

Name: Lohith karri

Student ID: 23040177

GitHub Repository: <https://github.com/LohithKarri/Statistics-and-Trends>

Introduction

This exploratory analysis of crime across several U.S. states uses high-quality U.S. crime datasets. The research objective would be to look for any patterns, trends, and insights into crime rates, types of crime, and geographic disparities between the states. The approach here will be to use Python to explore data through statistical methods and visualizations, where findings are communicated effectively to help interpret regional differences in crime. This would be the basis for further detailed analysis, including correlations with socioeconomic factors and changes over time, to make more data-driven policy recommendations. The project will represent crime rate disparities among states using persuasive graphs and charts to help understand the difference more clearly. Analysis on this topic is very significant, serving as a good start to further explorations by correlating crime data with demographic and economic factors in better strategies for preventing crimes and developing policies. In this project, the dataset has been selected from Kaggle based on us crime data.

Distribution of Violent Crimes

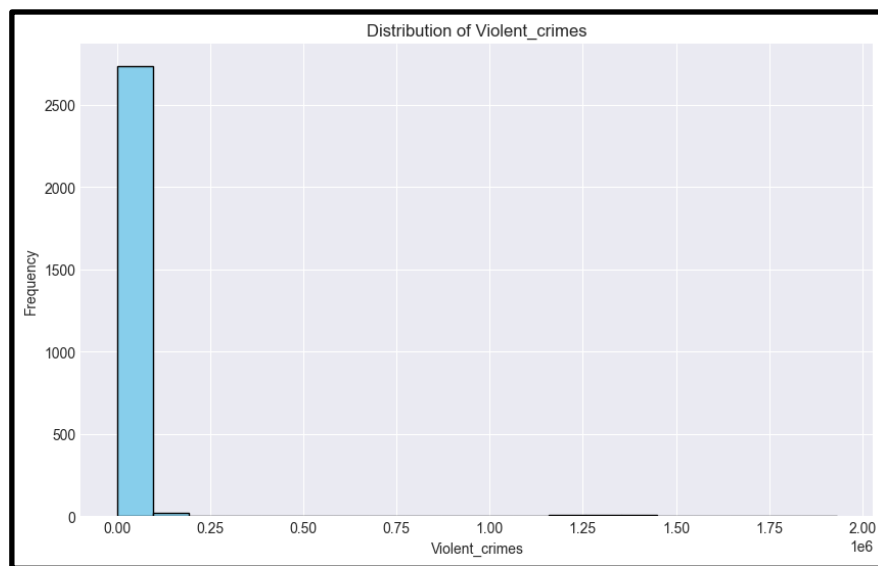


Figure 1: Distribution of Violent Crimes

This code part generates a histogram that shows the distribution of violent crimes from the dataset. The `plt.figure(figsize=(10, 6))` statement initially codes the size of the figure for the graphical results at 10 by 6 inches. Histogram data is created using the `plt.hist()` function, with the `'violent_crimes'` from the DataFrame `df`, and the density of 20 possible bins of progressive range.

The color parameter defines the fill color to be in the form of sky blue while edgecolor places a black boundary around each bin for best perspective. Lastly, plt.title(), plt.xlabel() and plt.ylabel() state the histogram and plt.show() presents the work.

