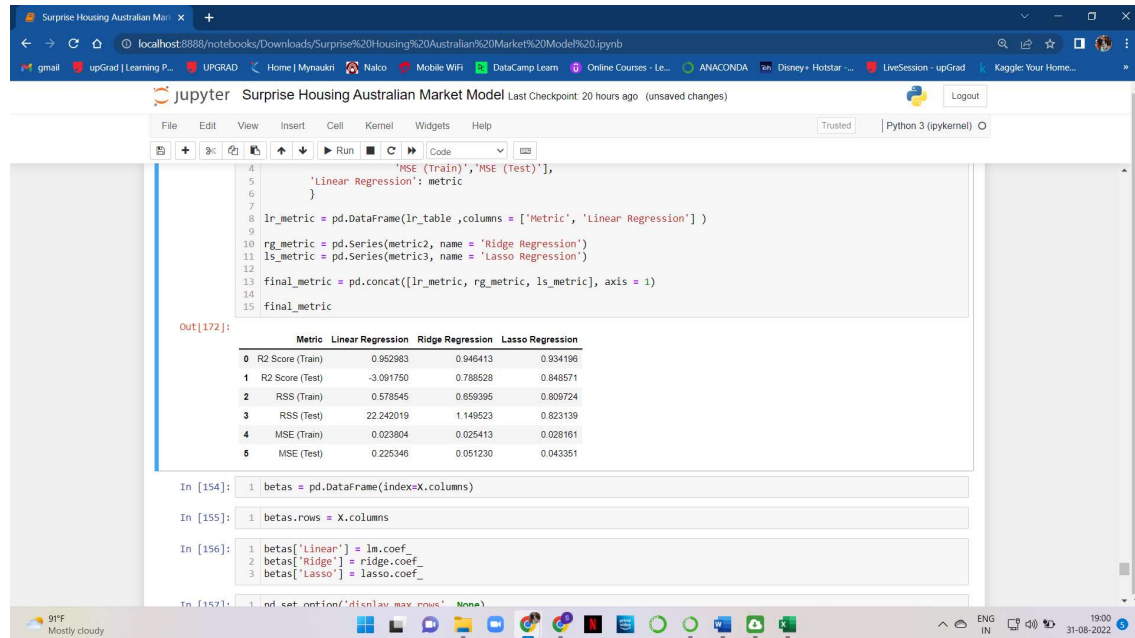


Question 1

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?

ANSWER:

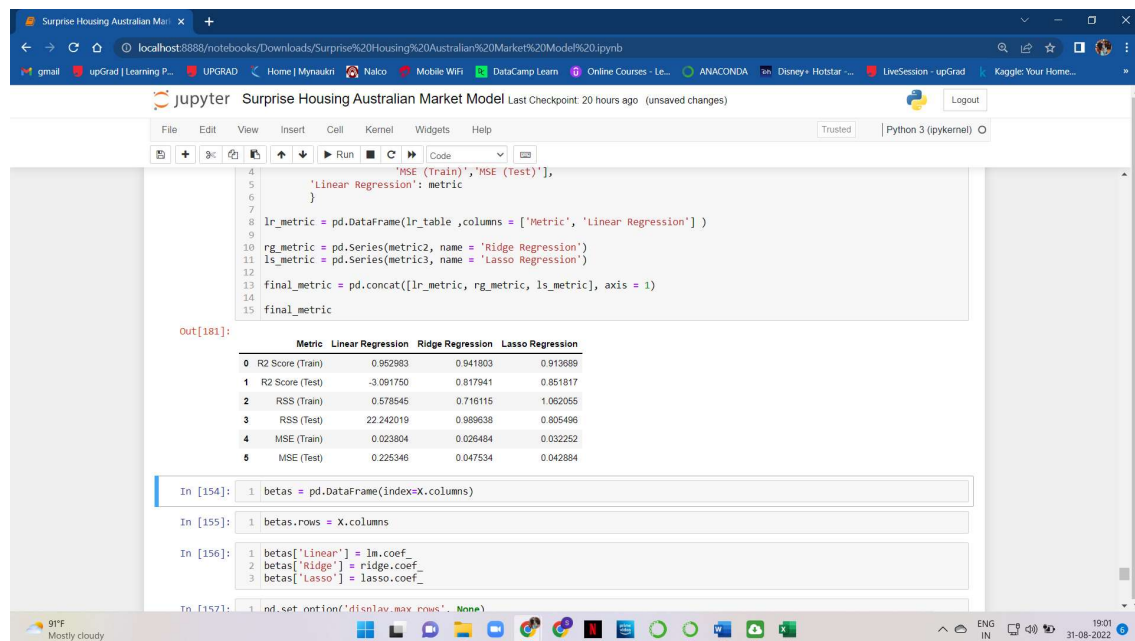
The optimum value of Alpha for Ridge is 0.2 and for Lasso is 0.0001.



The screenshot shows a Jupyter Notebook titled 'Surprise Housing Australian Market Model'. The code defines metrics for Linear, Ridge, and Lasso regression. The output shows the following metrics:

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	0.962983	0.946413	0.934196
1	R2 Score (Test)	-3.091750	0.788528	0.848571
2	RSS (Train)	0.578545	0.659395	0.809724
3	RSS (Test)	22.242019	1.149523	0.823139
4	MSE (Train)	0.023804	0.025413	0.028161
5	MSE (Test)	0.225346	0.051230	0.043351

The optimum value of Alpha for Ridge is 0.4 and for Lasso is 0.0002. R2 square goes up for ridge.



The screenshot shows the same Jupyter Notebook with updated alpha values. The output shows the following metrics:

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	0.962983	0.941803	0.913689
1	R2 Score (Test)	-3.091750	0.817941	0.851817
2	RSS (Train)	0.578545	0.716115	1.062055
3	RSS (Test)	22.242019	0.989638	0.805496
4	MSE (Train)	0.023804	0.026484	0.032252
5	MSE (Test)	0.225346	0.047534	0.042884

GrLivArea: Above grade (ground) living area square feet- is the most important variable.

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?

ANSWER:

I will go for Lasso as the R2 Square, RSS and MSE are better for lasso as compared to Ridge.

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?

ANSWER:

The five most important predictors are :

1. 1stFlrSF: First Floor square feet
2. 2ndFlrSF: Second floor square feet
3. BsmtFinSF1: Type 1 finished square feet
4. LotArea: Lot size in square feet
5. GarageArea: Size of garage in square feet

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?

ANSWER:

From Lasso we can see that there are only 20 variables which are important and we will only use them to do the prediction. This way the model will be more robust and generalisable.

This will increase the bias as we have brought down the variables from 290 to 20 but this will also make sure the model is generalisable and is not overfitting.