 15 incidents
9. San Antonio: 15 incidents
10. Indianapolis: 14 incidents

- State-Level Observations:
  States like California (CA) and Texas (TX) account for a disproportionately high number of incidents, indicating a need for state-level analysis of contributing factors.

Age Analysis

- Mean Age of Deaths: 36.25 years
- Mode Age of Deaths: 25 years
- Median Age of Deaths: 34 years
- Maximum Age of Death: 91 years (Victim: Frank W. Wratny, Race: White)
- Minimum Age of Death: 6 years (Victim: Jeremy Mardis, Race: White)

These findings highlight the wide age range of victims and raise critical questions about the circumstances leading to these deaths.

This analysis provides an initial understanding of the geographic, demographic, and socio-economic factors involved in police-involved fatalities, setting the stage for deeper investigations into systemic disparities.

3. Handling Missing Values and Data Cleaning

3.1 Managing Null Values and Duplicates

Missing values were handled by removing rows with null entries across all datasets to ensure completeness and reliability of the data.

This approach was necessary for consistent and accurate analysis.

Duplicated rows were removed to eliminate redundancy and maintain the integrity of the datasets.

U.S. Census Data:
Data on education, poverty rates, and median income required minimal cleaning, as it was already well-structured. Missing values and duplicates, if present, were removed.

Fatal Police Shootings Dataset:
Several missing values were identified in columns like "armed" or "mental illness," which were dropped to maintain data quality.

3.2 Data Cleaning Process

Standardization of Formats: Ensured consistent formatting of key columns, such as state and city names, to facilitate smooth merging and comparison across datasets.

City Name Uniformity: City names were cleaned to remove redundant suffixes (e.g., "city," "town") and standardize the naming convention across datasets.

Datetime Formatting: The date column in the police shootings dataset was formatted to a standard datetime format, enabling temporal analyses such as trend identification over time.

Consistency Across Datasets: Verified and standardized state abbreviations to ensure compatibility during data integration and analysis.

3.3 Impact on Data Quality

Enhanced Consistency: Uniform formatting across datasets eliminated potential mismatches, enabling seamless integration and analysis.

Facilitated Analysis: Proper formatting, such as standardized datetime and city names, allowed for more straightforward and efficient analysis of trends and patterns.

Minimized Errors: Removing inconsistencies and addressing missing values reduced the risk of inaccuracies in findings.

4. Exploring Data Distribution and Analysis

4.1 Univariate Analysis

1. What is the distribution of the age of victims in police-involved fatalities?

- Mean Age of Deaths: 36.25 years
- Mode Age of Deaths: 25 years
- Median Age of Deaths: 34 years
- Max Age of Deaths: 91 years (Frank W. Wratny)

- Min Age of Deaths: 6 years (Jeremy Mardis)

Insight: The mean and median ages suggest a tendency for police fatalities to involve individuals in their 30s, while the mode age indicates that many victims are young adults. The extremes of age highlight the diverse range of victims.

2. How are police fatalities distributed by race?

- Deaths by Race:
- White (W): 1,168 deaths
- Black (B): 592 deaths
- Hispanic (H): 401 deaths
- Asian (A): 36 deaths
- Native American (N): 29 deaths
- Other (O): 28 deaths

3. What is the gender distribution of fatalities?

Gender Disparity of Deaths:

- Male (M): 2,160 deaths
- Female (F): 94 deaths

Insight: The data shows a significant gender disparity, with males constituting the overwhelming majority of victims.

4. What is the fleeing status of victims?

Fleeing Status:

- Not fleeing: 1,528 deaths
- Car: 360 deaths
- Foot: 278 deaths
- Other: 88 deaths

Insight: The majority of victims were not fleeing, but a significant number were fleeing in a car or on foot.

5. How common is body camera usage in fatal police shootings?

Body Camera Usage:

- False: 2,002 deaths (88.82%)
- True: 252 deaths (11.18%)

Insight: Body cameras were used in a minority of fatal police shootings, which raises concerns about transparency and accountability in such incidents.

