Question 1

What is the optimal value of alpha for ridge and lasso regression?

- Optimal value of alpha for Ridge = 3
- Optimal value of alpha for Lasso = 0.0001

What will be the changes in the model if you choose double the value of alpha for both ridge and lasso?

- When doubled alpha for Ridge, there was no change to the metrics data (R2_score, RSS, MSE and RMSE)
- When double alpha for Lasso, R2 score dropped slightly on train and test dataset, RSS, MSE and RMSE went up by a very small value.

What will be the most important predictor variables after the change is implemented?

- Ridge predictor variables – No change
- Lasso – With alpha 0.0001, important predictor variables are –
    1. GrLivArea
    2. OverallQual_Excellent
    3. TotalBsmtSF
    4. OverallQual_VGood
    5. OverallQual_VExcellent
- When alpha is raised to 0.002, the 5th important predictor changed to LotArea
    1. GrLivArea
    2. OverallQual_Excellent
    3. TotalBsmtSF
    4. OverallQual_VGood
    5. LotArea
- The order of important predictors after the 5th variable changed.

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 will choose the Lasso regression's alpha as 0.0001 as the R2 score was bit higher for train and test datasets. Also, residual sum of squares, mean square error and root mean square error were low as compared to Ridge regression.
- Lasso also performs **feature elimination** by itself choosing significant features.

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?

Five most important predictor variable now are –

- **Neighborhood_NoRidge**
- **LotArea**
- **GarageArea**
- **Neighborhood_Crawfor**
- **Neighborhood_StoneBr**
- **BsmtFullBath**

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?

- A model is considered to be robust when it performs well also on the test data (unseen data) similar to training data. That is if data changes, the performance is still same. To

ensure this is happening, there model should have optimal value of bias and variance. The model should not overfit or underfit.

- Regularization and Bias Variance Trade-off ensures the model is not too simple and not too complex thus making the model more robust. Too complex model has high 8 accuracy but not robust.

- So a balance should be strike between model accuracy and complexity by Regularization methods like Ridge, Lasso.