This group assignment has one main question consisting of several components. It has two parts of group assignment as follows:

Group assignment 1: Solution Development (40%)
- Part 1 is the group component
- Part 2 is the individual component.

Group assignment 2: Report (30%)

Objective:

The main objective of this assignment is to find an optimal model for a problem for which historical data is given in a dataset. This is an end to end project that addresses optimality in hyper parameter selection and model parameter selection using different models.

Group assignment 1: Solution Development (40 marks)

You need to submit your codes using python notebook. Each section should consist of necessary introduction, brief description of the chosen methods together with a justification of your choice, clear presentation of the results with appropriate graphics and/or tables and critical analysis.

Note: You need to zip the python notebook when submitting in Moodle.

Part 1: Group Component (10 marks)

(a) **Dataset Selection and Model Selection (5 marks)**
Choose a medium or large size dataset (at least 1000 samples), from the internet resources. Then, suggest any four different models that may be suitable for your dataset.

(i) Linear Regression
(ii) Logistic Regression
(iii) Non-Linear Regression
(iv) Artificial Neural Network (ANN)
(v) Convolutional Neural Network (CNN)
(vi) Recurrent Neural Network (RNN)

(b) **Data Preprocessing and Exploratory Data Analysis (5 marks)**
Perform Data Preprocessing and Exploratory Data Analysis (EDA) on your dataset. Use suitable tools for this purpose. Critically analyse the dataset.

Part 2: Individual Component (30 marks)

In this component each member of the group picks one of the suggested models from part (c).

(d) **Meta parameter selection (10 marks)**
List and explain meta parameter of your model. You may use grid search or random search methods to choose meta parameters of your model.

(e) **Model tuning (10 marks)**
Fine tune your selected model with the feature engineering/processing or regularization. Briefly explain your methods, apply them on your data, and present your results both analytically as well as graphically. Compare and critically analyse your results. Then, choose your final model and build it.

(f) **Model Evaluation (10 marks)**
Select evaluation metrics and briefly describe them. Then, perform evaluation and present your results.

Group assignment 2: Report (30 marks)

Write a report to present the introduction, findings and discussion, and conclusion of the analysis. A maximum of 2000 words must be written with a word processor. Conduct the analysis using Python. Your report should cover the following requirements:

(a) **Introduction (5 marks)**
Describe your selection from your readings about this dataset. Explain why you think your selection of dataset is suitable for your assignment. Justify your model selection based on your dataset.

(b) **Findings and Discussions (20 marks)**
Discuss the Exploratory Data Analysis (EDA) based on your dataset. Discuss and critically analyse your proposed models based on the Meta parameter selection, model tuning and model evaluation.

(c) **Conclusions (5 marks)**
Critically analyse and compare the models proposed by your group members. Then, suggest on how your results may be improved further.