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?

Ans)

- The optimal value of alpha for ridge regression is 250.
- The optimal value of alpha for lasso regression is 430.
- When a new lasso model with double the alpha(500) was built, it reduced r2_score to 0.921 on the training set and increased r2_score to 0.814 on the testing set.
- Also of the total 333 coefficients, the magnitude of 57 coefficients increased, 152 decreased and 124 remained the same. **Meanwhile, the magnitude of the top five predictors decreased.**
- When a new ridge model with double the alpha(860) was built, it reduced r2_score to 0.899 on the training set and increased r2_score to 0.871 on the testing set.
- Also of the total 333 coefficients, the magnitude of 101 coefficients increased, 230 decreased and 2 remained the same. **Meanwhile, the magnitude of the top five predictors decreased.**
- The top five important predictor variables of new lasso model are:
    - **['HouseAge', 'OverallQual_10', 'OverallQual_9', 'PoolQC_Gd', 'GrLivArea']**
- The top five important predictor variables of new ridge model are:
    - **['GarageCars_3', '1stFlrSF', 'OverallQual_10', 'GrLivArea', 'OverallQual_9']**

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?

Ans)

- I've obtained the following scores on train and test sets based on best models identified by lasso and ridge regression.

	Model	Mean_train_r2_score	Mean_train_mae_score	Mean_test_r2_score	Mean_test_mae_score
0	Best Lasso Model	0.936097	12942.276654	0.755141	18630.461155
1	Best Ridge Model	0.914599	14565.562860	0.865000	17046.066498

- The best lasso model found seems to be overfit as the r2 score on the training set(0.936) is much higher than the testing set(0.755).
- Also the standard deviation of best lasso model found to vary much between train and test sets during cross validation.

std_train_score	std_test_score
511.53322	1600.720115

- On other hand, r2 score of best ridge model on training set(0.914) is slightly higher than testing set(0.865)
- And Also the standard deviation of best ridge model does not differ much between train and test sets during cross validation.

std_train_score	std_test_score
584.364148	617.963468

- Hence, the best ridge model is chosen and applied.

Question 3)

After building the model, you realized 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?

Ans)

- The top five important predictor variables in lasso model are:
  - **['HouseAge', 'PoolQC_Gd', 'GrLivArea', 'PoolQC_NoPool', 'PoolArea']**
- Removed above five variables and found the best alpha for lasso model as 310.
- The top five important predictor variables now are:
  - **['OverallCond_3', 'Fence_NoFence', 'GarageArea', 'MiscFeature_NoMisc', 'isRemodelled']**

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?

Ans)

- A model is said to be robust and generalisable when it does not overfit.
- When a model is overfit, it exhibits high variance. Thus a small change in data affects the model prediction heavily. Such a model identifies all patterns from training data but fail to recognize patterns from testing data.
- So in order to reduce overfitting, model complexity can be reduced. However making a model too simple can result in high bias or increase in model error.
- Hence to achieve a non overfitting model a balance between model accuracy and complexity should be found.
- This can be achieved by using regularization techniques like Ridge Regression and Lasso Regression.