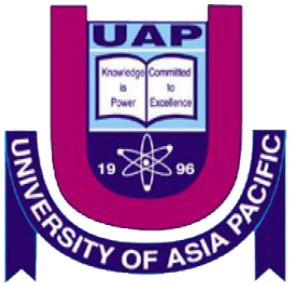


DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Course Title : Artificial Intelligence

Course Code : CSE 404

Experiment Name : Implementation of multivariable Linear regression Using a public data set.

Submitted To

Dr. Nasima Begum
Associate Professor
Department of CSE

Submitted By

Asikur Rahman Sumon
ID: 19201055
Section: A2

Date of Submission : 11-04-23

Implementation of multivariable Linear regression Using a public dataset.

Problem description: Implementation of a Linear regression model with a dataset and the dataset must be multivariant. At the basis of other parameter we have to predict another parameter.

Objective: There are several approach in Machine Learning to predict a data at the basis of other data. In this project we are going to implement "Linear Regression"- model to predict data.

For this approach, I'm going to use a Car prize dataset which is about (205, 24) in size .

	car_ID	symboling	CarName	fueltype	aspiration	doornumber	carbody	drivewheel	enginelocation	wheelbase	...	enginesize	fuelsystem	borera
0	1	3	alfa-romero giulia	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	
1	2	3	alfa-romero stelvio	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	
2	3	1	alfa-romero Quadrifoglio	gas	std	two	hatchback	rwd	front	94.5	...	152	mpfi	
3	4	2	audi 100 ls	gas	std	four	sedan	fwd	front	99.8	...	109	mpfi	
4	5	2	audi 100ls	gas	std	four	sedan	4wd	front	99.4	...	136	mpfi	

5 rows × 26 columns

Here is my care type data:

```
car_ID              int64
symboling           int64
CarName            object
fueltype           object
aspiration         object
doornumber         object
carbody            object
drivewheel         object
enginelocation     object
wheelbase         float64
carlength         float64
carwidth          float64
carheight         float64
curbweight          int64
enginetype         object
cylindernumber     object
enginesize          int64
fuelsystem         object
boreratio         float64
stroke            float64
compressionratio  float64
horsepower          int64
peakrpm             int64
citympg             int64
highwaympg          int64
price             float64
```

