The company wants to know the following things about the prospective properties:

- **Which variables are significant in predicting the price of a house, and**
- **How well those variables describe the price of a house.**

Also, determine the optimal value of lambda for ridge and lasso regression.

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?

Ridge alpha = 8

Lasso alpha = 0.001

-doubling the alpha values from the optimal values created would effectively halve the coefficients, still the top variables would be most important since tehre were very few positive coefficients after doing RFE ranking, qualscore8/9 would be well used, central air, and neighborhoodcrafor/stoneBR

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?

I would choose to apply ridge since I am not going for features selection and I am simply basing this off of test/train accuracy scores and R2 values, scores were more similar in ridge.

Ridge

Train 0.944985893757463

Test 0.9030147008370966

R2 = 0.055014106242537

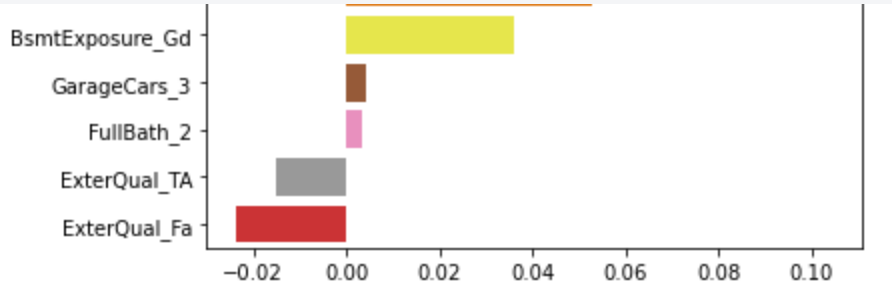
Lasso

Train - 0.9304808925131675

Test 0.8820107408934552

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?



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?

Making sure there is still a low R2 score and the accuracy between the train and test are similar to demonstrate high variance and low bias.