

**TASK**

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

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

**Introduction**

For this task, I have chosen the UsArrests dataset. Once the dataset has been loaded, the data will be cleaned, any missing data will be handled appropriately.

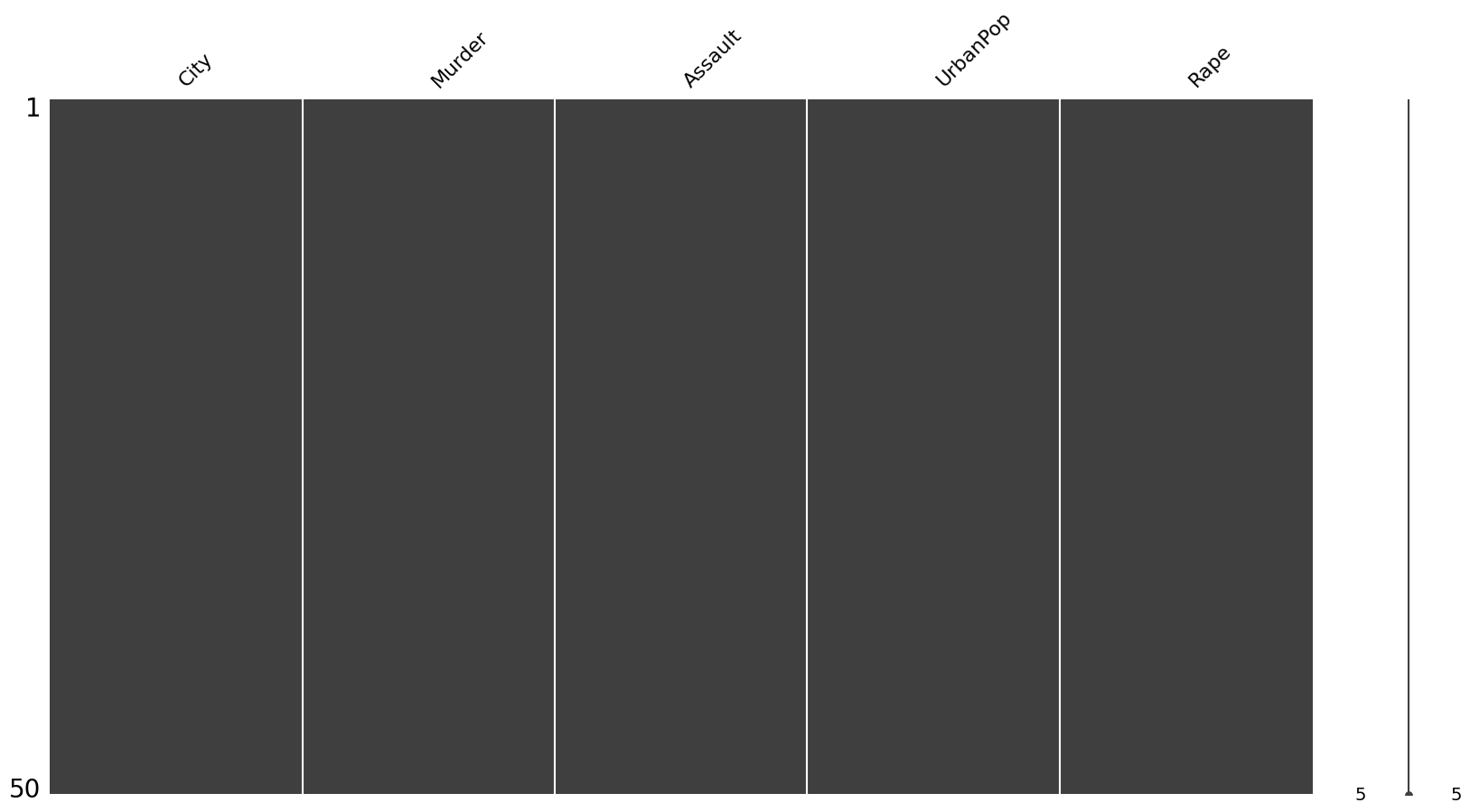
Following this, the data will be analysed using visualisations followed by assumptions and findings from the data.

**DATA CLEANING**

Using the .head() function, I believe all columns will be required for the handling of the data. No columns will be dropped.

**MISSING DATA**

Using the missingno.matrix function, I have managed to identify that there aren’t any rows that are missing values.

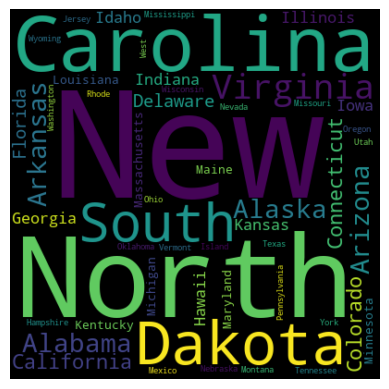


No data needs to be added / removed due to missing data

I have also performed a print statement to advise of the data types we are working with, any conversions would be completed at this stage, however upon review of the types, I feel there is no need to convert any of the types.

