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
- The optimal value of alpha for Ridge Regression is 10.
- The optimal value of alpha for Lasso Regression is 0.001.
- With doubling the value of alpha for both Ridge and Lasso, the effect of regularisation will be double, resulting in a lower value of coefficients and reduced r2 scores as well. In the case of Lasso regression, there will be fewer number of features selected.
- On doubling the value of alpha, the top 5 predictor variables are:
  - Ridge (alpha=20):
    - OverallQual
    - Neighborhood_NoRidge
    - GrLivArea
    - 2ndFlrSF
    - GarageCars
  - Lasso (alpha=0.002):
    - OverallQual
    - GrLivArea
    - GarageCars
    - Neighborhood_NoRidge
    - TotRmsAbvGrd

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
I will choose Lasso regression model (alpha = 0.001) over Ridge regression model (alpha = 10). After creating derived metrics and dummy variables, there was a total of 293 features used to build both models.

While Ridge regression model provides slightly better r2 scores compared to lasso regression, as shown below:

Ridge regression
- train: 0.88765
- test: 0.86208

Lasso regression
- train: 0.82164
- test: 0.80607

The Lasso regression model considers only 40/293 features for prediction, while the Ridge regression model uses all 293 features for prediction.

This makes the Lasso regression model much simpler, with only a slight decrease in r2 score.

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
After removing the top 5 variables and rebuilding the Lasso regression model using the remaining features, the next top 5 variables are:
- GarageArea
- TotRmsAbvGrd
- BsmtFinType1_GLQ
- 2ndFlrSF
- GarageType_Attchd

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
For a model to be robust and generalisable, it should not be impacted by the outliers in the training data. The model should also not give too much weightage to a few features. We can make a robust model using regularisation which will help us achieve a good balance between variance and bias.

It will ensure that our model performs well for both training data and test data while keeping the model as simple as possible.