Question 1

What is the optimal value of alpha for ridge and lasso regression?

The Optimal Value of Alpha for Ridge regression is **3.51**
Training R2 0.8955817207861989 Testing R2 0.8533533529371365

Where as The Optimal Value of Alpha for Ridge regression is **0.91**
Training R2 0.7744864089923619 Testing R2 0.7620538116080745

What will be the changes in the model if you choose double the value of alpha for both ridge and lasso?

If we choose double the value of Alpha for Ridge Regression
Training R2 0.8922527428889842 Testing R2 0.8506121932581238

And If we choose double the value of Alpha for Lasso Regression
Training R2 0.7325575624641615 Testing R2 0.7495500881007091

What will be the most important predictor variables after the change is implemented?

Followings are the Important Predictor variable along with their respective Coefficients

Selected Features and Coefficients (Descending Order):
YearBuilt: 0.001989726925053266
GarageArea: 0.0004429035741300149
GrLivArea: 0.0002706436062495077
TotalBsmtSF: 0.0001945505754640916
WoodDeckSF: 0.00018002207249729084
2ndFlrSF: 4.66606990448493934e-05
MasVnrArea: 4.6102728234679326e-05
BsmtFinSF1: 3.707307759808199e-05
LotArea: 8.788746769066721e-07

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?

When Comparing the results of Ridge and Lasso Regressions we can observe that the optimal value of Alpha for Lasso regression is **0.91** where we get the **Training R2 0.7744864089923619 Testing R2 0.7620538116080745,** where as for Ridge regression the optimal value of Alpha is **3.51** where we get the **Training R2 0.8955817207861989 Testing R2 0.8533533529371365.**

It can be clearly seen that we are getting better results with Ridge regression than Lasso Regression.
So I will choose **Ridge Regression**

Question 3

After building the model, you realized 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 Dropping the five most important predictor variables in Lasso model, here are the five most important predictor variables along with their coefficients for Alpha = 0.91

Training R2 0.687775403325745
Testing R2 0.7090012850314982

Fireplaces: 0.003021068701455231
Electrical: 0.0003572919656798884
CentralAir: 0.00032240171403235363
GarageArea: 0.00027263158576844478
BsmtFinSF2: 0.000236316118762784

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?

To make sure The model is robust and generalizable:
1. Cross-Validation: Techniques like k-fold cross-validation are used to assess the model's performance. Cross-validation helps estimate how well The model is likely to perform on unseen data. It provides a more accurate evaluation of your model's generalization performance compared to a single train-test split.
2. Feature Engineering and Selection: Carefully engineer and select features. Remove irrelevant features and focus on those that are most informative for the task. Avoid overfitting by reducing dimensionality and noise in the data.
3. Regularization: Appropriate regularization techniques to prevent overfitting. Regularization adds a penalty to the model's complexity, which can lead to more generalizable models.
4. Data Preprocessing: Normalize or standardize the data and handling missing values appropriately. Clean, preprocessed data can lead to more generalizable models.
5. Outlier Detection: Identify and handle outliers in the data. Outliers can have a disproportionate influence on the model and lead to poor generalization.
6. Validation Set: Using a separate validation set during model development to fine-tune hyperparameters and make decisions about the model's architecture. The test set should only be used for the final evaluation.

Implications for Model Accuracy:
Robustness and generalizability often come at the cost of absolute model accuracy, at least in terms of the training data. Here's why:
1. Bias-Variance Trade-off: As we make our model more robust and less prone to overfitting (reducing variance), it may become less expressive and have slightly higher bias. This trade-off is necessary to ensure the model can generalize well to unseen data.
2. Simpler Models: Simpler models (with fewer parameters) are often more robust and generalizable. However, they might not capture complex relationships in the data, leading to lower accuracy on the training set.
3. Generalization Focus: The goal of robustness and generalization is to make predictions that are accurate on new, unseen data. While it may result in slightly lower accuracy on the training data, it can help ensure that the model doesn't overfit and performs better in real-world scenarios.