

# Economic Determinants of Gun Violence in the United States

## 1. Introduction

Gun violence remains a critical concern in the United States, with significant social and economic implications. This project aims to analyse the relationship between gun violence incidents and state-level economic indicators such as GDP and GDP per capita. By examining correlations and trends, the project seeks to identify insights that may inform policymakers and researchers.

### 1.1. Main Analytical Questions

1. What are the long-term trends in gun violence incidents across U.S. states, and how have they changed over time?
  2. How does state-level GDP or GDP per capita correlate with gun violence rates?
  3. Which states have the highest incidence of gun violence, and what economic factors are associated with these states?
  4. What is the correlation between poverty and unemployment rates and gun violence incidents at the state level?
- 

## 2. Datasets

### 2.1. Gun Violence Data

- **Source:** Kaggle - Gun Violence Data
- **Metadata URL:** <https://www.kaggle.com/datasets/jameslko/gun-violence-data>
- **Data URL:** <https://www.kaggle.com/datasets/jameslko/gun-violence-data/download?datasetVersionNumber=1>
- **Data Type:** Zipped CSV
- **Description:** Records of gun violence incidents, including state, date, location, and key statistics such as fatalities and injuries.
- **License:** Public dataset for research use, governed by Kaggle terms requiring attribution.
- **Quality:** Comprehensive and detailed, though missing values in some fields like participant demographics required imputation.

### 2.2. U.S. GDP by State

- **Source:** Kaggle - U.S. GDP by State (1997-2020)
- **Metadata URL:** <https://www.kaggle.com/datasets/davidbroberts/us-gdp-by-state-19972020/data>
- **Data URL:** <https://www.kaggle.com/datasets/davidbroberts/us-gdp-by-state-19972020/download?datasetVersionNumber=1>
- **Data Type:** Zipped CSV
- **Description:** Annual GDP values by state from 1997 to 2020, providing economic context for gun violence analysis.
- **License:** Open data with non-commercial use allowed and attribution required.
- **Quality:** Well-structured dataset with minimal inconsistencies; years beyond 2020 are not included.

## 2.3. GDP per Capita by State

- **Source:** Kaggle - GDP per Capita in U.S. States
  - **Metadata URL:** <https://www.kaggle.com/datasets/solorzano/gdp-per-capita-in-us-states>
  - **Data URL:** <https://www.kaggle.com/datasets/solorzano/gdp-per-capita-in-us-states/download?datasetVersionNumber=1>
  - **Data Type:** Zipped CSV
  - **Description:** GDP per capita data from 2013 to 2017, offering insights into individual-level economic prosperity.
  - **License:** Public dataset for research use with attribution required.
  - **Quality:** Accurate and complete for the specified years but limited in temporal scope.
- 

## 3. Data Pipeline

### 3.1. Overview

The pipeline employs an **Extract, Transform, Load (ETL)** process, implemented in Python, to download, clean, validate, and integrate data from multiple sources. It ensures consistency and readiness for analysis.

### 3.2. Stages

1. **Extraction:**
    - Automated download of datasets using the Kaggle API through kagglehub.
    - Data extracted directly in-memory from ZIP files containing CSVs.
  2. **Transformation:**
    - **Cleaning:**
      - Missing values in numerical fields were imputed with column means, and categorical fields with the mode.
      - Rows with entirely null or zero values in critical columns were removed.
    - **Normalization:**
      - Standardized column names across datasets for consistency (e.g., state to State).
      - Removed unnecessary columns to reduce noise (e.g., incident\_url, latitude, longitude in the gun violence data).
    - **Integration:**
      - Merged datasets on state names and aligned data by year for comparative analysis.
    - **Validation:**
      - Ensured non-negative values for GDP and GDP per capita.
      - Verified the completeness of state-level records.
  3. **Loading:**
    - Data was loaded into a SQLite database (us\_gun\_violence\_with\_economy.db) with three tables:
      - Gun\_violence\_data: Processed gun violence data.
      - US\_State\_GDP\_data: State GDP data.
      - GDP\_Per\_Capita\_data: GDP per capita data.
-

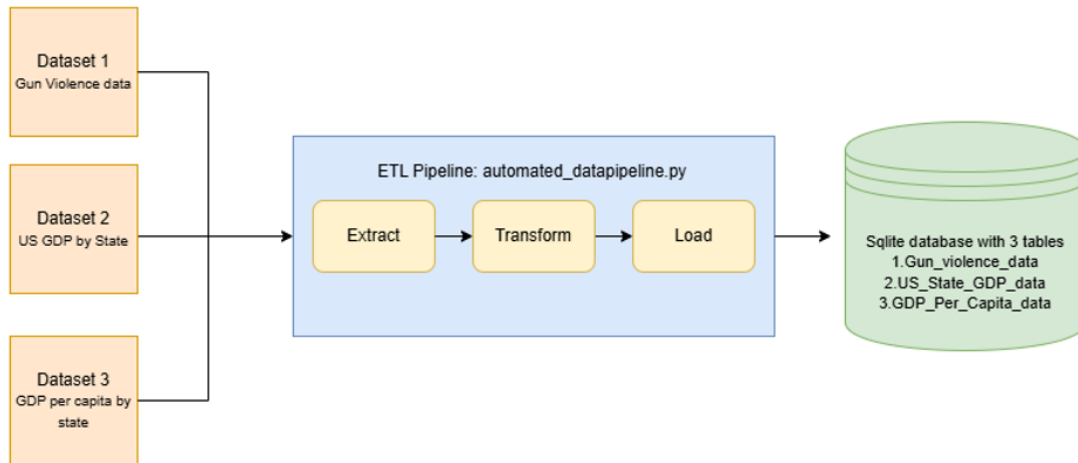


Figure: ETL Pipeline Flow chart

## 4. Results and Limitations

### 4.1. Output Data

- **Format:** SQLite database containing:
  - **Gun\_violence\_data:** Columns include Date, State, City/County, No\_of\_Killed, No\_of\_Injured, and other details.
  - **US\_State\_GDP\_data:** Annual GDP values for each state from 1997 to 2020.
  - **GDP\_Per\_Capita\_data:** GDP per capita values for each state from 2013 to 2017.
- **Quality:**
  - Accurate and consistent after applying imputation and validation.
  - Structured for querying and efficient analysis.

### 4.2. Limitations

- **Temporal Scope:**
  - GDP data ends in 2020, while gun violence data extends beyond this period.
  - GDP per capita is limited to 2013–2017, restricting long-term correlation analysis.
- **Spatial Resolution:**
  - Analysis is state-level, with no granularity at county or city levels.
- **Economic Indicators:**
  - Poverty and unemployment data were not included in this iteration but could enhance analysis if integrated.

### 4.3. Reflections

The pipeline successfully integrates three datasets and produces clean, validated data for analysis. However, limitations in temporal and spatial granularity could impact the precision of insights. Future work could incorporate additional datasets to address these gaps.

---