6. What is the prevalence of signs of mental illness in victims?

Signs of Mental Illness:

- False: 1,681 deaths (74.58%)
- True: 573 deaths (25.42%)

Insight: A significant proportion of victims (25.42%) showed signs of mental illness, suggesting that mental health may play a role in some police encounters.

7. What is the weapon type distribution in deaths by police shootings?
Weapon Distribution:

- Gun: 1,247 deaths
- Knife: 324 deaths
- Unarmed: 165 deaths
- Vehicle: 158 deaths
- Toy weapon: 102 deaths

Insight: The distribution shows that the majority of deaths involve firearms, followed by knives and unarmed individuals. This suggests that weapon presence plays a significant role in police shootings.

4.2 Bivariate Analysis

1. What is the correlation between high school completion rates and deaths caused by police shootings?

Correlation between High School Completion Rates and Deaths:

- Correlation Coefficient: -0.43

Insight: The negative correlation between high school completion rates and deaths caused by police shootings suggests that higher educational attainment may be associated with lower death rates, though the relationship is weak and may be influenced by other factors.

2. What is the correlation between median household income and deaths caused by police shootings?

Correlation between Median Household Income and Deaths:

- Correlation Coefficient: 0.21

Insight: The weak positive correlation indicates that states with higher median incomes may experience slightly more deaths caused by police shootings. However, the correlation is not strong, suggesting that income alone may not be a significant determinant in the number of deaths.

3. What is the correlation between deaths caused by police shootings and poverty levels per state?

Type: Bivariate
Correlation between Deaths and Poverty Levels:

- Correlation Coefficient: 0.24

Insight: The positive but weak correlation suggests that states with higher poverty levels tend to have a slightly higher rate of deaths caused by police shootings. However, the correlation is not strong, indicating that poverty levels are not the sole factor contributing to these deaths.

4.3 Multivariate Analysis

1. What are the racial disparities in deaths caused by police shootings?
Racial Disparity of Deaths:

- White: 80.97% of population, 0.52% of deaths
- Black: 7.83% of population, 0.26% of deaths
- Asian: 1.84% of population, 0.02% of deaths
- Hispanic: 7.75% of population, 0.18% of deaths
- Native American: 3.61% of population, 0.01% of deaths

Comparative Proportions of Population Deaths:

- White: 0.0064
- Black: 0.0332
- Asian: 0.0109
- Hispanic: 0.0232
- Native American: 0.0028

Insight: Black individuals are disproportionately affected by police shootings, with a significantly higher death rate compared to their proportion in the population. Other racial groups, such as Hispanic and Asian, also experience disparities, though to a lesser extent.

5. Feature Engineering and Visualization

5.1. Process of Feature Engineering

Feature engineering in the given code involves several steps to clean, transform, and create new features from the raw data, enhancing its suitability for analysis. Below is an outline of the process:

Data Cleaning:

- Removing duplicate and missing data using .dropna() and .drop_duplicates().
- Handling invalid or placeholder values (e.g., replacing {x} and - with 0 in percent_completed_hs).

Feature Transformation:

- Converting string columns to numeric (e.g., poverty_rate, Median Income).
- Parsing and formatting datetime columns (e.g., deaths_by_police['date']).
- Standardizing textual data (e.g., trimming whitespace and converting city names to lowercase).

Feature Creation:

- Extracting year_month from the date column for time series analysis.
- Calculating the death_rate and completion_rate as proportions for comparative analysis.

- Creating aggregated statistics by grouping data (e.g., mean poverty rates, death counts per state, racial proportions).
- Deriving new composite features, such as the share of non-white races.

