

⇒ Power BI → Extract data from multiple sources

↳ Modelling (summary based) + ERDs
↳ Shareable / Interactive dashboards.

↳ For statistic based KPIs.
↳ Think of Python Pandas API with filter, group, cols and func as params.

→ Report ; data, Relationships

↳ Extraction of data from multiple CSV raw files
↳ In premise server.

Power BI → support multiple servers / source data

↳ After securing connection → Query Editor.

↳ Transform data → New Pane.

⇒ POWER QUERY

Three major tabs
Remove duplicates Choose columns → HOME
Header Remove cols / Row

Reshaping tools

Test specific tools

Transform

1. get data

↳ Browse for raw files.


Add-columns



↳ Preview Panel.

↳ If we click transform-data

Name →

Applied steps → ^{define path} source.
promote headers.
Changed datatype.

⇒  Remove product-sz col.

↳ Apply and Close (Top ~~right~~ ^{left}  )

Text-Specific tools.

Split

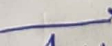
Format → case, strip.

Extract → length, ~~for~~ slicing, split('delimited') then slice.

Transform → Apply on the same col.
selected

Add col → New col for the transform.

↳ Customer data lookup.

↳ Transform  capitalizing first word of name-base cols

↳ Add col → Merge Initial, last + first
↳ join ($a \& b = 1$)
↳ separator space

extract elements before @ → split('@')[0]
extract element after @ before . → Chain two splits
re.find(r'@(\.+)\.', sent)

Transform replace.

⇒ Number Column transformation
STATISTIC → Sum, Mean
↳ Returns a statistic value + value counts + unique
↳ we ~~never~~ usually use it in query editor

standard + scientific

↳ Numeric transformation - logs, trigonometric transformation

↳ Edit Product data

↳ Already uploaded

↳ 0.9 is the discount ratio

↳ Product price $\times 0.9 \rightarrow$ discount price

Date & Time Columns.

we will create date-time columns.

↳ Calendar CSV

↳ Query Editor

start of the week = Sunday
↳ can be changed

↳ Has a parameter for this

pd. date_range (start, end, period, freq='D')
↳ Num of obs

↳ resample (smaller freq). fill()

↳ $\frac{92}{4} = 23$ at resampled time.

⇒ Index columns

↳ Rownum

⇒ Adding conditional columns

↳ If / Else → Apply on a series / col.

Case/when



Transaction table

↳ Sales

Add index column

↳ Condition Column.
↳ If, Else

From add column

↳ Index column

