**Data Science Toolbox: Python Programming  
PROJECT REPORT**  
**Violent crim**e rates across different demographics and geographies in **California** from 2000 to 2013  
  
Submitted by  
Priyanka Bisht  
Registration No: 12310565  
B.Tech CSE - Section K23GX  
Course Code: INT375

Discipline of CSE  
Lovely School of Computer Science and Engineering  
Lovely Professional University, Phagwara

# DECLARATION

I, Priyanka Bisht, student of B.Tech CSE under CSE Discipline at Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.  
  
Date: 12-4-2025  
  
Signature: Priyanka   
Registration No. 12310565  
Name: Priyanka Bisht

# ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my guide, Madam Gargi Sharma, for her invaluable support, guidance and encouragement throughout the project. I would also like to thank the faculty members and staff of the Department of CSE for their support and help.

# TABLE OF CONTENT

1. Introduction  
2. Source of dataset  
3. EDA process  
4. Analysis on dataset  
5. Conclusion  
6. Future scope  
7. References

# 1. INTRODUCTION

The purpose of this project is to perform an exploratory data analysis (EDA) on a dataset containing information about violent crime rates across different regions in California between 2000 and 2013. The analysis aims to explore the relationship between crime rates and factors like population size, race/ethnicity, and geographical distribution. Additionally, outlier detection and trends over time were analyzed to better understand the underlying patterns and correlations within the data.

# 2. SOURCE OF DATASET

The dataset used for this project is provided in the file named hci\_crime\_752\_pl\_co\_re\_ca\_2000-2013\_21oct15-ada (1). It includes details about commercial buildings across different states in the USA.

