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?

Ans: The optimal value of alpha for Lasso Regression is 0.001, and for ridge regression the optimal value is 2.0.
If we double the values of alpha, we may end up underfitting the data.

The important predictor variables in that case would be: Overall Condition, basement full bath, Garage area and so on.

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?

Ans: I would choose the model with ideal complexity, and would pick the model with better R squared score, and the model with the least difference between R squared values of train and test datasets.

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?

Ans: Will drop the five predictor variables from the training data, and train the model again based on the features available in the new incoming data, and build a new lasso regression model.

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?

Ans: We must choose an optimum value of lambda to make our regression model more generalisable and robust. By compromising a little on the bias, we can bring down the variance by using the optimum lambda in our cost function.

A low lambda value implies that, the problem of overfitting is not solved, and an high lambda value means that the model would end up underfitting.