

## 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?**

### Ridge Regression –

**Objective =  $RSS + \alpha * (\text{sum of square of coefficients})$**

Here,  $\alpha$  (alpha) is the parameter which balances the amount of emphasis given to minimizing RSS vs minimizing sum of square of coefficients.

The optimal value of alpha for the Ridge Regression is = **4.0**

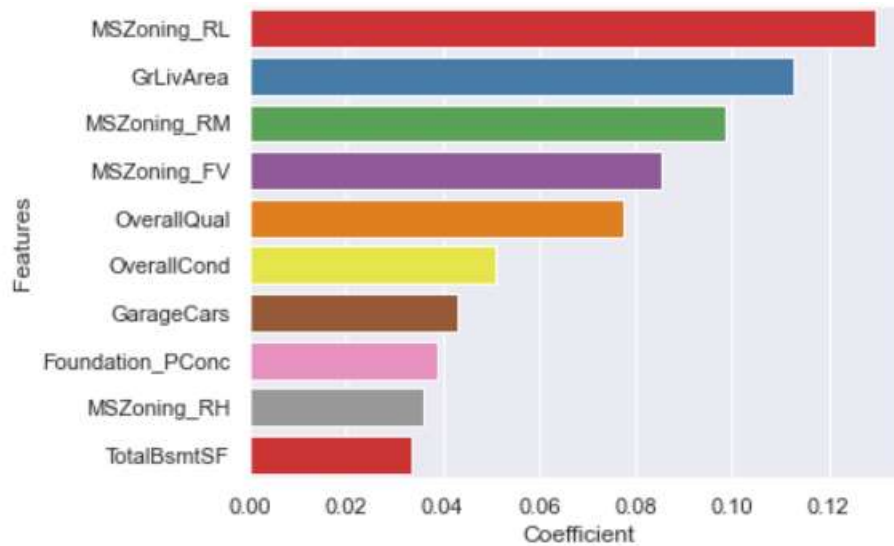
The R2 score for Training data = **91.34%**

The R2 score for Testing data = **90.60%**

The Mean squared error is = **0.0138**

The top 10 predictor features for the model –

	Features	rfe_support	rfe_ranking	Coefficient
10	MSZoning_RL	True	1	0.1296
4	GrLivArea	True	1	0.1128
11	MSZoning_RM	True	1	0.0985
8	MSZoning_FV	True	1	0.0852
0	OverallQual	True	1	0.0774
1	OverallCond	True	1	0.0508
6	GarageCars	True	1	0.0431
14	Foundation_PConc	True	1	0.0388
9	MSZoning_RH	True	1	0.0361
3	TotalBsmtSF	True	1	0.0334



## Lasso Regression

**Objective =  $RSS + \alpha * (\text{sum of absolute value of coefficients})$**

Here,  $\alpha$  (alpha) works similar to that of ridge and provides a trade-off between balancing RSS and magnitude of coefficients.

The optimal value of alpha for the Ridge Regression is = **0.0001**

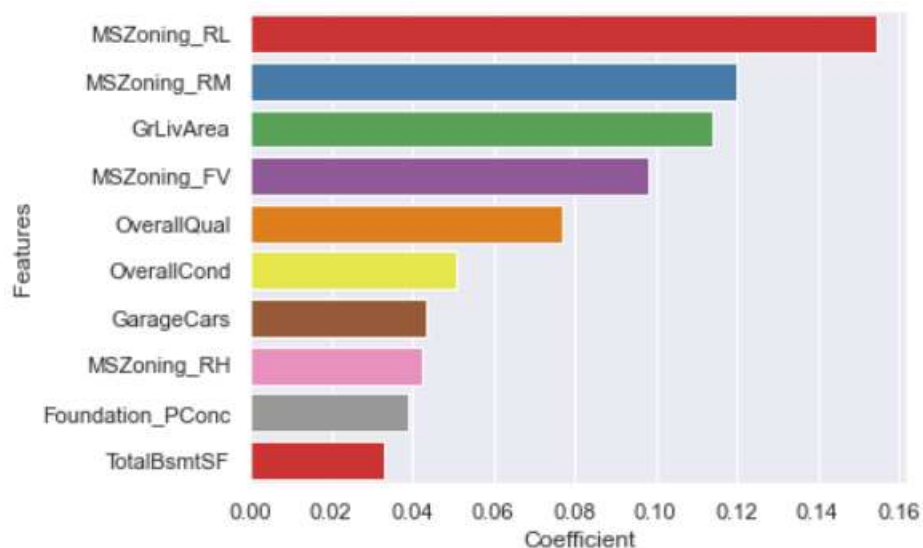
The R2 score for Training data = **91.36%**

