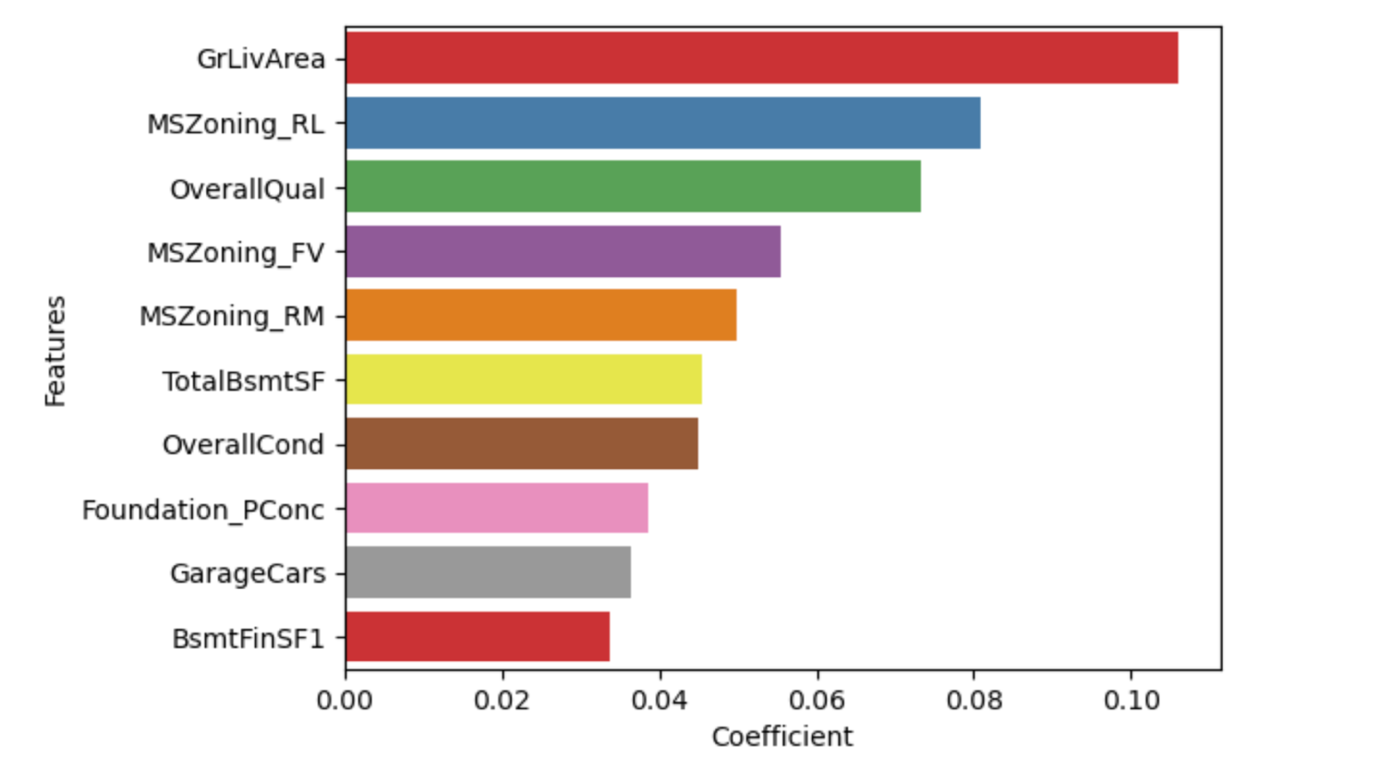
Assignment Questions

Q1: 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?

