DA5401: Assignment 3

Consider the provide multi-variate regression dataset (**Assignment3.csv**) having five predictor variables and one dependent (or output) variable. Your goal is to learn the regression function that minimizes the loss to the lowest.

Task 1 [5 points]

Fit OLS on the data directly and evaluate the baseline SSE loss. You will observe that the loss is very high, but that's ok. You will strive hard to apply creative ways to reduce the loss.

Task 2 [10 points]

Perform EDA on the dataset to understand the predictor features and how are they influencing each other. Also, study how each individual predictor influence the output variable. You may use correlation study to estimate the influence. Add necessary visualization and its representive interpretations to substantiate your inferences. The outcome of this step is figure out the requires features and their respective transformation.

Task 3 [5 points]

Fit OLS on the selected and transformed features and check if the loss has reduced from the baseline estimation.

Task 4 [10 points]

Install 'lazypredict' package and use the LazyRegressor class to build the regression models.
Compare the RMSE reported by all the regression models from LazyRegressor against your OLS losses.
Infer the reasons for why different techniques report different performance metrics.