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?

Answer-1:

During cross validation the optimal value of alpha turned out to be as below

Lasso Regression: 1000

Ridge Regression: 100

Doubling the regularization parameter (alpha) increases regularization strength, resulting in reduced model complexity, increased MSE, and decreased R². However, the consistency of top predictors suggests their strong influence and stability even with higher regularization, highlighting their importance in the predictive model.

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?

Answer-2:

Performance Metrics:

Ridge Regression:

MSE : 937,541,405.21

R²: 0.875

Lasso Regression:

MSE: 1,070,955,420.91

R²: 0.850

Doubling Alpha Values:

Ridge Regression (doubled alpha):

MSE: 1,460,555,988.27

R²: 0.803

Lasso Regression (doubled alpha):

MSE: 1,385,871,689.41

R²: 0.809

Decision

1. **Better Performanc**e: Ridge Regression demonstrated superior performance compared to Lasso Regression in terms of R^2 and MSE with the optimal alpha values. This indicates a better overall fit and predictive power for the dataset.

2. **Feature Retention**: Ridge Regression retains all features, which is advantageous if you believe that all features contribute valuable information to the model. This is particularly useful in a new market where understanding the influence of all variables can provide deeper insights.

3**. Handling Multicollinearity**: Ridge Regression is more effective at handling multicollinearity. This is beneficial when there are correlated features, as it distributes the impact more evenly among them.

Conclusion

For this scenario, I would select Ridge Regression with an optimal alpha value of 100. This choice offers a better fit and preserves all features, which is crucial for comprehending the dynamics of the new market and ensuring that no potentially important information is overlooked.

Question 3

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?

Answer-3:

According to the new Ridge Regression model, the top five most important predictor variables, aside from the initial top five identified by Lasso, are:

YearBuilt

FullBath

TotRmsAbvGrd

Fireplaces

BsmtFinSF1

Based on the new Ridge Regression model, the top five most important predictor variables, excluding the original top five identified by Lasso, are:

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?

Answer-5:

Cross-Validation MSE Scores: A compilation of the Mean Squared Error (MSE) scores for each fold during the cross-validation process.

Mean Cross-Validation MSE: The average of these MSE scores across all folds, providing an estimate of the model's generalization error.

Standard Deviation of Cross-Validation MSE: The variability of the MSE scores across different folds, indicating the consistency of the model's performance.

By examining these metrics, we can ensure that your Ridge Regression model is both robust and generalizable.