**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?**

- The optimal value of alpha for ridge regression: 6.0
- The optimal value of alpha for lasso regression: 0.0003

**When we double the alpha value for ridge regression to 12.0**

```
ridge2 = Ridge(alpha = 12.0)

ridge2.fit(x_train, y_train)

y_pred_r_train = ridge2.predict(x_train)
y_pred_r_test = ridge2.predict(x_test)

metric_ri = displayR2_RSS_MSE(y_pred_r_train, y_pred_r_test)
```

```
R2_Train : 0.9125823024570312
R2_Test : 0.8182215783150338

RSS_Train : 0.9737862162812543
RSS_Test : 0.9486011360791857

MSE_Train : 0.001041482584258026
MSE_Test : 0.0023655888680278944
```

Decrease in R-squared value

Increase in RSS and MSE

**When we double the alpha value for lasso regression to 0.0006**

```
lasso2 = Lasso(alpha=0.0006)

lasso2.fit(x_train, y_train)

y_pred_l_train = lasso2.predict(x_train)
y_pred_l_test = lasso2.predict(x_test)

metric_la = displayR2_RSS_MSE(y_pred_l_train,y_pred_l_test)
```

```
R2_Train : 0.8781679476514
R2_Test : 0.7916522959043728

RSS_Train : 1.3571436518332924
RSS_Test : 1.0872515393885585

MSE_Train : 0.0014514905367200989
MSE_Test : 0.0027113504722906696
```

Decrease in R-squared value

Increase in RSS and MSE

After the changes in alpha value

- Most important predictor variable for ridge is 1stFlrSF.
- Most important predictor variable for lasso is GrLivArea.

**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?**

| | metric | Linear Regression | Ridge Regression | Lasso Regression |
|---|---|---|---|---|
| 0 | R2 Score (Train) | 9.401289e-01 | 0.922543 | 0.907804 |
| 1 | R2 Score (Test) | -6.063252e+14 | 0.808717 | 0.800493 |
| 2 | RSS (Train) | 6.669316e-01 | 0.862828 | 1.027010 |
| 3 | RSS (Test) | 3.164076e+15 | 0.998202 | 1.041119 |
| 4 | MSE (Train) | 7.132959e-04 | 0.000923 | 0.001098 |
| 5 | MSE (Test) | 7.890464e+12 | 0.002489 | 0.002596 |

If we observe the above table which contains the R-Squared value, RSS and MSE for Ridge and Lasso regression, we can see that Ridge regression seem to perform better than Lasso regression. Though Lasso does feature elimination, the values corresponding to all metrices suggests that features eliminated by lasso were not noisy. So, application of Ridge regression is beneficial in this case.

**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?**

The five most important predictor variables in lasso model

OverallQual_10 0.048007
RoofMatl_WdShngl 0.052658
TotalBsmtSF 0.065120
OverallQual_9 0.070272
GrLivArea 0.251718

The five most important predictor variables in lasso model after eliminating the previous top 5

FullBath_3 0.036959
Neighborhood_NoRidge 0.038371
BsmtFinSF1 0.042905
2ndFlrSF 0.095920

1stFlrSF 0.221502

**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?**

To make sure that if a model is robust and generalisable we have to perform some techniques like cross validation to test the model's performance on data that was not used during training.

Such robust and generalisable model can be used with higher degree of confidence on unseen data. The accuracy of such a model is likely to be higher.