**USA Gun Violence Data Analysis**

**Abstract**

**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. Furthermore, the dataset could include information about the victims, suspects, and firearms utilized.

Furthermore, 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:**

**Dataset link:**

**OPEN DATASET:**

<https://www.kaggle.com/code/thewiredbear/gun-violence-characteristic-analysis/input>

**Dataset Details:**

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 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.

Detailed Dataset Columns

**Dataset Columns:**

|  |
| --- |
| incident\_id |
| date |
| state |
| city\_or\_county |
| address |
| n\_killed |
| n\_injured |
| incident\_url |
| source\_url |
| incident\_url\_fields\_missing |
| congressional\_district |
| gun\_stolen |
| gun\_type |
| incident\_characteristics |
| latitude |
| location\_description |
| longitude |
| n\_guns\_involved |
| notes |
| participant\_age |
| participant\_age\_group |
| participant\_gender |
| participant\_name |
| participant\_relationship |
| participant\_status |
| participant\_type |
| sources |
| state\_house\_district |
| state\_senate\_district |

**Dataset Records:**

A screenshot of a computer

Description automatically generated

**Methodology:**

**Data Collection:**

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

**Data Preprocessing:**

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.

**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.

Overall, 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.