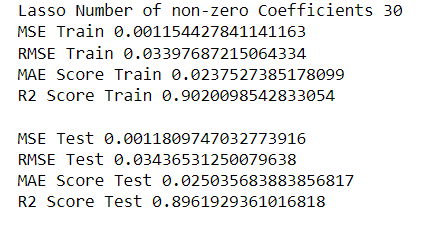
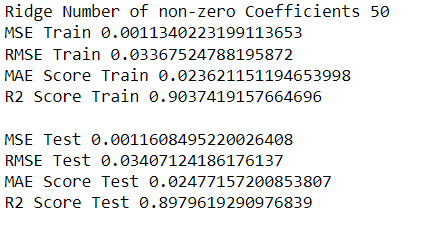
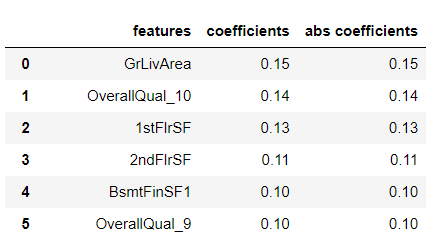
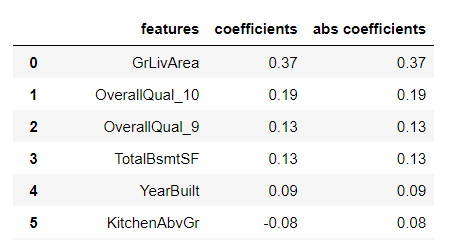
**Question 1**

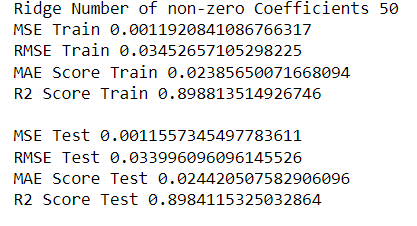
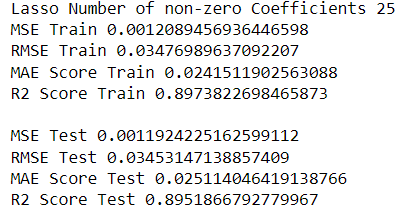
What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

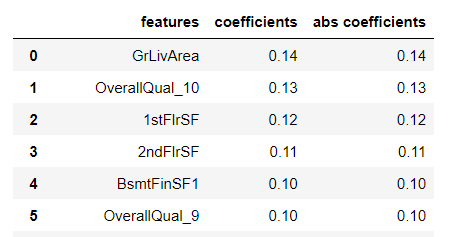
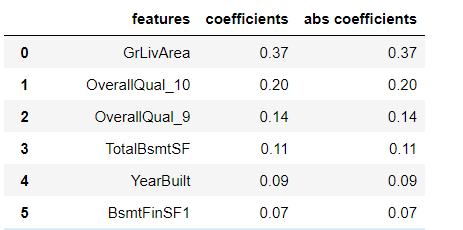
* The Optimum alpha Value (regularization parameter) for Ridge regression is 2
* The Optimum alpha Value (regularization parameter) for Lasso regression is 0.0001



* After increasing the alpha Value to 4 for the Ridge regression
* After increasing the alpha Value to 0.0002 for the Lasso regression

* After doubling the alpha values, a very small value of r2 score decreased on the Train and Test dataset, with small variation in MSE,RMSE,MAE values, but this can be a negligible effect.
* Top five features of Ridge and Lasso regressions do not change but the sixth feature in Lasso regression is changed.
* Few coefficients of top five features are changed.

**Question 2**

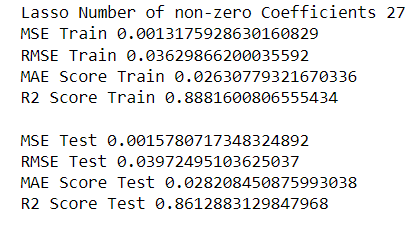
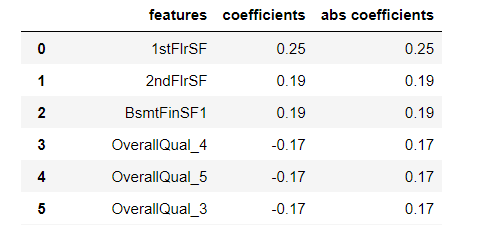
You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

* The Optimum alpha Value (regularization parameter) for Ridge regression is 2.
* The Optimum alpha Value (regularization parameter) for Lasso regression is 0.0001.
* Both Ridge and Lasso regressions performed similarly on the Train and Test datasets, however, I would like to take Lasso as the best-performed model because of fewer number of predictor variables.
* 30 predictor variables using Lasso regression can achieve the same result as Ridge regression with 50 predictor variables.
* A model should be as simple as possible and it should be robust, Lasso is performing best on both train and test datasets using fewer predictor variables.

**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 Optimum alpha Value (regularization parameter) for Lasso regression is 0.0001.
* GrLivArea, OverallQual\_10, OverallQual\_9, TotalBsmtSF, YearBuilt are the top five features of Lasso regression.
* After removing the top five features, The r2 score decreased from 0.90 to 0.88 on the training dataset and decreased from 0.89 to 0.86 on the test dataset.
* 1stFlrSF, 2ndFlrSF, BsmtFinSF1, OverallQual\_4, OverallQual\_5 are the top five features of Lasso regression.

**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?

* A model should be as simple as possible and it should be robust.
* Considering bias and variance tradeoff, underfit model has high bias and low variance, and overfit model has high variance and low bias.
* A good model should have low bias and low variance, it should not be complex and it should utilize a minimum number of variables to make predictions.
* A robust model should perform well on both train and test datasets, so test data is very important for analyzing the robustness of the model.
* Regularization helps to overcome overfitting and a robust model performs well on the unseen dataset.
* Sometimes due to the regularization effect the accuracy on the training dataset may decrease a little bit but the model will do well on the unseen dataset.
* So, it is very important to remember that model should not overfit on the train dataset and it should be flexible to make predictions if we make any changes to the training dataset.
* As complexity increases, bias reduces and variance increases, and we aim to find the optimal point where the total error is least.