A: Ridge Regression: The optimal value of Alpha for Ridge regression is

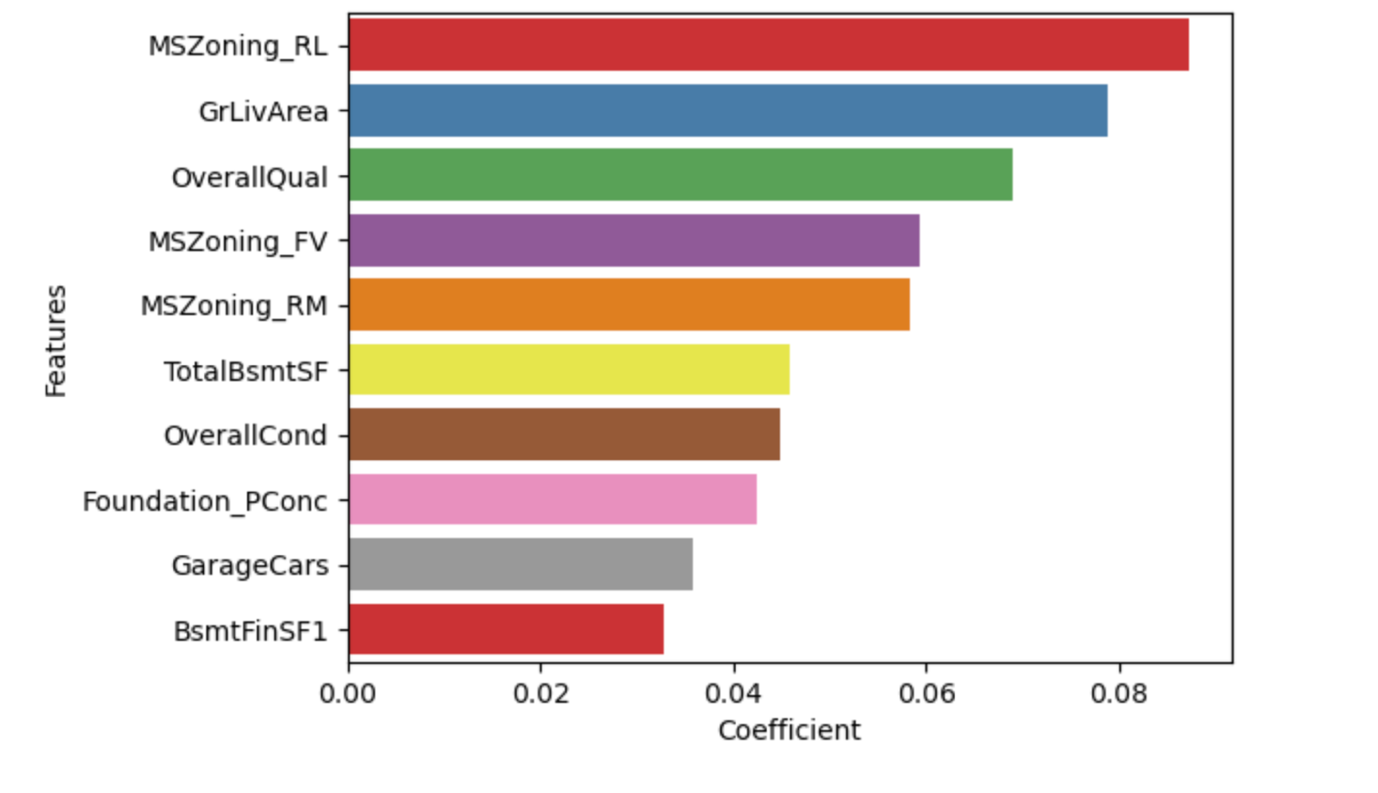
Lasso Regression: The optimal value of Alpha for Lasso Regression is 1. In the model here the lasso regression value obtained is 0.0004. When the value of alpha is doubled the 10 topmost predictor variables obtained are as follows:



The 10 topmost predictor variables are:

1. GrLivArea
2. MSZoning\_RL
3. OverallQual
4. MSZoning\_FV
5. MSZoning\_RM
6. TotalBsmTSF
7. OverallCond
8. Foundation\_PConc
9. GarageCars
10. BsmtFinSF1

Ridge Regression: The optimal value of alpha for ridge regression is 1. In the model that is implemented here the value of alpha obtained in 5. By doubling the alpha value the 10 topmost predictor variables are :



The variables are:

1. MSZoning\_RL
2. GrLivArea
3. OverallQual
4. MSZoning\_FV
5. MSZoning\_RM
6. TotalBsmtSF
7. OverallCond
8. Foundation\_PConc
9. GarageCars
10. BsmtFinSF1

Q2: 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?

Since the mean squared error of lasso is slightly lower than ridge it is better to consider lasso regression for the prediction. (Ridge - 0.013710 Lasso - 0.0135357)

Q3: 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: The five most important predictor variables are

1. MSZoning\_RL
2. GrLivArea
3. MSZoning\_RM
4. OverallQual
5. MSZoning\_FV

Q4: 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?

Ans: By keeping the model as simple as possible it will be more robust and generalisable. The simpler the model the bias would be more and variance would be less. Hence the model is more generalisable.

Bias: Bias is defined as error in the model , When the model is weak to learn from data. High bias indicates that the model is unable to learn from data.

Variance: Variance is defined as error in the model when the model tries to over learn the data. When a model has high variance the model performs exceptionally well on training data but poorly on testing data.

The Balance between bias and variance is a must to avoid to underfitting.