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?

➔ Alpha for Ridge is 0.1 and Lasso is 0.001.
Before Alpha Change:

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	0.868823	0.865196	0.865245
1	R2 Score (Test)	0.782129	0.815009	0.818566
2	RSS (Train)	1.362252	1.399917	1.399414
3	RSS (Test)	1.026917	0.871940	0.855177
4	MSE (Train)	0.038501	0.039030	0.039023
5	MSE (Test)	0.051053	0.047043	0.046589

After Alpha double:

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	0.868823	0.861638	0.863186
1	R2 Score (Test)	0.782129	0.816349	0.823289
2	RSS (Train)	1.362252	1.436871	1.420791
3	RSS (Test)	1.026917	0.865627	0.832913
4	MSE (Train)	0.038501	0.039541	0.039319
5	MSE (Test)	0.051053	0.046872	0.045978

We can see that Lasso Regression improved the drop in R2 score of Test but the R2 score for both Ridge and Lasso dropped very slightly. Residual Sum of Squares decreased for Ridge and Lasso both so that is a good sign. MSE for Train remains the same but for Test it reduces slightly. So, all in all the model has improved.

Going ahead with Lasso, the most important predictor vars are:
YearBuilt, 1stFloorSF, KitchenQual_TA(typical)

```
YearBuilt 0.1154357334248914
BsmtFinSF1 0.0792662312960738
1stFlrSF 0.31323733578954754
2ndFlrSF 0.1918528207652361
GarageArea 0.061715207033654 74
PoolArea 0.06695505151577316
MSSubClass_DUPLX -0.04552994014945036
Neighborhood_Crawfor 0.04253215294833564
KitchenQual_Fa -0.09779085176501812
KitchenQual_Gd -0.089511241882842 55
KitchenQual_TA -0.1112203244943943
Functional_Mod -0.0440407542988804
PoolQC_Gd -0.0
```

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?

➔ I choose Lasso Regression because it has least drop in Test performance. It also has lower Residual sum of squares and lower value of Mean Square error. Also, since it has reduced some of the columns coeff to 0 the model is simpler.

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?

➔ Most important predictor variables now are
1. GarageArea
2. PoolArea
3. BsmntFinSF1

```
BsmtFinSF1 0.14085884390631614
GarageArea 0.27217749104919714
PoolArea 0.2470844981727765
MSSubClass_DUPLX -0.026562397461622676
Neighborhood_Crawfor 0.05337665414280459
KitchenQual_Gd 0.0342992039004626
Functional_Mod -0.02292124194472596
PoolQC_Gd -0.03484256164279923
```

Question 4:

How can you make sure that a model is robust and generalisable? What are the implic ations of the same for the accuracy of the model and why?

➔ Model can be made more robust and generalisable by reducing the number of features . If using regularization, we can increase the value of penalty. However, this leads to dr op in accuracy of the model as the model is not able to learn properly from the test dat a.