Assignment-based Subjective Questions

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?

A1: 1. Ridge Regression Optimal alpha Value - 500

2. Lasso Regression Optimal alpha value -0.001

The values of Metrics like R^2 score, RSS and MSE changes is observed

Ridge:

1.R^2 Score in Train & Test data decreased

2. RSS value in Train & Test data increased

3. MSE Value Train & Test data increased

Lasso:

1. R^2 Score decreased on Train data and Increased on Test data

2. RSS Value Increased on Train data and Decreased on Test data

3.MSE remained same on the Train and Test data

Important predictor variable after change is implemented is '**GrLivArea**'

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?

A2: Lasso is better option to go since R^2 score better results were achieved in Lasso and also Lasso regression method reduce coefficient of the variables to zero which makes easier for removing the unwanted features

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?

A3: Important variables after excluding five most important predictor variables with coefficients

FullBath	0.053
GarageArea	0.047
MSZoning C (all)	0.039
BedroomAbvGr	0.039
HalfBath	0.036

Q4: 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?

A4: for robust and generalised model

1. huge data for training the model so that model understand wide range of patterns

2. Regularization techniques to reduce the overfitting of the model.

The implications of these techniques for the accuracy of the model are that they can potentially reduce the accuracy of the model on the training data. However, they can also significantly improve the accuracy of the model on the test data and in the real world. This is because these techniques help to prevent overfitting and make the model more robust to noise and variations in the data.