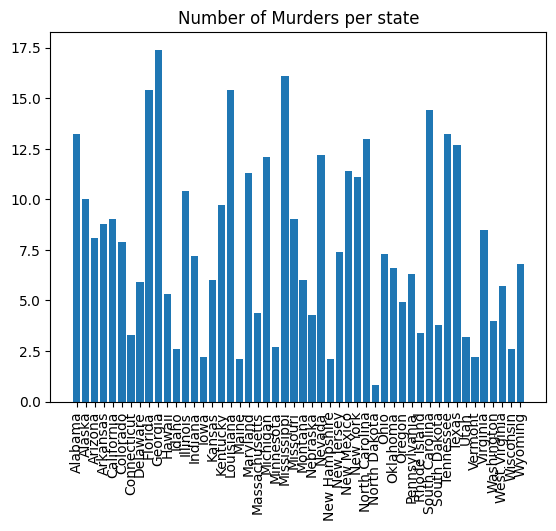
**DATA STORIES AND VISUALISATIONS**

There are 50 states / cities in the list which has been visualised by a wordcloud.



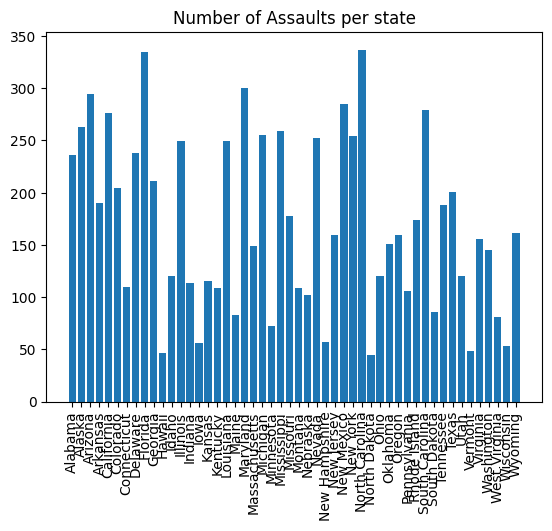
From the data we can see that North Dakota has the lowest Murder rate at 0.8, followed by New Hampshire at 2.1. The state with the highest Murder rate is Georgia at 17.4.

This can be seen by the bar chart that has been plotted



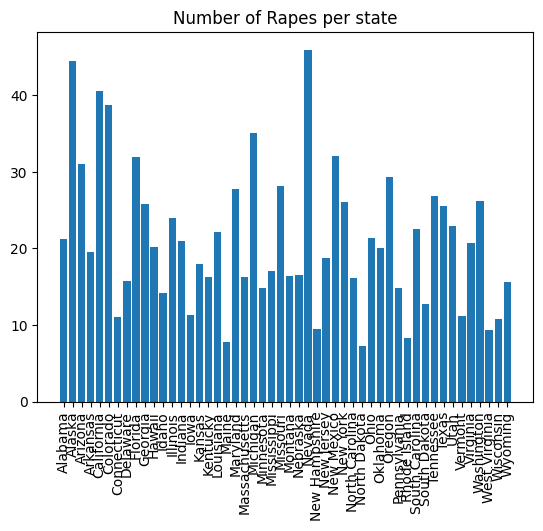
From the data we can see that North Dakota has the lowest Assault rate at 45, followed by Hawaii at 46. The state with the highest Assault rate is North Carolina with a rate of 337.

This can be seen by the bar chart that has been plotted.



From the data we can see that North Dakota has the lowest Rape rate at 7.3, followed by Maine at 7.8. The state with the highest Rape rate is Nevada with a rate of 46.

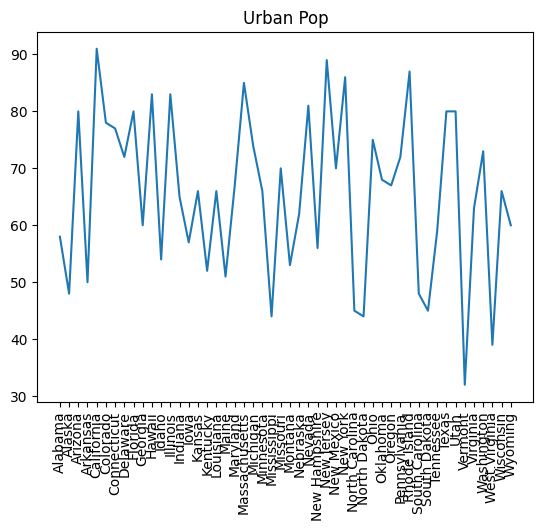
This can be seen by the bar chart that has been plotted.



