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?

Answer:
Optimal Value for Ridge is 0.9
Optimal Value for Lasso is 10

If we double the value in Lasso coefficient of less important gets zero and ridge lowers the coefficient

Important predictor variables are MSZoning, GrLivArea, OverallQual

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?

Answer:
For Ridge, score(train) : 0.89 & score(test) : 0.79
For Lasso, score(train) : 0.77 & score(test) : 0.76

Lasso regression is better than ridge, ridge lowers the coefficient when alpha value is doubled

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?

Answer:
5 most predictor variables are
   1. GrLivArea
   2. GarageArea
   3. TotalBsmtSF
   4. OverallCond
   5. OverallQual

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:

<u>Outlier Handling</u>: Detecting and handling outliers appropriately. Outliers can have a significant impact on model accuracy and robustness.

<u>Cross validation:</u> cross-validation to assess model performance across different subsets of data.

<u>Bias-Variance Trade-off:</u> Understand the bias-variance trade-off. Reducing model complexity may lead to higher bias but lower variance, making the model more robust.