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?

[Answer]

- **Optimal Values of Alpha**

  Optimal Value of Alpha for Ridge Regression: 10

  Optimal Value of Alpha for Lasso Regression: 100

- **If the alpha is doubled for both Ridge and Lasso**

  There is a slight dip in the model performance for both Ridge and Lasso.

  However, on the test data prediction, there is a slight improvement in the r2_score

- **Important Predictor Variables with 1 * Alpha and 2 * Alpha**

	Ridge Regression		Ridge Regression (2* Alpha)
0	GrLivArea	0	FullBath_3
1	FullBath_3	1	TotRmsAbvGrd_10
2	OverallQual_10	2	OverallQual_10
3	Neighborhood_NoRidge	3	GrLivArea
4	TotRmsAbvGrd_10	4	Neighborhood_NoRidge
5	OverallQual_9	5	Fireplaces_2
6	2ndFlrSF	6	OverallQual_9
7	1stFlrSF	7	1stFlrSF
8	Fireplaces_2	8	2ndFlrSF
9	BsmtQual_TA	9	BsmtExposure_Gd

	Lasso Regression		Lasso Regression (2* Alpha)
0	GrLivArea	0	GrLivArea
1	OverallQual_10	1	OverallQual_10
2	OverallQual_9	2	OverallQual_9
3	FullBath_3	3	FullBath_3
4	Neighborhood_NoRidge	4	Neighborhood_NoRidge
5	TotRmsAbvGrd_11	5	OverallQual_8
6	LotArea	6	Fireplaces_2
7	OverallQual_8	7	GarageCars_3
8	Fireplaces_3	8	TotRmsAbvGrd_10
9	Neighborhood_NridgHt	9	Neighborhood_NridgHt

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?

[Answer]

The model performance stats are as below for Ridge and Lasso. Both Lasso and Ridge have decent and almost similar metrics for the Trained Model and either could be preferred for predictions. Although, the R2 Score is a bit less for Lasso, it is still a decent score. Since Lasso does feature selection by reducing the coefficients of less important features to 0, hence, it could be selected if a simpler model is more preferred.

Ridge Regression	Lasso Regression
====> Training Data R2_Score: 0.8963204096948155 ====> Training Data RSS: 661548237390.5647 ====> Training Data MSE: 647941466.59213 ====> Training Data RMSE: 25454.694392039553 ====> Test Data R2_Score: 0.8134040877947518 ====> Test Data RSS: 525960235845.0811 ====> Test Data MSE: 1200822456.2673085 ====> Test Data RMSE: 34652.88525169742	====> Training Data R2_Score: 0.9014422235238655 ====> Training Data RSS: 628867486040.4008 ====> Training Data MSE: 615932895.2403533 ====> Training Data RMSE: 24817.99539125498 ====> Test Data R2_Score: 0.7088181565792092 ====> Test Data RSS: 820757910660.6879 ====> Test Data MSE: 1873876508.357735 ====> Test Data RMSE: 43288.29528126206

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?

[Answer]

The top 5 predictors are

```
0           1stFlrSF
1           2ndFlrSF
2            LotArea
3    TotRmsAbvGrd_11
4       Fireplaces_3
```

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?

[Answer]

The first and important step to obtain a more robust and generalisable model, the training data should be reliable, accurate, cleansed, free of redundant variables and outliers, and scaled properly. Once the model is built, a robust model should perform consistently on trained and test/new data. Since, real world scenarios, the test data is unseen and unpredictable, the model should be able to perform decently even if there are unpredicted changes in data in one or more predictor variables.

A more generalisable model is also simple rather than complex and an optimum Bias-Variance trade-off should be selected.