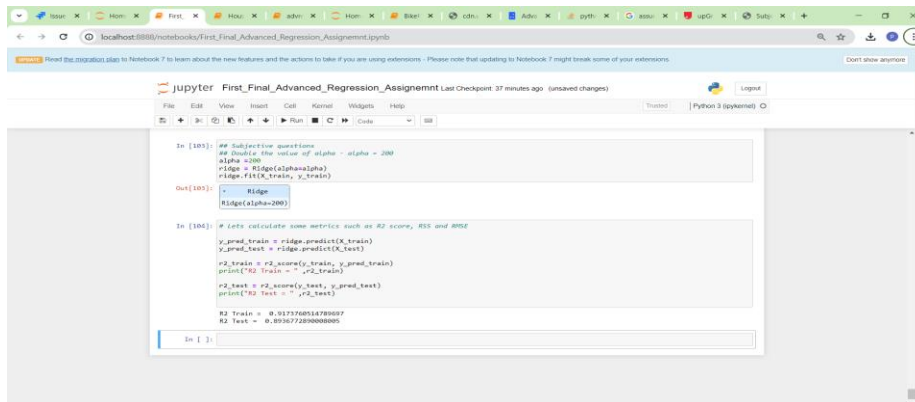


### Question 1:

*What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose to double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?*

Ans: The optimal value for **alpha** in ridge regression is **100** and for lasso it is 500. When we double the value of it in ridge regression – 200 and 1000,

In ridge regression, the difference between the r2 values for the train and test set comes down making the model perform better.



```
In [103]: # Subjective questions
# Double the value of alpha - alpha = 200
alpha = 200
ridge = Ridge(alpha=alpha)
ridge.fit(X_train, y_train)

Out[103]:
Ridge(alpha=200)

In [104]: # Lets calculate some metrics such as R2 score, RMSE and MSE
y_pred_train = ridge.predict(X_train)
y_pred_test = ridge.predict(X_test)
r2_train = r2_score(y_train, y_pred_train)
print("R2 Train = ", r2_train)
r2_test = r2_score(y_test, y_pred_test)
print("R2 Test = ", r2_test)

R2 Train = 0.9173708147898897
R2 Test = 0.8936772800000005
```

For lasso regression, it is very much close to the best value of alpha. So the original model itself is a good one.

The top 5 predictor variable based on the model is **GrLivArea**, **OverallQual**, **TotalBsmtSF**, **BsmtFinSF1**, **Neighborhood\_NridgHt**.

### Question 2:

*You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one would you choose to apply to and why?*

In this case, both lasso and ridge regression models are good, because both have good train and test r2 scores. Ridge performs slightly better because the difference between the test and train r2 is less. However, in general cases, lasso performs better, because it removes the features that are less important or not required for the model and hence makes the model much simpler.

### 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?*

After removing them and building the model, the remaining most important variables are Neighborhood, Exterior\_1<sup>st</sup>, FullBath, BstmtQual, KitchenQual

*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?*

Choosing the right predictor variables, Cross Validation, Regularization etc are mechanisms for the model to be robust and generalizable. Model is robust and generalizable if it performs well on the training set and on the unseen test set. Accuracy and reliability of the model is very critical for it to perform well on the test set.