## 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:

The optimal value of alpha (regularization parameter) for each model based on our cross-validation is:
Ridge Regression: 100
Lasso Regression: 1000

Doubling the alpha value will increase the regularization strength for both models. This means: Increased regularization, resulting in higher MSE and lower R², indicating reduced model performance.

The top predictors remained consistent, showing their strong influence even after increasing the regularization strength.

## 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 : 937,547,405.28
R²: 0.878
Lasso Regression:
MSE: 1,070,959,420.92
R²: 0.860

Doubling Alpha Values:
Ridge Regression (doubled alpha):
MSE: 1,470,555,988.37
R²: 0.808
Lasso Regression (doubled alpha):
MSE: 1,385,881,689.46
R²: 0.819
Decision
Given the above analysis, I would recommend Ridge Regression for the following reasons:

Better Performance: Ridge regression outperformed Lasso regression in terms of R² and MSE with the optimal alpha values, indicating a better overall fit and predictive power for the dataset.

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

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

Conclusion
I would choose Ridge Regression with the optimal alpha value of 100 for this scenario. It provides a better fit and retains all features, which can be valuable for understanding the dynamics of the new market and ensuring that no potentially important information is discarded.

## 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:
Based on the new Ridge regression model, the top five most important predictor variables (excluding the initial top five from Lasso) are:

YearBuilt
FullBath
TotRmsAbvGrd
Fireplaces
BsmtFinSF1
These variables now have the largest absolute coefficients and are the most significant in predicting house prices in the updated Ridge regression model.

## 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 list of MSE scores for each fold in the cross-validation process.

Mean Cross-Validation MSE: The average of the 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 the folds, indicating the consistency of the model's performance.

By following these steps, we can ensure that your Ridge regression model is robust and generalizable