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 Value of Alpha for Ridge: 10
Optimal Value of Alpha for Lasso: 0.001

The R2 score for Ridge changed from
Train => Original : 91.0 Doubled : 90.3
Test => Original : 87.9 Doubled : 87.3

The R2 score for Lasso Change from
Train => Original : 89.8 Doubled : 88.0
Test => Original : 87.2 Doubled : 85.7

Number of predictors for lasso changed from 87 to 65

Most Important Predictor for Ridge Changed will be same 'LotFrontage' though the coefficient value is less than that of optimal
Most Important Predictor for Lasso Changed will be same 'LotFrontage' though the coefficient value is less than that of optimal

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:

Optimal Value of Alpha for Ridge: 10
Optimal Value of Alpha for Lasso: 0.001

r2_score in train dataset:
r2_score for ridge: 0.91
r2_score for lasso: 0.9
r2_score in test dataset:
r2_score for ridge: 0.88
r2_score for lasso: 0.87

I would apply Lasso as the number of predictors will be less with little to no significant effect on the accuracy

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:

The five most important predictor variable are:
1. BsmtUnfSF
2. TotalBsmtSF

3. 1stFlrSF
4. 2ndFlrSF
5. LowQualFinSF

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:

1. Use models that are more robust to outliers
2. Use appropriate Data Transformation
3. Use more robust error metrics
4. Model explainability is more important than accuracy though generalisable models have problems is attaining higher accuracy