

**TASK**

**Exploratory Data Analysis on the Us Arrests Dataset**

[](http://www.hyperiondev.com/portal/)

**Introduction**

The "UsArrests" dataset is a collection of crime statistics from cities across the United States. It includes the following columns:

"City": The name of the city

"Murder": The number of murders in the city

"Assault": The number of assaults in the city

"UrbanPop": The percentage of the population living in urban areas

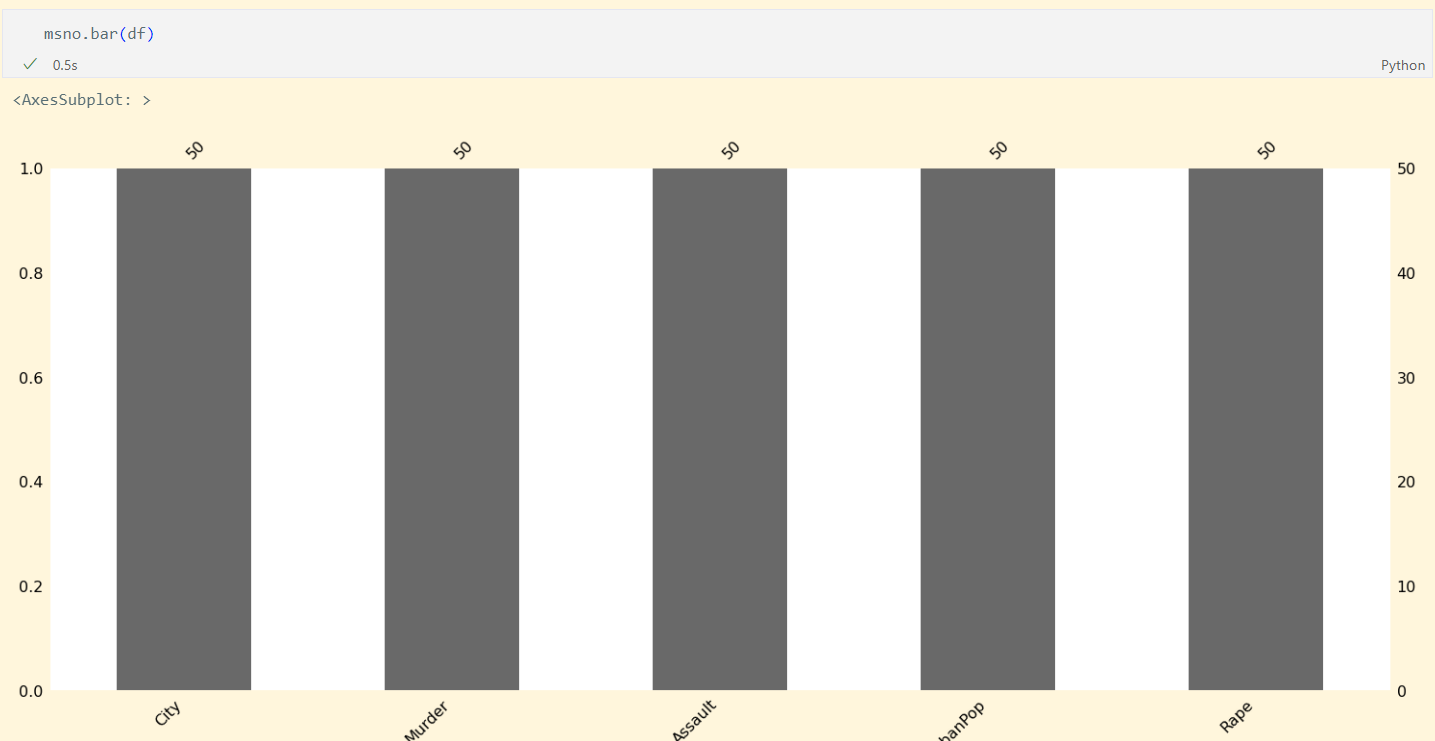
"Rape": The number of rapes in the city

In summary, the "UsArrests" dataset provides information on the crime rates in different cities across the United States, including the number of murders, assaults, and rapes, as well as the urban population of each city

