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 alpha value for ridge regression is 10
- Optimal alpha value for lasso regression is 100

After doubling the alpha value for ridge

Changes in ridge regression:

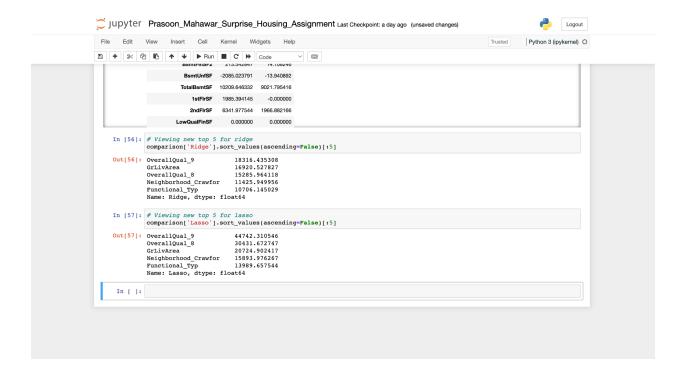
- R2 score of train set decreased from 0.94 to 0.93
- R2 score of test set remained same at 0.89

After doubling the alpha value for lasso

Changes in lasso regression:

- R2 score of train set decreased from 0.94 to 0.92
- R2 score of test set increased from at 0.87 to 0.88

Most important predictor variables after doubling the alpha value



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:

- Lasso regression is used when there are too many variables and we want to select features from it.
- Ridge regression is used when we want to reduce the coefficient magnitude.

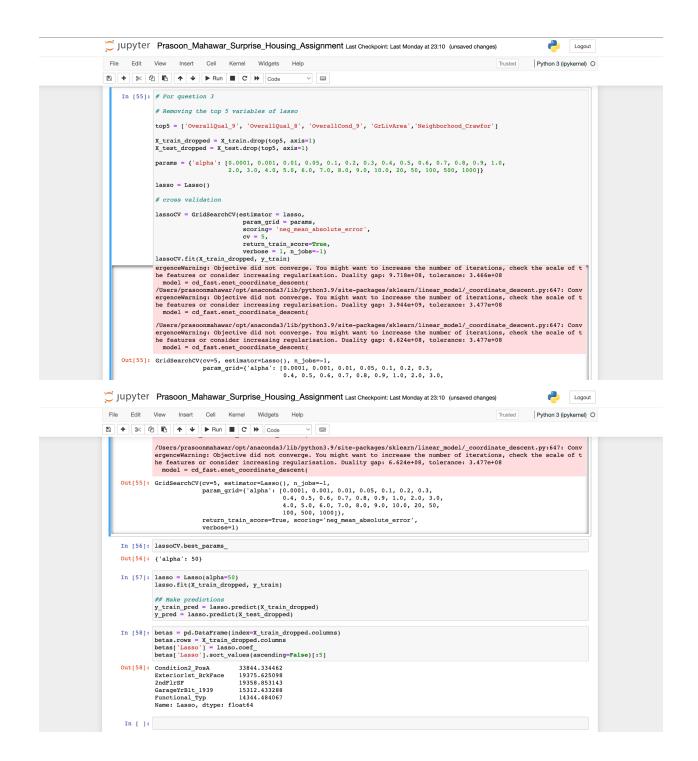
So as per the problem statement we need to tell variables are significant in predicting the price of a house so we will apply Lasso regression.

Question 3

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?

Answer 3

After performing the analysis as shown in the image below



The new five most important variable will be:

Condition2_PosA, Exterior1st_BrkFace, 2ndFlrSF, GarageYrBlt_1939, Functional_Typ

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 4

To make sure a model is robust and generalizable, we have to take care it doesn't overfit. This is because an overfitting model has very high variance and a smallest change in data affects the model prediction heavily. Such a model will identify all the patterns of a training data, but fail to pick up the patterns in unseen test data. In other words, the model should not be too complex in order to be robust and generalizable.

Now a complex model will have a very high accuracy. So, to make our model more robust and generalizable, we will have to decrease variance which will lead to some bias. Addition of bias means that accuracy will decrease.

In general, we have to strike some balance between model accuracy and complexity. This can be achieved by Regularization techniques like Ridge Regression and Lasso.