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:
Optimal value for alpha in ridge regression is 1.
Optimal value for alpha in lasso regression is 10.

When we double the alpha in ridge regression:
R2 score on train set with ridge using lambda 1: 0.9246055837608781
R2 score on test set with ridge using lambda 1: 0.8510688618176352
R2 score on train set with ridge using lambda 2: 0.9173300737069547
R2 score on test set with ridge using lambda 2: 0.8535493109107403

There is not much difference in terms of r2_score.

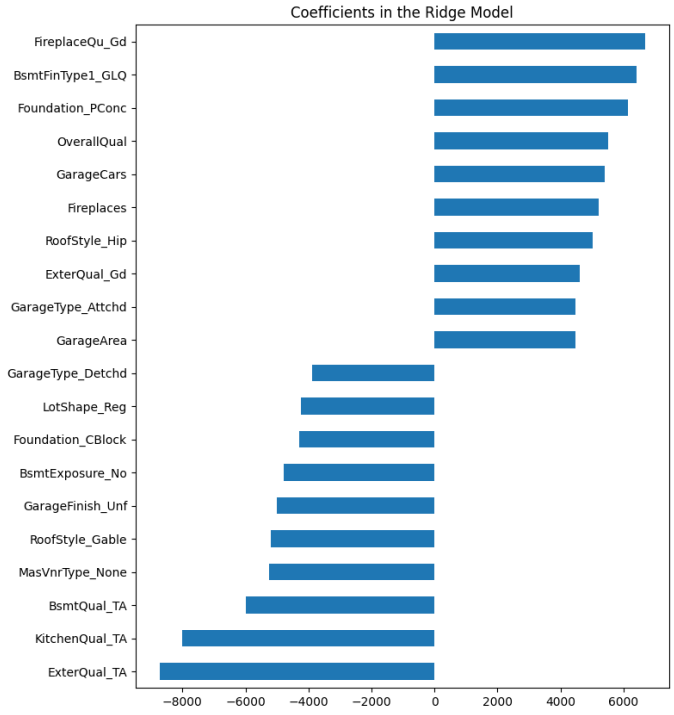
When we double the alpha in lasso regression:
R2 score on train set with lasso using lambda 10: 0.9407466231176879
R2 score on test set with lasso using lambda 10: 0.8217411397047484
R2 score on train set with lasso using lambda 20: 0.929889446463755
R2 score on test set with lasso using lambda 20: 0.8463608748459823

The model seems to be performing better when doubling lambda to 20.

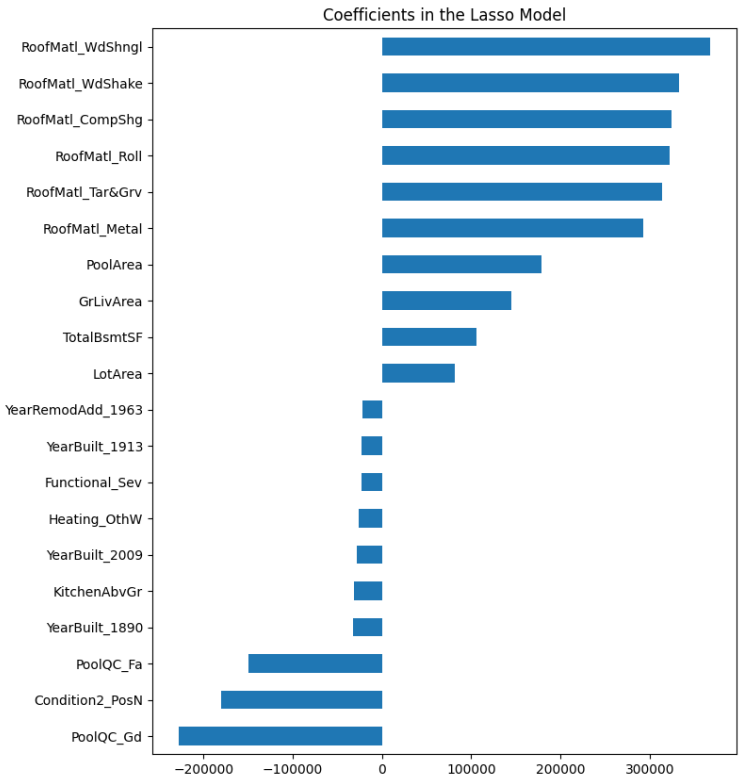
The most important predictor variables after doubling the alpha are:
PoolArea, GrLivArea, RoofMatl, PoolQC_Gd.

