Alejandra Castro

Final Project for the

CareerFoundry Data Analytics Certificate

analysis of gun violence in the united states

1. Project overview

**Motivation**: The United States is the country most afflicted by the phenomenon of mass shootings. Between 2000 and 2022, according to the Rockerfeller Institute of Government, there were 109 mass shootings compared to 35 shootings in 35 other countries of similar economic and cultural characteristics. Mass shootings are undeniably a concern for public health and safety, and while they are tied to a plethora of factors, it is important to understand them individually, as well as their relationships among one another. This analysis seeks to understand the relationship between the purchasing of guns and mass shootings to elucidate on preconceptions that feel intuitive but might not be entirely true: For example, the fact that more guns signify more mass shootings. Often, the access to guns is not the only factor that pushes an individual to commit this type of crime, so this analysis will try to elucidate on this specific relationship.

**Objective**: Determine the relationship between mass shootings and gun sales per state and county in the United States.

**Scope**: The data covers the 50 states of the United States and their corresponding counties in a span between 2015 and 2025.

1. Hypothesis: Initial Questions

The primary directive of this study is to identify the relationship between mass shootings and gun sales trends. To this end, the initial hypothesis is as follows:

There is a positive correlation between gun sales and mass shootings. The more guns that are sold, the more mass shootings there are. Likewise, the states and counties with the most guns sold are the ones with the most mass shootings.

1. Data Overview

The analysis is performed with 5 datasets. Three belong to the U.S. Census Bureau and two to independent organizations that track gun violence in the United States.

|  |  |
| --- | --- |
| Dataset name | US Census Population 2010-2020 |
| Summary | This dataset describes United States population by county and state between 2010 and 2020 It was published by the U.S. Census Bureau in 2020. It contains information from the NCHS and the FSCPE. The dataset describes United States population by county and state. According to the Demographic Analysis Statement of Methodology, the dataset includes administrative records of births provided by the National Center for Healthcare Statistics, and the Federal-State Cooperative for Population Estimates. This dataset was merged with a second population dataset called “US Census Population 2020-2024”. |
| Source (Data owner) | US Census Bureau (Via NCHS and FSCPE) |
| Type of data collection | Administrative |
| Date of publishing | 2020 |
| Relevant links | https://www2.census.gov/programs-surveys/popest/datasets/2010-2015/counties/ |

|  |  |
| --- | --- |
| Dataset name | US Census Population 2020-2024 |
| Summary | This dataset describes United States population by county and state between 2020 and 2024 It was published by the U.S. Census Bureau in 2024. It contains information from the NCHS and the FSCPE. The dataset describes United States population by county and state. According to the Demographic Analysis Statement of Methodology, the dataset includes administrative records of births provided by the National Center for Healthcare Statistics, and the Federal-State Cooperative for Population Estimates. This dataset was merged with a second population dataset called “US Census Population 2010-2020”. |
| Source (Data owner) | US Census Bureau (Via NCHS and FSCPE) |
| Type of data collection | Administrative |
| Date of publishing | 2024 |
| Relevant links | https://www2.census.gov/programs-surveys/popest/datasets/2020-2024/counties/ |

|  |  |
| --- | --- |
| Dataset name | US Cities |
| Summary | This dataset describes United States’ cities and states. It provides longitudinal and latitudinal coordinates and was published by the Simple Maps, utilizing data from the US Census Bureau, and the US Geological Survey. It was updated in January 2025. |
| Source (Data owner) | Simple Maps |
| Type of data collection | Administrative |
| Date of publishing | 2025 |
| Relevant links | https://simplemaps.com/data/us-cities |

|  |  |
| --- | --- |
| Dataset name | Mass Shootings 2015-2025 |
| Summary | This dataset describes the occurrence of mass shootings (defined as 4 or more people shot) between 2015 and 2025, compiled by the Gun Violence Archive, an independent organization that provides data regarding gun violence in the United States. The Gun Violence Archive is the collector and proprietary of this dataset. The collection is done utilizing automated queries and manual search through over 7500 sources from local and state police, media, and governmental data aggregates. Each incident is verified by a two-check validation process. |
| Source (Data owner) | The Gun Violence Archive |
| Type of data collection | Administrative |
| Date of publishing | 2025 |
| Relevant links | https://www.gunviolencearchive.org/methodology |

|  |  |
| --- | --- |
| Dataset name | Gun Sales 2015-2025 |
| Summary | This dataset describes positive the estimated gun sales between 2015 and 2025 in the United States, compiled by The Gun Violence Data Hub, a data archive operated by The Trace, a non-governmental organization that investigates gun violence in the United States. The data is derived from the FBI's NICS firearm checks. The dataset is derived from NICS data, a background check system run by the FBI which checks fire arm transfers conducted by licensed firearms dealers. The NICS data includes four categories of transfer checks: handgun, long gun, multiple, and other. Multiples only apply when a transaction involves two types of guns. The category 'other' includes frames, receivers, and other firearms that are neither handguns nor long guns. The seasonality adjustment is performed using the U.S. Census Bureau's X13-ARIMA-SEATS software. |
| Source (Data owner) | The Trace Gun Violence Data Hub |
| Type of data collection | Administrative |
| Date of publishing | 2025 |
| Relevant links | https://datahub.thetrace.org/dataset/gun-sales/#methodology |

1. Data Limitations

* Population data only spans to 2024, which does not make it possible to analyse 2025 trends readily. A time series will be needed.
* Gun sales data is undercounted, as described in the data profile and contains some missing values.

Ethical concerns:

* To limit confirmation bias, demographic characteristics were not considered in this study.
* Precise addresses were omitted to maintain anonymization.

1. Exploratory Analysis: Descriptive Statistics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Victims\_Killed** | **Victims\_Injured** | **Suspects\_Killed** | **Suspects\_Injured** | **Suspects\_Arrested** |
| **count** | 64.528.000.000 | 64.528.000.000 | 64.528.000.000 | 64.528.000.000 | 64.528.000.000 |
| **mean** | 0.968014 | 4.227.855 | 0.062794 | 0.042245 | 0.666966 |
| **std** | 1.599.781 | 6.625.384 | 0.248153 | 0.224177 | 1.099.940 |
| **min** | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| **25%** | 0.000000 | 3.000.000 | 0.000000 | 0.000000 | 0.000000 |
| **50%** | 1.000.000 | 4.000.000 | 0.000000 | 0.000000 | 0.000000 |
| **75%** | 1.000.000 | 5.000.000 | 0.000000 | 0.000000 | 1.000.000 |
| **max** | 60.000.000 | 439.000.000 | 3.000.000 | 5.000.000 | 14.000.000 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **count** | **Handguns\_Sold** | **Long\_Guns\_Sold** | **Total\_Guns\_Sold** | **Population** |
| **mean** | 64.295.000.000 | 64.295.000.000 | 64.295.000.000 | 4,81E+10 |
| **std** | 28.222.587.791 | 17.935.452.181 | 46.158.010.063 | 3,77E+11 |
| **min** | 22.367.401.347 | 12.210.584.514 | 33.291.325.242 | 1,45E+12 |
| **25%** | 0.000000 | 0.000000 | 0.000000 | 2,15E+08 |
| **50%** | 11.658.000.000 | 9.705.000.000 | 22.042.000.000 | 1,83E+10 |
| **75%** | 22.265.000.000 | 14.994.000.000 | 37.121.000.000 | 4,41E+10 |
| **max** | 37.070.000.000 | 22.969.000.000 | 60.362.000.000 | 2,08E+11 |
|  | 171.600.000.000 | 102.546.000.000 | 248.724.000.000 | 3,13E+13 |