

Homework -1 (100 points) Due date 02/09/2024 (Friday)

[Late submission will get a 10% penalty for every day.]

1. **(30 points)** Explore the dataset given ([assign1_data1.csv](#)).
 - i. List the pre-processing tasks that you need to perform.
 - ii. How do you plan to solve each of these tasks? Please briefly explain.

2. **(40 points)** Write a python program to complete the following tasks.
 - i. Load the dataset ([assign1_data1.csv](#))
 - ii. Based on your analysis, clean the dataset using appropriate methods. Make a comment about the method you used to solve the problems.
 - iii. Describe the data using *boxplot* or *pie-chart*.

3. **(30 points)** Write codes to execute the following tasks based on the given dataset ([assign1_data2.csv](#)).
 - i. evaluate Pearson correlation between the features and the target variable.
 - ii. rank the features based on their correlation with **target variable (quality)**.
 - iii. evaluate and show the **correlation matrix for all features**.
 - iv. From the correlation matrix, **list out all the redundant features**.
(Use threshold value ≥ 0.65)

Submission of this assignment:

- Pdf file: written/typed answer for question-1 and the report outputs for question-2 and question-3 ([.pdf](#) file).
- Python file: python codes for question-2 and question-3 ([.py](#) file).

[Submit both pdf and python files to the canvas]