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 lambda for Ridge Regression = **10**
- Optimal value of lambda for Lasso = **0.001**

Changes in Ridge Regression metrics:

● R2 score of train set remained same at 0.93

● R2 score of test set increased by 1% to 0.93

Changes in Lasso metrics:

● R2 score of train set remained same at 0.91

● R2 score of test set decreased by 1% to 0.91

Most Important Predictor after the change for Ridge Regression and Lasso regression after the change are shown in the below table:

ridge	lasso
GrLivArea	GrLivArea
OverallQual_8	OverallQual_8
OverallQual_9	OverallQual_9
Neighborhood_Crawfor	Functional_Typ
Functional_Typ	Neighborhood_Crawfor
Exterior1st_BrkFace	TotalBsmtSF
OverallCond_9	Exterior1st_BrkFace
TotalBsmtSF	CentralAir_Y
CentralAir_Y	YearRemodAdd
OverallCond_7	Condition1_Norm

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 model we will choose to apply will depend on the use case.

As we have too many variables here in this case our primary goal is feature selection, so we will use Lasso.

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?

Top 5 Predictor for both old and new feature are as follow,

Top5 Feature After dropping	Old Top5 Feature
2ndFlrSF	GrLivArea
Exterior1st_BrkFace	OverallQual_8
1stFlrSF	OverallQual_9
MSSubClass_70	Functional_Typ
Neighborhood_Somerst	Neighborhood_Crawfor

R square for new model decreased by 1% as it drop to 92% from 93%

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?

- We have to first check based on our domain knowledge that we are not using any irrelevant feature.

- While training model if there is huge difference in Train & Test R square, it means model is overfitting and hence we have to use regularisation techniques.

- To avoid overfitting and generalise we have to make sure we have diverse and representative dataset helps the model learn patterns that are applicable to a broader range of scenarios.

- Applying regularization techniques, such as L1 or L2 regularization, helps prevent the model from becoming too complex and reduces overfitting by penalizing model. This improves the model's ability to generalize to unseen data.

Enhancing the robustness and generalizability of a model ensures that its accuracy on the test set provides a better estimation of its performance on novel data, mitigating the risk of overfitting. By achieving these qualities, the model becomes more adaptable and trustworthy in handling various

scenarios beyond its training environment. This is particularly advantageous in dynamic settings where the distribution of data may evolve over time.

Viewing the matter from an accuracy standpoint, a highly complex model may exhibit impressive accuracy on the training data. However, for the sake of achieving a robust and generalizable model, we must address the issue of high variance, which can be accompanied by bias. The introduction of bias may result in a decrease in accuracy.

In essence, we need to strike a delicate balance between model accuracy and complexity.