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:-

Optimal alpha for Ridge model = 8.0

Optimal alpha for Lasso model = 0.001

After we double the optimal alpha values in the Ridge and lasso the changes will be given below.

Ridge:-

Slight variance in all major metrics.

	Metric	Ridge Regression(8.0)	Ridge Regression(16.0)
0	R2 Score (Train)	0.944788	0.938484
1	R2 Score (Test)	0.891202	0.893457
2	RSS (Train)	56.371515	62.807945
3	RSS (Test)	46.029516	45.075613
4	MSE (Train)	0.234972	0.248024
5	MSE (Test)	0.324176	0.320800

Important predictor variables

Ridge (8.0)		Ridge (16.0)	
OverallQual#9	1.472521	OverallQual#9	1.366305
GrLivArea	1.344647	GrLivArea	1.311168
OverallCond#9	1.325805	OverallQual#8	1.278206
OverallQual#8	1.301208	OverallCond#9	1.213175
Neighborhood#Crawfor	1.261499	Neighborhood#Crawfor	1.212318
Functional#Typ	1.219009	Functional#Typ	1.191169
Exterior1st#BrkFace	1.202558	TotalBsmtSF	1.173744
Neighborhood#NoRidge	1.186365	Exterior1st#BrkFace	1.163299
TotalBsmtSF	1.184958	Neighborhood#Somerst	1.148000
Neighborhood#Somerst	1.174010	Neighborhood#NoRidge	1.140915
Name: Ridge, dtype: float64		Name: Ridge_double, dtype: float64	

Lasso:-

Slight variance in all major metrics.

	Metric	Lasso Regression(0.001)	Lasso Regression(0.002)
0	R2 Score (Train)	0.946970	0.933316
1	R2 Score (Test)	0.866429	0.888422
2	RSS (Train)	54.143170	68.084440
3	RSS (Test)	56.510451	47.205653
4	MSE (Train)	0.230281	0.258233
5	MSE (Test)	0.359193	0.328292

Important predictor variables

Lasso(0.001)		Lasso(0.002)	
OverallQual#9	1.927326	OverallQual#9	2.010569
OverallCond#9	1.636737	OverallQual#8	1.601804
OverallQual#8	1.529917	OverallCond#9	1.411231
GrLivArea	1.349434	GrLivArea	1.375315
Neighborhood#Crawfor	1.344126	Neighborhood#Crawfor	1.316831
Exterior1st#BrkFace	1.296611	Exterior1st#BrkFace	1.251548
Functional#Typ	1.241809	Functional#Typ	1.244309
Neighborhood#Somerst	1.225900	Neighborhood#Somerst	1.228373
BsmtExposure#Gd	1.187232	OverallQual#7	1.180440
Neighborhood#NoRidge	1.185924	Neighborhood#NoRidge	1.173666
Name: Lasso, dtype: float64		Name: Lasso_double, dtype: float64	

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 choose Lasso, because it will help in the feature selection as well and as per in the problem statement we need to suggest the significant variable to predict the price of the house.

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:-

Top 5 predictors in our first model.

- OverallQual#9
- OverallCond#9
- OverallQual#8
- GrLivArea
- Neighborhood#Crawfor

Slight variance in the test data metrics.

	Metric	Lasso Regression(0.001)	Lasso new Regression(0.001)
0	R2 Score (Train)	0.946970	0.941332
1	R2 Score (Test)	0.866429	0.856574
2	RSS (Train)	54.143170	59.900405
3	RSS (Test)	56.510451	60.679579
4	MSE (Train)	0.230281	0.242216
5	MSE (Test)	0.359193	0.372207

New top predictor variables.

```
2ndFlrSF                1.364334
Exterior1st#BrkFace     1.315712
Functional#Typ          1.231544
1stFlrSF                1.219985
Neighborhood#Somerst    1.186105
TotalBsmtSF             1.173693
PoolQC#NoPool           1.164978
BsmtExposure#Gd         1.154839
GarageQual#Gd           1.132193
BsmtCond#TA             1.118044
Name: Lasso_new, dtype: float64
```

Question 4

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

ANS:-

To make sure that a model is robust and generalizable we need to follow the below steps.

1. Need to perform proper EDA on the input data
2. Extract the relevant feature information and relationship within fields.
3. Based on the relationship, we find a best model to fit the data
4. Using the hyper parameter tuning, will find optimal point to fit the data.
5. Will test the model using the test data to find the performance of the model based results we may re-train the model
6. Use regularization techniques to avoid over fitting.

The implication of the robust and generalizable in model accuracy, the model will perform on test data with similarly like how did perform on the train data. Model accuracy will not change drastically on the test data.