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 Alpha Lasso = 0.0010064630116739318
- Optimal Alpha Ridge = 10
- If I choosing a double valor or more great, the both model is underfitting. Example if choose in lasso alpha = 0.2, the R2 = 0.55
- If alpha in lasso = 0.2 , the most important predictor is ‘2ndFlrSF’ , with a value of 0 .3313002564676522.

Lasso

```
# I use the polynomial grade = 2 to fit lasso
from sklearn.linear_model import Lasso
model = Lasso(alpha=0.2)
model.fit(x_poly2, y_train)
y_train_predict= model.predict(x_poly2)
r2_score(y_train, y_train_predict)
print(list(zip(model.coef_, X)))
```

```
[(0.0, '1stFlrSF'), (0.3313002564676522, '2ndFlrSF'), (0.03513011705490791, 'OverallQual_3'), (-0.0, 'OverallQual_4'), (-0.0, 'OverallQual_5'), (-0.0, 'OverallQual_6'), (-0.0, 'OverallQual_8'), (0.0, 'OverallQual_9'), (0.0, 'OverallQual_10'),
(0.0, 'OverallCond_5'), (0.0, 'OverallCond_6'), (-0.0, 'OverallCond_7'), (-0.0, 'OverallCond_8'), (-0.0, 'OverallCond_9'), (0.0, 'RoofMatl_WdShngl'), (0.0, 'KitchenAbvGr_2'), (-0.0, 'TotRmsAbvGrd_3'), (-0.0, 'TotRmsAbvGrd_12'), (0.0, 'GarageC
ars_3')]
```

- If alpha in ridge = 20 , the most important predictor is ‘2ndFlrSF’ , with a value of 0. 38390996901796337

```
# I using the polynomial grade = 2 to fit ridge
from sklearn.linear_model import Ridge
model = Ridge(alpha= 20)
model.fit(x_poly2, y_train)
y_train_predict= model.predict(x_poly2)
r2_score(y_train, y_train_predict)
print(list(zip(model.coef_, X)))
```

```
[(0.0, '1stFlrSF'), (0.3830996901796337, '2ndFlrSF'), (0.2097353619702791, 'OverallQual_3'), (-0.15160618776013252, 'OverallQual_4'), (-0.1959471859578885, 'OverallQual_5'), (-0.17088873518157697, 'OverallQual_6'), (-0.0960965831379628 3, 'Ove
rallQual_8'), (0.10431249116014664, 'OverallQual_9'), (0.11411033971987374, 'OverallQual_10'), (0.12833779902622727, 'OverallCond_5'), (0.16634426751208323, 'OverallCond_6'), (0.10673673005408509, 'OverallCond_7'), (0.123229870988270 08, 'Over
allCond_8'), (0.12463437858923815, 'OverallCond_9'), (0.08031341353261022, 'RoofMatl_WdShngl'), (0.03982078193647217, 'KitchenAbvGr_2'), (-0.13548359051441 01, 'TotRmsAbvGrd_3'), (0.03377127112778924, 'TotRmsAbvGrd_12'), (0.01834728363686865,
'GarageCars_3')]
```

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?

R: In the case of the assigned problem we will implement the Lssso model with its optimal alpha because it is the one that presents the best performance of its R2 and we ensure that it is not overfitting.

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?

R:

Five most important with the final model:
2ndFlrSF
OverallQual_3
OverallQual_9
OverallCond_5
OverallCond_6
RoofMatl_WdShngl

```
## Evaluate the model wit n=2 and alpha = 0.0010064630116739318
lasso_best = Lasso(alpha=model.alpha_)
lasso_best.fit(x_poly2, y_train)
print(list(zip(lasso_best.coef_, X)))
```

```
[(0.0, '1stFlrSF'), (0.4242279199637594, '2ndFlrSF'), (0.2168663145405 4084, 'OverallQual_3'), (-0.4679539885344 1195, 'OverallQual_4'), (-0.5242886152676851, 'OverallQual_5'), (-0.4157348594076617, 'OverallQual_6'), (-0.2037888439475 5184, 'Ov
erallQual_8'), (0.1992036507177 2307, 'OverallQual_9'), (0.0, 'OverallQual_10'), (0.1079258184593 4138, 'OverallCond_5'), (0.348447589362209, 'OverallCond_6'), (0.1144840890738 6817, 'OverallCond_7'), (0.1780762672421573, 'OverallCond_8'), (0.16
6990422356 5187, 'OverallCond_9'), (0.2339572908922 4457, 'RoofMatl_WdShngl'), (0.0, 'KitchenAbvGr_2'), (-0.1852737957815 1149, 'TotRmsAbvGrd_3'), (0.0229536251621 37493, 'TotRmsAbvGrd_12'), (0.0, 'GarageCars_3')]
```

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?

R: It is robust because when performing the prediction in Lasso with test we had similar R2 error results to training, we also ensured that it is a model that has no overfitting and can be generalized because the best features have been used for this model., and I ensured that all the variables which are important are added to the model and ensure the number of independent parameters are much lesser than the number of data points.