DSS 740: Analytics with Machine Learning

Homework 2

1. Load and explore the Student Performance Dataset.
2. Create a table outlining the issues with the original data and how you treated it.
3. Create two tables that provide descriptive statistics of the original data and pre-processed data.
   a. What differences do you notice?
   b. Highlight the stats where the difference is significant (>20%), moderate (10% - 20%), and negligible (< 10%).
   c. How might these differences impact your model performance?
4. Perform univariate and bivariate graphical analysis on the pre-processed. For each of the graphs generated, state your conclusion. You may want to analyze the data statistically as well to get a better picture (e.g., reviewing the descriptive stats, along with the graphical representation).