Dataset Link: https://catalog.data.gov/dataset/violent-crime-rate-9a68e

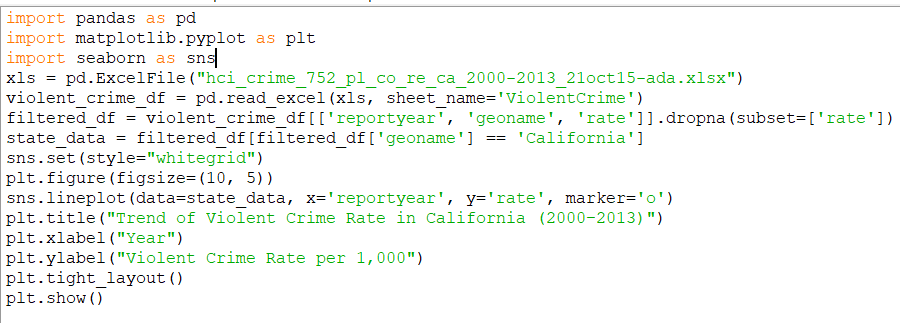
# 3. EDA PROCESS

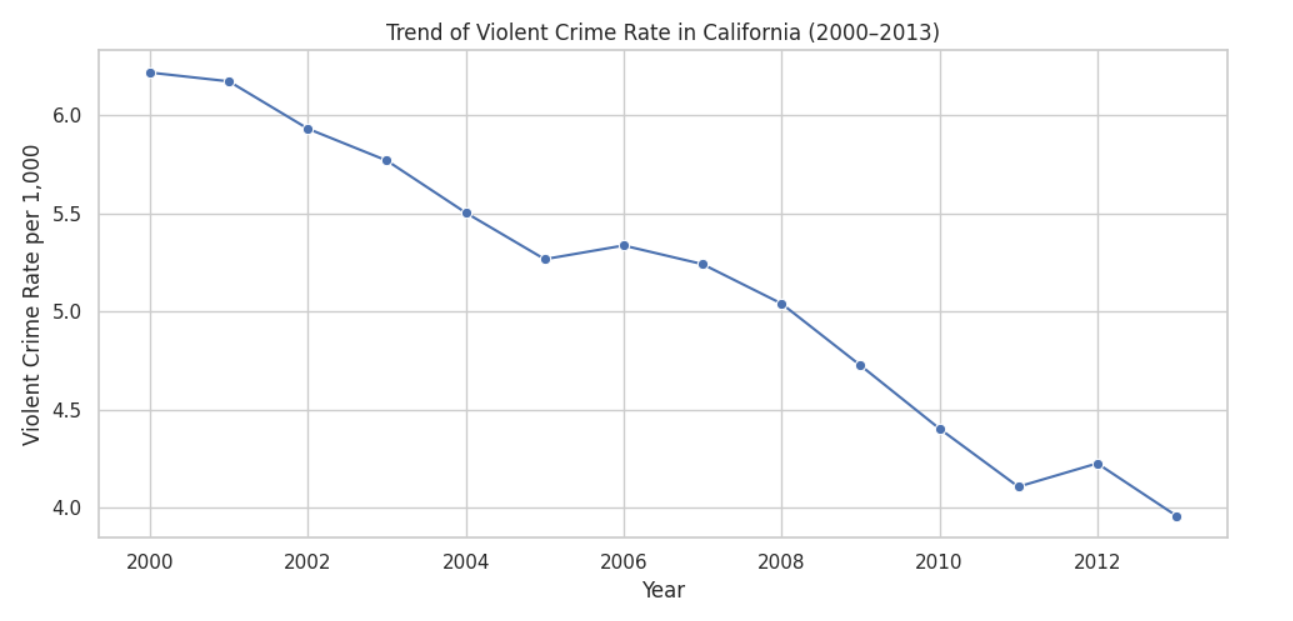
Exploratory Data Analysis (EDA) involves summarizing the main characteristics of the dataset using statistical graphics, plots, and information tables. The analysis was done using Python with pandas, matplotlib, and seaborn libraries.

# 4. ANALYSIS ON DATASET

## 4.1 Identify Crime Rate Trends Over Time

This analysis Visualizes how violent crime rates have changed in California or individual counties over the years.

