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:

- ➔ The optimal value of alpha for ridge and lasso regression are:
  Ridge Expression: 1
  Lasso Regression: 10
- ➔ If we double the value of alpha for lasso then alpha would be 2 and double the value of lasso would be 20. So, on doubling the value of alpha in case of Ridge expression, the R2 score of training data has decreased but increased in the case of test data. On doubling the value of alpha in case of Lasso expression the R2 score in case of training data has decreased but increased in the case of test data. Therefore, on doubling the value of alpha in case of Ridge and Lasso predictor values remain the same the only difference is that coefficients of these predictor values has changed.
- ➔ The most important predictor values after change are:
  - Lot size in square feet (LotArea)
  - Rates the overall material and finish of the house (OverallQual)
  - Rates the overall condition of the house (OverallCond)
  - Original construction date (YearBuilt)
  - Type 1 finished square feet (BsmtFinSF1)
  - Total square feet of basement area (TotalBsmtSF)
  - Above grade (ground) living area square feet (GrLivArea)
  - Total rooms above grade (does not include bathrooms) (TotRmsAbvGrd)
  - Pave road access to property (Street_Pave)
  - Roof material_Metal (RoofMatl_Metal)

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:

As the value of R2 score is slightly higher for lasso in case of test data set. So, I will choose Lasso regression to solve this problem.

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:

The five most important predictor variables after the changes are:

- First Floor square feet(11stFlrSF)
- Above grade (ground) living area square feet(GrLivArea)
- Pave road access to property (Street_Pave)
- Roof material_Metal(RoofMatl_Metal)
- Type of roof (Shed)(RoofStyle_Shed)

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:

In order to make the model more robust and generalized, we should ensure that the model shows equal to more accuracy than the training score. The model should be accurate enough for the data except. data which is provided. If the model is not robust and generalized it will predict the wrong output and may lead to loss in business.