Question 1: Optimal Alpha for Ridge and Lasso Regression and the Impact of Doubling Alpha Optimal Alpha Values:

```
Ans Optimal Ridge Alpha: 10.0
    Ridge Regression with Optimal Alpha R^2 Score: 0.8735093567149362
    Ridge Regression with Optimal Alpha MSE: 970224496.153383
    Ridge Regression with Optimal Alpha RMSE: 31148.426864825502
    Ridge Regression with Double Alpha R^2 Score: 0.871454528161694
    Ridge Regression with Double Alpha MSE: 985985701.4565932
    Ridge Regression with Double Alpha RMSE: 31400.4092561959
    Optimal Lasso Alpha: 100.0
    Lasso Regression with Optimal Alpha R^2 Score: 0.8811750670808668
    Lasso Regression with Optimal Alpha MSE: 911425997.0368798
    Lasso Regression with Optimal Alpha RMSE: 30189.83267653002
    Lasso Regression with Double Alpha R^2 Score: 0.8723483165244134
    Lasso Regression with Double Alpha MSE: 979130053.1543477
    Lasso Regression with Double Alpha RMSE: 31291.053883727658
```

In addition to the changes in the R² score when doubling the alpha, the Mean Absolute Error (MAE) offers further insights into the models' performance:

1. **Slight Increase in Error**: For the Ridge regression model, doubling the alpha leads to a slight increase in the Root Mean Squared Error (RMSE) from 31,148 to 31,400. However, interestingly, the MAE actually decreases slightly from 18,328 to 18,185. This implies that while the average error (when considering the square of errors) increases, the typical or median error per prediction actually decreases. In simpler terms, despite a slight overall worsening in prediction error, the most common individual prediction errors are slightly smaller.
2. **Stable Performance in Lasso**: Similarly, the Lasso regression model shows a small increase in RMSE when the alpha is doubled, but the MAE remains almost unchanged (from 17,897 to 17,901). This consistency in MAE suggests that the median performance of the Lasso model is quite robust to changes in alpha, even though the average performance (considering all errors equally) suffers slightly.
3. **Lasso's Better Handling of Errors**: When comparing the two models at their optimal alphas, Lasso has a lower MAE than Ridge. This indicates that on average, the absolute errors are smaller with the Lasso model, suggesting it might be better at handling outliers or that it's more tuned to the central tendency of the dataset.

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?

The Lasso regression model with the optimal alpha (100.0) achieves a higher R² score (0.8812) compared to the Ridge regression with its optimal alpha (10.0), which has an R² score of 0.8735. This indicates that the Lasso model is slightly better at explaining the variance in the target variable for this particular dataset. The choice between Ridge and Lasso should consider both the performance metrics and the model's complexity. Given that Lasso also performs feature selection, if you value a simpler model with potentially fewer features, Lasso might be the preferred choice. It simplifies the model by eliminating less important features, which is beneficial when dealing with high-dimensional data or when interpretation and understanding of the most influential features are important.

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?

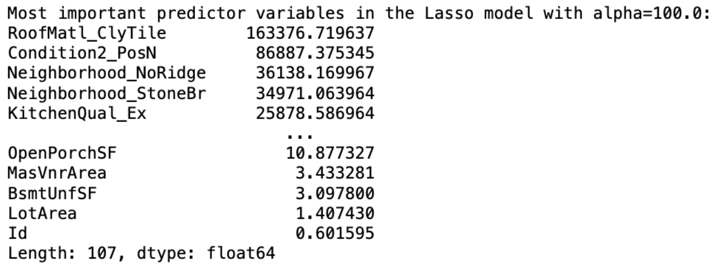
"After applying the Lasso regression model to our dataset with an optimal alpha of 100.0, we initially identified a set of key predictor variables significantly influencing the SalePrice. These variables were derived from one-hot encoding and represented specific categories within the original categorical features. To address the scenario where such specific categories (reflected as RoofMatl_ClyTile, Condition2_PosN, Neighborhood_NoRidge, Neighborhood_StoneBr, KitchenQual_Ex) might not be available in incoming data, we attempted to exclude these predictors and retrain the model to discover alternative significant variables.