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:
The variables we get after ridge regression are:



The variables we get after lasso regression are:

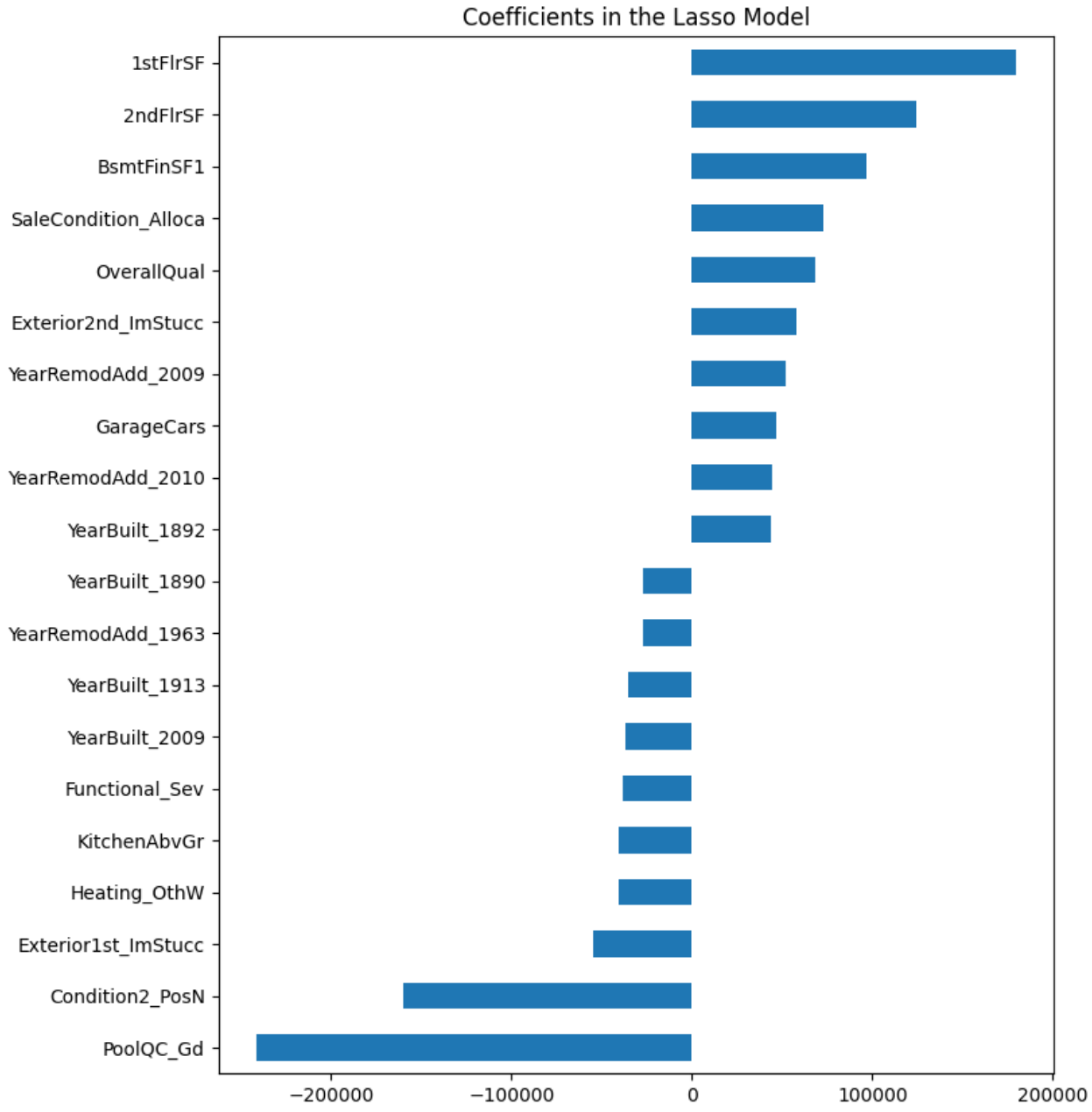


Now it is best to discuss with the business stakeholders and see which variables are important to them to keep and choose the model accordingly.

In my opinion, ridge regression did better in terms of the quality of the variables chosen with respect to the housing market.

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:
After dropping the five most important predictor variables, the result is as below:



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:
To make a model more robust, it needs to be tested on seen and unseen data. To do this we could use cross-validation such as K-fold to make it strong.

Also, regularisation should be performed to penalise for using more unwanted predictor variables in the model.

Next, instead of relying on a single metric such as accuracy or r2_score, it's best to analyse the model with different metrics and make sure that they are making sense.