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 for ridge and lasso are as follows :

Ridge : {'alpha': 1.0}

Lasso : {'alpha': 50}

Final Metric before double the value of alpha for both ridge and lasso

```
In [70]: final_metric
```

Out[70]:

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	9.232448e-01	9.211053e-01	9.186033e-01
1	R2 Score (Test)	8.860438e-01	8.919656e-01	8.863443e-01
2	RSS (Train)	3.324131e+11	3.416785e+11	3.525143e+11
3	RSS (Test)	2.605382e+11	2.469992e+11	2.598511e+11
4	MSE (Train)	2.000039e+04	2.027721e+04	2.059624e+04
5	MSE (Test)	2.701479e+04	2.630351e+04	2.697915e+04

Final Metric after double the value of alpha for both ridge and lasso

```
final_metric
```

Out[79]:

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	9.232448e-01	9.189607e-01	9.145543e-01
1	R2 Score (Test)	8.860438e-01	8.921016e-01	8.859047e-01
2	RSS (Train)	3.324131e+11	3.509664e+11	3.700498e+11
3	RSS (Test)	2.605382e+11	2.466882e+11	2.608561e+11
4	MSE (Train)	2.000039e+04	2.055097e+04	2.110229e+04
5	MSE (Test)	2.701479e+04	2.628694e+04	2.703127e+04

For both ridge and lasso R2 score of train and test has been decreased, Also there is increase in RSS and MSE for both test and train data.

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 regression model R2 score is slightly better than Lasso, However in Lasso Model insignificant features are being assigned to zero. Which will help us in finding less complex but similar efficiency . As it is always advisable to use simple yet robust 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?

After excluding the five most important predictor variables, below are the new five most important predictor variables

BsmtUnfSF
TotalBsmtSF
OverallCond
OverallQual
GarageType_Detchd

In [106]:
```
# Sorting the coefficients in ascending order
# Chose variables whose coefficients are non-zero
pred = pd.DataFrame(para[(para['Coeff'] != 0)])
pred = pred.sort_values((['Coeff']), axis = 0, ascending = False)
pred
```

Out[106]:

	Variable	Coeff
6	BsmtUnfSF	187833.879200
7	TotalBsmtSF	117484.698895
3	OverallCond	62377.344930
2	OverallQual	36662.057240
40	GarageType_Detchd	35843.032469
41	GarageQual_Gd	33907.357097
30	BsmtExposure_Gd	29347.027666
5	BsmtFinSF2	26790.330873

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?

A model can be considered to be robust and generalisable if the model is following the below behaviours .

Perform at the similar accuracy for the test data as of train data.
It should handle outliners efficiently

If the model is not robust, It cannot be trusted for predictive analysis.