Population vs. Violent Crimes

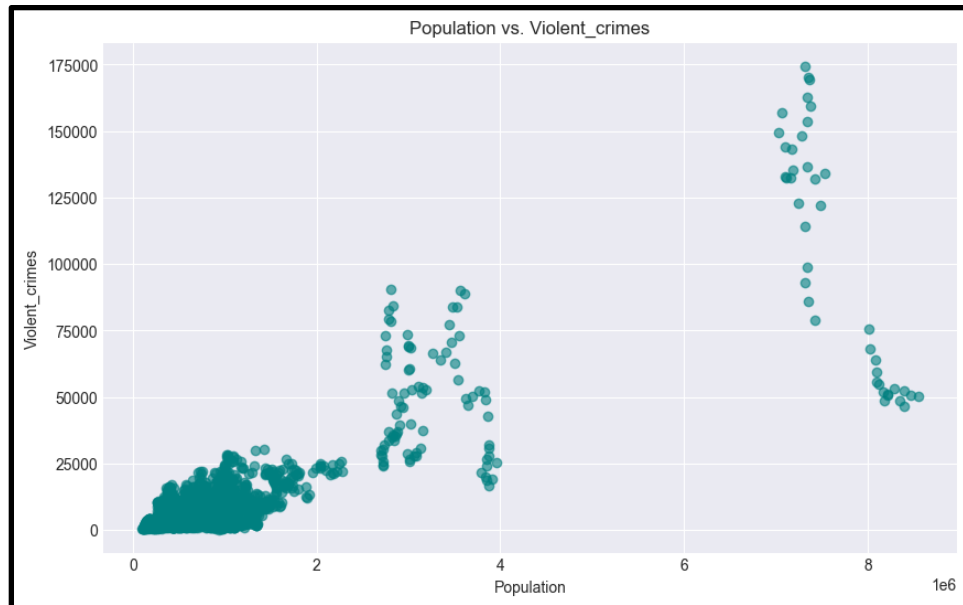


Figure 2: Population vs. Violent Crimes

It shows a scatter plot necessary for statistical exploration of the connection between population and violent crimes. The figsize parameter plt.figure(figsize=(10, 6)) which sets the figure size to 10 by 6 inches. Thus, using the plt.scatter() function, a scatter plot is formed; on the x-axis identified by the label *'population' and on the y axis 'violent_crimes'*. Another parameter, alpha, is assigned a value of 0.6 to make the points hazy, also to allow one to see data points that are overlapping. The point color parameter is set differently than the color parameter and is set as teal this time. Last, the title of the plot is provided with plt.title(), the abscissa is provided with plt.xlabel() and the ordinate is provided with plt.ylabel() whereas the plot is displayed with plt.show().

Heatmap of Correlations between Variables

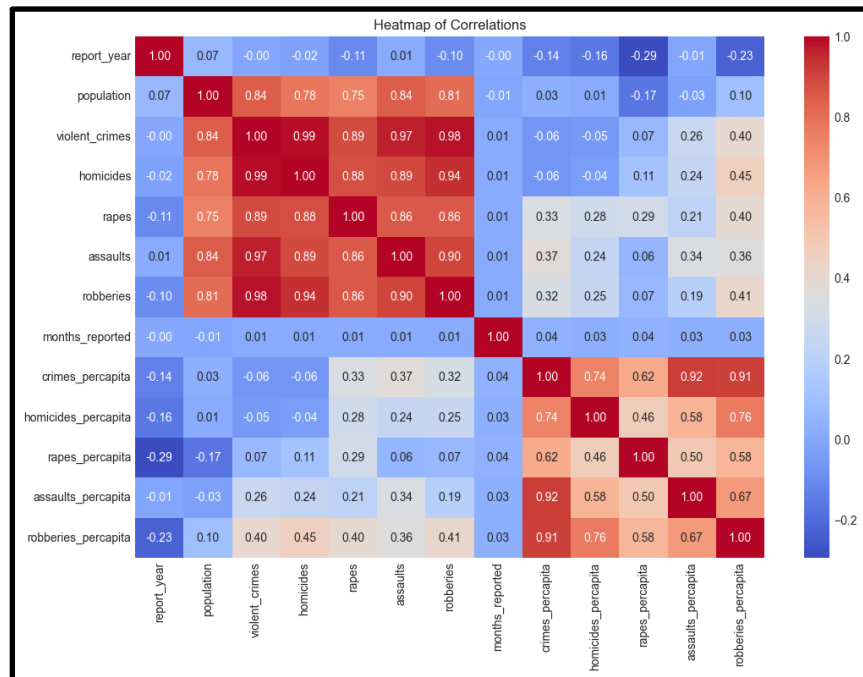


Figure 3: Correlations between Variables

This code creates a heatmap representing the correlation matrix of several variables, which are *report_year*, *population*, *violent_crimes*, *homicides*, *rapes*, *assaults*, *robberies*, and *months_reported*. The command `plt.figure(figsize=(12, 8))` is used to set the dimension of the figure to be 12 by 8 inches. The `sns.heatmap()` function in the Seaborn library will be employed to generate the heatmap through referring to `corr_matrix`. The parameter `annot=True` displays the correlation coefficient on the map. Finally, `cmap='coolwarm'` defines the color palette while `fmt=".2f"` formats the annotation to two decimal places. Lastly, `plt.title()` is used to set title and `plt.show()` for visualizing the heatmap.

Conclusion

In conclusion, this exploratory analysis of crime across various U.S. states lights up the importance of understanding crime trends and patterns in data-driven methodologies. Through high-quality datasets and good statistical tools and visualizations in Python, this report helps successfully illustrate how violent crimes are distributed, how population and crime rates relate to each other, and how there's a correlation between the different variables related to crime. This histogram is used to show the frequency of violent crimes, and the scatter plot indicates the relation of population size with the incident of crime. A more advanced heatmap further defined multiple

interdependencies among variables that policymakers could use as critical inputs. These findings also not only help understand better regional crime disparities but also open up the path towards further analyses that can make use of socioeconomic factors. This study thus emphasizes the use of data that informs crime prevention strategies and policy development for safer communities in the nation.

Reference

Dataset **link:** <https://www.kaggle.com/code/zikazika/analysis-of-world-crime/input?select=report.csv>

