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 optimal value of alpha for ridge regression is 1.
   The optimal value of alpha for lasso regression is 100.

   Before doubling the r2_score of ridge is : 0.9464859951916192
   Before doubling the r2_score of lasso is : 0.936874485574512

   After doubling the r2_score of ridge is : 0.9451845230280257
   After doubling the r2_score of lasso is :0.9302156693873652

   The most important predictor variables are ‘`Neighborhood_StoneBr`’, ’`SaleType_CWD`’, ’`BsmtExposure_Gd`’, ’`Neighborhood_Crawfor`’, ’`Neighborhood_NoRidge`’

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?

   Since the r2_score of ridge is a bit higher than lasso, i will choose ridge regression.

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 new five predictor variables are ‘`BsmtQual_Gd`’, ‘`Functional_Typ`’, ‘`SaleType_New`’, ‘`OverallQual`’, ‘`KitchenQual_Gd`’

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?

   A model can be said to be robust and generalisable if the predicted values are accurate enough for all the datas, which means for the testing data and for other real time datas which are not seen.

   The implications of the same accuracy of the model is that we have to treat the outliers. Outliers impact a lot in model accuracy.