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?**

   For Ridge Regression, from the below plot, we observe that when there's an increases in the alpha value, there's a decrease in the Negative Mean Absolute error for both test and train set. When I set the alpha value as 10, the Mean Squared Error came out be low (0.013743). Hence, I proceeded to build the model with alpha as 10.

   For Lasso Regression, from the below plot, we observe that when there's an increases in the alpha value, there's a decrease in the Negative Mean Absolute error for both test and train set. When I set the alpha value as 0.0004, the Mean Squared Error came out be low (0.013556).
   Since the MSE is low in case of Lasso, the variables predicted by are significant variables for predicting the price of a house.

   When we double the alpha values the model will try to become more and more generalised and we'll get more error for the test and train data.

   Some of the important variable post making the changes for Ridge regression: MS_ZoningRL, MS_ZoningFV, GrLivArea, Neighborhood_StoneBr.
   Some of the important variable post making the changes for Lasso regression: GrLivArea, GarageArea, Fireplaces, LotArea

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

The optimal lambda value in case of Ridge is :10

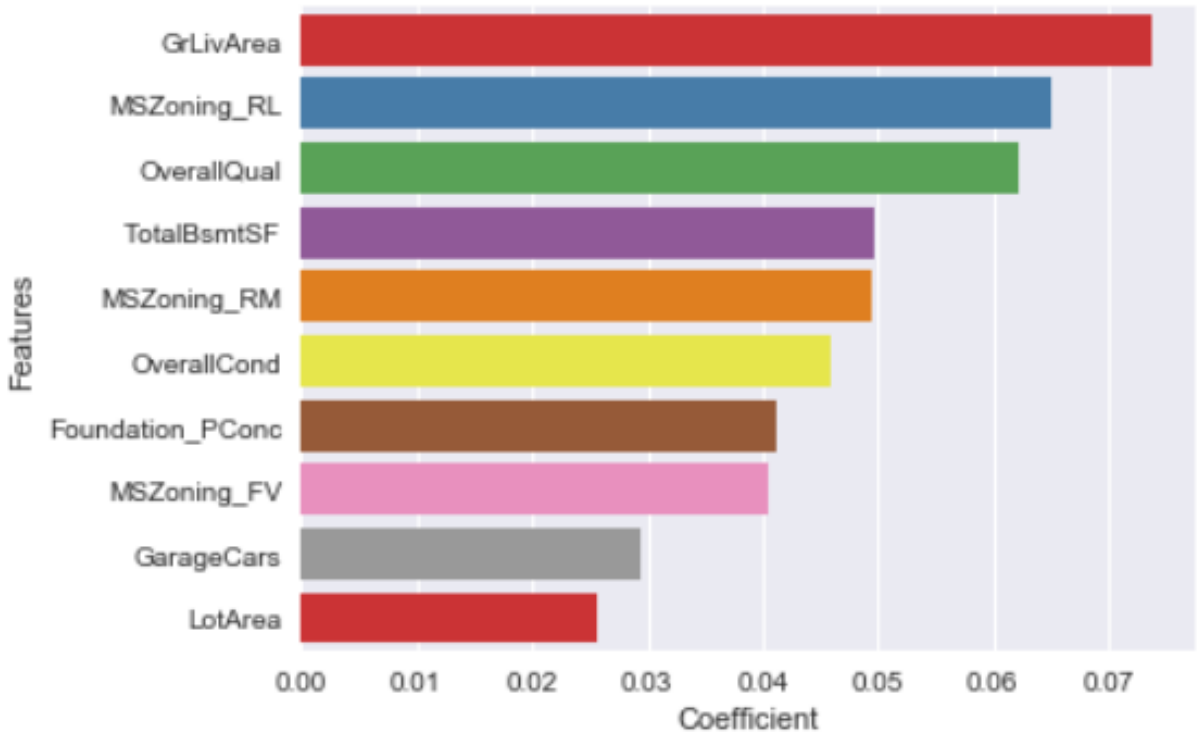
The optimal lambda value in case of Lasso is :0.0004

