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?

Ans) The optimal value of alpha for ridge and lasso regression is 2 and 0.002 respectively. There was not any substantial difference in the model performance with the r2 scores being very similar. The most important predictor variables for both cases are mentioned below.

Ridge regression: GrLivArea, MSZoning_RL, OverallQual, MSZoning_RM, TotalBsmtSF

Lasso regression: GrLivArea, OverallQual, MSZoning_RL, MSZoning_RM, 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?

Ans) Both types of regression are providing with similar results in terms of test/train set model predictions. I would be using lasso regression in this case because it is gives a slightly better model accuracy than linear and ridge regression.

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?

Ans) The most important variables after doing this experiment are: 1stFlrSF, FullBath, HalfBath, OverallCond, d_BsmtQual

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?

Ans) A robust model can be made by adding some noise to the test data as well as treat the outliers in the dataset properly. This might result in a slightly lower model accuracy in case of adding more data in the test set but it will lead to lesser overfitting and better model performance.