

Assignment - 02

Feature Transformation

Dataset :

PRICE
110
105
115
120
110
130
150
100
105

Normalization :-

$$\text{Formula, } X_{\text{new}} = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}}$$

X_i = absolute value / Actual value

X_{\min} = minimum value

X_{\max} = maximum value

for 110,

$$X_{\text{new}} = \frac{110 - 100}{150 - 100} \Rightarrow 0.2$$

for 105,

$$X_{\text{new}} = \frac{105 - 100}{150 - 100} \Rightarrow 0.1$$

for 115,

$$X_{\text{new}} = \frac{115 - 100}{150 - 100} \Rightarrow 0.3$$

for 120,

$$X_{\text{new}} = \frac{120 - 100}{150 - 100} \Rightarrow 0.4$$

for 110, X_{new}

$$= \frac{110 - 100}{150 - 100} \Rightarrow 0.2$$

for 130, X_{new}

$$= \frac{130 - 100}{150 - 100} \Rightarrow 0.6$$

$$\text{for } 150, X_{\text{new}} = \frac{150-100}{150-100} \Rightarrow 1$$

$$\text{for } 100, X_{\text{new}} = \frac{100-100}{150-100} \Rightarrow 0$$

$$\text{for } 105, X_{\text{new}} = \frac{105-100}{150-100} \Rightarrow 0.1$$

$X_{\text{new}}\text{-Value} =$

price
0.2
0.1
0.3
0.4
0.2
0.6
1
0
0.1

Standardization

$$X_{\text{new}} = \frac{X_i - X_{\text{mean}}}{SD}$$

[SD = Standard Deviation]

$$SD = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

N = the size of the population
 μ = population mean
 x_i = Each value from the population

population mean

$$\mu = \frac{110+105+115+120+110+130+150+100+105}{9} = 116.1$$

Now, $x_i - \mu$:

$$= 150 - 116.1 \Rightarrow -6.1$$

$$= 105 - 116.1 \Rightarrow -11.1$$

$$= 115 - 116.1 \Rightarrow -1.1$$

$$= 120 - 116.1 \Rightarrow 3.9$$

$$= 110 - 116.1 \Rightarrow -6.1$$

$$= 130 - 116.1 \Rightarrow 13.9$$

$$= 150 - 116.1 \Rightarrow 33.9$$

$$= 100 - 116.1 \Rightarrow -16.1$$

$$= 105 - 116.1 \Rightarrow -11.1$$

$$SD = \sqrt{(-6.1)^2 + (-11.1)^2 + (-1.1)^2 + (3.9)^2 + (-6.1)^2 + (13.9)^2 + (33.9)^2 + (-16.1)^2 + (-11.1)^2}$$

$$= 14.68$$

for $x_1 = 110$,

$$x_{\text{new}} = \frac{110 - 116.1}{14.68} = -0.41$$

for $x_1 = 105$,

$$x_{\text{new}} = \frac{105 - 116.1}{14.68} = -0.75$$

for $x_1 = 115$,

$$x_{\text{new}} = \frac{115 - 116.1}{14.68} = -0.07$$

for $x_1 = 120$,

$$x_{\text{new}} = \frac{120 - 116.1}{14.68} = 0.26$$

for $x_1 = 110$,

$$x_{\text{new}} = \frac{110 - 116.1}{14.68} = -0.41$$

for $x_1 = 130$,

$$x_{\text{new}} = \frac{130 - 116.1}{14.68} = 0.94$$

for $x_1 = 150$,

$$x_{\text{new}} = \frac{150 - 116.1}{14.68} = 2.30$$

for $x_1 = 100$,

$$x_{\text{new}} = \frac{100 - 116.1}{14.68} = -1.09$$

for $x_1 = 105$,

$$x_{\text{new}} = \frac{105 - 116.1}{14.68} = -0.75$$

X_{new}

price
-0.41
-0.75
-0.07
0.26
-0.41
0.94
2.30
-1.09
-0.75

□ Log Transformation :-

X_{new}

$$\text{for } 110, = \log 110 \Rightarrow 2.04$$

$$\text{for } 105, = \log 105 \Rightarrow 2.02$$

$$\text{for } 115, = \log 115 \Rightarrow 2.06$$

$$\text{for } 120, = \log 120 \Rightarrow 2.07 \rightarrow$$

$$\text{for } 110, = \log 110 \Rightarrow 2.04$$

$$\text{for } 130, = \log 130 \Rightarrow 2.11$$

$$\text{for } 150, = \log 150 \Rightarrow 2.17$$

$$\text{for } 100, = \log 100 \Rightarrow 2$$

$$\text{for } 105, = \log 105 \Rightarrow 2.02$$

price
2.04
2.02
2.06
2.07
2.04
2.11
2.17
2
2.02

□ Max Absolute Scales :-

$$X_{new} = \frac{x_i}{\max(\text{price})}$$

x_i = ~~each~~ original value

$\max(\text{price})$ = the maximum value of
the price column.

The maximum value of the price column = 150,

$$X_{new} = \frac{110}{150} \Rightarrow 0.73$$

$$\frac{105}{150} \Rightarrow 0.7$$

$$\frac{115}{150} \Rightarrow 0.76$$

$$\frac{120}{150} \Rightarrow 0.8$$

$$\frac{110}{150} \Rightarrow 0.7$$

$$\frac{130}{150} \Rightarrow 0.86$$

$$\frac{150}{150} \Rightarrow 1$$

$$\frac{100}{150} \Rightarrow 0.66$$

$$\frac{105}{150} \Rightarrow 0.7$$

price
0.73
0.7
0.76
0.8
0.73
0.86
1
0.66
0.7

Robust Statistics :-

Price
110
105
115
120
110
130
150
100
105

\Rightarrow Outliers \Rightarrow

Price
100
105
105
110
110
115
120
130
150

formulae

$$X_{\text{new}} = \frac{X_i - X_{\text{median}}}{X_{75} - X_{25}}$$

$$\text{for } 110, \quad X_{\text{new}} = \frac{110 - 110}{120 - 105} \Rightarrow 0$$

$$\text{for } 105, \quad X_{\text{new}} = \frac{105 - 110}{120 - 105} \Rightarrow -0.33$$

$$\text{for } 115, \quad X_{\text{new}} = \frac{115 - 110}{120 - 105} \Rightarrow 0.33$$

$$\text{for } 120, \quad X_{\text{new}} = \frac{120 - 110}{120 - 105} \Rightarrow 0.66$$

$$\text{for } 110, \quad X_{\text{new}} = \frac{110 - 110}{120 - 105} \Rightarrow 0$$

$$\text{for } 130, \quad X_{\text{new}} = \frac{130 - 110}{120 - 105} \Rightarrow 1.33$$

$$\text{for } 150, \quad X_{\text{new}} = \frac{150 - 110}{120 - 105} \Rightarrow 2.66$$

$$\text{for } 100, \quad X_{\text{new}} = \frac{100 - 110}{120 - 105} \Rightarrow -0.66$$

$$\text{for } 105, \quad X_{\text{new}} = \frac{105 - 110}{120 - 105} \Rightarrow -0.33$$

Price
0
-0.33
0.33
0.66
0
1.33
2.66
-0.66
-0.33