5.2. Impact of Feature Engineering on Predictive Power

Feature engineering significantly affects the predictive power of machine learning models by transforming raw data into informative and meaningful features. Here's how it applies to this dataset:

- Improved Data Quality:

By cleaning and standardizing the data, errors and inconsistencies are minimized, ensuring that the model learns from reliable inputs.

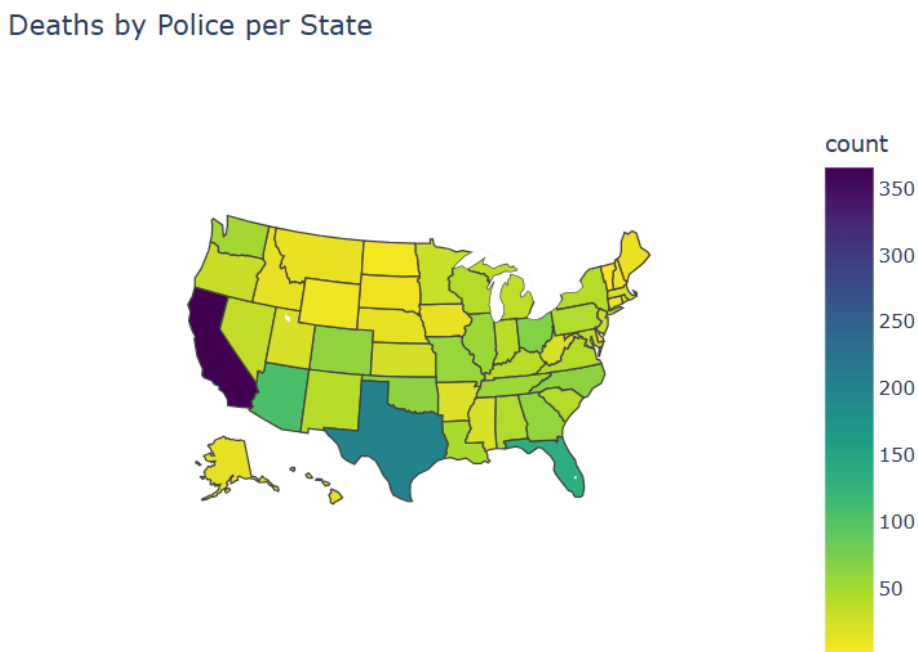
- Enhanced Feature Interpretability:

Features like death_rate, completion_rate, and rate_of_median_income provide a clearer relationship between variables, improving the model's ability to detect patterns.

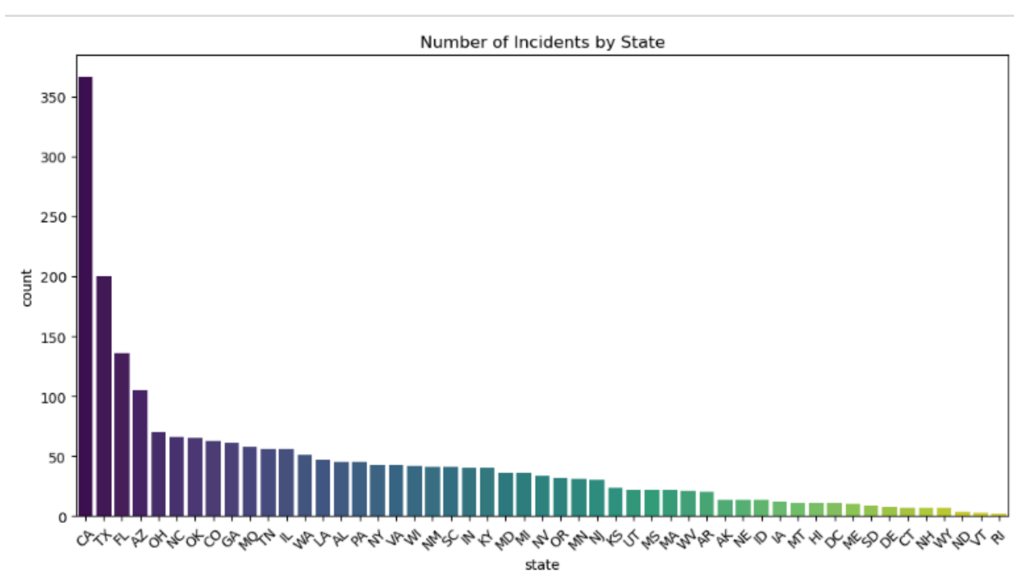
- Better Representation of Trends and Relationships:

Time series features, such as year_month and monthly death counts, capture temporal trends, which could be predictive in understanding seasonality or changes over time.

