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 :

   a. Optimal Value of alpha for Ridge and Lasso Regression are:

      Ridge :5

      Lasso: 0.0001

   b. Changes in model if double the alpha

      Before change

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	9.556120e-01	0.934305	0.942468
1	R2 Score (Test)	-2.039740e+20	0.898149	0.890590
2	RSS (Train)	2.208701e+00	3.268937	2.862749
3	RSS (Test)	4.267765e+21	2.131036	2.289192
4	MSE (Train)	4.651099e-02	0.056584	0.052952
5	MSE (Test)	3.121499e+09	0.069752	0.072294

      Fig Metric before change

      After change

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	9.556120e-01	0.927094	0.935696
1	R2 Score (Test)	-2.039740e+20	0.898234	0.896663
2	RSS (Train)	2.208701e+00	3.627724	3.199694
3	RSS (Test)	4.267765e+21	2.129251	2.162124
4	MSE (Train)	4.651099e-02	0.059608	0.055981
5	MSE (Test)	3.121499e+09	0.069723	0.070259

      Fig. Metric after change

      Here, we can observe little change in the metrics before and after change.

   c. Important change in predictor

      Before change:

      Ridge:

```
GrLivArea      0.173057
OverallQual    0.133591
TotalBsmtSF    0.099498
1stFlrSF       0.081954
2ndFlrSF       0.070622
```

      Lasso:

```
GrLivArea          0.344184
OverallQual        0.185885
TotalBsmtSF        0.169214
Condition2_PosA    0.133804
YearBuilt          0.076961
```

After Change

Ridge:

```
GrLivArea      0.138779
OverallQual    0.112784
TotalBsmtSF    0.085572
1stFlrSF       0.083005
GarageArea     0.064877
```

Lasso:

```
GrLivArea      0.353034
OverallQual    0.197835
TotalBsmtSF    0.136142
YearBuilt      0.075306
OverallCond    0.074624
```

Here, we can observe that top 3 predictor remains same.

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 are:

   Ridge: 5

   Lasso: 0.0001

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	9.556120e-01	0.934305	0.942468
1	R2 Score (Test)	-2.039740e+20	0.898149	0.890590
2	RSS (Train)	2.208701e+00	3.268937	2.862749
3	RSS (Test)	4.267765e+21	2.131036	2.289192
4	MSE (Train)	4.651099e-02	0.056584	0.052952
5	MSE (Test)	3.121499e+09	0.069752	0.072294

Fig. Metric for optimal value

From the above fig we can say Ridge perform better comparing R2 square value other parameters like RSS and MSE have mixed result.

While Lasso regression helps in feature reduction by making coefficients to zero for lesser important factor.

Due to this Lasso is preferred.

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.
   Ridge:

BsmtFinSF1	0.123410
TotRmsAbvGrd	0.110085
GarageArea	0.104573
LotArea	0.092571
BsmtUnfSF	0.090794

   Lasso:

Condition2_PosA	0.204785
BsmtFinSF1	0.189472
BsmtUnfSF	0.142441
GarageArea	0.136435
TotRmsAbvGrd	0.120643

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.
   The trade-off between bias and variance.
   Simple model are more generalized, more bias and less variance.
   Accuracy in training and test data should be high.
   No outliers in training data makes model more robust
   Regularization helps in achieving more robust model
   If accuracy is not good , model cant be used for predictive analysis.