Time buckets & boundary rules

Turn individual samples into reporting intervals while making the treatment of boundary values explicit.

Synthetic example · independently invented data

One-page print edition · A3 landscape

01 Inspect the inputs

02 Apply the method

03 Check the result

INVENTED INPUTS

Eight samples on one timeline

Minute	Value
0	2
2	1
5	4
7	3
9	2
10	5
14	1
15	2

Invented elapsed minutes and sample values

Minute 0 is the start of the demonstration. Values are arbitrary units; no date or operational timeseries is used.

SELECTED PYTHON

The working calculation

Pandas is imported as pd. The lists minutes and values contain the inputs at left.

```
samples = pd.Series(
    values, index=pd.to_timedelta(minutes, unit='min')
)
totals = samples.resample(
    '5min', closed='left', label='left'
).sum()
assert totals.sum() == samples.sum()
```

The code shown here is executed to calculate this example's results.

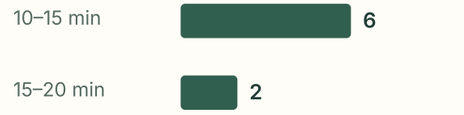
COMPUTED RESULT

Four reporting buckets

20

units retained

Input total = output total



The first bucket contains minutes 0 and 2. The sample at minute 5 belongs to the next bucket.

What this demonstrates

A reporting interval needs an explicit start, width and boundary convention. Resampling keeps the aggregation separate from presentation.

Check the boundary

These bins are left-closed: [0, 5), [5, 10), [10, 15), [15, 20). A boundary sample appears exactly once.

Thomas Eapen
nowlabs.net/tangerine-woa

New illustrative material explaining the approach. Values, configurations and results shown here are not historical workplace records.