**Team 116 – Progress Update**

Our team is slightly ahead of schedule with the project, using the Employee Performance dataset (a pre-approved dataset) as the sole data source for this project. Data collection is complete as of the first week of October 2024. We have not faced any challenges downloading and analyzing the data in R.

Based on the TA's feedback, we've refined our approach to focus on developing multiple models using various variable selection methods. We have decided not to pursue separate models for different departments or education levels, as the data has shown minimal variation between these groups, making this approach less effective. We have ensured that the assumptions for Multiple Linear Regression are met, and we created visualizations (box plots, scatter plots, residual plots, and a correlation matrix) to better understand our features. Next steps include using different feature selection techniques (LASSO, Forward, and Backward) to select features for potential models. We'll compare these models by fitting them on a training set and evaluating their performance on a test set, using metrics such as R-squared and RMSE.