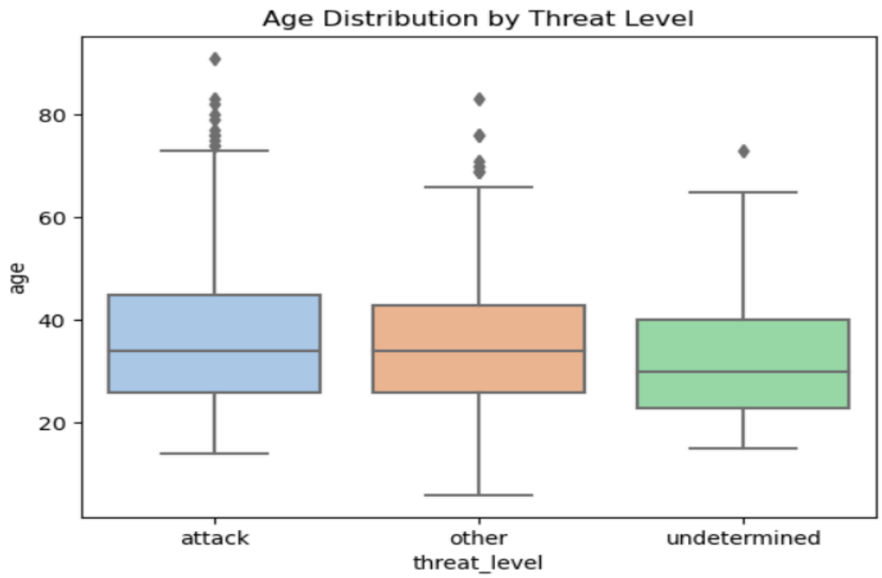
5.3 Visualization



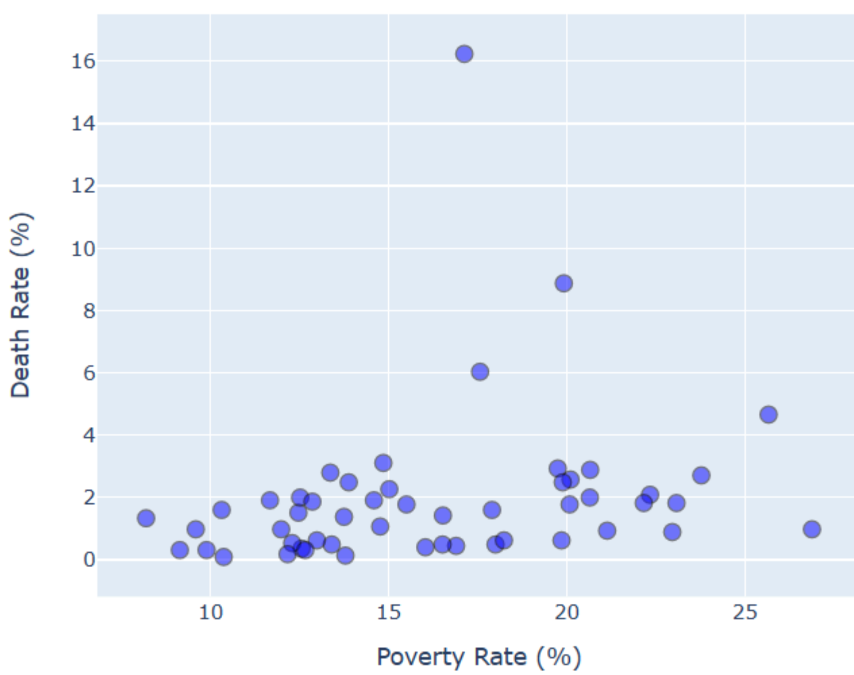
5.2.1 Choropleth map of the United States depicting state-wise deaths caused by police shootings



