

*Term Project*

*Data Science and Machine Learning*

**INSTRUCTOR:** *Reena Shaw*

**CLASS:** *AML 1114*

* *Group task*
* *Due date:*
* *Presentation Date: Presentation is during the lecture period and it should be done within 20-minutes*
* *Total marks:* 
  + *Report and Python Files: 70 marks*
  + *Presentation: 30 marks*

**Instructions:**

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

First, complete the following tutorial that shows how to apply exploratory analysis to a dataset:

<https://www.kaggle.com/pmarcelino/comprehensive-data-exploration-with-python>

Now, using Python, apply a similar approach to 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:

1. Use Python for the analysis.
2. Are there any missing values in the dataset? How to handle that?
3. Apply the steps described in the above tutorial

* Deliverables

1. Completed Python files (Jupyter ntbk of the EDA)—only 1 submission per group. Mention all the group-members at the topmost cell in your Jupyter ntbk.
2. A .pdf report describing your observations and how do you like to use the dataset for further analysis: regression or classification?
3. PowerPoint slides