**DATA SCIENCE**

# **Stage II**

## Member task:

In this stage, we will do statistical analysis of the data. We make use of the merged data in this stage.

Determine the moments of the normalized death value we get the mean, median, mode and standard deviation values.

To visualize the normalized death value, we will make use of plotly histogram.

In histogram, we show the mean and median line using black and red color. Below is the histogram plot with max value as 63 and min value 1.

Chart, histogram

Description automatically generated

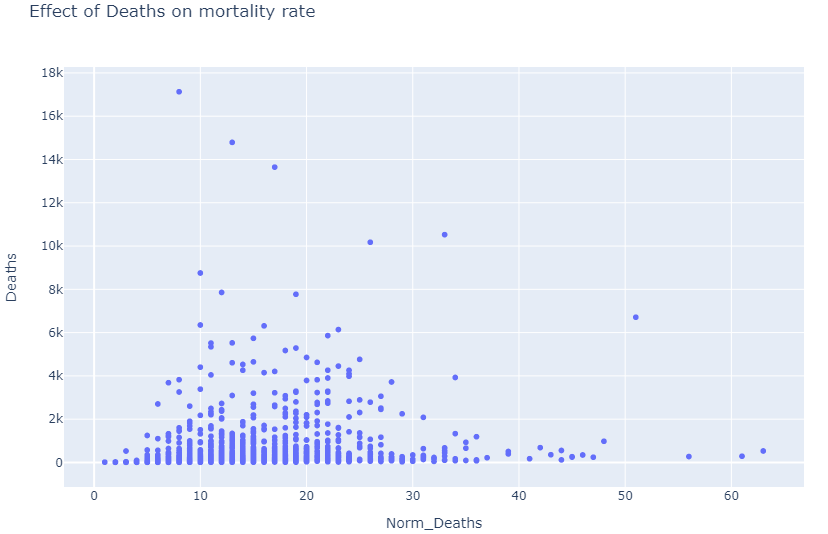
Next, we codify the normalized death data value to some categorized label value. In order to do so, we determine some range for different labels and add one column named “label” to main dataset.

Range are:

* Very Low (v\_low) - [1.999 - 8.0]
* Low (low) - [8.0 - 11.0]
* High (high) - [11.0 - 16.0]
* Very High (v\_high) - [16.0 - 64.0]

In stage I, we determine some variables that could impact the mortality rate. In this stage, we will visualize those factors using scatter plot to check if they really correlate with mortality rate or not.

To do that we make use of plotly dynamic scatter plots. We will use different parameters on Y axis and will have normalized death value on X axis.



Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated



Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

From the plots, it seems clear that, there are few parameters that greatly correlate with mortality rate.

Those factors are namely:

* Opioid dispensing rate
* Unemployment
* Drug overdose rate
* Insufficient sleep value

Lastly, in this stage, I have created an iPython widget which shows all the 10 plots dynamically by selecting different values from the dropdown. Its super easy to build plus helpful to view and analyze different plots.