Train In Data

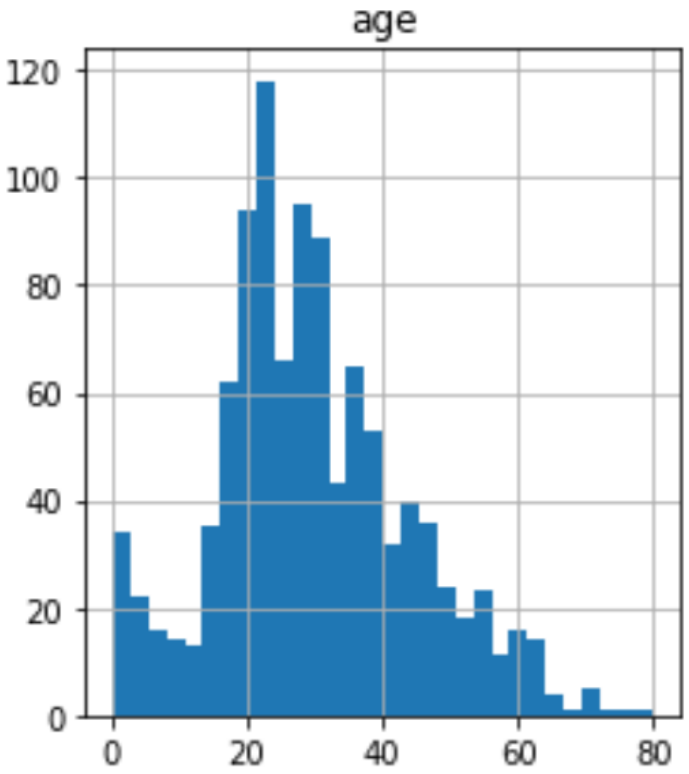
Equal-frequency discretisation

Equal-frequency discretisation: definition

Equal frequency discretisation divides the scope of possible values of the variable into N bins, where each bin carries the same amount of observations.

Interval boundaries correspond to the quantiles.

Equal-frequency discretisation: definition



Intervals = 10

Each interval contains 10% of total observations

Intervals:
0-16; 16-20; 20-22; 22-25; …
50-74

Equal-frequency discretisation: definition

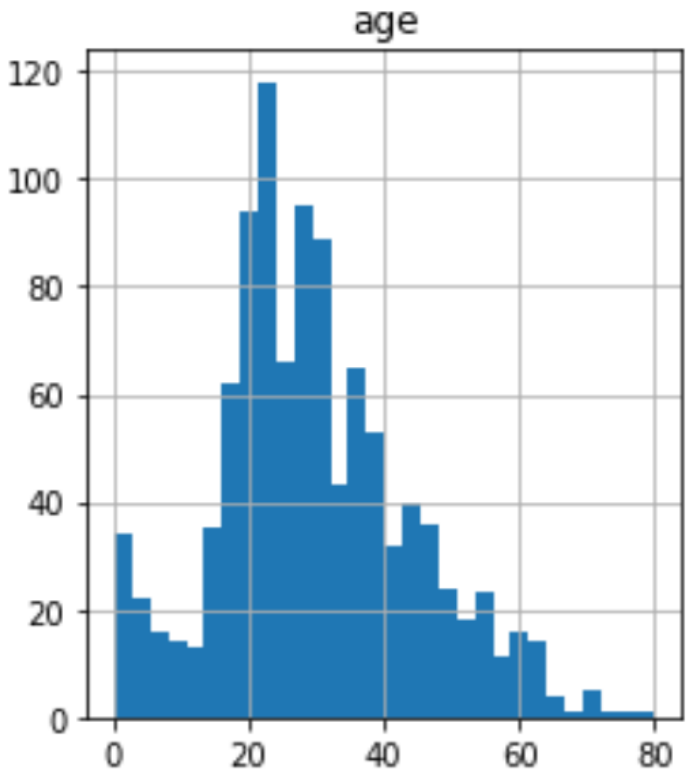
age	Age_disc
38.0	(36.0, 42.0]
21.0	(20.0, 22.25]
42.0	(36.0, 42.0]
34.0	(31.0, 36.0]
25.0	(22.25, 25.0]
4.0	(0.167, 16.0]
48.0	(42.0, 50.0]
52.0	(50.0, 74.0]
57.0	(50.0, 74.0]
32.0	(31.0, 36.0]

