Clean durations, then pivot

Convert readable duration strings into numbers before summarising by group and activity.

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

Six records, one deliberate fault

Group	Activity	HH:MM
A	Review	00:35
A	Review	00:25
A	Build	01:10
B	Review	00:40
B	Build	invalid
B	Build	00:50

Six independently invented duration records

The text “invalid” is deliberate. It is excluded and counted before the pivot, rather than treated as a zero duration.

SELECTED PYTHON

The working calculation

Pandas is pd. rows contains the input records; every duration is expressed as hours:minutes.

```
data = pd.DataFrame(rows)
valid = data['duration'].str.fullmatch(r'\d+:[0-5]\d',
na=False)
clean = data.loc[valid].copy()
clean['minutes'] = pd.to_timedelta(
    clean['duration'] + ':00'
).dt.total_seconds() / 60
summary = clean.pivot_table(
    index='group', columns='activity', values='minutes',
    aggfunc='sum', fill_value=0, margins=True,
    margins_name='Total'
)
```

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

COMPUTED RESULT

A reconciled summary

220

minutes in accepted records

5 accepted · 1 excluded

Group	Build	Review	Total
Group A	70	60	130
Group B	50	40	90
Total	120	100	220

Summed minutes, including row and column totals

Group A contributes 130 minutes and Group B contributes 90. The summary totals 220 minutes.

What this demonstrates

Cleaning establishes a trustworthy numeric measure. A pivot then exposes its distribution across two dimensions, with totals that can be checked.

Check the boundary

Only non-negative HH:MM values with minutes below 60 are accepted. Rejected rows remain a separate quality signal.

Thomas Eapen
nowlabs.net/tangerine-woa

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