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?
The best value of the lambda is 0.4 for Lasso and 0.001 for ridge. If we increase the value of lambda, the model becomes more and more generalized and it may lead to under fitting or the model is highly biased.

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 would choose the value of lambda from the lasso regression as the model has very less variance between the teat and the train data

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?

The important predictors are

1. Garage Area
2. MasVnrArea
3. SaleType
4. Foundation
5. Street

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 can be more robust and genralisable by making sure we have the highest possible accuracy on both test and train data or variance is low between the predicted values of test and train data. In order to achieve that we need to make sure we don't over fit the train data so the model doesn't perform on test data or we should not under fit so model will be having high bias or too genralised.

So choosing the value of lambda such that the bias and variance are optimal while losing the R2 on train data but the diff between predicted and test will be appropriate