The R2 score for Testing data = **90.58%**

The Mean squared error is = **0.0138**

The top 10 predictor features for the model –

	Features	rfe_support	rfe_ranking	Coefficient
10	MSZoning_RL	True	1	0.154383
11	MSZoning_RM	True	1	0.120059
4	GrLivArea	True	1	0.114074
8	MSZoning_FV	True	1	0.098017
0	OverallQual	True	1	0.076969
1	OverallCond	True	1	0.050589
6	GarageCars	True	1	0.043607
9	MSZoning_RH	True	1	0.042382
14	Foundation_PConc	True	1	0.038795
3	TotalBsmtSF	True	1	0.033118



## Model performance after doubling the value of optimal alpha

### Ridge Regression –

Optimal alpha = **4.0**

Current alpha = **8.0**

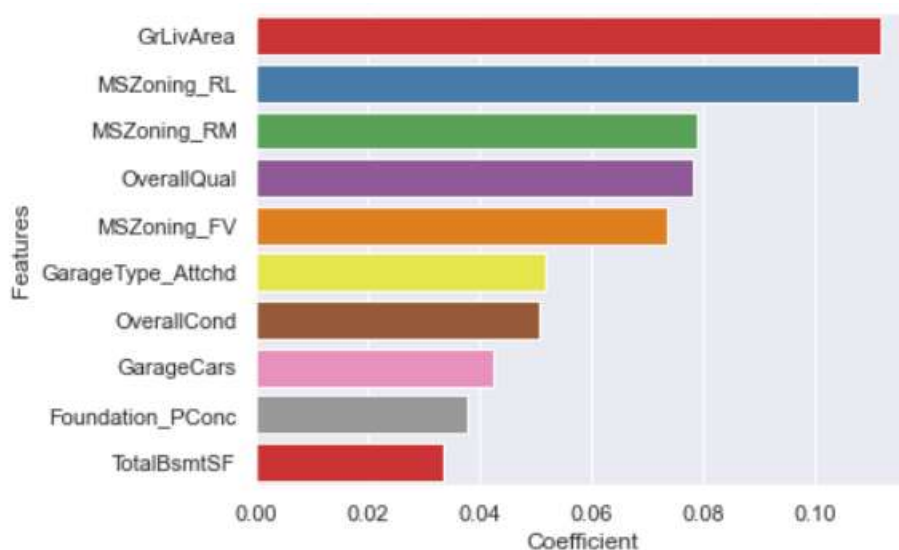
Training R2 Score = **91.29%**

Testing R2 Score = **90.63%**

The mean squared error = **0.0137**

The top 10 predictor features for the model –

	Features	rfe_support	rfe_ranking	Coefficient
9	GrLivArea	True	1	0.1117
22	MSZoning_RL	True	1	0.1078
23	MSZoning_RM	True	1	0.0791
2	OverallQual	True	1	0.0781
20	MSZoning_FV	True	1	0.0738
45	GarageType_Attchd	True	1	0.0520
3	OverallCond	True	1	0.0508
15	GarageCars	True	1	0.0425
43	Foundation_PConc	True	1	0.0380
7	TotalBsmtSF	True	1	0.0336



### Lasso Regression:

Optimal value of alpha = **0.0001**

Current value of alpha = **0.0002**

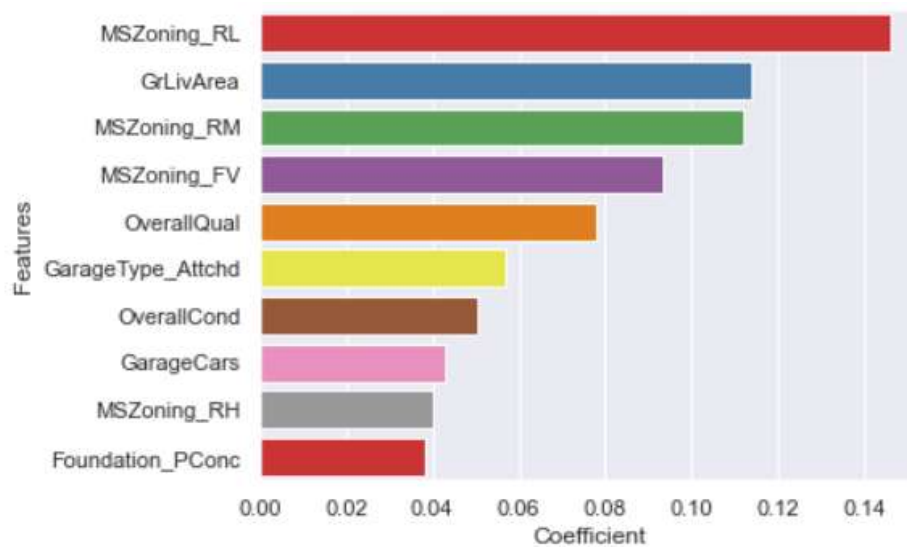
The R2 score for Training data = **91.35%**

The R2 score for Testing data = **90.70%**

The Mean squared error is = **0.0136**

The top 10 predictor features for the model –

	Features	rfe_support	rfe_ranking	Coefficient
22	MSZoning_RL	True	1	0.145832
9	GrLivArea	True	1	0.114062
23	MSZoning_RM	True	1	0.112114
20	MSZoning_FV	True	1	0.093458
2	OverallQual	True	1	0.077799
44	GarageType_Attchd	True	1	0.056797
3	OverallCond	True	1	0.050609
15	GarageCars	True	1	0.043088
21	MSZoning_RH	True	1	0.039912
42	Foundation_PConc	True	1	0.038163



## 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?**

The Lasso Regression would be the best option it would help in feature elimination and the model will be robust.