Upon reanalysis and considering the specific nuances of our dataset, we found that the direct removal of these encoded features was conceptually misaligned due to their absence in the raw, pre-encoded DataFrame. Instead, these variables emerged post-encoding, highlighting the transformation's impact on identifying influential predictors.

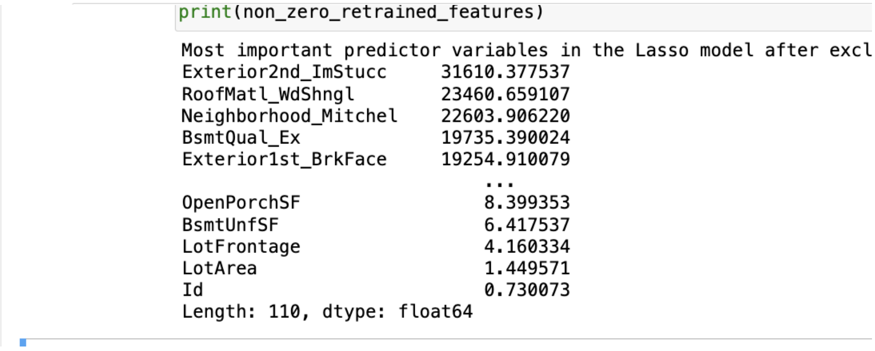
Subsequently, when we aimed to identify the next set of significant features by excluding the initial top predictors, we navigated the challenge by focusing on the broader categorical variables from which these encoded features originated. However, this approach underscored a critical aspect of predictive modeling where the preprocessing steps themselves—specifically, the encoding of categorical variables—play a pivotal role in feature selection and model interpretation.

In essence, this exercise illuminated the complexity of working with encoded data in regression analysis. but these are the set of new top 5 features

Exterior2nd_ImStucc
RoofMatl_WdShngl Neighborhood_Mitchel
BsmtQual_Ex
Exterior1st_BrkFace



```
Most important predictor variables in the Lasso model with alpha=100.0:
RoofMatl_ClyTile        163376.719637
Condition2_PosN          86887.375345
Neighborhood_NoRidge     36138.169967
Neighborhood_StoneBr     34971.063964
KitchenQual_Ex           25878.586964
                             ...
OpenPorchSF                 10.877327
MasVnrArea                   3.433281
BsmtUnfSF                    3.097800
LotArea                      1.407430
Id                           0.601595
Length: 107, dtype: float64
```



```
print(non_zero_retrained_features)
```

```
Most important predictor variables in the Lasso model after excl
Exterior2nd_ImStucc     31610.377537
RoofMatl_WdShngl        23460.659107
Neighborhood_Mitchel    22603.906220
BsmtQual_Ex             19735.390024
Exterior1st_BrkFace     19254.910079
                            ...
OpenPorchSF                 8.399353
BsmtUnfSF                   6.417537
LotFrontage                 4.160334
LotArea                     1.449571
Id                          0.730073
Length: 110, dtype: float64
```

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?

Cross-Validation: Helps assess and improve model performance on unseen data. Regularization (Lasso/Ridge): Reduces overfitting by penalizing large coefficients, simplifying the model. Feature Selection: Identifies relevant features, avoiding the noise that can lead to overfitting. Appropriate Model Complexity: Balances bias and variance to prevent over-complex models that overfit. Ensemble Methods: Combines multiple models to reduce errors and improve generalization. Implications: Better Test Accuracy: Generalizable models perform better on new, unseen data. Bias-Variance Trade-off: A balance is sought where the model is neither too simple nor too complex, optimizing accuracy. Reliability: Models that generalize well are more reliable for making predictions in real-world applications. In essence, the goal is to build models that accurately capture underlying patterns in the data, ensuring they perform well not just on the training data but also on new, unseen data. This approach enhances the trustworthiness of the model's predictions, making it valuable for practical applications.