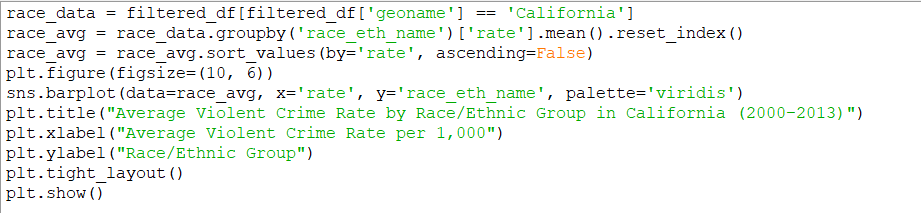


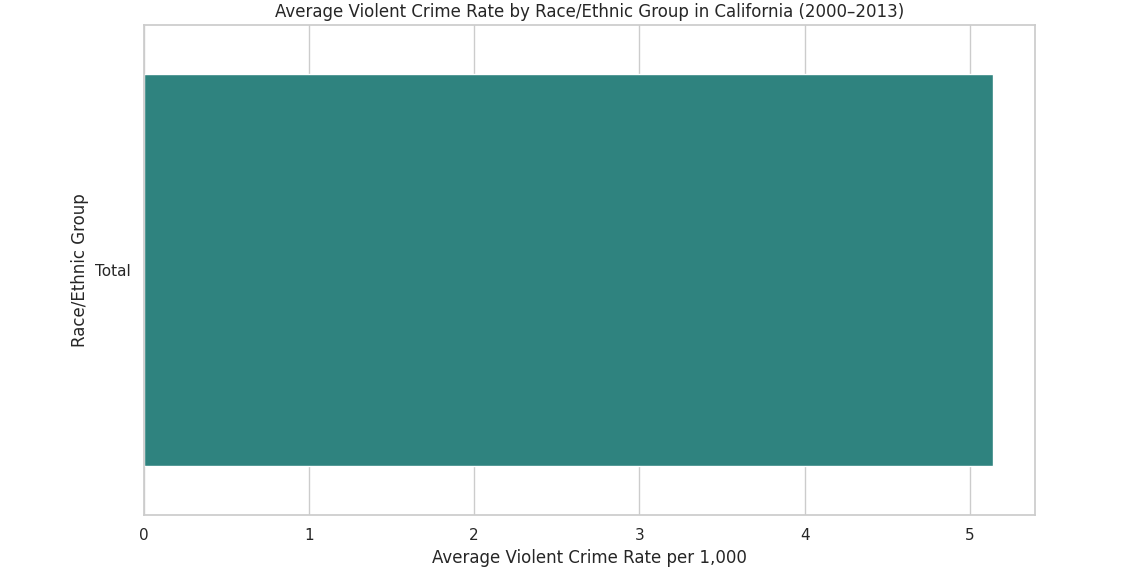
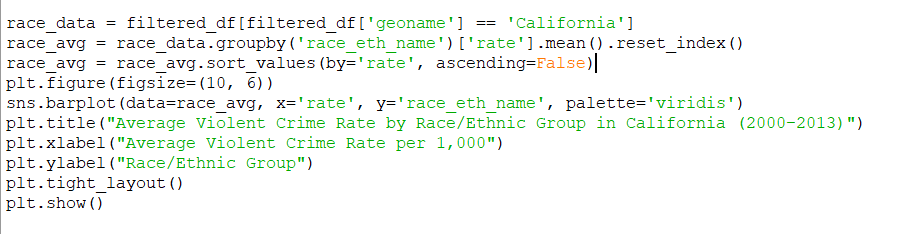
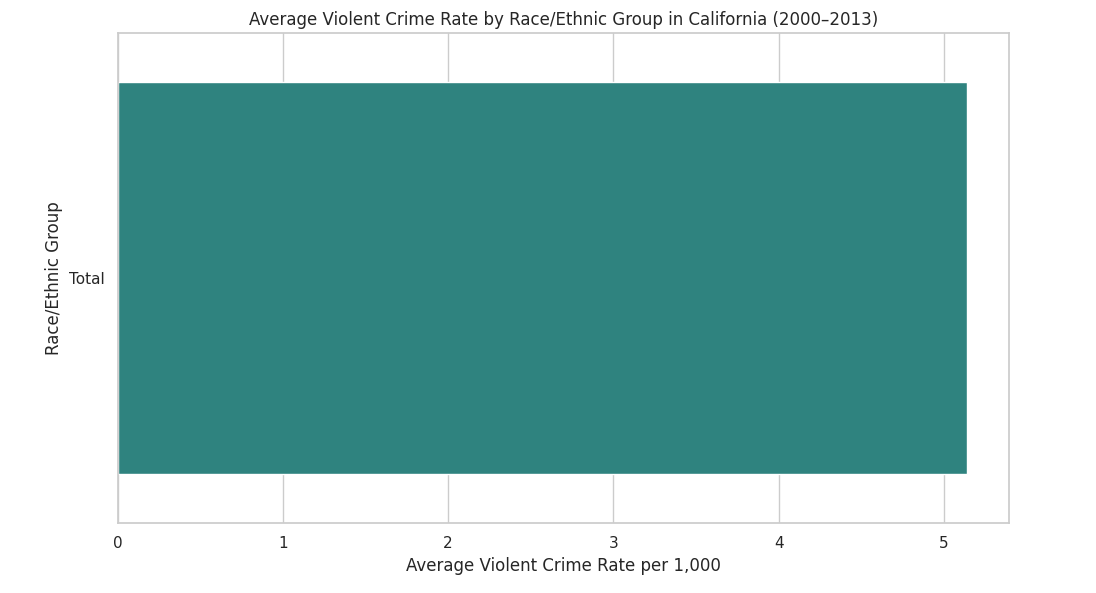


*Figure 1: This figure shows trend of violent crime rate in California(2000-2013)*

## 4.2 Compare Crime Rates Across Counties or Ethnic Groups

Understanding regional and demographic differences in violent crime rates by comparing means, medians, and distributions.

