

Data Cleanup & Audit

A synthetic supplier-price export reconstructed without inventing missing values.

180	150	30	5
RAW ROWS	DELIVERED ROWS	DUPLICATES REMOVED	AUDITABLE SHEETS

Executive result

A mixed-format supplier dataset becomes a deduplicated, import-ready table while the raw input, removed rows, normalization rules and unresolved values remain inspectable.

What this sample proves

Cleaning is not silent beautification. A buyer can see what changed, why it changed, which records were removed and which fields remain review-required.

Use boundary

All records are synthetic. The sample makes no supplier, price, availability or commercial claim and contains no client or personal data.

State transformation

01 Chaos

Whitespace, case drift, inconsistent country and currency labels, mixed date formats, duplicate business keys, invalid quantities and missing lead times.

02 Operation

Raw preservation → deterministic normalization → composite-key duplicate detection → validation rules → exception quarantine → reconciliation.

03 Usable state

A clean machine-readable table, QA log, transformation rules and delivery summary suitable for filtering, import and buyer review.

Uncompressed residual

Ambiguous supplier identity, unknown source authority and genuinely missing values are not guessed into completeness.

Observable proof

- Raw input kept unchanged
- Composite-key deduplication
- Validation and exception log
- Before/after counts reconcile

Portfolio integrity

Independently constructed self-directed technical work using public or synthetic inputs. Not a reproduced marketplace brief, confidential client engagement, paid result or claim of deployment.