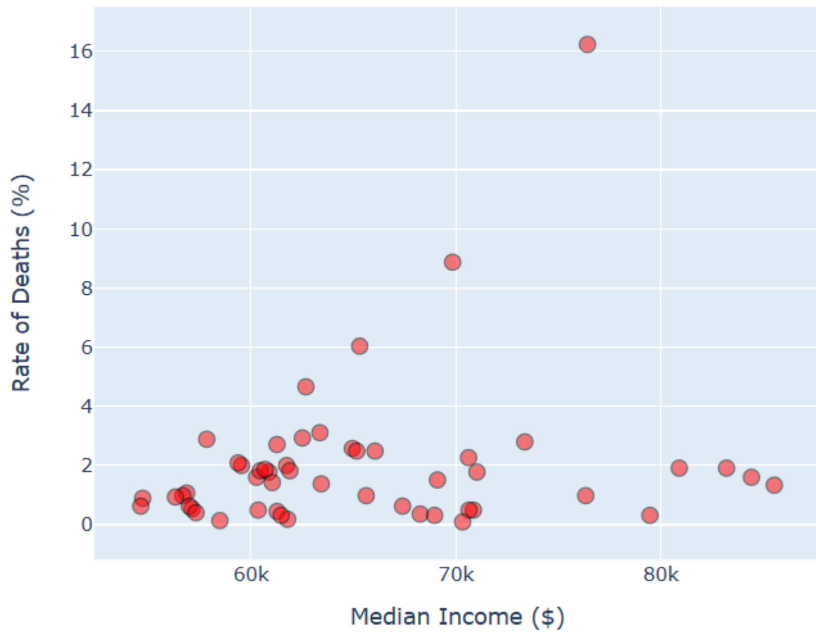
5.2.2 Bar plot illustrating state-wise deaths caused by police shootings.



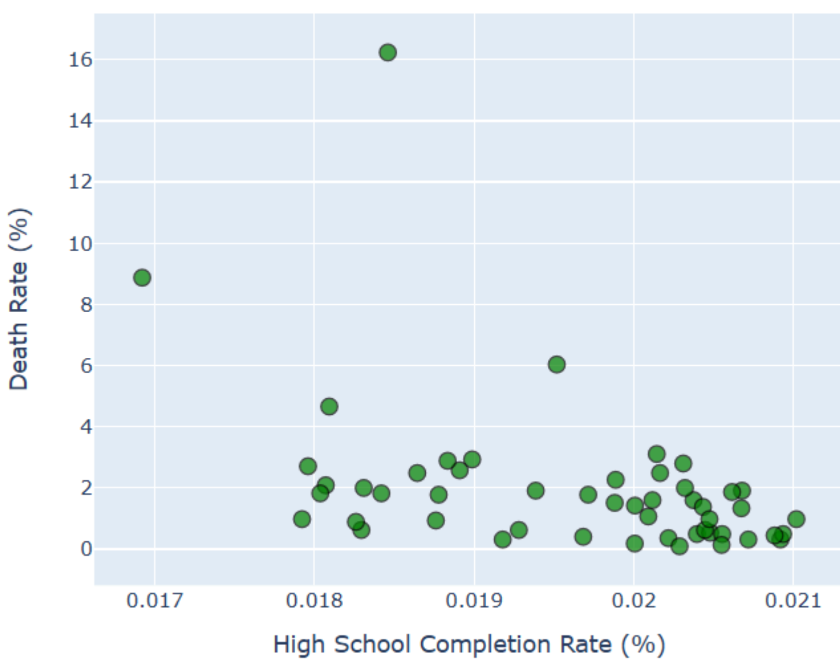
5.2.3 Box plot showing the distribution of age categorized by threat levels



