Report for Innomatics Hackathon Assignment 2022

In **data_2_var** there are two columns, but in the initial stage there is no columns name, that's why I named as "input" and "target".

I try to analyze this data set. Analysis is -----

- There are 2 columns and 1110 rows present.
- There is no null value present.
- N dim value is 2.

Columns-wise Statistical analysis →

For **input** column:

Mean is: 6.899714180454869

Mode is: -125.3666687

Median is: 15.739103114999999

Variance is: 6292.307993639906

Standard Deviation is: 79.32406944704681

For **target** column:

Mean is: -26.09331772561168

Mode is: -95.0

Median is: -46.036318765000004

Variance is: 4865.896797611318

Standard Deviation is: 69.7559803716593

Correlation is:

```
df.corr()
```

	input	target
input	1.000000	0.454607
target	0.454607	1.000000

Skewness is:

```
df.skew()
```

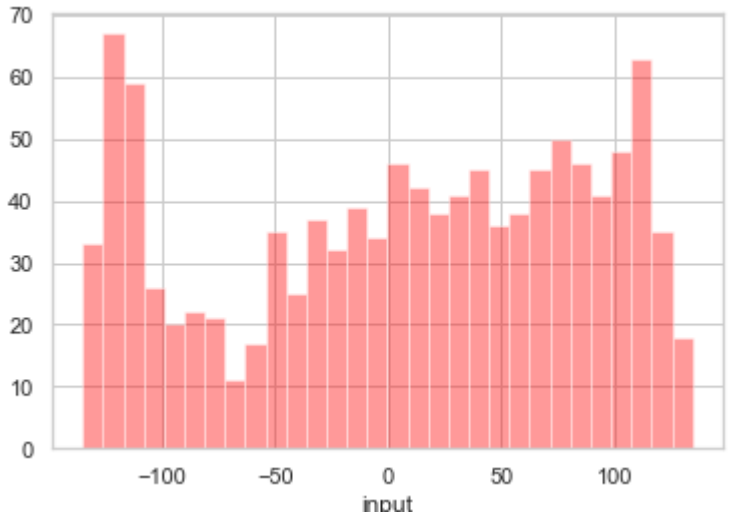
```
input    -0.269522
target    0.909210
dtype: float64
```

Covariance is:

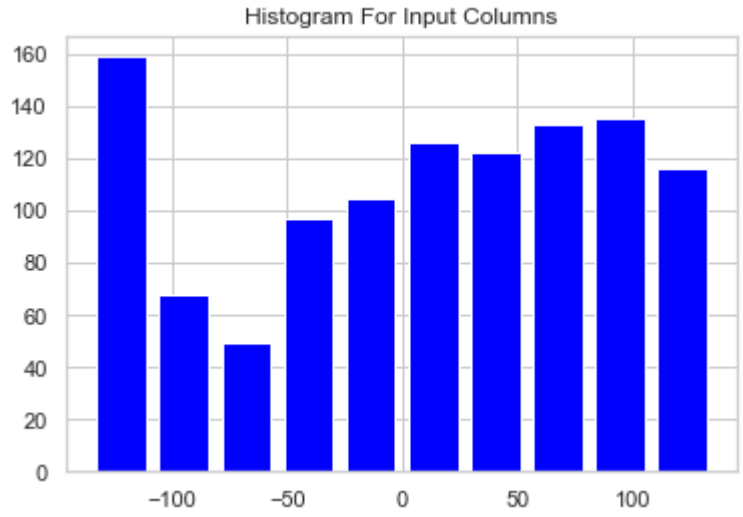
	input	target
input	6292.307994	2515.491940
target	2515.491940	4865.896798

Visualization

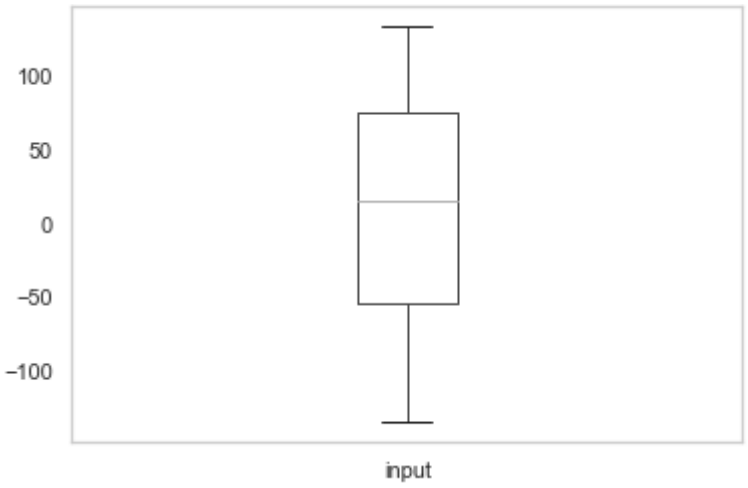
Distribution plot for “input” column:



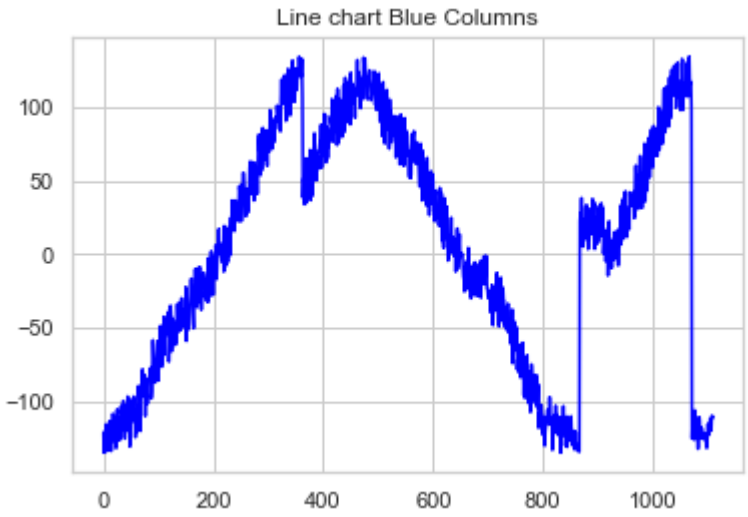
Histogram for “input” column:



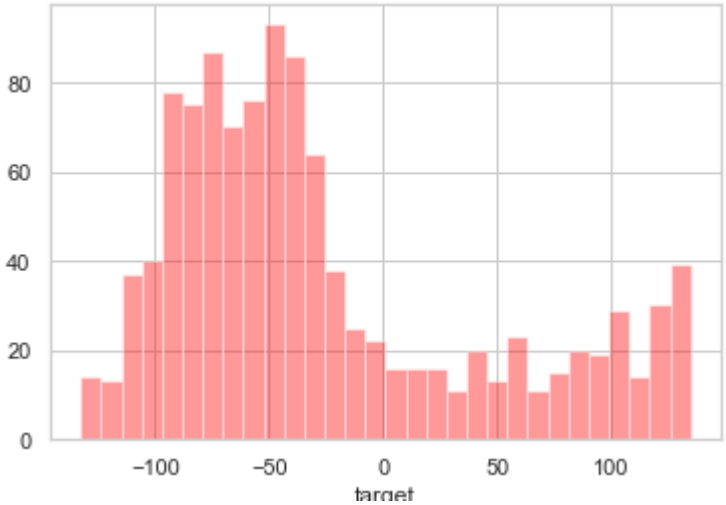
Box plot for “input” column:



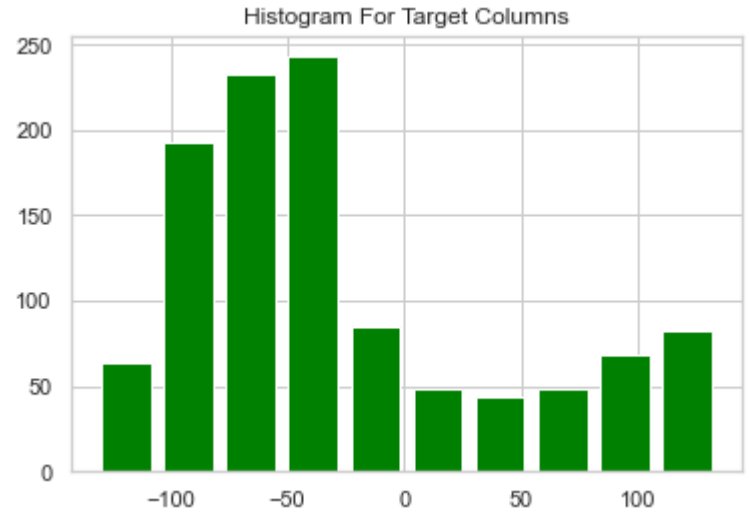
Line chart for “input” column:



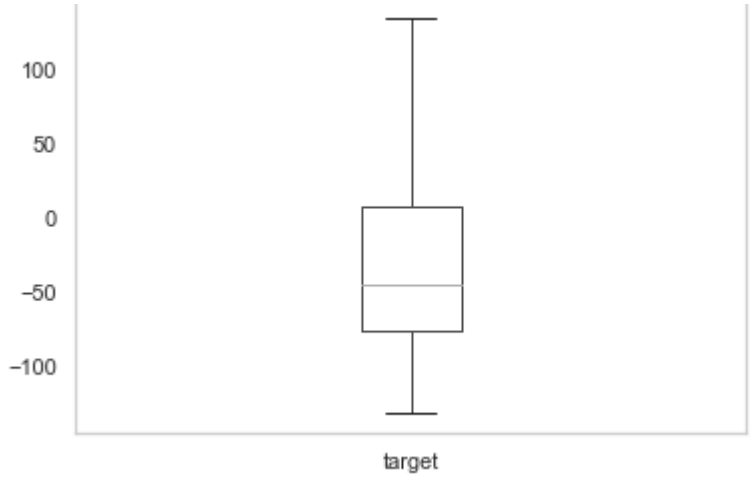
Distribution plot for “target” column:



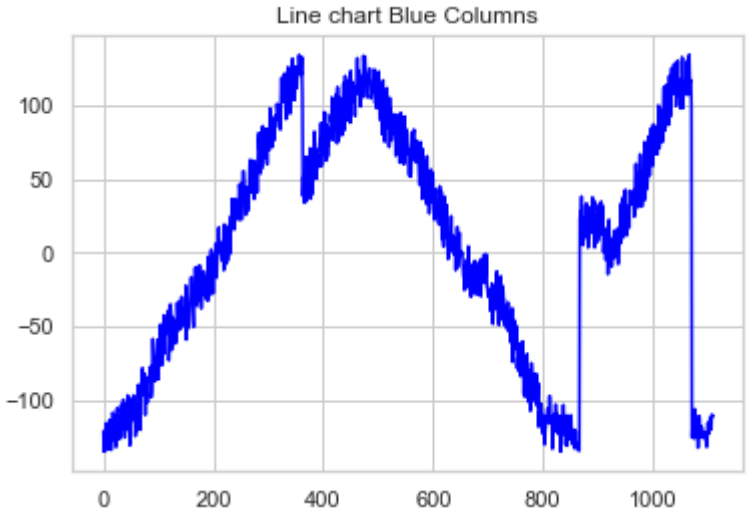
Histogram for “target” column:



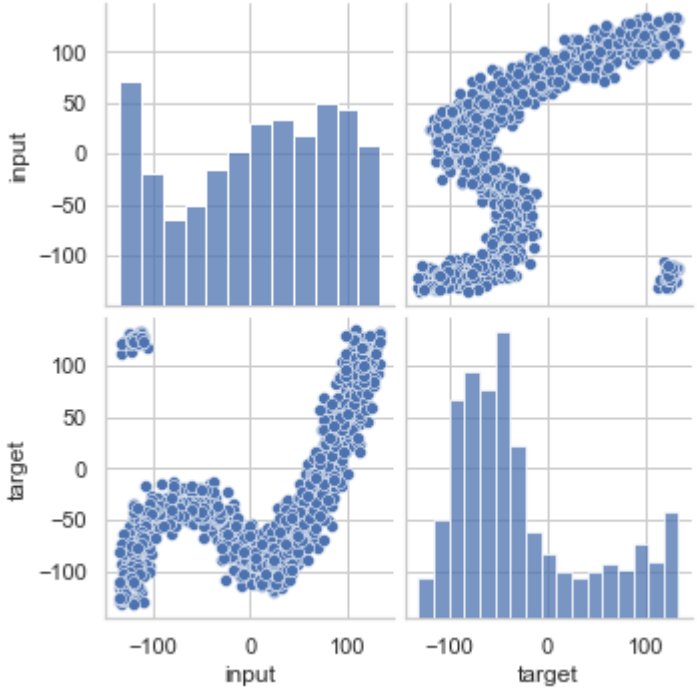
Box plot for “target” column:



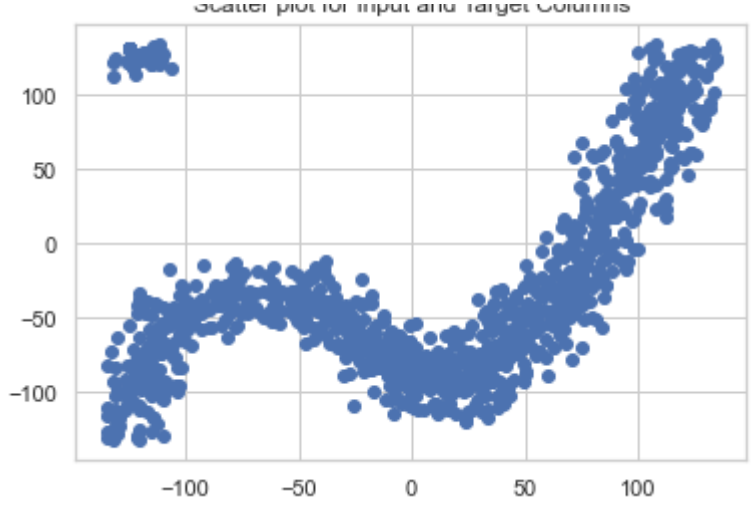
Line chart for “target” column:



Pair plot:



Scatter plot:



- I split the dataset into 70:30 and fit the linear regression model and get a 0.22452779262141676 score value.