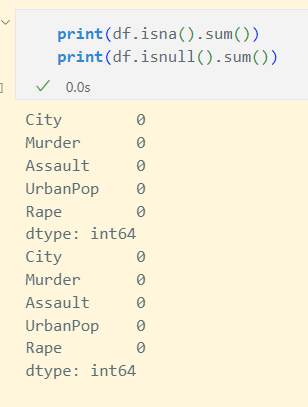
**DATA CLEANING**

**First I used the .head and .shape to look at the first 5 rows and then determine the length and width of the data set. After this I used missingno as msno to use msno.bar()**

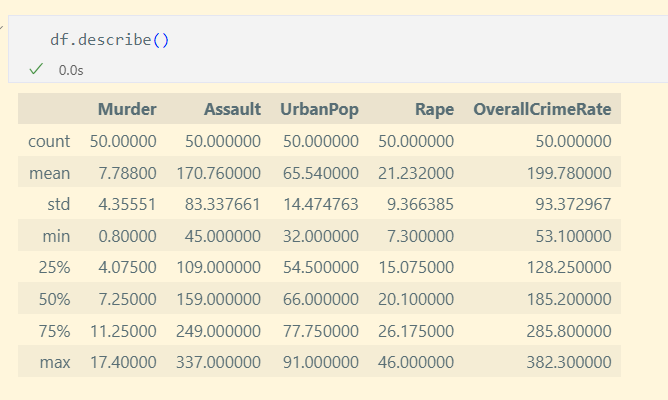
**The plot created showed no missing data.**

****

**Then I double checked the data for missing/erronious enteries with the .isna() and .isnull() methods**

****

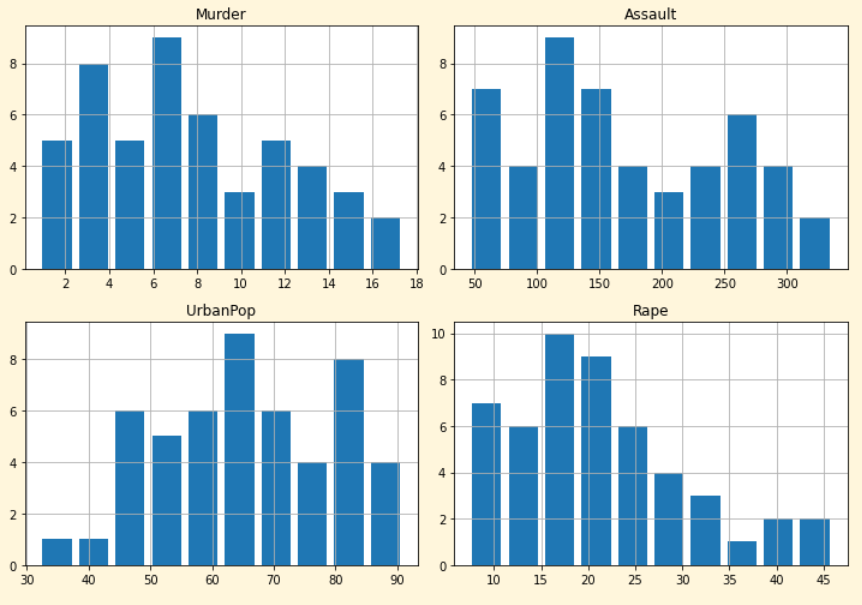
**Finally I used the . describe function to get an overview of the data set**

