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?

	Ridge	**Lasso**
Optimal Value	3	100
Doubling	6	200

	Ridge (R2 - Train)	**Lasso (R2 Train)**	**Ridge (R2 - Test)**	**Lasso (R2 Test)**
Optimal Value	0.948	0.941	0.902	0.907
Doubling	0.942	0.931	0.903	0.903

Slight decrease is observed while doubling the optimal values across R2 scores.

Top Features:

- OverallQual
- OverallCond
- YearBuilt
- Neighborhood_StoneBr
- Exterior1st_BrkFace
- TotalBsmtSF
- LotArea

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?

Lasso has sightly higher R2 score in test compared to Ridge, hence we shall consider Lasso.

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?

After dropping the top 5 predictor variables and building and executing the model again, these are the new top 5 important predictors.

- LotFrontage
- MasVnrArea
- BsmtFinSF1
- GrLivArea
- YearRemodAdd

Question 4

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Steps to be followed for more robust and generalizable model:

- Outliner/null handling
- Avoid overfitting.
- Achieving tradeoff between bias and variance.

If the models R2 score is significant on unseen data (there shouldn’t be significant drop), then the model shall be considered as robust and generalizable.