Problem Statement Part 2

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:

optimal values for alpha was
Ridge alpha: 1
Lesso alpha: 10

After Making it double:
Ridge alpha: 2:

- **0.8820877173152851 (new train r2 score)**
- **0.87108088253483 (new test r2 score)**
- **remember for alpha: 1 r2_train:0.884340040460635 r2_test: 0.869613280468847**
- **r2 for train data has been decrease slightly but r2 for test data has been increase significantly**

Lesso:

- **0.8854019697956436 (new train r2 score)**
- **0.8670105921065014 (new test r2 score)**
- **remember for alpha: 1 r2_train:0.8859222400899005 r2_test: 0.8646666084570094**
- **r2 for both train and test data has been decreased slightly**

Important Predictors are as follow:

- **LotArea**
- **OverallQual**
- **OverallCond**
- **YearBuilt**
- **BsmtFinSF1**

- **TotalBsmtSF**
- **GrLivArea**
- **TotRmsAbvGrd**
- **Street_Pave**
- **RoofMatl_Metal**

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: r2score of Ridge is slightly higher thant lesso for test_data so we will go with Ridge

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:

- **Five most important predictors are LotArea,OverallQual,YearBuilt,BsmtFinSF1,TotalBsmtSF**
- **0.7934006266062914 - train r2 score**
- **0.7753609527816951 - test r2 score**
- **R2score of training and testing data has decreased**

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: There is not much difference between the r2_score for train and test data so we can model is robust and generalisable

Below are parameter should be satisfied which we have already seen in model:

Model accuracy : > 75%
P-Value : < 0.05
VIF: <5