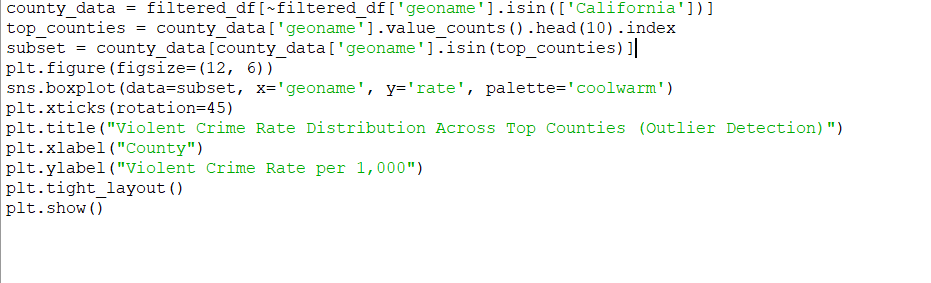


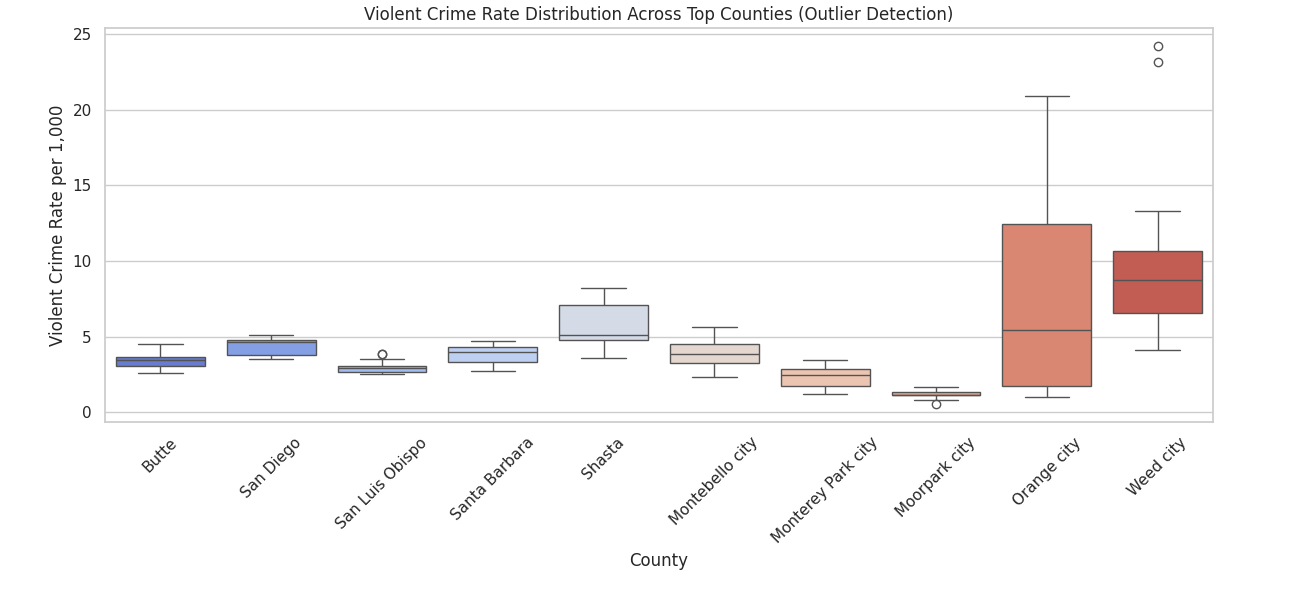
  
  
  
  
  
  
  
  


## *Figure 2: Compare Crime Rates Across Counties or Ethnic Groups*

## 4.3 Detect Outliers in Crime Rates Using Box Plot

Visualizing crime rate outliers across different counties in California using a box plot.

