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?

1. **Optimal value of alpha for ridge and lasso regression**

```
print(f'Ridge:',ridge_alpha)
print(f'Lasso:',lasso_alpha)
```

```
Ridge: 7.0
Lasso: 0.001
```

2. **Double the value of alpha for both ridge and lasso**

Overall, there is no much change to the model accuracy. R2 score remains almost the same for both Ridge and Lasso for test data.

- Ridge(alpha=14.0)

```
alpha = ridge_alpha*2
ridge = Ridge(alpha=alpha)
ridge.fit(X_train, y_train)
```

```
Ridge(alpha=14.0)
```

```
Ridge_train_score: 0.9157597425675179
Ridge_test_score: 0.859745217706103
```

- Lasso(alpha=0.002)

```
alpha = lasso_alpha*2
lasso = Lasso(alpha=alpha)
lasso.fit(X_train, y_train)
```

```
Lasso(alpha=0.002)
```

```
Lasso_train_score: 0.90959028138821
Lasso_test_score: 0.858079414210642
```

3. **Most important predictor variables after the change is implemented**

- **For Ridge**

Features	Coefficient
GrLivArea	0.115617
MSZoning_RL	0.112167
MSZoning_RM	0.084486
OverallQual	0.069402
MSZoning_FV	0.052552

- **For Lasso**

Features	Coefficient
GrLivArea	0.118059
OverallQual	0.075426
MSZoning_RL	0.048757
OverallCond	0.041605
GarageArea	0.035747

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?

Both Ridge and Lasso Regression models have similar R2_score values of approximately 86% for test data.

However, **Lasso Regression** is considered over Ridge Regression as Final model since Lasso regression helps by performing feature selection.

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?

After eliminating the five most important predictor variables realized in the lasso model from the incoming data, the five most important predictor variables now are as below.

Features	Coefficient
2ndFlrSF	0.107029
BsmtExposure_None	0.086516
TotalBsmtSF	0.074146
GarageArea	0.048560
BsmtQual	0.046335

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?

- A model can be made robust and generalizable by avoiding overfitting without missing to identify the underlying patterns in the data. When a model is overfit, it will perform well with training data but not with test data. Hence, bias will be low and variance will be high.
- So, we should make sure the model complexity is managed well such that its neither high nor low. If Model complexity is high, then the model will overfit. If low, then the model will underfit resulting in not identifying the patterns in the data.

- Here is where Regularization helps to manage the model complexity by shrinking the model coefficients towards zero and thus it will avoid overfitting.
- Ridge and Lasso are two methods of Regularization

Implications

- By doing so, the accuracy of the model would remain consistent for both training and test data and the model will perform well for both training and unseen (test) data
- In this, we compromise a little on the bias for higher reduction in variance by maintaining the tradeoff between bias and variance