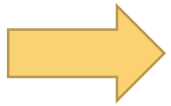
Intervals = 10

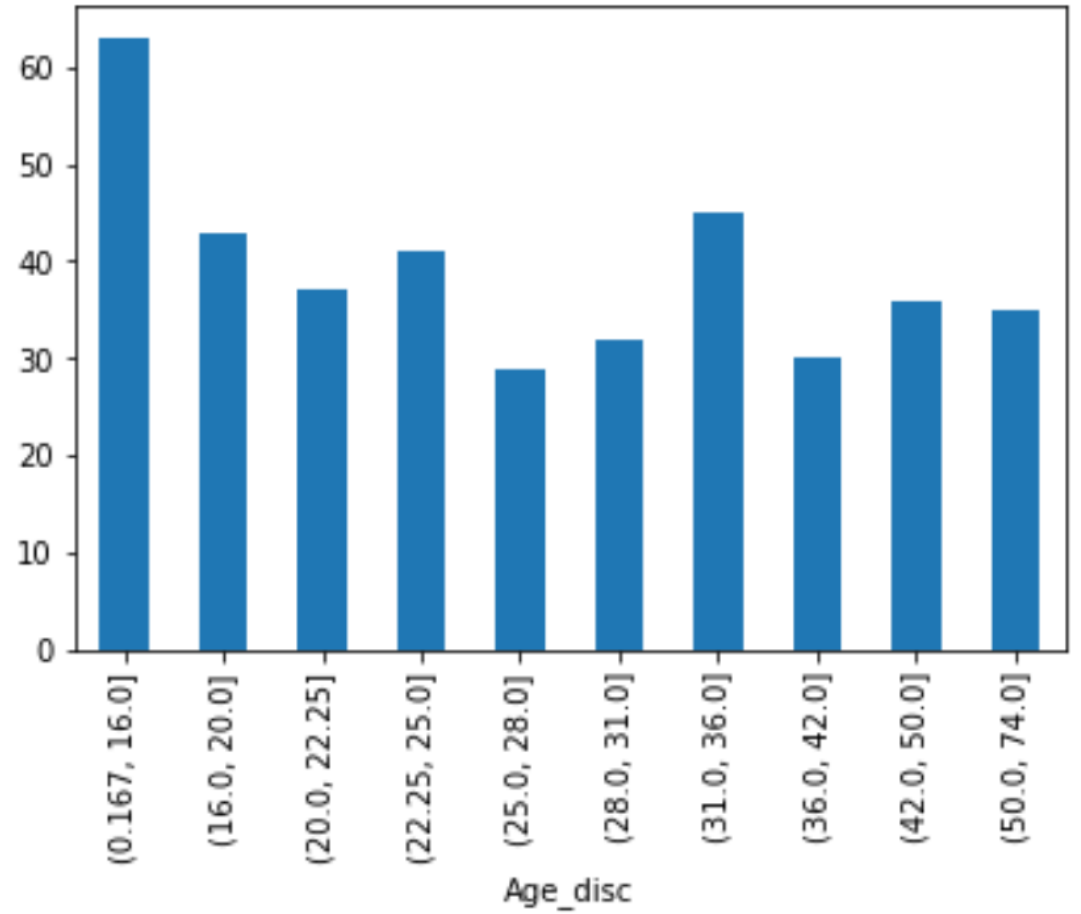
Each interval contains 10% of total observations

Intervals:
0-16; 16-20; 20-22; 22-25; …
50-74

Equal-frequency discretisation: definition







Equal-frequency discretisation: summary

- Improve value spread
- Handles outliers
- Creates discrete variable
- Good to combine with categorical encodings

Train In Data

THANK YOU

www.trainindata.com