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?

- Best alpha value for Lasso : {'alpha': 0.001}
- Best alpha value for Ridge : {'alpha': 0.9}

After the alpha is doubled: it will overfit the model and test score will be reduced.
Important predictor variables in that case become:
- MiscVal
- BsmtHalfBath
- LowQualFinSF
- BsmtFullBath

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?

- After creating model in both Ridge and Lasso it is found that the there is no significant difference in r2_scores
- However lasso will penalize more on the dataset
- lasso can also help in feature elimination

hence Lasso will be used.

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?

- EnclosedPorch
- BsmtFullBath
- Neighborhood_Gilbert
- TotRmsAbvGrd

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?

Model needs to be robust and generalisable so as to avoid overfitting and model performs better for both Training and test data sets. The model should not just perform better on training set but its accuracy should also be satisfactory for test data set. To ensure the same outlier treatment is an important aspect to achieve high accuracy in prediction by the model. Giving too much weightage on the outliers to achieve high accuracy in prediction by the model can impact the performance of the model on test data.

To make the model robust and generalisable outlier analysis needs to be done so that the relevant data generic to both training and test data is retained and the outliers which can impact the robustness can be removed.

A non- robust model can lead to inaccurate predictions and hence become unfit of predictive analaysis.