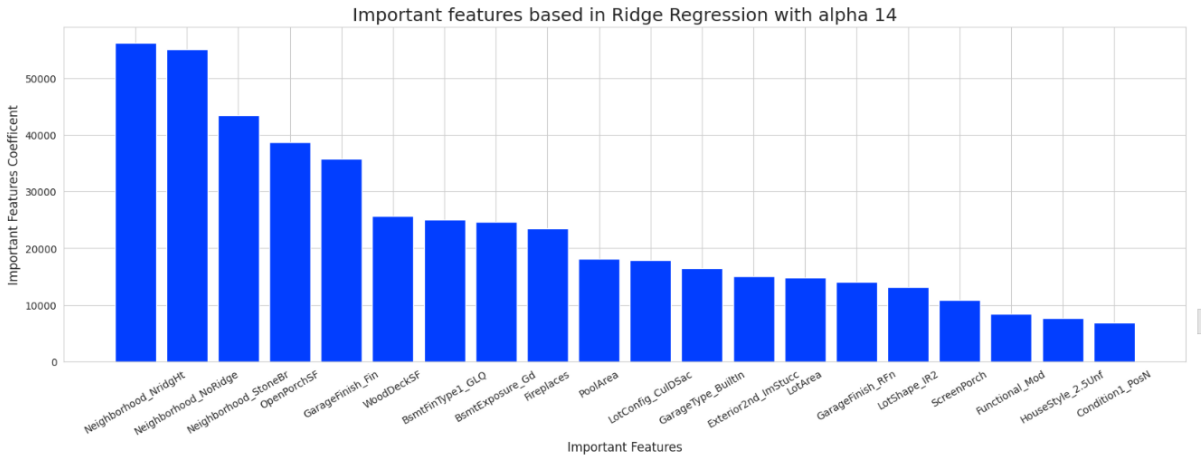
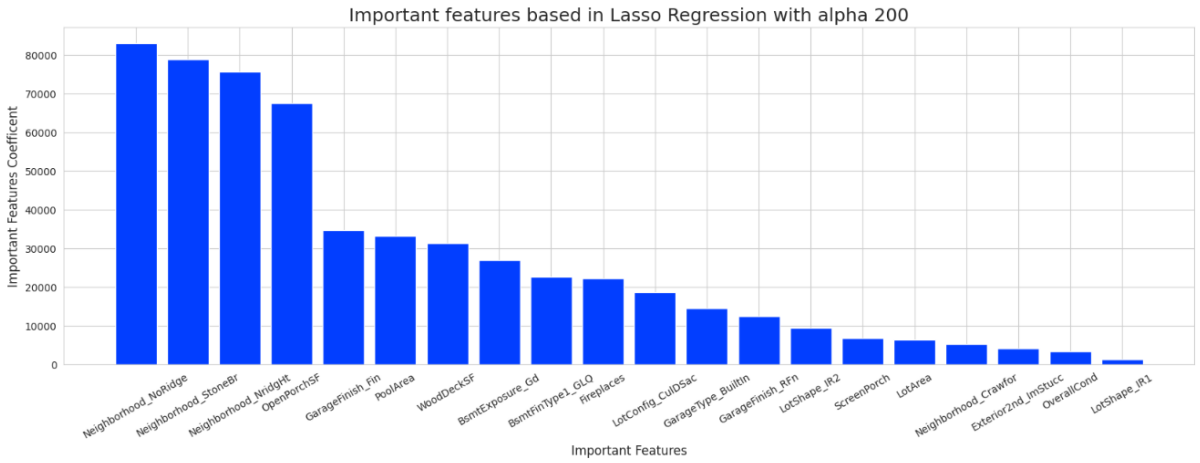
Question 1

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?

Answer:

The Optimal value of alpha for ridge regression is 7.0 and for lasso regression is 100. On doubling up the value of alpha in ridge and lasso we cannot see major change in R2-score but the important features for the model are getting changed. The most important predictor variable after implementing changes in Ridge regression is Neighbourhood Nridgeht i.e No ridge height in neighbourhood. The most important predictor variable in lasso is Neighbourhood No ridge.





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?

Answer:

We will choose the optimal value which we received using GridsearchCV. Because those are the only value which will provide the generalized 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?

Answer:

The consecutive variable which has ranked between 6 – 10 will be chosen as next top five important variable.

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?

Answer:

Model should show almost same accuracy on train and test data. Having high train accuracy and less test accuracy means our model is highly overfitted and cannot answer properly on unseen data. We need to have a bias variance trade-off. A good model is the one which has low-bias and low-variance.