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: -

For me the optimal value of alpha for ridge regression is 175 & 700 for lasso regression. If I double the value of alpha for both of these then my model will start underfitting gradually meaning that my training accuracy itself will be very low. So, after the change we found that basement quality is an important feature along with roof style.

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: -

We saw that the optimal value of the lambda for ridge and lasso are 175 and 700 respectively. So, in ridge we got r2 score of 94.80 & 90 for train and test data respectively. Similarly in lasso we got r2 score of 87.22 & 87.35 for train and test data respectively. In this case we have both the models performing really well hence we chose to do a cross validation which basically will help us to figure out a more stable model. After cross validation we found that ridge is having a higher r2 score with almost similar standard deviation as lasso so we concluded that we should go ahead with ridge regression.

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: -

After doing this we found that land contour, overall quality, Sale Type are proved to be very significant variables.

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: -

For making the model robust and generalisable it is very important that the model is not overfitting so first you need to make sure that the model is not overfitting. If you found that the model is overfitting you need to use some penalty for the variables which are not that of a great value. You need to use cross validation technique to ensure that the model is giving stable outcomes. Cross validation will tell you on different randomly selected sets of data how will the model perform.