Question1 :

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 1: I tried with different values of alpha for Ridge and Lasso.

Optimal value of alpha for Ridge is 5

Optimal value of alpha for Lasso is 100

Doubled the alpha values, i.e alpha=10 for Ridge & alpha=200 for Lasso.

For Ridge: Coefficient values increased and R2 score have decreased from 78.93 to 78.5

For Lasso: As the alpha is increased more features are removed, the co-efficient increased and one feature was removed, R2 score have decreased, 78.9 to 78.1

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 2:

After comparing the results of Ridge and Lasso, R2 of Ridge is better than R2 of Lasso. Lasso didn't remove any coefficients in the case of Lasso, so I will choose Ridge over Lasso for my model.

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 3:

After dropping these 5 important predictors [Neighborhood_NoRidge, Neighborhood_NridgHt, Neighborhood_Somerst, OverallQual, Neighborhood_Veenker], I recreated the model and found next set of important features are : **GarageType_BuiltIn, GarageType_Attchd, GarageType_Detchd, 1stFlrSF, 2ndFlrSF**

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 4:

A model is robust if the model is stable, i.e does not change drastically upon change in training set. The model is generalizable if it does not overfits the training data, and works well for new data.

In order to find the model is robust and generalizable, we can check R2 value, RMSE, k-fold cross validation, try to increase the data, decrease the features with different techniques and validate the variance and bais.

The robust and generalizable model will perform equally well on both training and test data. If model is not robust and generalizable the accuracy will drop in test data. It is because the ovefit model has very high variance and underfit model has very high bias.