Q1: 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?

Ridge alpha= 0.05 double alpha=0.1

Lasso alpha= 20 double alpha=40

After changes

**Lasso**

GrLivArea, OverallQual , Neighborhood\_NridgHt, LotArea, BsmtFinSF1

**Rigde**

OverallQual, Neighborhood\_NridgHt, 1stFlrSF, LotArea, GrLivArea

Q2: 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?

The r2\_score of lasso is slightly higher for the test dataset so we will choose lasso regression to solve this problem.

Q3: 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?

2ndFlrSF,1stFlrSF, TotalBsmtSF, BsmtQual\_Ex,YearBuilt

Q4: 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 sure a model is robust and generalizable, we must **take care it doesn't overfit**. This is because an overfitting model has very high variance and a smallest change in data affects the model prediction heavily. Such a model will identify all the patterns of a training data, but fail to pick up the patterns in unseen test data.