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?

Answer: -

Optimal value for Ridge regression is 8 and for Lasso regression is 0.001.

When we have double the value of alpha for both ridge and lasso regression the most important predictor variable observed is **OverallQual_9.**

Below are the table which I observed after implementing the steps: -

Ridge maximum coefficient	**Ridge most important column**
0.3955813193971312	OverallQual_9
Ridge maximum coefficient_double alpha	**Ridge most important column_double alpha**
0.3175015969948623	OverallQual_9
Lasso maximum coefficient	**Lasso most important column**
0.6832028240106427	OverallQual_9
Lasso maximum coefficient_double alpha	**Lasso most important column_double alpha**
0.7042354212019802	OverallQual_9

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?

Answer: -

In the current assignment we are having lot of variables to analyse, so I will prefer to choose Lasso regression since it helps in feature elimination and using it make model more robust.

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?

Answer: - Now, The five most important predictor variables are :-

1>Condition2_PosA

2>2ndFlrSF

3>Exterior1st_BrkFace

4>Functional_Typ

5>Neighborhood_Somerst

Note: -Please refer the “Surprise Housing_Project_V1.2.ipynb” file for steps by step execution

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?

Answer: -

A robust, generalized model has the ability to adapt and react appropriately to previously unseen, fresh data chosen from the same distribution as the model’s initial input

A robust and generalized model should not overfit or underfit at any circumstances, the best procedure to achieve a good generalized model is to perform cross validation technique

By generalization, we find the best variance bias trade-off between underfitting and overfitting so that a trained model obtains the best performance, where the validation error has its global minimum

As mentioned, a generalized model is neither too complex to overfit nor too simple to underfit. A complex model has a high accuracy, this often causes issues with robust and generalized model to previously unseen data, as a result we try to make the model simpler with less dip in accuracy, we achieve this using Regularization methods like Lasso or Ridge Regression