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?

Optimal value of alpha for ridge: 2 (Doubled the alpha values and output values are lesser change)

Optimal Value of alpha for Lasso: 20 (Doubled the alpha values and output values are lesser change)

Top Features: **Neighborhood_NoRidge, Neighborhood_NridgHt, Neighborhood_Crawfor, Neighborhood_StoneBr, OverallQual**

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?

Since Lasso is recommend for feature selection, It tends to small number of significant parameters and others are close to zero.

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 New 5 predictors are:

HouseStyle_2.5Fin
Exterior2nd_BrkFace
Neighborhood_Somerst
Neighborhood_NPkVill
KitchenQual

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?

To Make model robust and generelisable:

Accuracy >70%

VIF all features < 5 (In our case all feature VIF values are < 5 and highest value is 2.65)