Q1. 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?

A1. The suitable value of alpha for both ridge and lasso was found out to be 0.001. The most important predictor variable in this scenario is **GrLivArea ->** Above grade (ground) living area square feet. Which also makes sense intuitively, one pays for the amount of area the house occupies. If doubled the value of alpha for lasso, the predictor variable with highest impact still remains the same but the hierarchy of other variables is affected a bit.

Q2. 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?

A2. 0.001 because it has the highest r2 score.

Q3. 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?

A3. Overall Condition of the house, Masonry veneer area in square feet, Size of garage in square feet, Wood deck area in square feet.

Q4. 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?

A4. The model should have similar performance on both the test and the train sets. Since this is the case, we can conclude that the model is robust. While if this not the case, there could be further need for fine tuning of the model or understanding of the data set.