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) Optimal value for Ridge : 2.89
b) Optimal Value for Lasso: 0.001

If we double the value of both Ridge and Lasso it'll be 5.78 for Ridge and 0.002 for Lasso.

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?

According to our assignment scenario Lasso regression will be better selection.

As MSE for Lasso seems little lower in Ridge so it will perform well not only in test and train data but also in unseen data.

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?

Top 5 feature before building the model: OverallQual, GrLivArea, GarageCars, GarageArea, TotalBsmtSF

After recreating the model, top 5 predictor: 'TotRmsAbvGrd', '2ndFlrSF', '1stFlrSF,Full Bath, TotRmsAbvGrd

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?

The simpler the model is the robust it can be. A robust and generalisable model will work best on unseen data and not only test and train data.

The Implications are:

Overfitting: Overfitting models will work bust on Training dataset and give higher accuracy but it'll fail to perform well on unseen data.

Data Quality, Complexity, Features helps to get better accuracy in unseen data.