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?

- Optimal value of lambda for Ridge Regression = 6
- Optimal value of lambda for Lasso = 0.001
-

Changes in Ridge: Almost Same

Changes in Lasso: Increase in R2 for Train and Test

After changes implemented, most important predictor variables are GrLivArea, OverallQual_Very Good, OverallQual_Excellent, Functional_Typ, Neighborhood_Crawfor, TotalBsmtSF, Exterior1st_BrkFace CentralAir_Y, YearRemodAdd, Condition1_Norm

- 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?

If there are too many variables and our primary goal is feature selection, then we will use **Lasso.**

Ridge Regression will be used if we don't want to get very large coefficients and reduction of coefficient values is our prime goals then we will use Ridge Regression.

Question 3: After building the model, you realized 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?

- GrLivArea
- OverallQual_8
- OverallQual_9
- Functional_Typ
- Neighborhood_Crawfor
- Exterior1st_BrkFace
- TotalBsmtSF
- CentralAir_Y

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?

A model is robust and generalisable when the variance does not affect the model's performance to a great extent and model is able to adapt properly to new, previously unseen data, drawn from the same distribution as the one used to create the model.

From the prespective of **Accuracy**, a too complex model will have a very high accuracy. So, to make our model more robust and generalizable, we will have to decrease variance which will lead to some bias. Addition of bias means that accuracy will decrease.