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?

A:

In My Final Model, The optimal of Alpha in Ridge and Lasso is as below:

- Ridge - 6
- Lasso - 0.0001

The Mean Squared error in case of Ridge and Lasso are:

- Ridge - 9.23
- Lasso - 9.22

When we double these values, the Model Performance

- For Ridge, the MSE increases from 9.23 to 9.55, and Top Contributors change their positions between each other.
- For Lasso, the MSE reduces from 9.22 to 9.14. The Top Contributors remain same.

With the optimal Alpha, the Most Important Predictor Variables in Ridge are

- MSZoning_RL
- GrLivArea
- MSZoning_RM
- TotalBsmtSF
- OverallQual
- MSZoning_FV

With the optimal Alpha, the Most Important Predictor Variables in Lasso are

- GrLivArea
- MSZoning_RL
- OverallQual
- TotalBsmtSF
- OverallCond
- MSZoning_FV

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?

A:

Ridge and Lasso gave similar results.

We will choose to apply Lasso in this Business context, as the choice of Feature is crucial, and we need to have a simple model, with low number of features we can focus on to make the decision to select the house for investment.

While Ridge can be better in other use cases needing lot more co-related parameters.

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?

A:

The Initial five most Important Predictor variables in Lasso Model were

- GrLivArea
- MSZoning_RL
- OverallQual
- TotalBsmtSF
- OverallCond
- MSZoning_FV

If we create another model, excluding these important Predictor variables, then the next 5 important predictor variables are.

- LotArea
- FullBath
- FireplaceQu_Gd
- GarageCars
- MSSubClass_60

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:

Ensuring that a Lasso model is robust and generalizable involves various considerations and practices. Here are some strategies to enhance the robustness and generalizability of a Lasso model:

Cross-validation: Use cross-validation techniques, such as k-fold cross-validation, to assess the model's performance on multiple subsets of the data. This helps evaluate the model's stability and reduces the risk of overfitting to a specific dataset.

Feature Scaling: Ensure that features are appropriately scaled. Lasso regularization is sensitive to the scale of features, so standardizing or normalizing them can contribute to more stable and consistent results.

Hyperparameter Tuning: Tune the hyperparameters of the Lasso model, particularly the regularization strength (alpha). Grid search or randomized search can be used to find the optimal alpha value that balances bias and variance.

Feature Selection Stability: Assess the stability of feature selection. If the Lasso model consistently selects certain features across different subsets of the data or random seeds, it indicates more robust feature importance.

Outliers Handling: Be mindful of outliers in the data, as Lasso regression is sensitive to extreme values. Address outliers through robust methods or preprocessing techniques.

Validation Set: Split the dataset into training and validation sets. Train the model on the training set and evaluate its performance on the validation set to estimate how well it will generalize to new, unseen data.

Implications for Accuracy:

Sparsity in Feature Selection: Lasso regularization promotes sparsity by driving some coefficients to exactly zero. While this can lead to a more interpretable model with fewer features, it might sacrifice accuracy if important features are penalized too heavily. In our Business case, we need sparse features, to make faster decisions based on simple framework.

Regularization Impact: The strength of regularization (controlled by alpha) influences the trade-off between fitting the training data well and preventing overfitting. Too much regularization might result in underfitting, reducing model accuracy, while too little regularization could lead to overfitting.

Model Interpretability vs. Accuracy: Lasso is often favored for feature selection and interpretability. However, the trade-off is that the model may not capture complex relationships present in the data, potentially affecting accuracy on certain tasks.