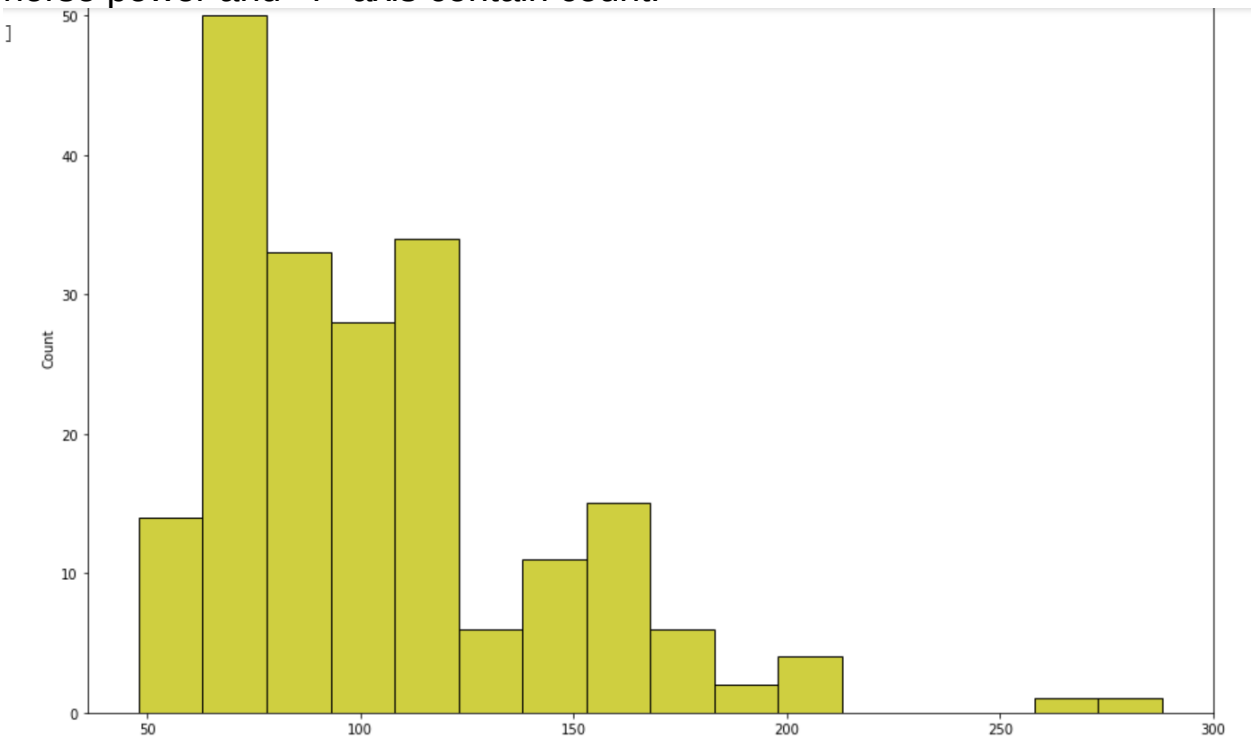
Car describe:
p

	count	mean	std	min	25%	50%	75%	max
car_ID	205.0	103.000000	59.322565	1.00	52.00	103.00	154.00	205.00
symboling	205.0	0.834146	1.245307	-2.00	0.00	1.00	2.00	3.00
wheelbase	205.0	98.756585	6.021776	86.60	94.50	97.00	102.40	120.90
carlength	205.0	174.049268	12.337289	141.10	166.30	173.20	183.10	208.10
carwidth	205.0	65.907805	2.145204	60.30	64.10	65.50	66.90	72.30
carheight	205.0	53.724878	2.443522	47.80	52.00	54.10	55.50	59.80
curbweight	205.0	2555.565854	520.680204	1488.00	2145.00	2414.00	2935.00	4066.00
enginesize	205.0	126.907317	41.642693	61.00	97.00	120.00	141.00	326.00
boreratio	205.0	3.329756	0.270844	2.54	3.15	3.31	3.58	3.94
stroke	205.0	3.255415	0.313597	2.07	3.11	3.29	3.41	4.17
compressionratio	205.0	10.142537	3.972040	7.00	8.60	9.00	9.40	23.00
horsepower	205.0	104.117073	39.544167	48.00	70.00	95.00	116.00	288.00
peakrpm	205.0	5125.121951	476.985643	4150.00	4800.00	5200.00	5500.00	6600.00

