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 is 10.0 and lasso is 0.001.

- When we double the value of alpha, the coefficients values will increase. For ridge regression the R-square value for training set has increased and R-square value for test set has decreased. For lasso regression the R-square value for training set has decreased and for test set increased. In Ridge regression the RSS, MSE, RMSE values of training set are improved and become worse for test set. In lasso regression the RSS, MSE, RMSE values of test set are improved and become little worse for train set.

- Important predictors variables of ridge regression after change is implemented is:

	Parameter	Coef
53	OverallQual_Other	0.501
35	Neighborhood_Crawfor	0.296
52	OverallQual_8	0.273
85	KitchenQual_Other	0.242
7	TotalBsmtSF	0.209
99	SaleCondition_Partial	0.203
68	ExterQual_Other	0.161
79	BsmtFinType1_No Basement	0.157
11	FullBath	0.155
66	MasVnrType_None	0.154

- Important predictors variables of lasso regression after change is implemented is:

	Parameter	Coef
53	OverallQual_Other	0.917
52	OverallQual_8	0.485
35	Neighborhood_Crawfor	0.394
79	BsmtFinType1_No Basement	0.316
99	SaleCondition_Partial	0.263
7	TotalBsmtSF	0.242
85	KitchenQual_Other	0.220
55	OverallCond_7	0.168
51	OverallQual_7	0.158
46	Condition1_Norm	0.153

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?

➔ Among the two regression model, we can choose lasso regression model, because it has good R2 value for test set. And also lasso regression has low RMSE value compared to ridge regression. And apart from that the lasso regression helps in feature elimination which cannot be performed by 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?

➔ Excluding the five most important predictor variables, the five most important predictor variables now is:

- OverallQual_Other (Coefficient: 0.917)
- OverallQual_8 (Coefficient: 0.485)
- TotalBsmtSF (Coefficient: 0.242)
- SaleCondition_Partial (Coefficient: 0.263)
- Neighborhood_Crawfor (Coefficient: 0.394)

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?

- In order to make sure that our model is robust we have to ensure that there are no significant outliers present in the dataset apart from that which are relevant to the dataset. Switching from mean squared error to mean absolute difference reduces the influence of outliers. Too much weightage should not be given to the outliers, so that the accuracy predicted by the model is very high.

- The model should not get overfitted, it should be generalizable. This can be achieved by using proper techniques like regularization. So that the test accuracy is not less than training accuracy.

- Confidence interval can be used. This would help standardize the predictions made by the model.