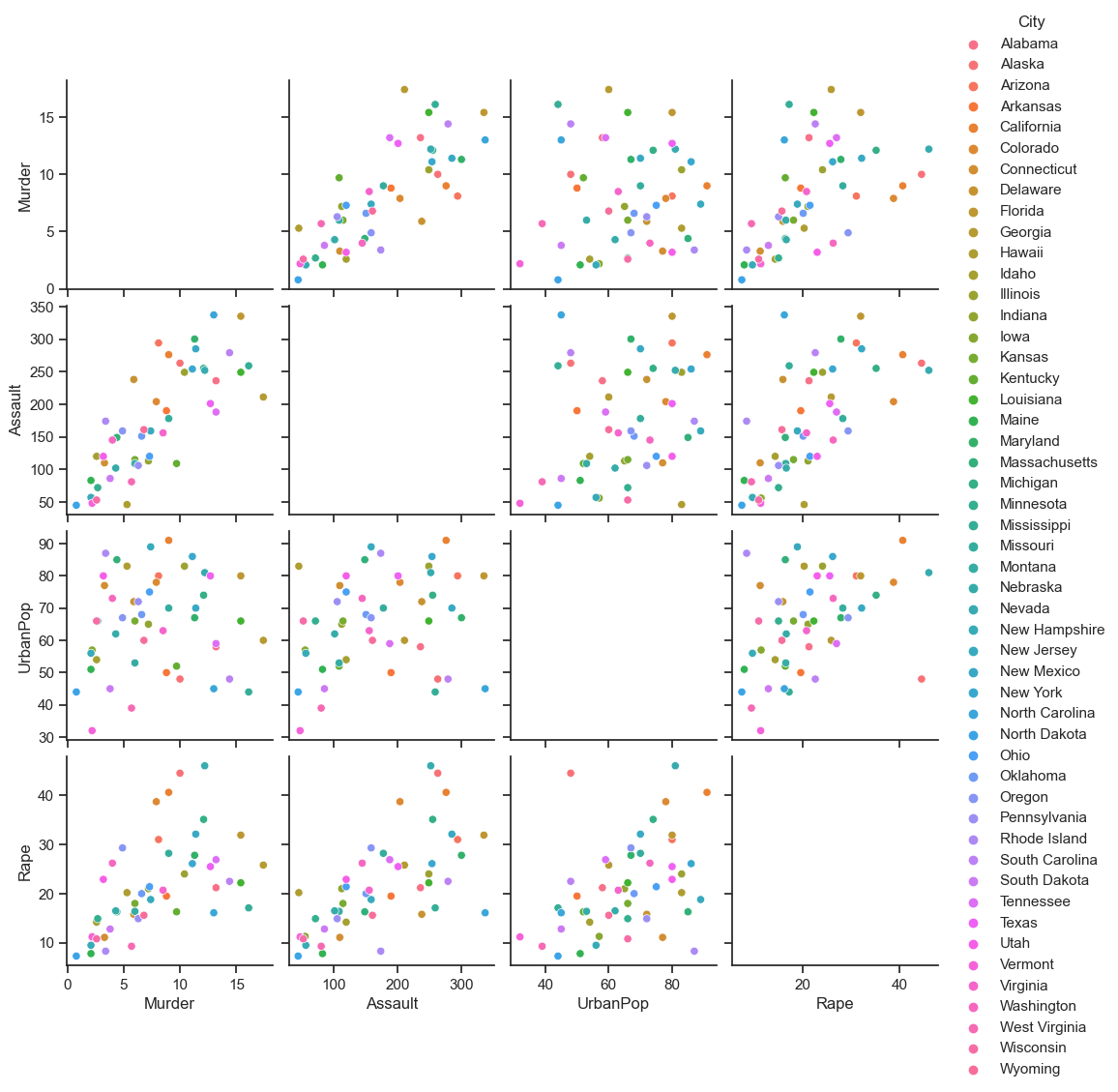
From the data we can see that California has the highest UrbanPop with a percentage of 91.

The state with the lowest UrbanPop is Vermont with a rate of 32 percent.

This can be seen by the line chart that has been plotted.



Finally a scatter matrix has been plotted to compare all Murders, Assaults, Rapes and UrbanPop.



When looking at this data, one pattern observed was;

* Rape vs Murder / Assault / UrbanPop all seem to plot similar values at a similar point within the graph.
* The Assault vs Murder graph follows a line of best fit at a 45 degree angle on the graph. This assumes that the more assaults mean more murder.

One assumption to make from this data is that North Dakota is one of the safest states/ cities to live due to the low levels of Murder/ Assault and Rape rates. North Dakota has one of the lower levels of UrbanPop which could explain the reasoning behind this.

One point to make from this data, although California has a higher UrbanPop, with a high rate of Assaults, the Murer and Rape rates are relatively low, when compared to another state such as Alaska which has a higher rate of all 3 crimes with a lower UrbanPop. This would assume that having a low UrbanPop doesn’t mean the levels of crime will be low.

**THIS REPORT WAS WRITTEN BY : Chris**

