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 Regression is 5.0 if we double the alpha value in ridge it may tends to increase in Y_test value and also in X_train value which leads to steady increase in the Ridge Regression.

The optimal value of alpha for lasso regression is 0.2 if we double the alpha value in lasso it may leads the model to try penalize more and try to make most of coefficient values to zero

The most important predictor variables after the change is implemented for both ridge and lasso are

- LotArea
- OverallQual
- OverallCond
- BsmtFinSF1
- TotalBsmtSF
- 2ndFlrSF
- GrLivArea
- BsmtFullBath
- FullBath
- HalfBath

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:I'll choose the lasso Regression as the lambda value increases lasso shrinks the coefficients to zero and it makes the variables exactly to zero. Lasso also makes the decision based on the variable selection so when the lambda value increase the lasso values became zero and neglected by model.

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 five most Important predictor variables are

- MSZoning_RL
- MSZoning_RM
- GrLivArea
- MSZoning_FV
- OverallQual

These predictor could be a most important when another model is build.

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: The Model should be simple, evethough the accuracy may decrease but the model will be robust and generalisable. it also defined the simple model will perform equally on both test and train dataset.