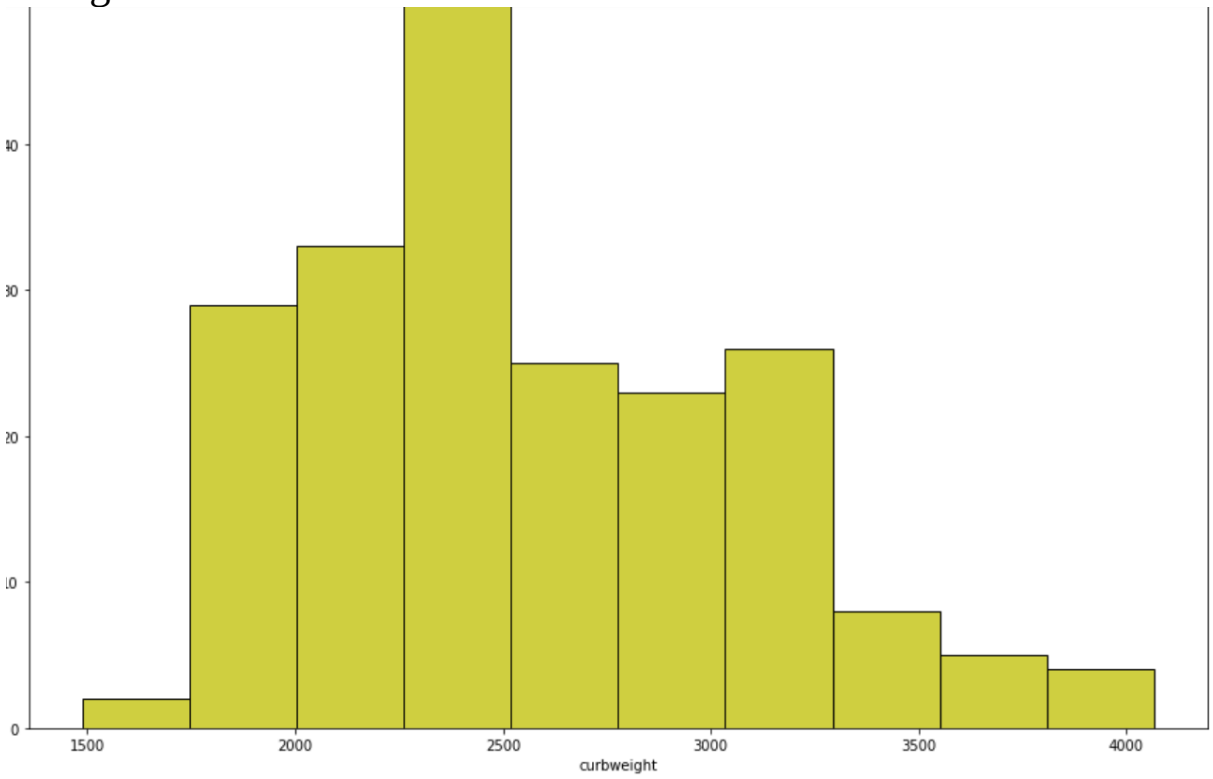
✓ 0s completed at 12:34 AM

Plot Some Data: Car horse power Distribution plot—Here “X” axis contain

horse power and “Y” axis contain count.

