Summarising your data analysis procedures in up to 20 slides for a 5-minutes presentation. The slides should include the following contents:

Hi Dr Tran.

This is my presentation for SIT718 Real World Analytics – Assement 2.

Our first task was to create scatterplots and histograms of the available data.

The pairing the scatter plot and histograms together displays a positive skew in the chlorides, sulfur dioxide and alcohol variables Citric acid also skews positive though not as strongly as three. pH levels appear to be normaly distributed. As does the quality variable though the stepped and clumped accumulation of data creates a distortion.

Reviewing the central tendancy, correlation and k-s results of the variables shows the data does not present as a normal distribution and transformations must be conducted to ascertain further insights.

This leads to part two of the assessment and the selection of four variables that after data transformation will assist in the prediction of a wine’s quality. I have chosen the variables based on their correlation coefficient and as such have excluded the pH variable as its coefficient is closest to zero.

-What kinds of the data distribution you have identified in the raw data.

-Explain the transformations applied for the selected four variables and the variable ofinterest.

-Include two tables - one with the error measures and correlation coefficients, and onesummarising the weights/parameters and any other useful information learned for yourdata.

-Explain the importance of each of the variables (the four variables that you haveselected).

-Which fitting function is the best fitting model on your selected data.

-Give your prediction result and comment on whether you think it is reasonable.

-Discuss the best conditions (in terms of your chosen four variables) under which ahigher quality wine will occur.

-Comment the implications and the limitations of the fitting model you used forprediction.