**Assignment 1**

All questions have to be answered.

If there are any indications of a significant correlation between your outcomes and any of your colleagues outcomes – the default policy would be to assign zero (0) to both of you. As such please resist any such effort.

Please submit your assignments as jupyter notebooks.

Question 1

Download any two (2) dataset of your choice (that you find interesting) from the UCI Machine Learning Repository (<https://archive.ics.uci.edu/ml/index.php>).

Read these datasets into a Pandas dataframe.

Extract at least three (3) insights from the dataset that can be of value to the appropriate Business/Agency that might find the Dataset useful.

You should explicitly ensure the following workflow for each of the three (3) insights

1. What is the specific Business Question that you are asking
2. What is the method and mechanism of working through the problem
3. What is the insight/outcome. Try and present your insight as a visualization and as a paragraph.

Question 2

Design two(2) programs that are compute intensive. They should be at least 25 lines efficient Python Code.

For each of the program - Explore the clock time to run this program using %%time operator.

Apply the *dask.delayed* object at appropriate places in each of the two(2) programs to achieve a \*significant\* speedup.