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: The optimal value in ideal world of alpha for ridge and and lasso regression is 1. But the model I created has values Ridge - 0.01551 Lasso - 0.01535 If we double the alpha or try to increase the alpha value, the coefficients will get half or decrease to 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?

Answer 2: I would use lasso because it helps in feature reduction (as the coefficient value of one of the feature became 0), Lasso has a better edge over Ridge.

Also sometimes it depeneds on business requirements. If accuracy is needed, we go for ridge, if robustness needed we go for lasso. As far as R squared is concerned, Ridge is better.

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: The next top 5 most important predictor variables would be – which will have highest coefficient value. It wil be foundation Pconc, GarageCars, BSMTSF, OverallCond, BSMTFinSF1

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: The most important thing as per me is accuracy. It has to be accurate not just in train data, but test data also. And it should be overfitting and not too much complex, with no unnecessary predictors.