RIDGE AND LASSO ASSIGNMENT
QUESTIONS PART 2

Q: - 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?

A: -

For the selected model the optimal value of alpha for ridge regression is 0.03 and optimal value of alpha for lasso regression is 0.00001 and the r2 values at these alphas is as per below table

	Ridge (alpha = 0.03)	Lasso (0.00001)
Updated Training r2	0.9436267223613428	0.9434445807145405
Updated Test r2	0.8250618663341869	0.831486924552079

If we double the value of alpha for ridge and lasso the values of r2 get updated as below

	Ridge (alpha = 0.06)	Lasso (0.00002)
Updated Training r2	0.9401358566820938	0.9373047055641524
Updated Test r2	0.8240643880785239	0.8310286622875408

As observed the r2 values reduce marginally for both training and test datasets when the alpha value is doubled

When the alpha value is doubles the top 5 variables for both ridge and lasso are as per below table

Ranking	Ridge	Lasso
1	RoofMatl_WdShngl	RoofMatl_WdShngl
2	RoofMatl_Membran	RoofMatl_Membran
3	PoolQC_Gd	RoofMatl_WdShake
4	RoofMatl_CompShg	RoofMatl_CompShg
5	RoofMatl_Metal	RoofMatl_Metal

RIDGE AND LASSO ASSIGNMENT
QUESTIONS PART 2

Q2: - 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?

A: - Both ridge and lasso show almost same performance in terms of r2 value of train and test datasets. However lasso can be selected and chosen to apply since it will be giving zero value to coefficients which are having insignificant effect on computing the target variable thus reducing the final number of predictor variables which play a role in the selected lasso regression model.

RIDGE AND LASSO ASSIGNMENT
QUESTIONS PART 2

Q3: - After building the model, you realised 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?

A: - After running the lasso regression model with removing the previous important predictor variables and keeping the same value as alpha as before, the below are the most important revised predictor variables

Ranking	Revised Lasso
1	PoolArea
2	PoolQC_NA
3	PoolQC_Fa
4	PoolQC_Gd
5	MiscFeature_TenC

RIDGE AND LASSO ASSIGNMENT
QUESTIONS PART 2

Q4: - 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?

a. Generalizable model: - An over fitting model is characterized with a high r2 score in the training set and a very low r2 score in the test set. An over fitting model is not generalizable since it becomes specific to the training dataset on which it has been trained. A generalized model is characterized with a high r2 score on test data and less difference between the training and the test dataset. When we try to generalize a model we need to accept a lower r2 score on the training dataset indicating a higher overall error. But this tradeoff becomes acceptable since this can lead to significant increase in the r2 score of the test data. An ideal model should perform on the testing data in the same way as it will perform on the training dataset.

b. Robust model: - A robust model is a model which should be able to make correct predictions of the target variable. A non-robust model can be characterized by a low r2 score on the training dataset itself which indicates that the model is very simple to capture the affecting variables properly. To build a robust model more number of features of the dataset should take part in predicting the target variable. Increased number of variables can start making the model more complex and then the model can start moving towards over fitting and thus reducing the model generalizable properties.

The below graph can help to understand the robustness and generalizable property of any model and its effect on the model accuracy

RIDGE AND LASSO ASSIGNMENT
QUESTIONS PART 2