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?

Optimal value of alpha.

Lasso : 0.001

Ridge: 0.220

After doubling the alpha, error metrics value increased and the R2 score value came down.

With optimized values:

	R2 Score	RSS	MSE	RMSE
Linear Regression	0.833599	112.804997	0.293763	0.541999
Lasso Regression	0.870691	87.659634	0.228280	0.477787
Ridge Regression	0.872305	86.565698	0.225432	0.474796

Lasso value after doubling:

```
{'R2 Score': 0.8431213576012969,
 'RSS': 106.3494394551403,
 'MSE': 0.2769516652477612,
 'RMSE': 0.5262619739709123}
```

Ridge Value after doubling:

```
{'R2 Score': 0.8682591663800185,
 'RSS': 89.30829330630252,
 'MSE': 0.2325736804851628,
 'RMSE': 0.4822589351014274}
```

In all the cases , the most important predictor variable is total SF.

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?

K-Fold cross validation is done using lassocv and ridgecv implementations.The implementations search the best possible alpha value across the given range of alphas and provide the output.

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?

The five most important variables are as below:

```
betas.sort_values(by='Lasso Q3 Coefficient',ascending=False)
```

:

	Lasso Q3 Coefficient
1stFlrSF	1.591217
Total_Bath	1.378566
TotRmsAbvGrd	0.866221
OverallQual__9	0.608409
LotArea	0.542984

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?

- By keeping the model simple so that is simple enough to understand the implications but not so simple that it underfits.
- Make sure the model is not overfitting on the training data.
- Make sure the assumptions of the linear regression are met.
    - Error terms should have constant variance.
    - Residuals should be normally distributed.
    - There should be linear relationship between predictor and target variables.
- Accuracy of the model is best measured by Adj. Rsq and F-statistic as this measures the performance of the model with respect to the number of features and overall model performance.