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 to double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Answer:
Optimal value of alpha for both models:

1. Ridge Regression: 1
2. Lasso Regression: 0.0002

if we double the value of alpha:

	Metric	Lasso Regression	Ridge Regression	Lasso Regression 2	Ridge Regression 2
0	R2 Score (Train)	0.923324	0.923490	0.921332	0.918540
1	R2 Score (Test)	0.887191	0.885687	0.886839	0.890671
2	RSS (Train)	13.053511	13.025251	13.392620	13.868027
3	RSS (Test)	6.836132	6.927326	6.857468	6.625267
4	MSE (Train)	0.112250	0.113282	0.114868	0.116889
5	MSE (Test)	0.125217	0.126049	0.125412	0.123270
6	RMSE (Train)	0.112250	0.113282	0.114868	0.116889
7	RMSE (Test)	0.125217	0.125217	0.015728	0.123270

Ridge:

- R2 score for train is decreased little and R2 score for test is increased little.
- MSE is almost same.
- OverallQual and 1stFlrSF are the top two feature in both cases with little coeff value decreased.

Lasso:

- Both R2 for train and test is almost same after doubling alpha
- MSE is almost same.
- GrLivArea become the top variable over OverallQual after doubling alpha with coeff increase significantly.

The most important predictor variable after the changes:
For Lasso:

	Features	Lasso_values
GrLivArea	GrLivArea	0.764058
OverallQual	OverallQual	0.544344
OverallCond	OverallCond	0.312768
1stFlrSF	1stFlrSF	0.234929
GarageCars	GarageCars	0.195244
MSZoning_RH	MSZoning_RH	0.174221
Neighborhood_StoneBr	Neighborhood_StoneBr	0.159521
MSZoning_RL	MSZoning_RL	0.150340
MSZoning_FV	MSZoning_FV	0.145780
BsmtFinSF1	BsmtFinSF1	0.134156

For Ridge :

	Features	Ridge_values
OverallQual	OverallQual	0.430157
1stFlrSF	1stFlrSF	0.372490
GrLivArea	GrLivArea	0.326685
OverallCond	OverallCond	0.289843
MSZoning_RH	MSZoning_RH	0.216384
2ndFlrSF	2ndFlrSF	0.180160
BsmtFinSF1	BsmtFinSF1	0.179739
MSZoning_RL	MSZoning_RL	0.177493
Neighborhood_StoneBr	Neighborhood_StoneBr	0.175155
GarageCars	GarageCars	0.175056

More details are available in Notebook. Please refer same if needed.

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?

Answer:
After analysis, Optimal value of alpha for both Ridge and Lasso models are:

1. Ridge Regression: 1
2. Lasso Regression: 0.0002

And below are the metrics:

	Metric	Linear Regression	RFE Regression	Lasso Regression	Ridge Regression
0	R2 Score (Train)	9.248783e-01	0.920096	0.923324	0.923490
1	R2 Score (Test)	-5.060041e+24	0.866831	0.887191	0.885687
2	RSS (Train)	1.278896e+01	13.603156	13.053511	13.025251
3	RSS (Test)	3.066354e+26	8.069985	6.836132	6.927326
4	MSE (Train)	1.122495e-01	0.115768	0.112250	0.113282
5	MSE (Test)	8.386252e+11	0.136048	0.125217	0.126049
6	RMSE (Train)	1.122495e-01	0.115768	0.112250	0.113282
7	RMSE (Test)	8.386252e+11	0.136048	0.125217	0.125217

As per metrics,

- Normal Regression is worst model due to negative R2 score
- Ridge and Lasso both have good performance over RFE.
- The MSE of Ridge and Lasso are:
  Ridge - 0.125217
  Lasso - 0.126049

Since Lasso has slightly good R2 score and MSE is less than Ridge and it helps in feature reduction, Lasso has a better edge over Ridge.

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?

Answer:
Top five variables in lasso models are :

1. OverallQual
2. GrLivArea
3. 1stFlrSF
4. MSZoning_RH
5. MSZoning_FV'

After dropping the five most important predictors, we get the following top 5 predictors

1. BsmtFinSF1
2. OverallCond
3. BsmtUnfSF
4. 2ndFlrSF
5. FullBath

More details are available in Notebook. Please refer same if needed.

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?

Answers:
A model is robust when variation in data does not affect the performance of model and it is generalized when it adapts the unseen data well.

To keep the model robust and generalized, model should be kept simple not as simple to underfit. And not so complex that it overfits data. We must verify the trade-off between Variance and Bias also.

Regularization can be done to make model more robust and generalisable. Regularization may reduce accuracy of training data a bit, but accuracy on test data will be significantly improved as – regularization makes sure that only important parameters/features are selected.