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?

Answer:

```
The optimal value of the alpha for the Ridge Regression is : 0.01

The optimal value of the alpha for the Lasso Regression is : 0.0001
```

Prior to doubling the alpha

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	0.904332	0.903956	0.898839
1	R2 Score (Test)	0.823469	0.825736	0.834294
2	RSS (Train)	15.353895	15.414286	16.235615
3	RSS (Test)	12.722314	12.558977	11.942193
4	MSE (Train)	0.122630	0.122871	0.126102
5	MSE (Test)	0.170430	0.169332	0.165122

After Doubling the Alpha

	Metric	Linear Regression	Ridge Regression2	Lasso Regression2
0	R2 Score (Train)	0.904332	0.903056	0.883158
1	R2 Score (Test)	0.823469	0.827610	0.840362
2	RSS (Train)	15.353895	15.558820	18.752219
3	RSS (Test)	12.722314	12.423878	11.504918
4	MSE (Train)	0.122630	0.123446	0.135523
5	MSE (Test)	0.170430	0.168419	0.162071

Observations.

- Ridge Regression after doubling the alpha
  No significant changes in the Ridge Regression after doubling the alpha.

- Lasso Regression after doubling the alpha

There is a slight decrease in the R2 score of training data but a slight increase in the R2 score of test data observed.

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?

We will go ahead with Lasso as the values provides better prediction rates on both test as compared to training data sets.

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?

TotalBsmtSF - Total square feet of basement area
GrLivArea - Above grade (ground) living area square feet
MSZoning_FV - Identifies the general zoning classification of the sale - Floating Village Residential.
MSZoning_RH - Identifies the general zoning classification of the sale - Residential Low Density.
MSZoning_RL - Identifies the general zoning classification of the sale - Residential High Density

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??

The model should be generalized so that the test accuracy is not lesser than the training score. The model should be accurate for datasets other than the ones which were used during training. Too much importance should not given to the outliers so that the accuracy predicted by the model is high. To ensure that this is not the case, the outliers analysis needs to be done and only those which are relevant to the dataset need to be retained. Those outliers which it does not make sense to keep must be removed from the dataset. If the model is not robust, It cannot be trusted for predictive analysis.