****

**MISSING DATA**

N/A

**DATA STORIES AND VISUALISATIONS**

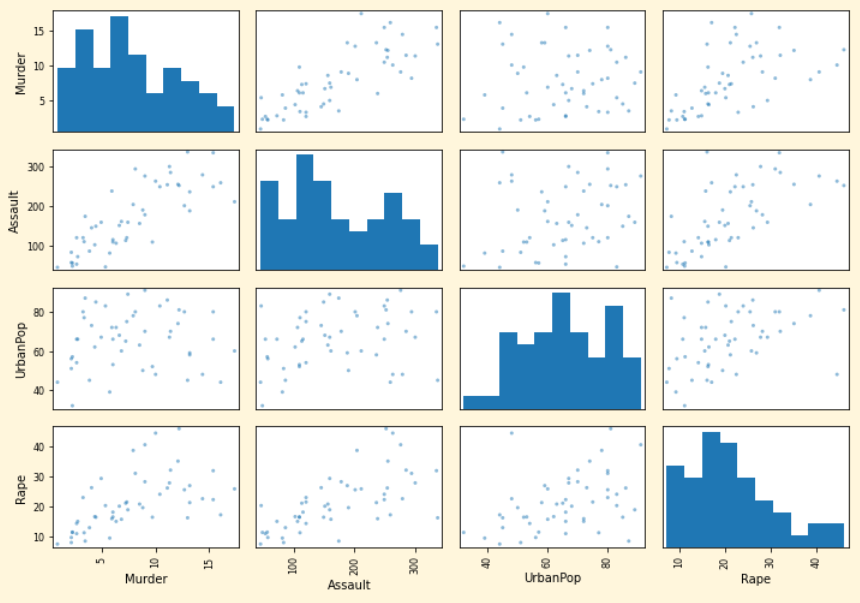


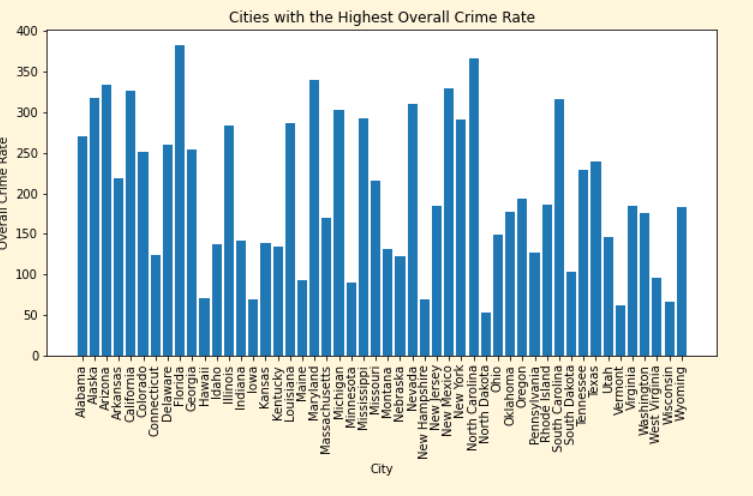
The histogram presents the distribution of data in the numerical columns of the "UsArrests" dataset. Each histogram represents the frequency of cities with a specific value in the corresponding column. For instance, the histogram for the "Murder" column displays the distribution of the number of murders in different cities.

The shape of the histogram offers insights into the distribution of the data. It can help determine whether the data is symmetrical, skewed to the left or right, or has multiple modes. These characteristics of the histogram can provide valuable information about the nature of the data and can be used to inform data analysis and modeling.

For example, a symmetrical histogram would suggest that the data is evenly distributed, while a skewed histogram would indicate that the data is not evenly distributed and has a skewed distribution towards one side. A multi-modal histogram, on the other hand, would indicate that the data has more than one peak and is not well described by a single central tendency or statistical model.

In summary, the histogram provides a visual representation of the distribution of data in the numerical columns of the "UsArrests" dataset. By examining the shape of the histogram, we can gain insights into the distribution of the data and inform data analysis and modeling. Below I charted the same data in a scatter format.



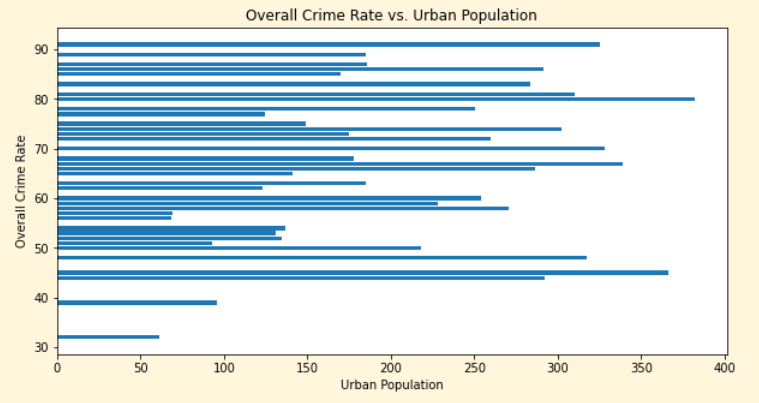


This chart provides information on the cities with the highest overall crime rates. It gives us a clear understanding of which cities are safer and which cities have higher crime rates.

