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?

The optimal value of alpha for ridge regression is 500 and lasso regression is 1000. If we choose double the value of alpha for both ridge and lasso, the model coefficient estimates would be shrinking more towards 0. The most important predictor variables are **GrLivArea, OverallQual, BsmtFinSF1, GarageArea, TotalBsmtSF** after the change is implemented.

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?

I will choose to apply the optimal value of lambda of 1000 for lasso regression since it also results in feature selection and helps me to choose the most important predictor variables as required in the assignment.

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 five most important predictor variables now are **MasVnrArea, LotArea, Fireplaces, OverallCond and Functional**.

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?

When we do regularisation, the model complexity is reduced and overfitting is prevented. Hence, ridge and lasso regression makes the model robust and generalisable. The accuracy of the model is reduced but the difference of accuracy on training and test sets is minimum. The accuracy is reduced since the model coefficients are penalised during regularisation and their magnitude shrunk towards 0.