

**Student ID:** s4761003

## **Peer Review Points -**

### **Group G**

- It would be better to describe the summary of data differentiating before and after data transformation. More visualisations, like a scatter plot with a trend line or a correlation matrix, are needed to establish some relation between the different factors before building a multi-linear regression model. An increase or decrease in the average value of each factor yearly can provide some more insights into factors affecting coral coverage and give an easier explanation for the conclusion.

### **Group L**

- An essential part of making the data fit for use is the removal of outliers. Various numerical columns of Stock Data can be used to visualise the boxplot for each column. In the same sense histogram to check the distribution of data can also be built. It would be beneficial to implement some form of outlier detection for the numerical columns, for example removing rows where data does not lie between the maximum and minimum value computed using Inter quartile range.
- A correlation heatmap matrix could be built to show the relationship between various features. It will give us a clear understanding of choosing other features that can reveal new trends like upward/downward patterns with respect to "close" value. This pattern can be plotted to gain additional valuable insights and other details that affect the value of "close".

### **Group M**

- While each step of the data science process is implemented in the slides, no topic or problem statement is clearly mentioned or chosen for the analysis. There is confusion about the goal of the analysis. Whether the subject is concerned about various factors affecting breast cancer or selecting the best classifier for finding breast cancer. Major Visualisations also talk about the performances of different classifiers rather than the factors affecting breast cancer. An explicit goal of the analysis needs to be chosen.
- More visualisations need to be added in the presentation that shows trends or insights from the data. For example, differences in features/factors affecting breast cancer patients at different ages. Difference in the factors affecting Males or Females differently. Etc. No clear conclusion is presented as a result of the analysis.