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 1: optimal value of alpha for ridge was 0.3 and optimal value of alpha for lasso was 0.001

With that below are the results:

```
TRAIN SET R2 score is 0.9107085657771093
TEST SET: R2 score is 0.8035244387952423
```

When I doubled alpha values to – 0.6 and 0.002 I get the below results:

```
TRAIN SET R2 score is 0.8717340922925314
TEST SET: R2 score is 0.805121221414439
```

Hence with the increase (doubling) in alpha lasso is performing better than ridge.
shown through R2 Score.

As per the final heatmap OverallQuall is having the highest correlation with Sales_price so as the most important predictor variable.

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 2: I will choose Lasso as it improves the performance as shown by R2 score values, also we know our data has high number of predictor variable and that will make the model very complex and to get rid of that we need to reduce the number of variables by making the coef's 0 (feature elimination) which is done with lasso. Also, it performs good with a R2 score which is less different than the training score.

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?

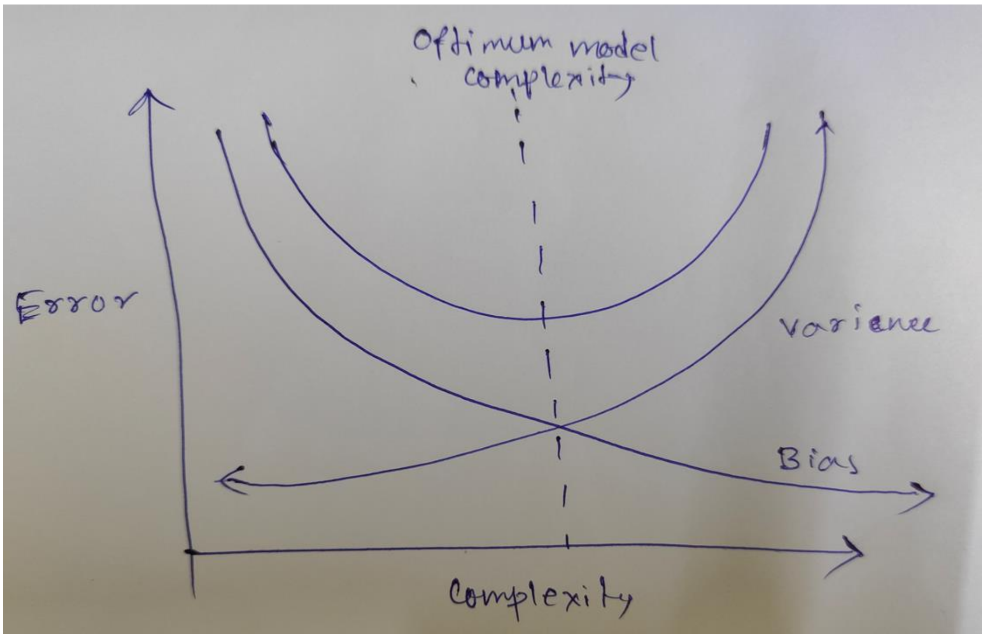
five most important Predictors now: ['LotArea','SaleType_New','BsmtFinSF1','ExterQual_Gd','Exterior2nd_VinylSd']

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?

A model should be robust by making sure it gives correct results means error should be towards 0, however the model should also be practical so that it can be used on the unseen data as in real life it will do that. To make sure it does not over fit to the training data we need to regularize the model so it slightly leans towards making error to make sure it does not over fit. It can be visible by checking the difference between model evaluation score of the training and test data. We can achieve this through ridge and lasso regression.

As you see in the below picture we should be considering a trade off point between error and fit of the data (bias and variance) which should be optimal complexity of the model , hence the model stay robust as well as generalisable.



Bias variance Trade off