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?

- The optimal value of alpha for Ridge and Lasso regression was selected using cross-validation.
- This value provides a balance between model complexity and prediction accuracy.
- Doubling alpha in Ridge regression increases regularization strength and shrinks coefficients, but does not remove any variables.
- Doubling alpha in Lasso regression has a stronger effect and removes weaker predictors by setting their coefficients to zero.
- After increasing alpha, the most important predictors are those with strong influence on house prices, such as overall quality, living area, neighborhood, and garage-related features.

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?

- Lasso regression is preferred for this problem.
- It performs regularization as well as feature selection.
- Less important variables are removed by setting their coefficients to zero.
- This results in a simpler and more interpretable model.
- A simpler model is more useful for business decision-making.

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?

- In the original Lasso model, important variables were identified based on non-zero coefficients.
- If the top five predictor variables are not available, the model would need to be rebuilt using the remaining features.
- After rebuilding the model, the importance of variables would change as the model relies on other correlated predictors.
- Based on the remaining features in the dataset, the new important predictors are likely to include overall quality, basement-related area features, garage size or capacity, neighborhood-related variables, and year built or remodeled.
- This indicates that the model can still make reasonable predictions even when some key variables are missing, although overall accuracy may reduce slightly.

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?

- A model can be made robust by evaluating it on data that was not used during training.
- Train-test split helps assess model performance on unseen data.
- Cross-validation ensures the model performs consistently across different subsets of data.
- Regularization techniques such as Ridge and Lasso help prevent overfitting.
- A generalisable model may show slightly lower accuracy on training data but performs better on new, unseen data.