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?

For Ridge regression, alpha = 0.1, for lasso, alpha = 20

Once we update alpha=0.2 for ridge and alpha=40 for lasso regression, the following table we notice –

R^2	Ridge Regression		Lasso Regression	
Population	Alpha=0.1	Alpha=0.2	Alpha=20	Alpha=40
Train	0.91	0.90	0.91	0.90
Test	0.83	0.85	0.82	0.85

So, in both the cases, R2 value in training data has increased and test data has increased.

The top 5 variables after the change in alpha values are –

LotArea, OverallQual, OverallCond, YearBuilt, MasVnrArea

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?

As we can see from the above table, the r2_score of ridge regression is slightly higher than lasso for the test dataset so we will choose lasso regression to solve this problem

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?

After removing the top 5 variables, we get the next 5 variables as –

1stFlrSF, BsmtFinSF1, GrLivArea, TotalBsmtSF, 2ndFlrSF

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 overfits when train accuracy is high, but test accuracy is low, and on the other hand the model underfits when both train & test accuracy is low. The model should have good predictive power for datasets other than the ones which were used during training.