Question 1

Answer:

Ridge Alpha 1 , lasso Alpha 10

Predictors are same but the coefficent of these predictor has changed

1. LotArea Lot size in square feet
2. OverallQual Rates the overall material and finish of the house
3. OverallCond Rates the overall condition of the house
4. YearBuilt Original construction date
5. BsmtFinSF1 Type 1 finished square feet
6. TotalBsmtSF Total square feet of basement area
7. GrLivArea Above grade (ground) living area square feet
8. TotRmsAbvGrd Total rooms above grade (does not include bathrooms)
9. Street_Pave Pave road access to property
10. RoofMatl_Metal Roof material_Metal

Question 2:

Answer:

The r2_score of lasso is slightly higher than lasso for the test dataset so we will choose lasso regression to solve this problem

Question 3:

Answer:

five most important predictor variables

1. 11stFlrSF First Floor square feet
2. GrLivArea Above grade (ground) living area square feet
3. Street_Pave Pave road access to property
4. RoofMatl_Metal Roof material_Metal
5. RoofStyle_Shed Type of roof(Shed)

Question 4:

Answer:

The model should be generalized so that the test accuracy is not lesser than the training score. The model should be accurate for datasets other than the ones which were used during training. Too much importance should not give to the outliers so that the accuracy predicted by the model is high. Those outliers which it does not make sense to keep must be removed from the dataset. If the model is not robust, It cannot be trusted for predictive analysis.