The Mean Squared error in case of Ridge - 0.013743
The Mean Squared error in case of Lasso - 0.013556

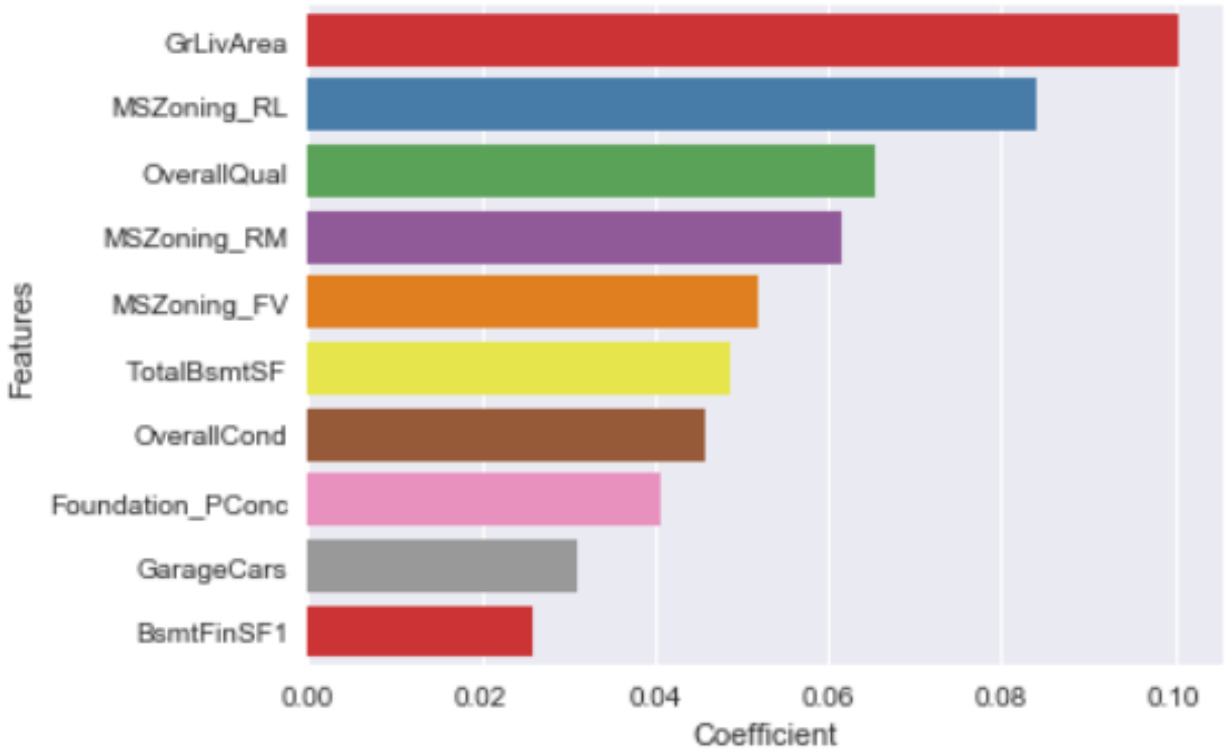
Since the MSE is low in case of Lasso, based on Lasso, the factors that generally affect the price are the Zoning classification, Living area square feet, Overall quality and condition of the house, Foundation type of the house, Number of cars that can be accommodated in the garage, Total basement area in square feet and the Basement finished square feet area

Therefore, the variables predicted by Lasso in the above bar chart as significant variables for predicting the price of a house.

Below are the variables predicted by ridge.



Below are the variables predicted by lasso.



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 variables now?**

   GrLivArea, GarageArea, OverallCond, OverallQuall, TotalBsmtSF

4. **How can you make sure that the model is robust and generalisable? What are the implications of the same for accuracy of the model and why?**

   The aim is to make sure that the model is as simple and general as possible. To achieve this we use the Bias-Variance tradeoff. Bias indicates the difference between the predicted value and the actual value. Variance is a measure of the extent to which the predictions vary for a given data point. To get the best results we need to be able to strike a balance between these two. That is, we don't want to underfit or overfit the model.