**Lasso** tends to do well if there are a small number of significant parameters and the others are close to zero (ergo: when only a few predictors actually influence the response). **Ridge** works well if there are many large parameters of about the same value (ergo: when most predictors impact the response).

The **Lasso's** penalty term is based on the sum of absolute coefficients, and the specification of a penalty coefficient is similar to that of **Ridge regression**; however, the **Lasso** is **more computationally intensive**.

### Question 3

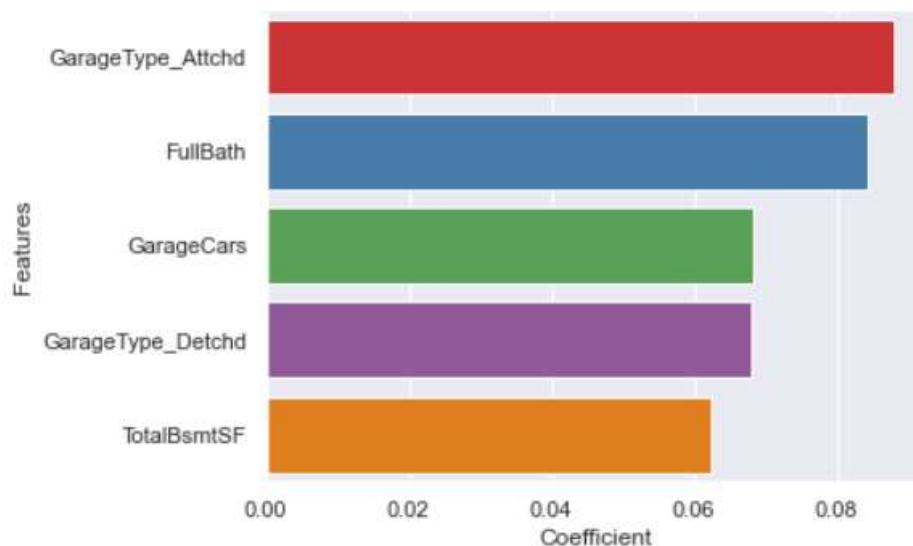
**After building the model, you realized that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?**

The top 5 predictor variables in the model with optimal alpha (0.0001) were –

- MSZoning\_RL
- MSZoning\_RM
- GrLivArea
- MSZoning\_FV
- OverallQual

If we drop them and build the model with same alpha – The top 5 predictor variables would now be –

	Features	rfe_support	rfe_ranking	Coefficient
39	GarageType_Attchd	True	1	0.088069
9	FullBath	True	1	0.084253
13	GarageCars	True	1	0.068380
42	GarageType_Detchd	True	1	0.067933
6	TotalBsmtSF	True	1	0.062501



#### 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?**

Per, Occam's Razor—given two models that show similar 'performance' in the finite training or test data, we should pick the one that makes fewer on the test data due to following reasons:-

- Simpler models are usually more 'generic' and are more widely applicable
- Simpler models require fewer training samples for effective training than the more complex ones and hence are easier to train.
- Simpler models are more robust.
  - Complex models tend to change wildly with changes in the training data set
  - Simple models have low variance, high bias and complex models have low bias, high variance
- Simpler models make more errors in the training set. Complex models lead to over fitting —they work very well for the training samples, fail miserably when applied to other test samples

Therefore to make the model more robust and generalizable, make the model simple but not simpler which will not be of any use. Regularization can be used to make the model simpler. Regularization helps to strike the delicate balance between keeping the model simple and not making it too naive to be of any use. For regression, regularization involves adding a regularization term to the cost that adds up the absolute values or the squares of the parameters of the model.

Also, Making a model simple leads to Bias-Variance Trade-off:

- A complex model will need to change for every little change in the dataset and hence is very unstable and extremely sensitive to any changes in the training data.
- A simpler model that abstracts out some pattern followed by the data points given is unlikely to change wildly even if more points are added or removed

Bias quantifies how accurate is the model likely to be on test data. A complex model can do an accurate job prediction provided there is enough training data. Models that are too naïve, for e.g., one that gives same answer to all test inputs and makes no discrimination whatsoever has a very large bias as its expected error across all test inputs are very high.

Variance refers to the degree of changes in the model itself with respect to changes in the training data.



Thus accuracy of the model can be maintained by keeping the balance between Bias and Variance as it minimizes the total error as shown in the below graph

