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?

A:

The Optimal Value of alpha:

Lasso: 0.001 (with r2 score on train and test: 0.9239996330020745, 0.8939313759890265)

Ridge: 20 (with r2 score on train and test: 0.9193230104030329, 0.8956248581858526)

After changing alpha to double:

Lasso: 0.002 (with r2 score on train and test: 0.917078506507462 0.8984228466486353)

Ridge: 40 (with r2 score on train and test: 0.9139612250069122, 0.8967355836930018)

Increasing alpha introduces bias which result in r2 score decreasing on train but performing slightly better on test.

Most important factors for Lasso:

- Neighborhood_Crawfor
- GrLivArea
- Neighborhood_Somerst
- YearBuilt
- SaleCondition_ord
- Exterior1st_BrkFace
- OverallQual
- Neighborhood_BrkSide
- OverallCond
- Functional_Typ

Most important factors for Ridge:

- GrLivArea
- OverallQual
- Neighborhood_Crawfor
- SaleCondition_ord

- OverallCond
- YearBuilt
- Functional_Typ
- 1stFlrSF
- Condition1_Norm
- TotalBsmtSF

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?

A: Lasso regression as it helps to reduce the number of feature with almost same result.

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?

A:

Next 5 predictor after removing first 5 are :

- 2ndFlrSF
- OverallQual
- YearBuilt
- 1stFlrSF
- OverallCond

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?

A: R2 score of Train and test prediction should be similar.

If model is not generalizable, then test score will most probably drop.

To ensure generalizability, remove outliers which may skew the predictions and apply Lasso/Ridge regularization.