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?

Optimal value of alpha: 500

Change in coefficients of the model in Ridge and Lasso for Alpha = 500 and Alpha = 100

	Alpha = 500	Alpha = 1000
Total number of coefficients	68	68
Number of coefficients for Lasso Model	32	24
Number of coefficients dropped as per Lasso Model	36	44
Average of Ridge Coefficients	864.23	811.83

As it can be seen that in the case of Lasso, the number of coefficients got decreased from 32 to 24 (25% reduction in number of coefficients)

Average of Ridge model coefficients got reduced from 864 to 812 (6% reduction)

Change in most important predictor variables (Top 10) is as follows for Alpha = 500 to 1000.

Alpha = 500	Alpha = 1000		Alpha = 500	Alpha = 1000
Ridge	**Ridge**		**Lasso**	**Lasso**
GrLivArea	GrLivArea		GrLivArea	GrLivArea
1stFlrSF	1stFlrSF		YearBuilt	GarageCars
2ndFlrSF	GarageCars		GarageCars	YearBuilt
GarageCars	TotRmsAbvGrd		LotArea	YearRemodAdd
FullBath	Fireplaces		YearRemodAdd	LotArea
TotRmsAbvGrd	TotalBsmtSF		TotalBsmtSF	BsmtFullBath
Fireplaces	FullBath		BsmtFullBath	Fireplaces
TotalBsmtSF	2ndFlrSF		FullBath	FullBath
YearRemodAdd	GarageArea		Fireplaces	LotConfig_CulDSac
GarageArea	YearRemodAdd		BsmtFinSF1	TotalBsmtSF

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?

Ridge based model will be selected as the drop in R-squared between Train and Test datasets are minimal in Ridge (89% to 86%) compare to Lasso (93% to 84%)

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?

Original Top 5 coefficients using Lasso

GrLivArea
YearBuilt
GarageCars
LotArea
YearRemodAdd

New Top 5 coefficients using Lasso

2ndFlrSF
1stFlrSF
GarageArea
TotalBsmtSF
Neighborhood_Crawfor

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?

To make a model robust and generalizable, the important factor is the model performance on the unseen data or the test data. Hence the R-squared value for training data and the test data or an unseen data can be checked and make sure that the R-squared is not dropping too much.

This test need to be done on a periodic basis as and when new data is available and the model parameters need to be checked and ensure that the model performance on unseen data is not dropping