Please limit your answers to less than 500 words per question.

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 The Optimal value of alpha for ridge is 0.61 and for lasso it is 0.61.

The R2 Score of the model on the test dataset for doubled alpha is 0.6120810835586621

The MSE of the model on the test dataset for doubled alpha is 0.06470834456376354

Updated scores are below with no major changes.

The R2 Score of the model on the test dataset for doubled alpha is 0.6124734734030226

The MSE of the model on the test dataset for doubled alpha is 0.06464289042843771

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 Although I have both the models still I would like to go ahead with Lasso since this can take care of the feature selection, Additionally I can also try to take some additional measures for better training accuracy.

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?

The firs first five important predictor variables are Fireplaces, HalfBath, Foundation, BsmtFullBath, RoofStyle.

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 I could understand that the model is robust and more generalisable by testing the model on test data if the accuracy is nearby with the training accuracy, Additionally I can also use methods to review the model performance and also create a retraining pipeline with newer data so that the performance does not degrades in near future.