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 1:

Optimal value of alpha for Ridge = 0.0001
Optimal value of alpha for Lasso = 0.0001

Changes in Ridge when alpha is doubled.

1. Accuracy

Alpha	0.0001
Train R2	0.99999999822542
Test R2	0.99999998001820
Alpha	0.0002
Train R2	0.999999992905675
Test R2	0.999999920372149

2. Top 5 Variables

Alpha	0.0001
Most important variable	PoolArea
Alpha	0.0002
Most important variable	PoolArea

Changes in Lasso when alpha is doubled.

1. Accuracy

Alpha	0.0001
Train R2	0.999943146873117
Test R2	0.999941615407897
Alpha	0.0002
Train R2	0.999775100426939
Test R2	0.999763073038845

2. Top 5 Variables

Alpha	0.0001
Most important variable	PoolArea
Alpha	0.0002
Most important variable	PoolArea

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 2:

For Ridge I will choose alpha = 0.0002 because the drop in accuracy is negligible. And the model is choosing variables that make more business sense. For e.g. MSZoning and Foundation_Stone

For Lasso I will choose alpha = 0.0002 because the drop in accuracy is negligible. And the model is choosing variables that make more business sense. For e.g. MSZoning, Basement Finishing, Foundation Stone, Exterior Quality

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?

Answer 3:

After the model change the five most important predictor variables are:

Variable	(
ScreenPorch	(
ExterQual_Fa	-(
MSSubClass	-(
BsmtQual_Fa	-(
ExterQual_TA	-(

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?

Answer 4:

To make the model robust and generalisable we have to strike balance between bias and variance. That means that the model should be simple without compromising with accuracy. In this case I will proceed with Lasso and alpha as 0.0002 because:

1. The accuracy difference between Lasso and Ridge is insignificant
2. Lasso is also choosing variables that make business sense i.e. MSZoning, Basement Finishing, Foundation Stone, Exterior Quality