According to the chart, North Dakota is considered to be the safest city, with a relatively low overall crime rate. This is an important piece of information for individuals looking to move to a new city or for policymakers looking to allocate resources to areas with high crime rates. By identifying North Dakota as a safe city, it can be a valuable resource for people who are looking for a safer place to live or work.

On the other hand, the chart also highlights Florida as the most dangerous city, with a relatively high overall crime rate. This information is important for individuals and organizations looking to reduce crime and increase public safety. By understanding the crime situation in Florida, they can prioritize their efforts and allocate resources to address the specific crime challenges in that city.

It is important to note that the chart represents only a snapshot of the crime situation and may not reflect long-term trends. Nevertheless, the information provides a valuable tool for individuals and organizations looking to reduce crime and increase public safety. By understanding the cities with the highest overall crime rates, they can make informed decisions and develop effective strategies to reduce crime and promote public safety.

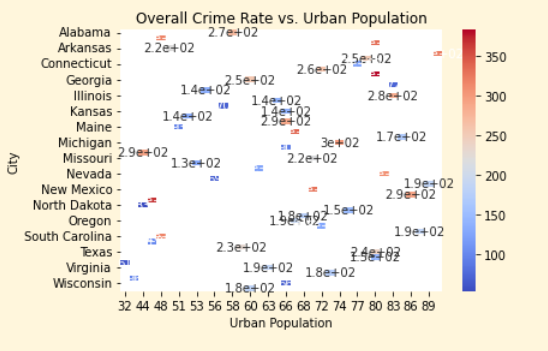


This data analysis is looking at the relationship between urban population and overall crime rate in various cities. By examining the chart, we can see a correlation between the two factors. In general, cities with lower urban populations tend to have lower overall crime rates. This suggests that urban population can have an impact on crime rate.

However, this correlation is not absolute and there are exceptions to the pattern. Some cities with smaller urban populations still have higher crime rates. This implies that other factors, such as socioeconomic status, policing strategies, and cultural norms, also play a role in determining crime rates.

It is important to consider these nuances when evaluating crime trends and devising solutions to reduce crime. While urban population may have some influence on crime, it is not the sole determining factor. To effectively address crime, a comprehensive and multi-faceted approach must be taken.

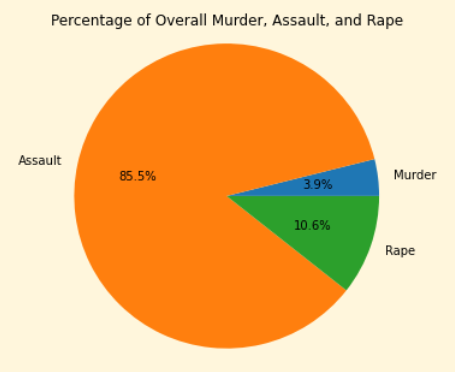
Below is another way to visualize the same data “Overall Crime Rate vs. Urban Population” some charts and plots can be harder for some people to read than others, for this reason I included the heat map below “Overall Crime Rate vs. Urban Population”

****

**The heat map created above is a representation of the relationship between "UrbanPop" and "OverallCrimeRate". Each cell in the heat map represents a unique combination of "UrbanPop" and "OverallCrimeRate". The color of each cell is determined by the value of "OverallCrimeRate", with warmer colors (e.g. red) representing higher values, and cooler colors (e.g. blue) representing lower values.**

**In this particular heat map, "UrbanPop" is represented on the x-axis and "City" is represented on the y-axis. The annotations inside each cell show the actual value of "OverallCrimeRate". By observing the pattern of colors in the heat map, you can see how "OverallCrimeRate" varies with "UrbanPop" across different cities.**

**For example, if you see a warm color (e.g. red) in a cell located at the intersection of a high "UrbanPop" value and a certain city, this suggests that the overall crime rate is high in that city with a high population. On the other hand, if you see a cool color (e.g. blue) in a cell located at the intersection of a low "UrbanPop" value and a certain city, this suggests that the overall crime rate is low in that city with a low population.**