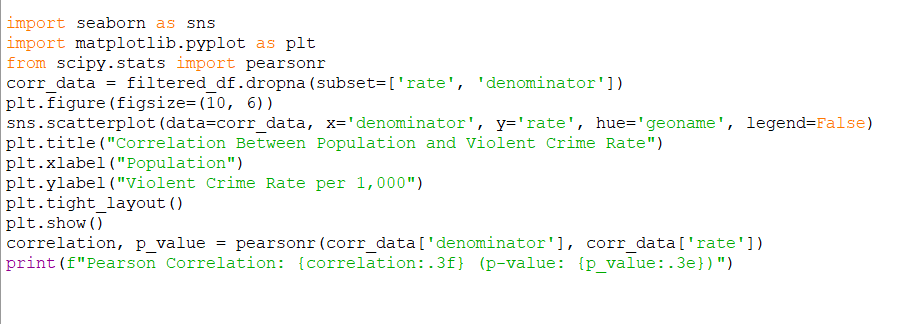


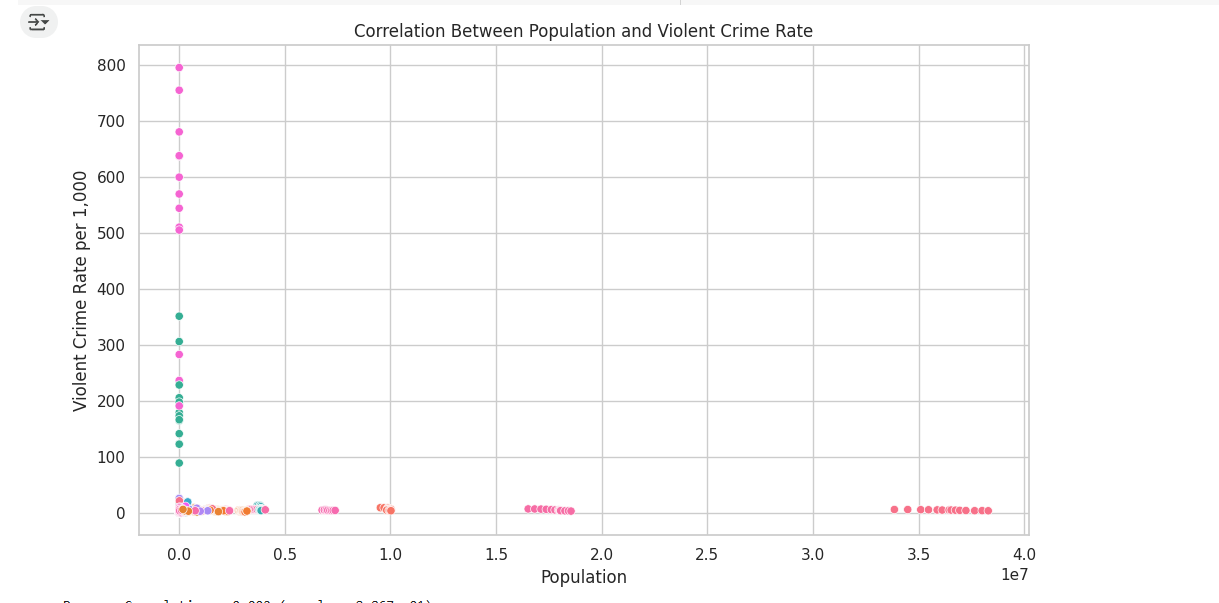


*Figure 3: crime rate outliers across different counties in California*

## 4.4 Analyze Correlation Between Crime Rate and Population

This objective explores whether there’s a relationship between population size and the violent crime rate using a **scatter plot** and **correlation coefficient**.