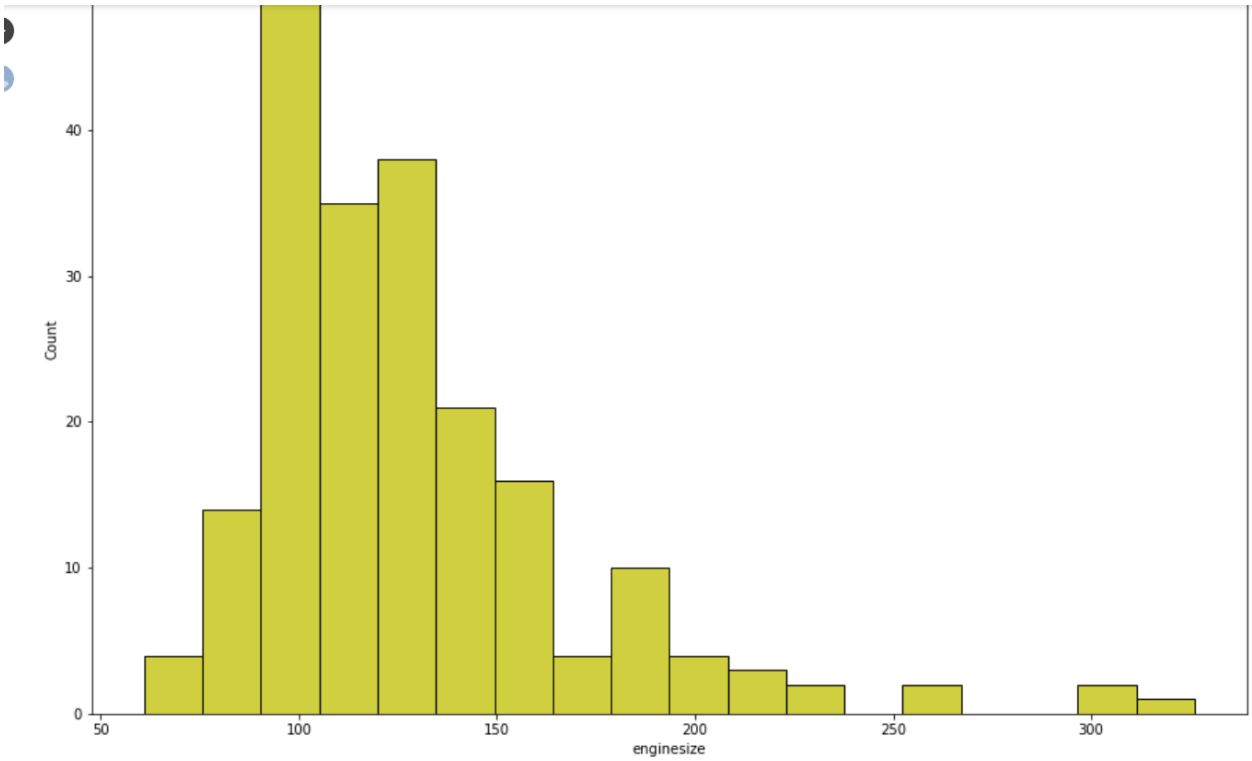


Now,I am ploting car curb weight plot: Here “X” axis contain car curb weight and “Y” axis contain count.

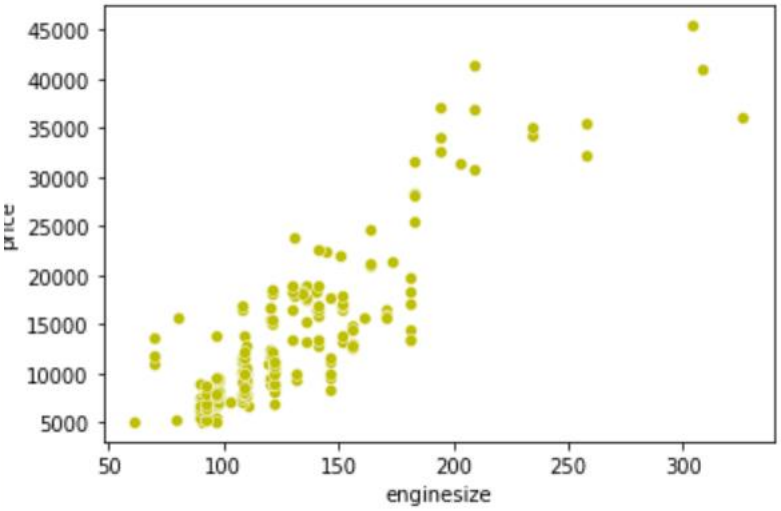


Now,I am ploting enginsizeplot: Here “X” axis contain enginsize and “Y”

axis contain count.

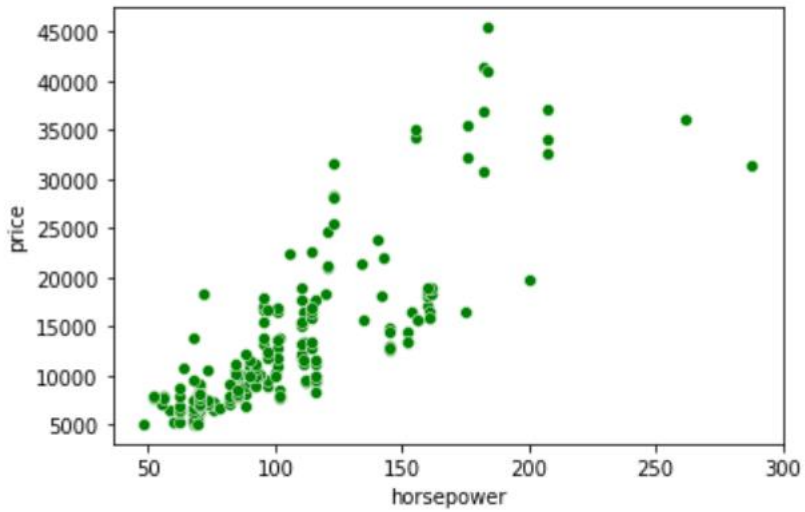


Now, I am showing scatterplot of enginesize.

