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?

Ans:

- Optimal Value for Ridge: 20.0
- Optimal Value for Lasso: 0.001

If we increase the alpha, we observe that it imposes a stronger penalty on the coefficients.

- In Ridge regression, raising the alpha results in smaller coefficient values without completely removing any coefficients.
- In Lasso regression, increasing the alpha can cause certain coefficients to become zero, effectively removing the corresponding features from the model.

#top 10 features of Lasso after the change:

	Lasso (alpha=0.002)
GrLivArea	0.1295476168551743
OverallQual	0.09985847950401355
OverallCond	0.0577299479658135
GarageArea	0.040826788864928534
MSZoning_RL	0.0405256498180674
BsmtFullBath	0.03234490920969306
Foundation_PConc	0.030593820566961288
SaleType_New	0.029049451670817857
MSZoning_FV	0.027110501366675945
Neighborhood_NridgHt	0.026670850940743326

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?

Ans: Lasso Regression produced same R2 score on test data but lesser RMSE value, than Ridge Regression. Hence, Choosing Lasso as the final 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:

	Lasso
TotalBsmtSF	0.07798200871262419
Fireplaces	0.07533535222726824
BedroomAbvGr	0.06854225048590654
HouseStyle_2Story	0.04809349453239714
Foundation_PConc	0.04776511940896496

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?

Ans: To ensure a robust and generalizable Machine Learning model:

1. Use a diverse dataset.
2. Split data into training and testing sets.
3. Implement cross-validation techniques.
4. Apply regularization techniques.
5. Perform feature engineering.

Having a robust and generalizable model improves accuracy and reliability, as it can make accurate predictions on unseen data and handle different scenarios effectively. This makes the model less affected by noise or outliers, and it can be easily applied to new datasets without significant loss in performance.