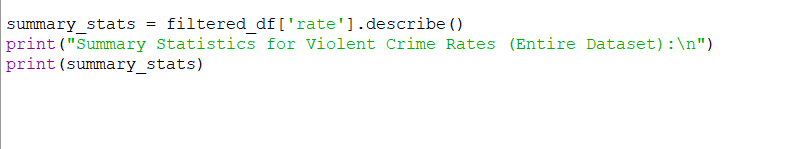


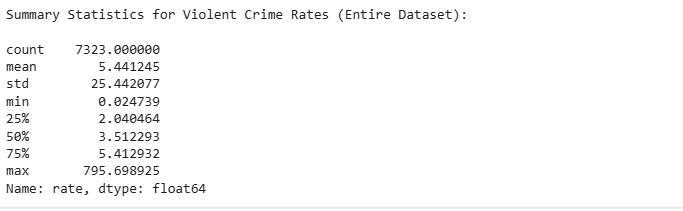
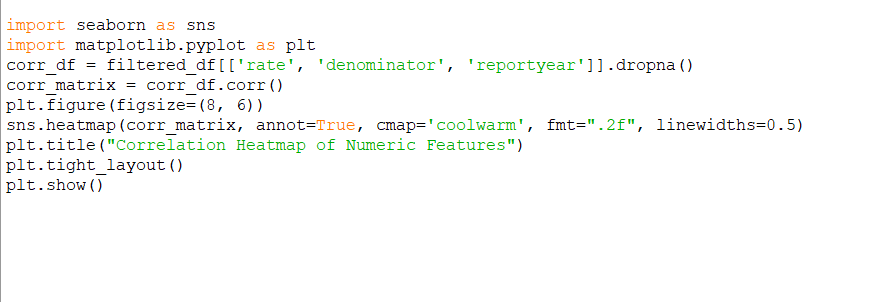
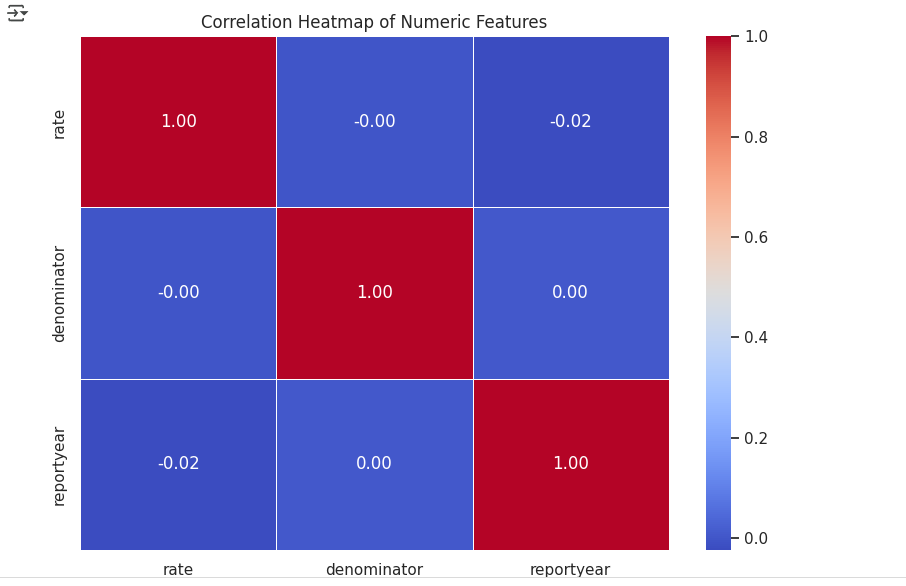


*Figure 4: Correlation Between population size and the violent crime rate*

## 4.5 Generate Summary Statistics for Violent Crime Rates

Gives a quick descriptive statistics like mean, median, min, max, and standard deviation for crime rates in your dataset.



*Figure 5: statistics*

# 5. CONCLUSION

The primary objective of this project was to perform a comprehensive exploratory data analysis (EDA) on a dataset containing information about violent crime rates in California, spanning from 2000 to 2013. Through a variety of visualizations and statistical techniques, several important insights were extracted regarding the trends in crime rates, regional differences, outlier behaviors, and the relationship with population size.

The line plot depicting the trend of violent crime rates in California from 2000 to 2013 revealed fluctuations in crime over the years, highlighting periods of increase or decrease in violent crime incidents. Further examination using bar charts comparing crime rates across different ethnic groups in California showed notable disparities, indicating the influence of demographic factors on crime rates.

Box plots helped identify counties with unusually high or low violent crime rates, enabling the detection of outliers that might require additional investigation. The scatter plot analysis of the relationship between population size and violent crime rates showed that larger populations often correspond to higher crime rates, though with varying intensity across counties.

Finally, summary statistics for the entire dataset, as well as for specific years, counties, and ethnic groups, provided a detailed statistical overview, identifying central tendencies, variability, and any anomalies in the data.

Overall, this project demonstrated how EDA techniques, including visualization and statistical analysis, can uncover significant patterns and provide actionable insights for policymakers, law enforcement, and researchers interested in understanding crime dynamics.

# 6. FUTURE SCOPE

Future studies could explore deeper into the temporal aspect of crime data, examining how crime rates have evolved over longer periods or correlating with external factors such as economic downturns or legislative changes.

# 7. REFERENCES

<https://www.w3schools.com/python/>

<https://data.gov/>

GITHUB LINK: https://github.com/Prince087/Analysis-of-Commercial-Building-Dataset-in-the-USA