

*Term Project*

*Data Science and Machine Learning in Canada*

**INSTRUCTOR:** Click or tap here to enter text.

**CLASS:** *AML 1104*

* *Individual/Group task*
* *Time allowed: Refer to Instructor Announcement*
* *Total marks: 30*

**Instructions:**

*Follow the instructions below for exploratory data analysis and submit a report with the result and code by the due date.*

* Conduct an exploratory data analysis of UCI Machine Learning data set. The followings are just 2 examples, feel free to chose others…
* Use of one of the following UCI Machine Learning data sets:

1. Heart Disease Data Set [<https://archive.ics.uci.edu/ml/datasets/Heart+Disease> ]
2. Adult Data Set [ <https://archive.ics.uci.edu/ml/datasets/adult> ]

The expectations include the following from your experience in the course:

GROUPS OF 3-4 PPL. YOU CAN FORM THE GROUPS!

1. Use Python Jupyter Notebook for the analysis.
2. Load the data set into a data frame
3. Are there any missing values in the dataset? How to handle that.
4. Use a measure of central tendency for each feature.
5. Show the dispersion (standard deviation and IQR) of features.
6. Use Data visualization techniques to plot in graphs.

* Deliverables

1. Completed Jupyter notebook
2. A .pdf report describing your observations and how you like to use the dataset for further analysis; regression or classification?