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 Ridge is 1 and for Lasso it is 0.0001. With these alphas the R2 of the model was approximately 0.86.

:

	Ridge Co-Efficient
OverallQual	0.744589
LotArea	0.315264
Condition2	0.310771
SaleType	0.292043
TotRmsAbvGrd	0.288071
OverallCond	0.279143
FullBath	0.269663
GarageArea	0.250040
BsmtFullBath	0.197690
Fireplaces	0.174043
LotFrontage	0.153911
WoodDeckSF	0.148162
ScreenPorch	0.140376
MasVnrArea	0.123568
GarageCars	0.114842
3SsnPorch	0.104617
Heating	0.090333
SaleCondition	0.089355
BsmtHalfBath	0.086696
BsmtQual	0.085266

[237]:

	Ridge Doubled Alpha Co-Efficient
OverallQual	0.683949
TotRmsAbvGrd	0.279286
LotArea	0.272825
OverallCond	0.264664
FullBath	0.260146
SaleType	0.251837
Condition2	0.245109
GarageArea	0.233239
BsmtFullBath	0.193642
Fireplaces	0.183341
LotFrontage	0.154358
WoodDeckSF	0.145741
GarageCars	0.136393
ScreenPorch	0.133906
MasVnrArea	0.117482
3SsnPorch	0.092345
BsmtQual	0.090132
Heating	0.082531
HalfBath	0.079371
BsmtHalfBath	0.079097

Lasso Regression Model

1]:

	Lasso Co-Efficient
OverallQual	0.828367
Condition2	0.396963
LotArea	0.375448
SaleType	0.326164
TotRmsAbvGrd	0.298923
OverallCond	0.289127
GarageArea	0.276551
FullBath	0.275483
BsmtFullBath	0.199511
Fireplaces	0.159973
LotFrontage	0.151359
WoodDeckSF	0.145459
ScreenPorch	0.139219
MasVnrArea	0.125636
SaleCondition	0.109163
3SsnPorch	0.097613
BsmtHalfBath	0.086918
Heating	0.082362
BsmtQual	0.080287
GarageCars	0.078469

]:

	Lasso Doubled Alpha Co-Efficient
OverallQual	0.834039
Condition2	0.370796
LotArea	0.367510
SaleType	0.300736
TotRmsAbvGrd	0.296646
OverallCond	0.283590
GarageArea	0.274438
FullBath	0.268523
BsmtFullBath	0.197568
Fireplaces	0.159682
LotFrontage	0.152348
WoodDeckSF	0.141475
ScreenPorch	0.132206
MasVnrArea	0.119121
SaleCondition	0.093215
BsmtQual	0.081470
GarageCars	0.079645
BsmtHalfBath	0.077701
3SsnPorch	0.077078
HalfBath	0.073872

Overall since the alpha values are small, we do not see a huge change in the model after doubling the alpha.

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: The optimum lambda value in case of **Ridge** is **1** and **Lasso value is 0.0001**

Which one will we choose and why

Let check Mean Squared Error for both **Ridge** and **Lasso model.**

MSE value for Ridge is 0.024448793176475275
MSE value for Lasso is 0.024310505985110805 so clearly both are almost same.
Since Lasso helps in feature reduction (as the coefficient value of some of the features become zero), Lasso has a better edge over Ridge and should be used as the final model.

Question 3: After building the model, you realized 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 variables in the current lasso model lets remove it. Lets rebuild the lasso model excluding the following.

1. **OverallQual**
2. **Condition2**
3. **LotArea**
4. **SaleType**
5. **TotRmsAbvGrd**

The MSE of the new model without the top 5 predictors above is **0.03680235445753542**

Now the new top 5 important predictor variables are

1. **FullBath**
2. **OverallCond**
3. **GarageArea**
4. **Fireplaces**
5. **LotFrontage**

Question 4 How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?