Question 1

What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Ans. Optimal values for Ridge and Lasso both are 0.0001.

Increase in R squared for Ridge while decrease in R squared for Lasso.

1stFlrSF Is still the important variable.

Below are other important variable

'LotArea','1stFlrSF','2ndFlrSF','OverallQual','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?

We would be using Ridge since it gives lower Test error compared to Lasso.

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?

Ans. Earlier these are the important variables,

:'LotArea','1stFlrSF','2ndFlrSF','OverallQual','Neighborhood_NoRidge'

After removal of above below are the important variables:

OverallCond, BsmtFullBath, GarageCars, KitchenQual_TA, BsmtQual_NA

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?

We need to make sure that train and test accuracy doesn't differ much, for our model train and test both R2 error was ~84%. If the model is overfitting, we can make it more generalisable by using regularisaton methods. This would help increase the test accuracy.