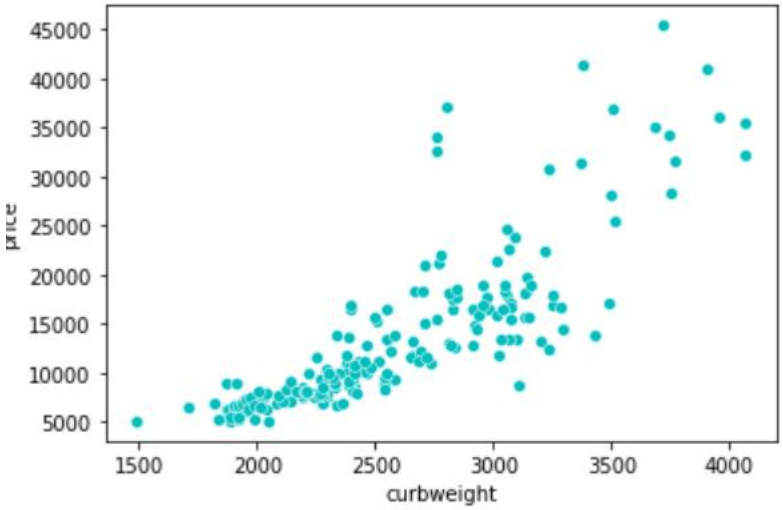


Now, I am showing scatterplot of horsepower.

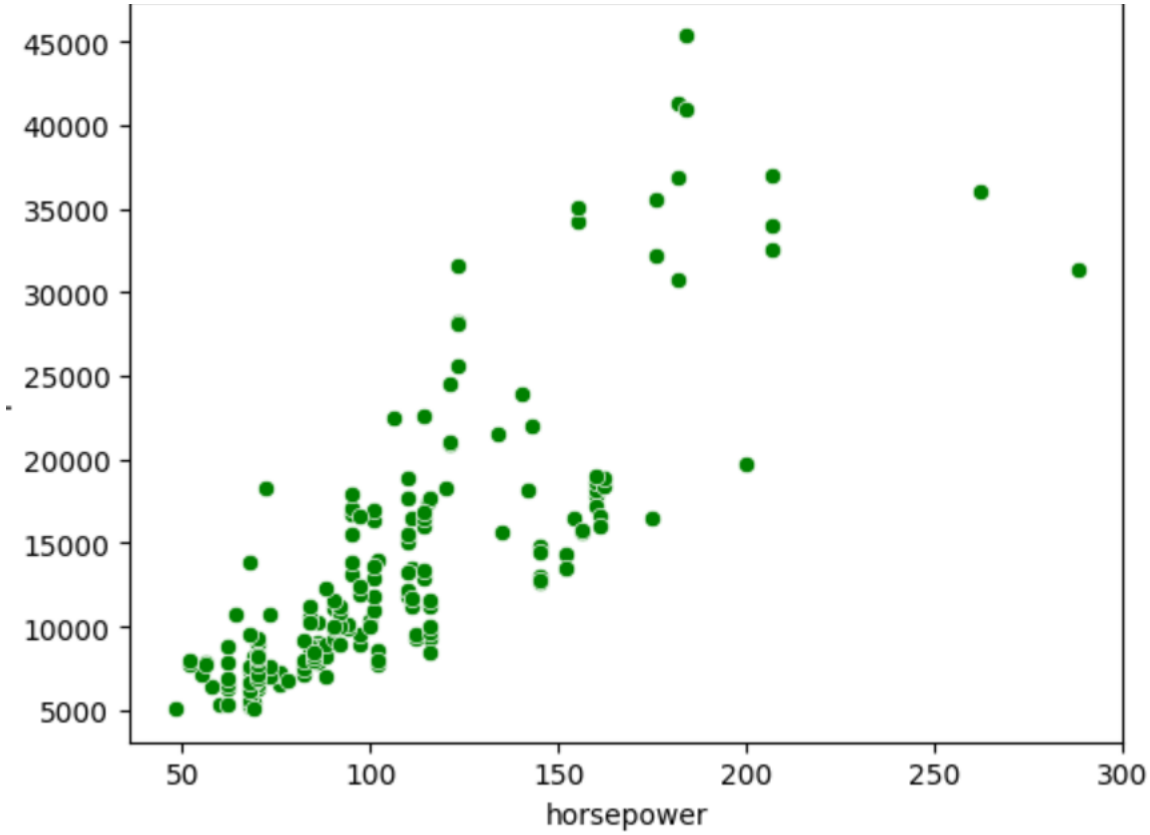
```
1   sns.scatterplot(x="horsepower", y="price", data=df,
```



