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:

Ridge Alpha value: 2

Lasso Alpha value: 0.001

The values of train & test r2 scores reduced on doubling the alpha values. The most important predicator variable.

GrLivArea remained as the most important predictor, however there was a significant reduction in its coefficient value.

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: Lasso regression seems more suitable for this model as Lasso regression was more accurate and consistent with the test & train r2 values. Lasso eliminated the coefficients of a few features and brought them to zero and was more accurate.

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?

Ans: The next 5 most important variables along with their lasso coefficients are found as below:

Neighborhood_Crawfor	0.023206
CentralAir_Y	0.021246
SaleCondition_Partial	0.017653
Neighborhood_StoneBr	0.017045
RoofStyle_Hip	0.013286

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?

Ans :The model should be generalised to keep the test accuracy consistent with the train accuracy. Keeping the features count optimum shall keep the model bias & variance balanced leading to higher test accuracy. The accuracy on the test data should not vary greatly with changes in the test dataset and the same is obtained by keeping the models generalised & robust.