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 for Ridge = 5

Optimal Value of Alpha for Lasso = 0.001

When alpha is doubled:

Effect on Ridge – Remains the same

Effect on Lasso – R2 Scores decrease a bit while RSS and MSE scores increase a bit

ORIGINAL ALPHA VALUE METRICS:

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	9.594928e-01	0.957152	0.952360
1	R2 Score (Test)	-1.550167e+22	0.866066	0.865326
2	RSS (Train)	6.501093e+00	6.876814	7.645798
3	RSS (Test)	1.117183e+24	9.652399	9.705798
4	MSE (Train)	7.979585e-02	0.082069	0.086536
5	MSE (Test)	5.050393e+10	0.148450	0.148860

WHEN ALPHA IS DOUBLED, METRICS VALUE:

:

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	9.594928e-01	0.957152	0.941276
1	R2 Score (Test)	-1.550167e+22	0.866066	0.864590
2	RSS (Train)	6.501093e+00	6.876814	9.424802
3	RSS (Test)	1.117183e+24	9.652399	9.758828
4	MSE (Train)	7.979585e-02	0.082069	0.096078
5	MSE (Test)	5.050393e+10	0.148450	0.149266

Important Predictor Variables –

In [415]:
```
## View the top 10 coefficients of Ridge regression in descending order
betas['Ridge'].sort_values(ascending=False)[:10]
```

Out[415]:
```
RoofMatl_CompShg    0.259595
RoofMatl_Tar&Grv    0.173271
RoofMatl_WdShngl    0.129206
MSZoning_RL         0.111530
RoofMatl_WdShake    0.103893
MSZoning_RM         0.082750
GrLivArea           0.072785
RoofMatl_Membran    0.058344
MSZoning_FV         0.057248
RoofMatl_Metal      0.056569
Name: Ridge, dtype: float64
```

In [416]:
```
## View the top 10 coefficients of Lasso in descending order
betas['Lasso'].sort_values(ascending=False)[:10]
```

Out[416]:
```
RoofMatl_CompShg    0.249690
RoofMatl_Tar&Grv    0.168176
GrLivArea           0.140652
RoofMatl_WdShngl    0.124946
RoofMatl_WdShake    0.103698
RoofMatl_Membran    0.054004
RoofMatl_Metal      0.053041
MSZoning_RL         0.051258
RoofMatl_Roll       0.050063
TotalBsmtSF         0.040081
Name: Lasso, dtype: float64
```

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?

The determination of which regression to choose mainly depends on our requirement.

If we have too many variables and one of our **primary goal is feature selection**, then we will use **Lasso**.

If we don't want to get too large coefficients and **reduction of coefficient magnitude** is one of our prime goals, then we will use **Ridge Regression**.

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?

```
In [436]: ## View the top 5 coefficients of Lasso in descending order
          betas['Lasso'].sort_values(ascending=False)[:5]

Out[436]: RoofMatl_CompShg    0.249690
          RoofMatl_Tar&Grv    0.168176
          GrLivArea           0.140652
          RoofMatl_WdShngl    0.124946
          RoofMatl_WdShake    0.103698
          Name: Lasso, dtype: float64
```

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 is robust when any variation in the data does not affect its performance much.

A generalizable model is able to adapt properly to new, previously unseen data, drawn from the same distribution as the one used to create the model.

To make sure a model is robust and generalizable, we have to take care it doesn't overfit. This is because an overfitting model has very high variance and a smallest change in data affects the model prediction heavily. Such a model will identify all the patterns of a training data, but fail to pick up the patterns in unseen test data.

In other words, the model should not be too complex in order to be robust and generalizable.

If we look at it from the prespective of Accuracy, a too complex model will have a very high accuracy. So, to make our model more robust and generalizable, we will have to decrease variance which will lead to some bias. Addition of bias means that accuracy will decrease.