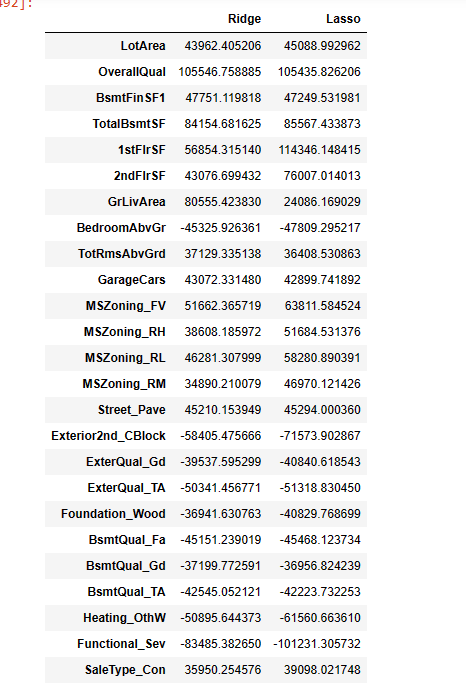
**Subjective Questions**

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?

A: The optimal value of alpha for Ridge Regression is 0.2 and Lasso Regression is 0.0001. When we doubled the alpha, we observed that the r2\_score for training set decreased slightly whereas we see a slight increase in testing set r2\_score. Post doubling the alpha in both Ridge and Lasso regression we see the same predictor variables whereas their coefficients are slightly different. Please find the below screenshot of Coefficients post doubling alpha for ridge and lasso



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?

A: Lasso regression will be the best choice to apply because the r2\_score is slightly higher when compared to r2\_score of ridge regression which helps in developing/achieving the robust model with more relevant features.

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?

A: The five most important predictor variables in the lasso model are

* 'LotArea'
* 'OverallQual'
* 'BsmtFinSF1'
* 'TotalBsmtSF'
* '1stFlrSF'

After dropping the above five predictor variables and performing lasso regression we found that the r2\_score decreased in both train and test sets. Now the most important top 5 predictor variables in Lasso Model are

* 'GrLivArea'
* 'MSZoning\_FV'
* 'MSZoning\_RL'
* 'MSZoning\_RH'
* 'GarageCars'

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: To call a model to be generalised model there should not be much difference between the accuracy of seen (training) and unseen(testing) datasets. If the accuracy score of training set is higher than the accuracy of testing set then the model will become overfitted model because the model memorized the data and the results will be not accurate. The splitting of training and testing dataset will also play a key role in make the model more generalised, because we split the training and testing data as 90 and 10 then we are feeding or training the model on the complete data and testing on only some portion of it which will land into overfitting model.

Robustness of the model is not completely dependent on high test score. It always goes with assumption as test scores are lesser or near by training score but not higher. The robustness of the model will always be calculated based on it performance on unseen data. Apart from this to make a model robust enough we should concentrate on outliers in the data and make sure that outliers which are really showing impact on the model should be retained and others should be dropped from the dataset. While performing outlier analysis to make the model robust we should not build the complex model the model should always be simple and generalised.