- 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:
  The optimal value of alpha (λ) for both Ridge and Lasso regression is determined using cross-validation where the value corresponding to the minimum validation RMSE is selected. This optimal alpha represents a balance between model complexity and generalisation ability.
    - ➔ If the value of alpha is doubled, the strength of regularisation increases, leading to greater coefficient shrinkage. In Ridge regression, doubling alpha results in Further reduction in the magnitude of coefficients No coefficients becoming exactly zero Increased bias and reduced variance Retention of all predictors in the model In Lasso regression, doubling alpha leads to More aggressive shrinkage Some previously important coefficients becoming exactly zero
    - ➔ Typically, variables related to overall quality, living area, basement area, and garage capacity continue to dominate, while weaker predictors are eliminated in the Lasso 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:
  Between Ridge and Lasso regression, Lasso regression is preferred for this problem. While both methods help reduce overfitting and improve generalisation, Lasso has the added advantage of automatic feature selection by shrinking some coefficients exactly to zero. This simplifies the model and makes interpretation easier, which is particularly important from a business perspective.

- 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:

  When the five most important predictors from the original Lasso model are unavailable, the model must be retrained after excluding those variables. Lasso regression is then applied again to identify the next most influential predictors from the remaining feature set.

After rebuilding the model, the new set of important predictors typically includes variables related to:

- *Location (Neighborhood)*
- *Garage size or area*
- *Number of bathrooms*
- *First floor living area*
- *Year of remodelling or construction*

These variables continue to capture critical pricing information such as usability, amenities, and property upgrades. Although the exclusion of the original top predictors may slightly reduce model performance, the retrained model remains robust and interpretable.

- 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 and generalisable by ensuring that it performs well on unseen data rather than only fitting the training data. This can be achieved through:
- Proper train–test splitting
- Cross-validation during hyperparameter tuning
- Use of regularisation techniques
- Avoiding data leakage
- Evaluating residual plots for assumption violations