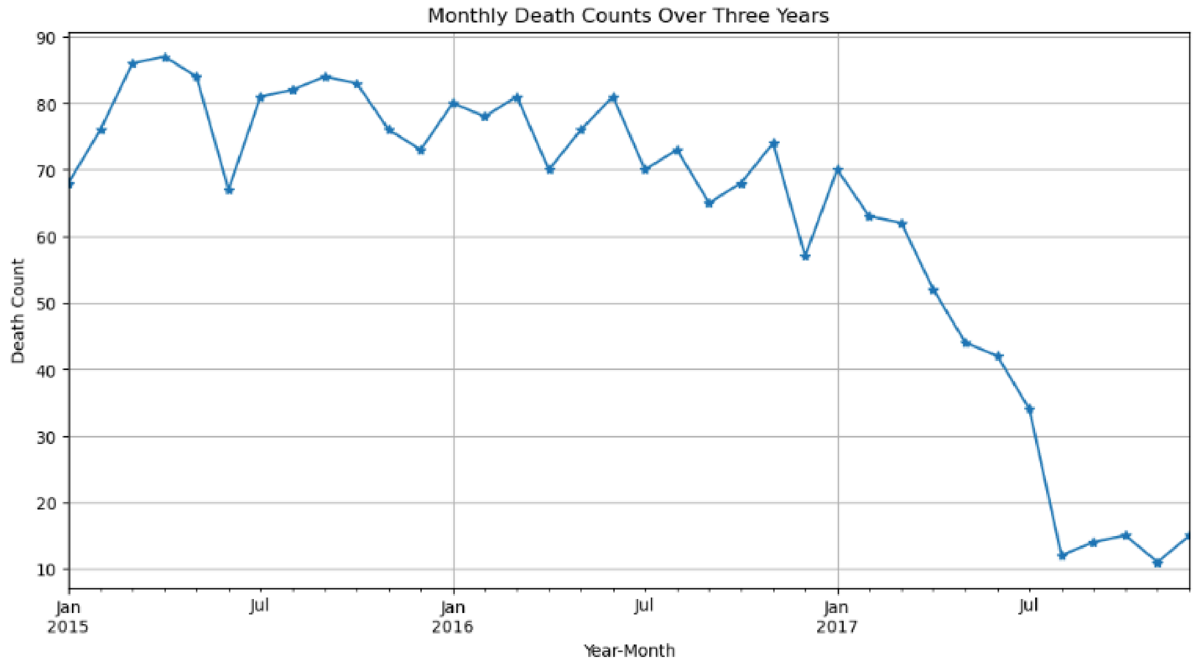
5.2.4 Scatter plot depicting the relationship between deaths caused by police in the U.S. and poverty rates by state.



5.2.5 Scatter plot illustrating the relationship between deaths caused by police and median income (in USD).



5.2.6 Scatter plot highlighting the correlation between high school completion rates and death rates.



5.2.7 Line Plot highlighting monthly death count over three years

6. Conclusion

The analysis of police-involved fatal shootings in the United States reveals significant systemic disparities, particularly in education and racial demographics, while also highlighting limited correlations with economic factors. States with higher high school completion rates tend to experience fewer fatal police encounters, underscoring the role of educational attainment in mitigating such incidents. However, the correlations between median income, poverty rates, and fatal police shootings were weak, suggesting these economic factors alone are not primary determinants of such outcomes. Racial disparities remain stark, with Black individuals disproportionately affected, alongside notable disparities for Hispanic and Native American groups. These findings call for systemic reforms focusing on education, economic equity, and addressing racial biases to reduce fatal police encounters and promote justice.