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 to double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Answer:

The optimal value of Alpha for Ridge is 100 and Lasso is 0.001.

When we double the Alpha values, the R2 values on train data reduces by 2-3% and there is slight increase in r2 for test set on both Ridge and Lasso.
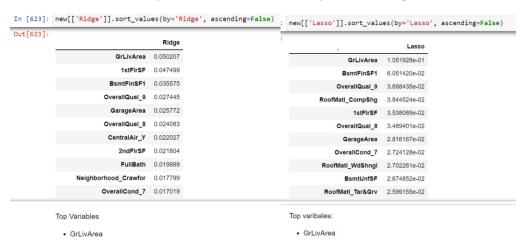
Ridge: 0.9363803974918815
0.871274992707527

Lasso: 0.9395737338603581
0.871395207236168

GrLivArea is the most important Predictor variables post the change.

In [623]:	new[['Ridge']].sort_values(by='Ridge', ascending=False)		new[['Lasso']].sort_values(by='Lasso', ascending=False)	
Out[623]:		**Ridge**		**Lasso**
	GrLivArea	0.050207	GrLivArea	1.051928e-01
	1stFlrSF	0.047499	BsmtFinSF1	6.051420e-02
	BsmtFinSF1	0.035575	OverallQual_9	3.888435e-02
	OverallQual_9	0.027445	RoofMatl_CompShg	3.844524e-02
	GarageArea	0.025772	1stFlrSF	3.538069e-02
	OverallQual_8	0.024083	OverallQual_8	3.489401e-02
	CentralAir_Y	0.022027	GarageArea	2.816167e-02
	2ndFlrSF	0.021804	OverallCond_7	2.724128e-02
	FullBath	0.019999	RoofMatl_WdShngl	2.702261e-02
	Neighborhood_Crawfor	0.017799	BsmtUnfSF	2.674852e-02
	OverallCond_7	0.017019	RoofMatl_Tar&Grv	2.599155e-02

Top Variables

- GrLivArea

Top varibales:

- GrLivArea

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:

We are choosing Lasso as it has slightly better r2 values on train and test sets.

Lasso R2 values:

0.9520222962389313
0.8711195459049577

Question 3

After building the model, you realized 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:

We dropped the five most significant variables and created a new Lasso model and noticed that the R2 values on train and test reduced.

```
0.9410175935049023
0.8664974761598184
```

The new significant variables now are:

2ndFlrSF, 1stFlrSF, BsmtFinSF1, BsmtUnfSF, OverallCond_7

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?

Answer:

We need to make sure the model doesn't have many outliers in the training data. And should make sure that the R2 values on test set is not very less compared to that on training set. The model should be able to predict well on unseen data.

The model should be robust and generalizable, if not it cant be used to efficiently predict the data accurately.