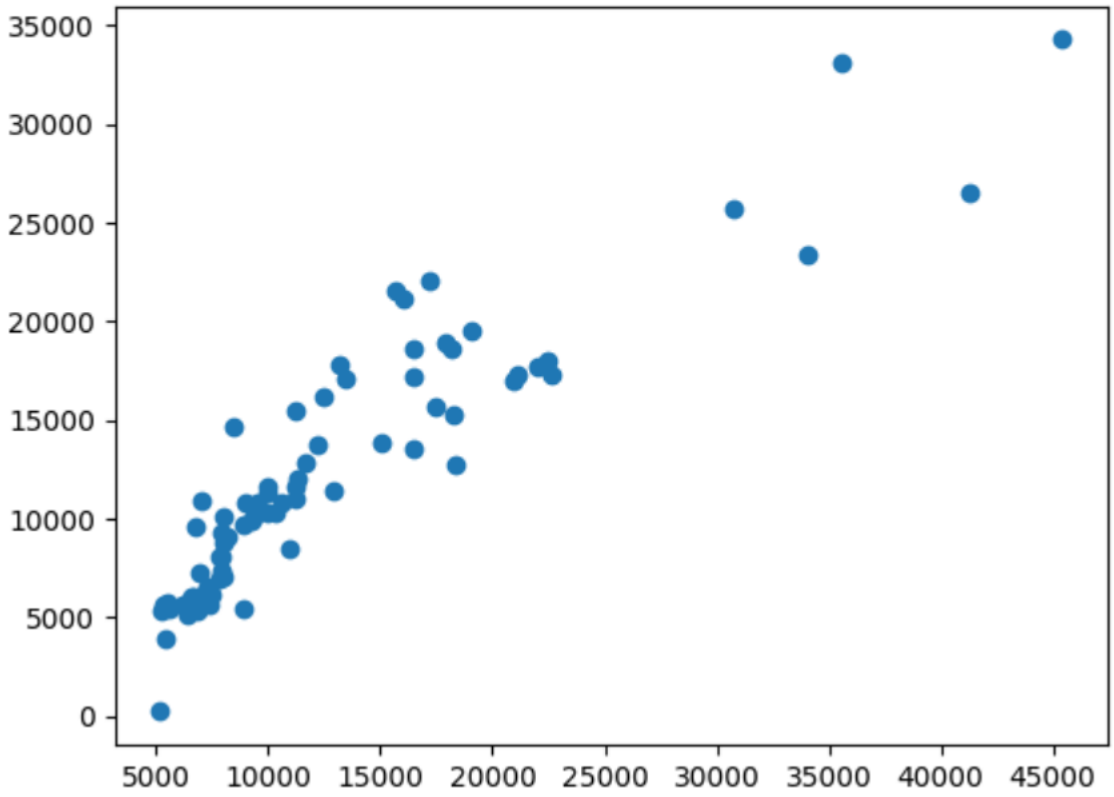
Now, I am showing scatterplot of Curbweight.



