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:- The optimal value of alpha for ridge and lasso regression is 8 & 0.01 respectively

When doubled the value of alpha for both ridge and lasso regression, we see a performance drop in both models.

The most important predictor variable after the change is **OverallQual**.

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?

Ans:- I would choose Ridge regression as performance in the Ridge regression model is better.
Train R^2 value: 0.858
Test R^2 value: 0.856

Following are the R^2 value for Lasso regression
Train: 0.806
Test : 0.795

Question 3

After building the model, you realized 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: - 'LotArea', 'Heating_OthW', 'Heating_Wall', 'HeatingQC_Fa', 'HeatingQC_Gd'

Question 4

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

Ans:- We can make sure that a model is robust and generalizable by checking it's R^2 values. If the train and test R^2 is 0.05 apart and R^2 is on the training dataset is good, we can say that we didn't overfit or underfit the model.