Problem Statement - Part II

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?

Lamda Optimal Values for

- Ridge = { **'alpha': 20**}
- Lasso = { **'alpha': 0.0001**}

After doubling the value of Alpha

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	0.923549	0.922000	0.923522
1	R2 Score (Test)	0.904093	0.901761	0.904288
2	RSS (Train)	11.535100	11.768768	11.539062
3	RSS (Test)	6.372939	6.527904	6.359985
4	MSE (Train)	0.107187	0.108268	0.107206
5	MSE (Test)	0.121599	0.123069	0.121476

Most important predictor variables

- GrLivArea
- OverallQual
- MSZoning_RL
- OverallCond
- TotalBsmtSF
- MSZoning_FV
- GarageCars
- BsmtFinSF1
- MSZoning_RM

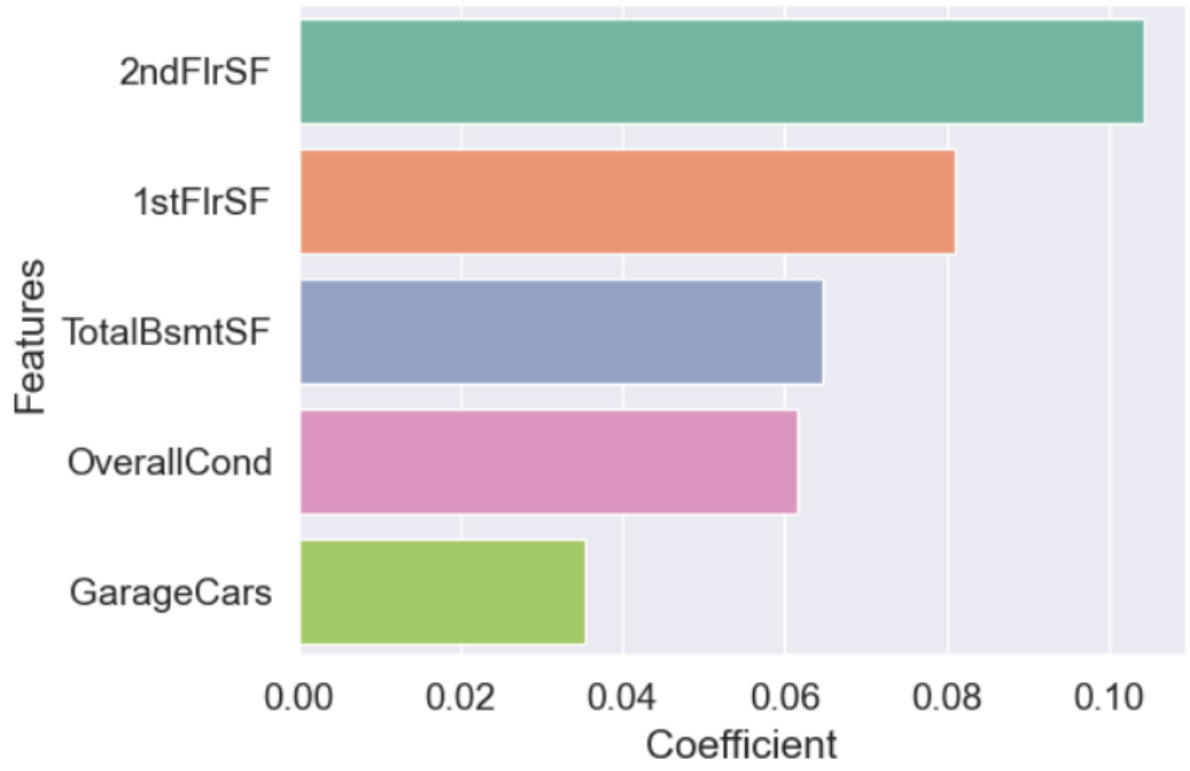
Question 2

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?

Lasso Regression seems to be best with higher test scores and less variance. So it is best to choose the lasso regression model. Because it gives better results with a simpler 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?



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?

If the model is not robust, then the prediction will be not considered trustworthy. We can as the model is generalizable when we got the same accuracy for seen and unseen data.