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) :

In case of Ridge Regression optimal alpha = 20

In case of Lasso Regresson optimal alpha = 100

If we double the alpha for Ridge regression , there is a slight decrease in the model accuracy i.e, R2 score .

If we double the alpha in case of Lasso Regression also , there is a slight decrease in the model R2 score .

With Ridge regression the following are the most important features after doubling the alpha

	Features	Ranks
20	GrLivArea	16962.583591
55	RoofMatl_WdShngl	14487.311320
104	Neighborhood_StoneBr	14035.554273
75	Functional_Typ	13553.651125
18	2ndFlrSF	13127.063625
98	Neighborhood_NridgHt	13043.966726
2	OverallQual	13010.086236
97	Neighborhood_NoRidge	12063.337559
48	SaleType_New	10142.518312
15	TotalBsmtSF	9983.015756

In case of Lasso regression , the following are most important features after the doubling the alpha

	Features	Ranks
55	RoofMatl_WdShngl	59886.128018
20	GrLivArea	30305.181592
104	Neighborhood_StoneBr	26043.820916
97	Neighborhood_NoRidge	22442.728492
98	Neighborhood_NridgHt	21910.055437
48	SaleType_New	21658.052290
75	Functional_Typ	15428.818270
2	OverallQual	12044.355658
15	TotalBsmtSF	9876.692164
88	Neighborhood_Crawfor	9614.586422

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 will go with Lasso regression , since the Lasso regression will help in feature selection as well along with regularization thereby simplifying the 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 next 5 variables based on RFE ranks would be part of the important predictor variables

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 go by the R2 score and any R2 score 0.85 is good for prediction.