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?

Optimal value of alpha:¶
- For Ridge regression : 1.0
- For Lasso Regression : 0.0001

Actual r2 values:
Ridge regression train r2: 0.93
Ridge regression test r2: 0.75

Lasso:
Train: 0.93
Test: 0.74

For Ridge regression aplha is 1.0 and now doubling it and making it 2.0

- Ridge Regression train r2: 0.93
- Ridge Regression test r2: 0.76

For Lasso regression alpha is 0.0001 and not doubling it and making it 0.0002

- Lasso Regression train r2: 0.93
- Lasso Regression test r2: 0.75

The most important predictor variables after changing the alpha value are:

	Feature	Coef	mod
0	LotFrontage	10.174821	10.174821
14	BsmtFullBath	0.822796	0.822796
3	OverallCond	0.568885	0.568885
9	CentralAir	0.492011	0.492011
2	OverallQual	0.484268	0.484268
73	Exterior1st_CBlock	-0.479688	0.479688
33	MSZoning_RH	0.379703	0.379703
35	MSZoning_RM	0.297522	0.297522
36	Street_Pave	0.246404	0.246404
20	GarageQual	0.240831	0.240831

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?

I did not find a drastic difference in the test R2. values from lasso and ridge. However lasso has allowed me to eliminate the variables that are not important. So I will choose the lasso for this model.

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?

Ans:

	Feature	Coef	mod
0	LotArea	10.783831	10.783831
10	FullBath	0.786474	0.786474
1	ExterQual	0.528539	0.528539
6	1stFlrSF	0.506503	0.506503
28	MSZoning_RH	0.501818	0.501818

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 ensure that a model is robust and generalizable, you can follow these practices:

- **Use Cross-Validation:** Split your data into training and validation sets multiple times using techniques like k-fold cross-validation. This helps evaluate the model's performance on different subsets of data and ensures it generalizes well.
- **Regularization:** Use regularization techniques like Ridge or Lasso Regression to prevent overfitting and improve the model's ability to generalize to unseen data.
- **Feature Engineering:** Create meaningful and relevant features that capture the essence of the data, helping the model learn important patterns effectively.
- **Hyperparameter Tuning:** Optimize hyperparameters using techniques like grid search or random search to find the best combination for model performance on unseen data.
- **Evaluate on Test Data:** Finally, assess the model's performance on a completely separate test dataset to ensure it generalizes well and performs accurately on new, unseen data.

Implications for model accuracy: When a model is robust and generalizable, it tends to perform well not only on the training data but also on unseen test data. By prioritizing generalization, the model's accuracy on new data might not be as high as its accuracy on training data. However, a

generalizable model is more reliable and trustworthy in real-world scenarios, where predictions need to be accurate for new instances.