**Stage-2 Member task**

**Raveena Arasikere Rakesh**

1. In stage-2 the first task was to identify the type of distribution across the dataset. Here, we understood whether the data is continuous or discrete by plotting a histogram of the normalized death rate.

**Chart, histogram

Description automatically generated**

From the above histogram we can see that the data is **discrete** and then we calculate the measure of centers, median and mean using the mean () and median () default python functions. And their values are,

* **Mean**: 10.140316205533598
* **Median:** 9.0

1. Then we create a histogram for the entire US dataset, also plot the mean and median as the vertical lines in on the histograms.

Chart, histogram

Description automatically generated

Chart, histogram

Description automatically generated Chart, histogram

Description automatically generated

1. Next, we understand the mortality to other variables relationship, by codifying the mortality rates into different quantiles. And the quantile distribution is as follows,

0.1,7.0,10.0,15.0---‘Very\_low','Low','High','Vey\_high'

(Where, 'Very\_low' < 'Low' < 'High' < 'Vey\_high')

1. Once we have divided the mortality rates into different quantiles, we then compare these against different variables, we have chosen 4 variables (Total female population raw value’, ‘Total male population raw value','Opiod\_Dispensing\_Rate',

'Population') for comparison and plotted the graphs against all 4 variables.

Chart, scatter chart

Description automatically generated

* When we compare the opiod mortality rate against the **total male population**, we can see that the highest male population falls under the very\_high quantitle, which means that the opiod mortality rate is higher where the male population is higherChart, scatter chart

  Description automatically generated
* When we plot the moratlity rate vs **the opiod dispensing rate** graph we can see that the highest opioid dispensing rate falls under the very\_high quantile range, meaning that the death rate increases as the opioid dispensing rate increases. Here, our initial hypothesis holds true.

Chart, scatter chart

Description automatically generated

* According, to the graph plotted above we can see that the highest population rate belong to the very\_low quantitle range, meaning that the opioid mortatlity rate is not influenced by the population and our hypothesis was wrong.

Chart, scatter chart

Description automatically generated

* When we compare the opiod mortality rate against the total female population, we can see that the highest female population falls under the very\_high quantitle, which means that the opiod mortality rate is higher where the female population is higher.