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?

For optimal value we use a set of values of in form of list and tried with every value using loop finally we choose value where shrinkage is optimal

Ridge:

$$J(\theta) = MSE(\theta) + \alpha \frac{1}{2} \sum_{i=1}^{n} \theta_i^2$$

Cost function   mean squared error   regularization term

Lasso:

$$J(\theta) = MSE(\theta) + \alpha \sum_{i=1}^{n} |\theta_i|$$

For Ridge it is = 100

For lasso it is = 0.01

Double in lasso decrease our r2 value and it is 85 and 86 in ridge

We lost columns in feature selection top 5 now are

```
OverallQual
KitchenQual
GarageCars
Fireplaces
CentralAir
```

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?

We used this feature to get optimal penalty in assignment

print(model_cv.best_params_)

print(model_cv.best_score_)

Grid search give us value with highest mean score provided by sklearn lib

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?

```
PavedDrive'
'GarageArea'
WoodDeckSF
TotalBsmtSF'
'ExterQua
```

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?

We try to make our model simple not complex find perfect mix of bias and variance

And thus we use regularization , Since it reduces error thus results in better accuracy