Now, I am ploting scatterplot of horse power and prize. Wher prize contain “Y” axis. And horse power contain “X” axis.

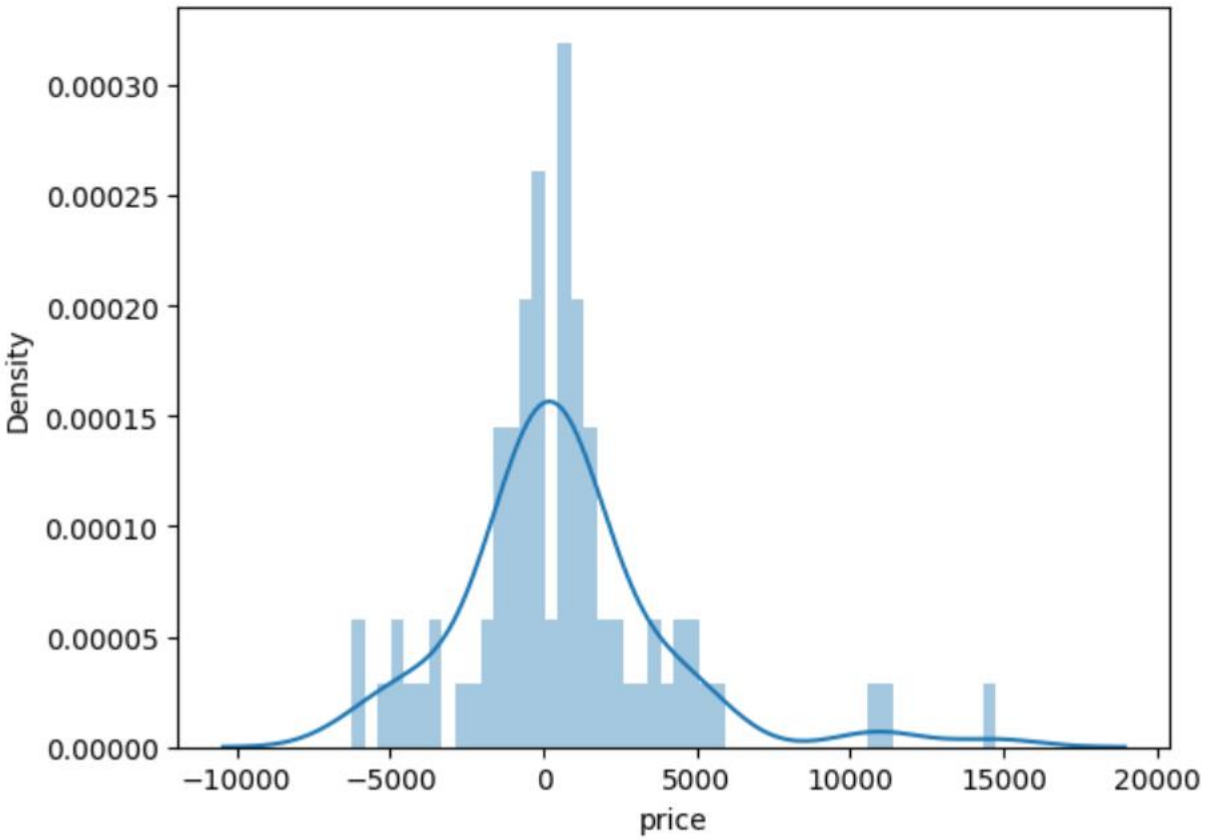



```
<matplotlib.collections.PathCollection at 0x7f891a4be700>
```



My testing predictions:

```
sns.distplot((y_test-predictions),bins=50);
```



Conclusion:

In conclusion, we have successfully implemented a Multivariable Linear Regression model to evaluate the performance of Higher Education students. We normalized the data, trained the model using the hypothesis function, cost function, and gradient descent algorithm, evaluated the performance of the model using the cost function, and finally, made a prediction using the model. Further improvements can be made by tuning the hyper parameters and adding more features to the model