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?

   The optimal value of alpha for ridge and lasso are 5 and 100 respectively. If the value of alpha is too high, the underlying pattern in the data will not be learned, meaning the model may underfit. While a low alpha value will lead to overfitting where the training data will be memorized rather than learning the pattern. When we doubled the alpha for ridge and lasso, both the R2s decreased. Following are my top 5 predictors,

   1. Condition1_RRNe
   2. GarageYrBlt_1932.0
   3. SaleType_ConLI
   4. GarageYrBlt_1924.0
   5. GarageYrBlt_2010.0

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?

   Post choosing the optimal alpha value for ridge and lasso, I built 2 separate models using those alpha values and then looked at the R2 and MAPE of train and test. I found out that ridge regression was performing better than lasso as the R2 and MAPE were in a similar range for both train and test sets

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?

   Following are the top 5 predictors,

   1. SaleType_CWD
   2. GarageYrBlt_1915.0
   3. Exterior2nd_Other
   4. MiscVal
   5. Condition1_PosA

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?

   To make sure that the model is robust and generalizable, we can run cross-validation where the model will be trained on multiple train sets and tested on multiple test sets. This way, there is more generalization. When we do this, more often than not we get similar ranges in train and test R2 and MAPE.