1. What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose to double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

   Ans:
   The optimal value of alpha for both Ridge and Lasso is 500. If we double the alpha in Ridge the R-Square of Training data increases by 2% but on Testing data it drops by 1% so not that many changes in the model performance in Lasso the R-Square on Training data drop by 4% and on Testing data performance increases by 1%. Still, Ridge performed better in this situation.

   The most important variables in Ridge are:
   1. GrLivArea
   2. OverallQual
   3. 1stFlrSF
   4. Neighborhood_NoRidge
   5. GarageCars
   6. TotRmsAbvGrd
   7. RoofMatl_WdShngl
   8. TotalBsmtSF
   9. FullBath
   10. OverallQual

   The most important variables in Lasso are:
   1. GrLivArea
   2. OverallQual
   3. GarageCars
   4. RoofMatl_WdShngl
   5. SaleType_New
   6. Neighborhood_NoRidge
   7. BsmtExposure_Gd
   8. Neighborhood_NridgHt
   9. Neighborhood_Crawfor
   10. BsmtFinType1_GLQ

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?

   Ans:
   As we can see Lasso Regression R-Square is good on training data (93%) but its accuracy drops a little on testing data (84%).
   Compare to Lasso Regression Ridge's R-Square is a little bit low (89%) but its accuracy has not dropped on testing data (86%) as compare to Lasso Regression
   In this scenario, Ridge is performing ideally as it is well-balanced between biased and variance.

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?

   Ans:
   After excluding the 5 most important variables and rebuilding the model the 6 most important variables are:
   - 2ndFlr
   - 1stFlr
   - Neighborhood
   - BsmtExposure
   - HouseStyle
   - FullBath

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?

   Ans:
   We can make sure that the model is not overfitting by checking the training and testing data model performance. The model should be simpler as possible. It should not be more biased and more variance.

More biased means there are some errors in the model. If more error is present in the model will be underfitting.
More variance means the model memorizes the training data and performs well on training data but fails to explain new unseen data.

So, we build the simpler model in a way that the model will be slightly biased But there will be less variance and will be more robust and generalized.
With increasing some errors in the model and decreasing the variance model performs good in both training and testing data and it's called bias-variance trade-off.