Problem Statement - Part II

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?

Ans:-
The optimal value of alpha for Rigde is 10 and Lasso is 20.
After Doubling the alpha value for Ridge and Lasso is 100 and 200.

```
#Calculate Ridge alpha = 10 now take 20
alpha = 20
ridge = Ridge(alpha=alpha)

ridge.fit(X_train_new, y_train)
print(ridge.coef_)
y_pred_train = ridge.predict(X_train_new)
y_pred_test = ridge.predict(X_test_new)

r2_train_lr = r2_score(y_train, y_pred_train)
print(r2_train_lr)

r2_test_lr = r2_score(y_test, y_pred_test)
print(r2_test_lr)
```

```
[ 34603.73049923  26484.20572298  21934.80303394 -14304.04419664
  -8398.61636418 -14084.45998769   6439.47929329  -2876.14510134
   9465.42507825  -6539.25981152  -7202.30214512  14420.26024717
  32995.37881477  34748.86476891  14096.97942932   6558.96718207
  11853.54141095  -3044.04198866  -7855.03007219  -3216.04996194
  -1162.42112534  -1805.76458379  -3266.94517238  -1988.18036166
   6476.44947657   8889.17346424   8779.01595969 -12398.61383132
  -9816.20324019 -12917.09642516]
0.7915156423777195
-251403.06787068438
```

These are some of the Top Featured Elements:-

1. Neighborhood_NoRidge
2. Neighborhood_NridgHt
3. 2ndFlrSF
4. OverallQual

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?

Ans:-

We will choose Lasso Regression as observed in Jupyter Notebook as it gives Feature Selection.
It helps us remove the unwanted features from the model. Thus maintaining the accuracy of the model.

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?

Ans:-

Next top 5 features after droping 5 main predictors:-

1. 1stFlrSF
2. MSSubClass_90
3. MSSubClass_120
4. Exterior2nd_AsphShn
5. HouseStyle_1Story_

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?

Ans:-

Following are the steps to make the model Robust and Generalized:-

1. Model Accuracy: It should be more than 70-80%.
2. P-value of all the features are < 0.05.
3. VIF < 7.