## Stage\_2 Member Task Report

**Task 1: Create a histogram of Normalized Deaths variable for entire US as a distribution.**

* Add a vertical line to the histogram representing the mean Normalized Opioid Death for US
* Add another line for median to compare

For this task I used stage 1 data to create the histogram of Norm\_Deaths column.

Chart, histogram

Description automatically generated

**Inference from the above diagram:**

* From the above histogram the mean is a black dashed line and median is red dashed line
* The mean is 12 and median is 11
* The histogram is ploted to understand the distribution of the data and tells how frequently every value in a data set is repeated
* The data is right skewed or positive skew, this means most of the data is near the lower range and as the no. of deaths increases the values become less frequent
* When the distribution of data is right skewed the mean is on the right side and mode is on the left. To get both mean and median symmetrical, the deaths data near the tail should be added to get a symmetrical distribution or normal distribution.
* The mean gets pulled towards the outliers and median resists this pull and hence median is known to be robust compared to mean. So, it is always better to get the median value of deaths to understand the data more clearly.
* Death is a discrete variable we cannot have deaths as decimals thus it’s a discrete random variable.

**Task 2: Understand variable to mortality relationships for 2019 data**

To perform this task I used cut() function of pandas that will segment the data into the range of values we need and assign the labels to those ranges. A new column named death category was created to store the variables based on the values.

Thus from the above result we see that for the Norm\_deaths value the range and labels assigned are as follows:

Very Low - [1 - 8.0]

Low - [8.0 - 11.0]

High - [11.0 - 16.0]

Very High - [16.0 - 64.0]

**Task 3 : Plot a second variable to Normalized Mortality in a scatter plot to observe any trends and describe the trends. Describe any trends you observe**

**Chart, scatter chart

Description automatically generated**

**Chart, scatter chart

Description automatically generated**

Rest images can be see in the notebook on github.

From the above scatterplots we are trying to co-relate the 2 variables. The scatterplot shows how the two variables are related whether there is positive correlation or negative correlation.

* The Deaths variable with Drug overdose deaths raw value, Insufficient sleep raw value and Injury deaths raw value are positively correlated. We see a positive increasing slope and hence both the variables are positively co-related. The rest other graphs show a slight increase in their slopes and are positively related. These variables show a strong relation compared to others.
* From the reports and the research, it was observed that with the intake of drugs there were many overdose cases that were reported which led to death. So as read through the reports, there were other drugs that was being mixed along with opiod. So knowingly unknowingly people consumed the drugs which led to the overdose, and we see there was increase in the number in the graph.
* With variable Injury death and firearm fatalities and motor vehicle accidents are also correlated. With the intake of drugs people were not having control over their mind and also not alert while driving the vehicles, this again led to the increase in deaths value and so we see a very strong correlation with Norm\_Death values them. The rest variables also follow the name trend but sexually transmitted infections surprisingly shows no only a slight increase not much. So I infer that though syringes were exchanged for intake of drugs the deaths were not because of this act. Even if they were very few deaths occured because of it.
* When comes to data points the values are more close and not scattered from each other this tells us how that the correlation is high as there is high possibility of a line in the graph. The scatter plot also shows outliers as we move to the higher norm\_death values