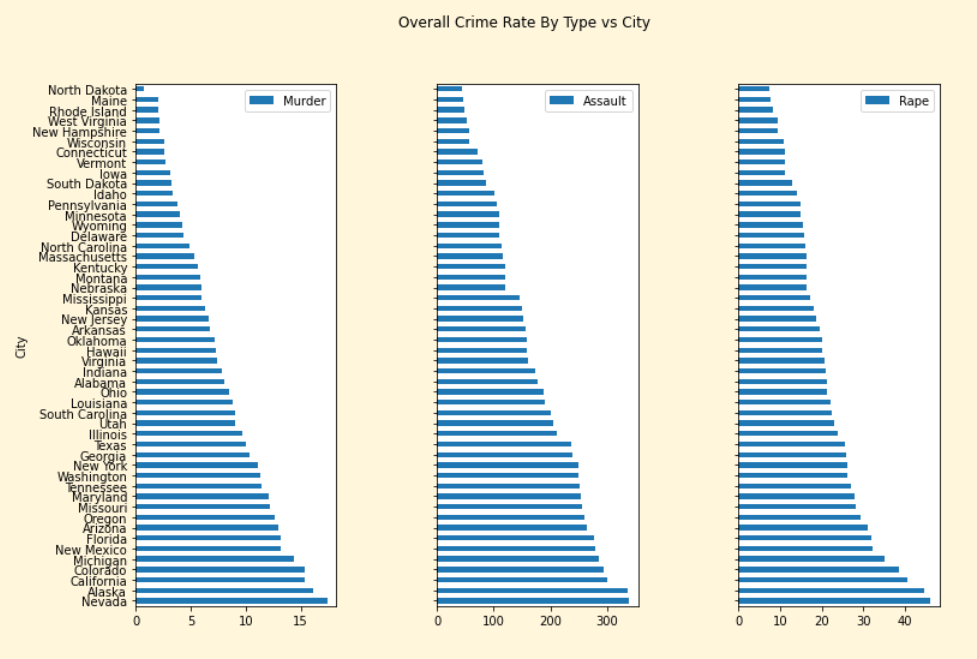


The pie chart provides a visual representation of the distribution of crime rates across three different categories: Murder, Assault, and Rape. By examining the chart, it is immediately clear that Assault is the most prevalent crime category, accounting for 85.5% of the overall crime rate for the given data set.

In comparison, Murder and Rape have much smaller proportions, representing 3.9% and 10.6% of the overall crime rate respectively. This information gives us a clearer understanding of the composition of crime in the dataset and allows us to focus on the areas that require the most attention.

The high proportion of Assault in the data set highlights the need to focus on reducing the rate of this particular crime category. By reducing the rate of Assault, it would likely result in a reduction of the overall crime rate.

It is important to note that the pie chart represents only a snapshot of the crime situation and may not reflect long-term trends. Nevertheless, the information provides a valuable tool for individuals and organizations looking to reduce crime and increase public safety. By understanding the distribution of crime rates across categories, they can prioritize their efforts and allocate resources to the areas that require the most attention.

****

The charts above provide valuable insight into the correlation between various crime types and cities. The data indicates that there is a relationship between the various types of crimes and that when one type of crime is low, the others tend to be low as well. This suggests that crime is not isolated to one particular category but rather is interconnected in some way.

Additionally, the charts reveal that there is no instance of high crime in one category while the others are low. This further underscores the interconnectedness of crime and highlights the importance of considering multiple crime categories when evaluating crime trends and devising solutions to reduce crime.

The charts also highlight the relative safety of cities. North Dakota is identified as the safest city while Nevada is considered to be the most dangerous. This information can be useful for individuals looking to move to a new city or for policymakers looking to allocate resources to areas with high crime rates.

It is important to note that the charts provide a snapshot of the situation at a specific point in time and may not necessarily reflect long-term trends. Nevertheless, the information can be a valuable tool for individuals and organizations looking to reduce crime and increase public safety.

**THIS REPORT WAS WRITTEN BY : Samuel Blackhurst**

