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 values of alpha for Ridge and Lasso Regression are

	Ridge	Lasso
Original Alpha	20.0	0.0001

After Doubling the Alpha values for both Ridge and Lasso, We get this

	Ridge	Lasso
Doubled Alpha	40.0	0.0002

Ridge Top 3 predictors after doubling Alpha

	Ridge (alpha=20.0)	Lasso (alpha=0.0001)	Ridge (alpha = 40.0)	Lasso (alpha = 0.0002)
GrLivArea	0.095499	0.123556	0.087723	0.125120
OverallQual	0.074746	0.075611	0.073916	0.076125
2ndFlrSF	0.054217	0.038563	0.053961	0.035799

Lasso Top 3 predictors after doubling Alpha

	Ridge (alpha=20.0)	Lasso (alpha=0.0001)	Ridge (alpha = 40.0)	Lasso (alpha = 0.0002)
GrLivArea	0.095499	0.123556	0.087723	0.125120
OverallQual	0.074746	0.075611	0.073916	0.076125
TotalBsmtSF	0.049188	0.062122	0.044042	0.059742

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: The optimal values of Lambda or alpha for ridge and lasso regression are found to be as follows

	Ridge	Lasso
Original Alpha	20.0	0.0001

However, As we build the models with both ridge and lasso, we found out that the ridge regression is slightly better in terms of test r2_score and also has lower RMSE when compared to Lasso. Hence Choosing a ridge regularization gives better results.

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: In the lasso model when the top 5 predictors are removed and rebuilt, then the top 5 predictors in the new model are

	Lasso
2ndFlrSF	0.146058
1stFlrSF	0.117215
BsmtFinSF1	0.060119
YearRemodAdd	0.055234
GarageCars	0.037679

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: We can make sure that the model is robust and generalisable by striking a balance between bias and variance of the model. This can be achieved by using Ridge and Lasso Regularisation techniques.

In general, when the model tends to overfit in the train dataset, the bias of the model is very less, whereas the variance is very high. This results in a higher train accuracy and lower test accuracy. By using either of the above mentioned regularisation techniques we can allow for a little increase in bias to significantly reduce the variance, so that the final model becomes robust and stable.