Question 1

Q. What is the optimal value of alpha for ridge and lasso regression?

A. Alpha for ridge : 4 Alpha for lasso : 100

Q. What will be the changes in the model if you choose double the value of alpha for both ridge and lasso?

A. As alpha value increased r2score dropped by ~1% in both test and train data for both regression.

Q. What will be the most important predictor variables after the change is implemented?

A. Neighborhood_NoRidge

Question 2

You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

A. I will use lasso as it provides feature selection options as well.

Question 3

After building the model, you realised that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?

A. MSSubClass_160, OverallQual, Foundation_Slab, MSSubClass_80, HalfBath

Question 4

How can you make sure that a model is robust and generalisable? What are the implications of the same for the accuracy of the model and why?

A. Model accuracy : As high as possible but not ideal. I was able to generate 81% accuracy which can be considered good.

P-Value : it should be less than standard cutoff 0.05. I was able to choose features below 0.05 or around it.

VIF : should be less than standard cutoff. I was able to choose features below 5.