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 to 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 the alpha for Ridge is 1.0 and for the Lasso is 0.00001 with these alphas the R2 of the model was approx. 0.78, After doubling the alpha value the model accuracy was remains same.

Since the alpha value was quite small, doubling it does not do any significant change in both the models and the R2 and MSE remains almost the same. The most important predictor variables also remain the same, but the Evaluates the height of the basement - Excellent attribute gains higher importance than Alley in the Ridge Model.

Ridge Regression Model

	Ridge Doubled Alpha Co-Efficient
TotRmsAbvGrd	0.208272
GarageArea	0.173356
LotArea	0.132928
OverallCond	0.068787
LotFrontage	0.057228
OpenPorchSF	0.054851
CentralAir_Y	0.051812
Neighborhood_StoneBr	0.048193
BsmtQual_Ex	0.046360
Alley_Pave	0.043235

	Lasso Doubled Alpha Co-Efficient
TotRmsAbvGrd	0.223099
GarageArea	0.191003
LotArea	0.165313
OverallCond	0.076941
SaleType_Con	0.057506
OpenPorchSF	0.053775
Neighborhood_StoneBr	0.053233
LotFrontage	0.051198
SaleType_CWD	0.051091
CentralAir_Y	0.049670

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?

Both Ridge and Lasso values for this dataset seems to be similar and there is little bit significant change in the lambda value. And Mean Square Error also seems to be aprrox same for both the models.

To choose the best one for this prediction, we can go for the Lasso. Since Lasso helps in feature elimination.

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?

Initially, when we build the model, we found that we have few important predictor variables as below:

1. Total Rooms above grade
2. Total Garage Area
3. Lot Area
4. Overall condition
5. Sale Type

If we remove those features and rebuild the model with the same dataset, we will obtain the below top 5 predictor features from the model.

	Lasso Co-Efficient
LotFrontage	0.268143
OpenPorchSF	0.114755
BsmtQual_Ex	0.067629
CentralAir_Y	0.063115
KitchenQual_Ex	0.061314

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 two model which we created shows similar performances in the training and test data set, to pick the best one, if we go for the Occam's Razor method.

Based on the Occam's Razor theorem, we must choose the simpler model, since usually simpler model is more 'generic' and more robust and easier to train the model as well.

Therefore, to make the model more robust and generalisable, make the model simple but not simpler which will not be any use.

One more technique to make the model robust and generalisable adding the penalty to the cost function by using the method called regularization. Also, making a model simple leads to Bias-Variance Trade-Off.