

## 02 · Limpieza de ventas

Este archivo contiene el pipeline de limpieza de `ventas.csv` con:

- Estandarización de texto (trim, tildes fuera, mayúsculas)
- Coerción de tipos (fecha ISO→date UTC, unidades entero  $\geq 0$ , precio\_unitario Decimal  $\geq 0$ )
- Validación con causas y cuarentena
- Deduplicación (clave: fecha, id\_cliente, id\_producto) con "último gana" por \_ingest\_ts
- Cálculo de importe (Decimal) y QA rápida
- Trazabilidad: \_source\_file, \_ingest\_ts, \_batch\_id

NOTA: Es texto plano con bloques tipo "#%%%" para leer/ejecutar sin JSON.

-----  
NameError

Traceback (most recent call last)

Cell In[5], line 15

12 CENT = Decimal("0.01")

14 # Rutas (anclado a project/)

---> 15 ROOT = Path(\_\_file\_\_).resolve().parents[1]

16 DATA = ROOT / "data" / "drops"

17 OUT = ROOT / "output"

NameError: name '\_\_file\_\_' is not defined