Clustering:

1. Do the clustering methods generate the same clusters?

Yes, both the clustering methods K-means and hierarchical generate the same clusters

1. Does scaling effect the clustering?

Yes, Scaling affects clustering in a negative way for k-means.

This can be determined from the silhouette coefficient.

The silhouette coefficient for k-means without using scaled data is 0.59 and when using scaled data is 0.45

Yes, scaling affects the clustering in a negative way for hierarchical clustering

The silhouette coefficient for hierarchical without using scaled data is 0.59 and when using scaled data is 0.36

1. Does the clustering produce interesting groupings?

The clustering does not produce very interesting grouping since the clusters are too near which makes it difficult to classify it

Linear Regression:

1. Is the relationship significant?

Null hypothesis is that the relationship between predictor and outcome is not linear

From the above plot

1)From the p-values(0.00) our null hypothesis stands incorrect which means there is a significant relationship between

predictor and outcome variable.

2)Also, from the linear regression plots shown above, we can clearly infer from the slope values

that for each change in predictor variable there is a linear change in the outcome variable making

the model significant.

1. Are any model assumptions violated?

The model has 5 assumptions:

* Linear Relationship – This assumption is not violated since from the slope plot , each of the predictor is linearly related to the outcome variablel
* Multivariate normality – This assumption is not violated since the residual plot is normally distributed
* No or little multicollinearity – This assumption is not violated since from the heat map the correlation between residual sugar and alcohol is -0.45
* No autocorrelation – This model assumption is not violated since the Durbin-watson value is in the range of 1.5 – 2.5. The value is 0.87
* Homoscedasticity – This model assumption is not violated since the regplot shows equal variance across the regression lines

1. Is there any multi-colinearity in the model?

No, there is not much multi-colinearity in the model since the correlation between residual sugar and alcohol is -0.45

1. In the multiple regression models are predictor variables independent of all the other predictor variables?

Yes, the predictor variables are almost independent of the other predictor variables

1. In in multiple regression models rank the most significant predictor variables and exclude insignificant ones from the model

The most significant predictor is Alcohol since R2 is 0.986 and insignificant is residual sugar since its R2 is 0.6

1. Does the model make sense?

Yes, the model makes sense since since the R2  is 0.99.

1. Cross-validate the model. How well did it do?

Cross validation did not make any difference for my model. It shows R2 as 0.001 which was the same initially.

1. Does regularization help with creating models that validate better on out of sample data?

No regularization does not help with creating models that validate better on the test data since the RMSE remains the same

1. Is the relationship significant?

Yes, The relationship is significant since the score is 0.78 and the same can also be seen from the confusion matrix.

Also, the ROC curve shows that the curve stays far away from the line

1. Are any model assumptions violated?

* Binary logistic requires the dependent variable to be binary and ordinal logistic which is good wine quality(1) o bad wine quality(0)
* Absence of multicollinearity –This assumption is partially violated since from the heat map some of the predictors are greater than 1 which show that there is moderate correlation between some of the dependents while there is a lot of multicollinearity between residual sugar , density and alcohol
* The outcome variable should be independent

1. Cross-validate the model. How well did it do?

Cross validation the model did not help since it gives the same accuracy as it gave prior to cross validating it. Accuracy = 0.78