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?

- The Optimal value on RFE supported variables for Ridge is 13 and for Lasso it is 0.0005. If these values are doubled then for both ridge and lasso the r2 score on both Train and Test sets start decreasing, indicating the rise of variance and dip of bias which basically means the model starts underfitting.

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 will choose LASSO regression on RFE supported variables, because after hyperparameter tuning , both train and test sets have similar r2 score and Lasso additionally pushes the unimportant variables to 0 , therefore identifying the correct variables and coefficients is easier and efficient.

After building the model, you realized 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?

- After dropping the top 5 variables from the LASSO model , the accuracy of the model dropped by 3 % for train set and 4 % for test set when using alpha 0.0005 . The top 5 most important predictor variables now in my model are:
  1.First Floor square feet
  2.Second Floor square feet
  3.MasVnrArea: Masonry veneer area in square feet
  4.Good BsmtExposure: Refers to walkout or garden level walls
  5.FullBath: Full bathrooms above grade

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

- For a robust and generalizable model , the accuracy of the model on both train and test sets should be relatable. Hyperparameter tuning using either ridge or lasso can help maintain the balance between Bias and Variance in a regression model. A small compromise on the training bias will hugely impact(reduce) the variance on the test and the model can therefore be more generalizable. The accuracy of the trained model will decrease when using regularizations but only to improve the accuracy on test data.