



TASK

Exploratory Data Analysis on the UsArrests Dataset

[Visit our website](#)

Introduction

I was given UsArrests.csv to analyse. This dataset has details about crime in US cities. Most of the data in the csv file is numeric in nature. Using various data analysis tools I have explored this dataset and I am presenting the findings in this report.

This report includes findings on the different populations in the United States and the amount of crime incidents in each city.

DATA CLEANING

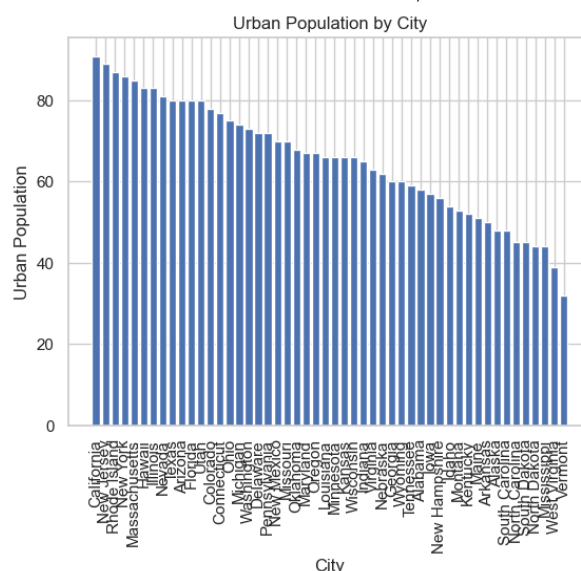
I used pandas to clean my data. Pandas was used to drop rows and columns I did not need, merge rows into different datasets and create new datasets. Matplotlib and Seaborn were used to create visualisations

MISSING DATA

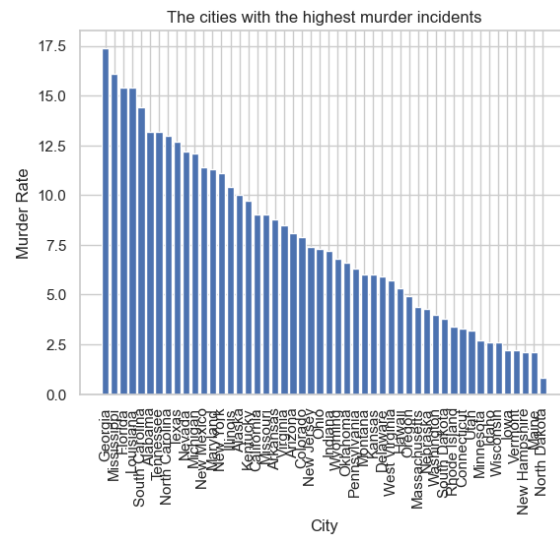
There was no missing data in this dataset

DATA STORIES AND VISUALISATIONS

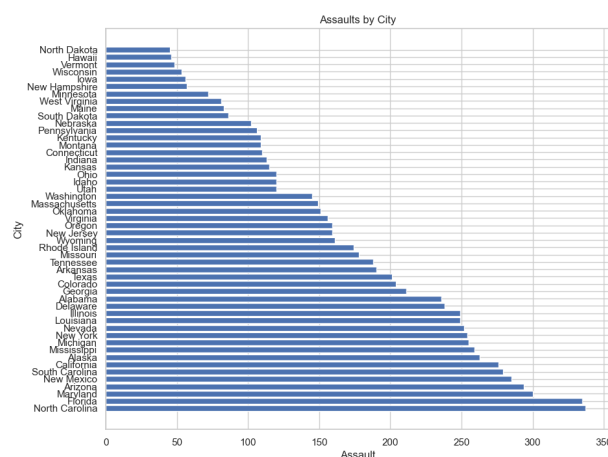
First I set out to explore which cities are the most populated within this dataset. After applying some data analysis methods I have discovered that the top 3 most populated cities are: California, New Jersey and Rhode Island.



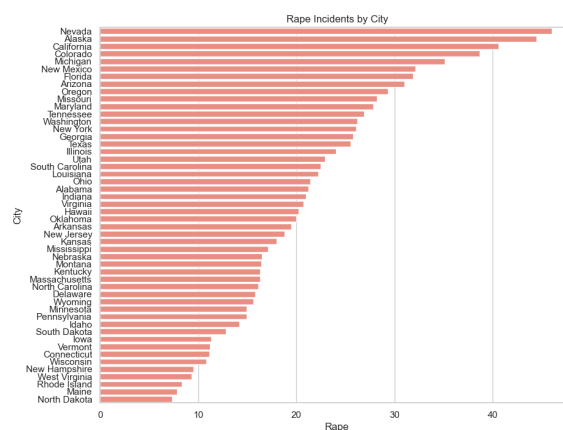
I then went on to explore the different crimes in the different US populations. I discovered that the top 3 cities with the most murder incidents are: Georgia, Mississippi and Florida.



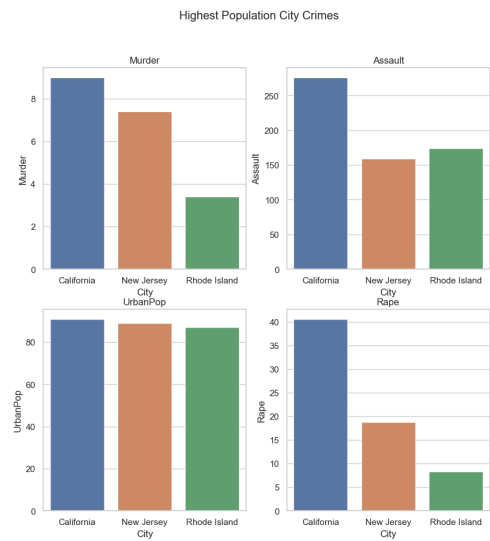
I discovered that the top 3 cities with the highest assault rates are: North Carolina, Florida and Maryland



I discovered that the top 3 cities with the highest rape incidents are: Nevada, Alaska and California.



Once I was done exploring the crime incidents I decided to explore the crime incidents in the top 3 most populated cities. What I discovered is that not only is California not the most populated city in the dataset, they also have the highest crime incidents out of all the top 3 most populated cities.



THIS REPORT WAS WRITTEN BY :Pule Molefe

