Q1 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? Ans: The Optimal value of alpha for ridge is 2 and for lasso it is 0.0001.

Answer The Optimal value of alpha for ridge is 2 and for lasso it is 0.0001.

```
The R2 Score of the model on the test dataset for doubled alpha is 0.825999
8671982054
The MSE of the model on the test dataset for doubled alpha is 0.00186229053
36132811
The most important predictor variables are as follows:
```

Out[144]:

	Ridge Doubled Alpha Co-Efficient
Total_sqr_footage	0.149028
GarageArea	0.091803
TotRmsAbvGrd	0.068283
OverallCond	0.043303
LotArea	0.038824
Total_porch_sf	0.033870
CentralAir_Y	0.031832
LotFrontage	0.027526
Neighborhood_StoneBr	0.026581
OpenPorchSF	0.022713
MSSubClass_70	0.022189
Alley_Pave	0.021672
Neighborhood_Veenker	0.020098

	Ridge Doubled Alpha Co-Efficient
BsmtQual_Ex	0.019949
KitchenQual_Ex	0.019787
HouseStyle_2.5Unf	0.018952
MasVnrType_Stone	0.018388
PavedDrive_P	0.017973
RoofMatl_WdShngl	0.017856
PavedDrive_Y	0.016840

```
The R2 Score of the model on the test dataset for doubled alpha is 0.823779
8637847479
The MSE of the model on the test dataset for doubled alpha is 0.00188605081
05446826
The most important predictor variables are as follows:
```

Out[145]:

	Lasso Doubled Alpha Co-Efficient
Total_sqr_footage	0.204642
GarageArea	0.103822
TotRmsAbvGrd	0.064902
OverallCond	0.042168
CentralAir_Y	0.033113
Total_porch_sf	0.030659
LotArea	0.025909

	Lasso Doubled Alpha Co-Efficient
BsmtQual_Ex	0.018128
Neighborhood_StoneBr	0.017152
Alley_Pave	0.016628
OpenPorchSF	0.016490
KitchenQual_Ex	0.016359
LandContour_HLS	0.014793
MSSubClass_70	0.014495
MasVnrType_Stone	0.013292
Condition1_Norm	0.012674
BsmtCond_TA	0.011677
SaleCondition_Partial	0.011236
LotConfig_CulDSac	0.008776
PavedDrive_Y	0.008685

Q2 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 : Given the slightly better performance of the Ridge regression in terms of R2 Score and MSE, and considering that the difference in performance is relatively small, Ridge regression appears to be the better choice for this particular dataset and problem. However, if feature selection or model simplicity is a priority, and given the very low alpha for Lasso, re-evaluating the Lasso model might be worthwhile. Remember, the choice can also depend on domain-specific considerations and the ultimate use of the model.

Q3 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?

```
The R2 Score of the model on the test dataset is 0.7330077964268464
The MSE of the model on the test dataset is 0.0028575670906482546
The most important predictor variables are as follows:
```

Out[146]:

	Lasso Co-Efficient
LotFrontage	0.146535
Total_porch_sf	0.072445
HouseStyle_2.5Unf	0.062900
HouseStyle_2.5Fin	0.050487
Neighborhood_Veenker	0.042532

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?
Answer Ensuring that a model is robust and generalizable involves several key strategies and considerations. The goal is to create a model that performs well not only on the training data but also on new, unseen data. Here's how you can achieve this:

Quality and Variety of Data:

Diverse Data: Use a diverse and representative dataset that covers the full spectrum of scenarios the model might encounter in real-world use. The data should include a variety of cases and outliers. Volume of Data: More data often helps the model learn patterns more effectively, leading to better generalization. Cross-Validation:

Use techniques like k-fold cross-validation to assess model performance. This involves splitting the data into several subsets, training the model on some subsets, and testing it on the others. This helps ensure that the model performs well across different data samples. Regularization:

Techniques like Lasso and Ridge regression penalize model complexity, helping to prevent overfitting. Overfitting occurs when a model is too closely fitted to the training data and fails to generalize to new data. Feature Engineering and Selection:

Properly selecting and engineering features can significantly improve model robustness. This includes removing irrelevant features and transforming features to better capture relationships. Model Complexity:

Choose the right level of model complexity. Simpler models are often more generalizable. Overly complex models may fit the training data very well but perform poorly on new data. Hyperparameter Tuning:

Optimize hyperparameters using techniques like grid search or random search with cross-validation. Ensemble Methods:

Use ensemble methods like Random Forests or Gradient Boosting. These methods combine multiple models to improve robustness and generalization. Evaluation Metrics:

Use appropriate evaluation metrics that align with your objectives and consider both accuracy and other aspects like precision, recall, F1 score, etc. Model Updating:

Regularly update the model with new data to ensure it remains relevant and adapts to changes over time. Domain Knowledge Incorporation:

Incorporate domain expertise into model development to ensure that the model respects known relationships and constraints.