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?

Optimal value of alpha for Ridge Regression was 500 and 300 for lasso regression.

The R2 square, RSS and MSE values decrease for both testing and training data for Ridge regression

The R2 square, RSS and MSE values decrease for training data but improve slightly for testing data.

In the case of Ridge the predictor variable remains the same.

While in lasso the predictor changes from 'RoofMatl_CompShg' to 'GrLivArea'

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?

I would chose to use Ridge since it is more robust to change in data as seen when tested with test dataset. For this dataset Ridge adapts better to the unseen data.

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?

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?