Remittance Stub Parsing

Automated extraction, validation, and reconciliation of retailer remittance advice — Cinderhaven Provisions case study

The PDF Pile

Every dollar a retailer pays arrives with a remittance advice. Each format buries deductions in different layouts, naming conventions, and reason codes. A single supplier managing four retail channels receives four incompatible PDF formats, each requiring manual keying before anyone can ask: did this deduction match an invoice?

15
Remittance stubs processed

4
Distinct retailer formats

155
Deduction line items extracted

$207,835.06
Total deductions identified

The 4 remittance formats — costco, keHE, unfi, walmart — produced 155 deduction entries totaling $207,835.06. Each format uses its own reason-code taxonomy, column layout, and arithmetic conventions. The pipeline normalizes all four into a single typed ledger in under a second per stub.

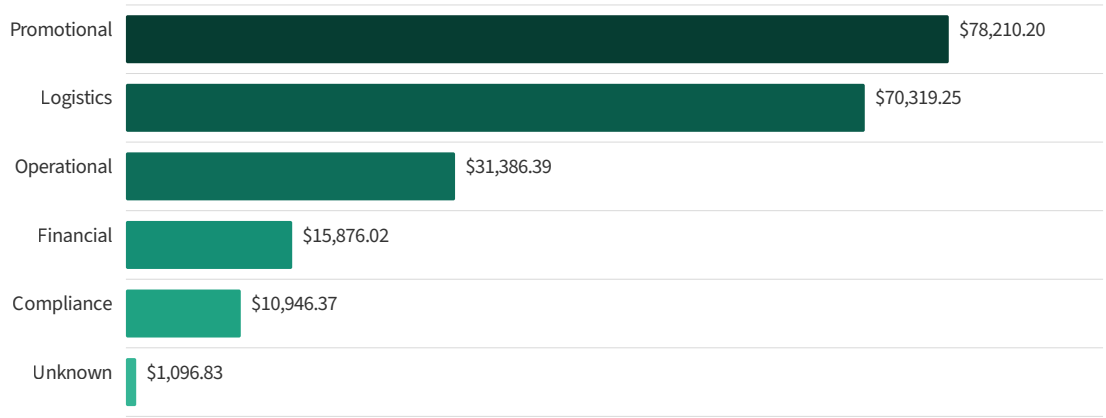
Source: synthetic Cinderhaven Provisions remittance stubs generated to match production-format PDFs from each retailer/distributor. All figures derived from automated extraction — no manual keying.

The Unified Ledger

Deductions from 4 formats collapse into six categories once reason codes are mapped. The category breakdown reveals where the money goes — and which categories carry the highest dispute potential.

CATEGORY	COUNT	AMOUNT	SHARE
Promotional	53	$78,210.20	37.6%
Logistics	57	$70,319.25	33.8%
Operational	22	$31,386.39	15.1%
Financial	13	$15,876.02	7.6%
Compliance	9	$10,946.37	5.3%
Unknown	1	$1,096.83	0.5%
Total	**155**	**$207,835.06**	**100.0%**

Deduction Category Distribution



Promotional deductions account for 38% of total deductions ($78,210.20), followed by logistics at $70,319.25. Against Cinderhaven's ~$3.6M/yr all-in trade spend and 3,357 chargebacks, these deductions represent one remittance cycle's contribution to the annual trade cost.

Cinderhaven canonical figures: ~$3.6M/yr all-in trade spend, 3,357 chargebacks (2,873 retailer + 484 distributor), 36-month window. Category assignment via per-retailer reason-code YAML configs.

Recovery Potential

Reconciliation matches each deduction's invoice number against Cinderhaven's accounts-receivable ledger. Unmatched deductions — those with no corresponding invoice or a partial amount mismatch — represent recoverable dollars, but only if disputed before the 90-day window closes.

$26,173.80
Within the 90-day dispute window

$181,164.64
Past the dispute window (forfeit risk)

> $26,173.80 in unreconciled deductions remain inside the 90-day dispute window. Every week of delay narrows the recovery margin.

> $181,164.64 in deductions have passed the 90-day dispute window. Without automated extraction, these dollars expired before anyone reconciled them.

STUB	RETAILER	MATCH STATUS	MATCHED	UNMATCHED	DAYS REMAINING
`costco_stub_01.pdf`	Costco	UNMATCHED	$0.00	$10,330.64	0 days
`costco_stub_02.pdf`	Costco	UNMATCHED	$0.00	$11,815.87	0 days
`costco_stub_03.pdf`	Costco	UNMATCHED	$0.00	$16,698.89	0 days
`costco_stub_broken_unmapped_code.pdf`	Costco	UNMATCHED	$0.00	$12,684.90	0 days
`keHE_stub_01.pdf`	KeHE	PARTIAL	$165.54	$10,165.10	0 days
`keHE_stub_02.pdf`	KeHE	UNMATCHED	$0.00	$14,868.66	0 days
`keHE_stub_03.pdf`	KeHE	UNMATCHED	$0.00	$13,086.90	90 days
`unfi_stub_01.pdf`	UNFI	PARTIAL	$165.54	$10,165.10	0 days
`unfi_stub_02.pdf`	UNFI	UNMATCHED	$0.00	$14,868.66	0 days
`unfi_stub_03.pdf`	UNFI	UNMATCHED	$0.00	$13,086.90	90 days
`unfi_stub_multipage.pdf`	UNFI	UNMATCHED	$0.00	$46,030.99	0 days
`walmart_stub_01.pdf`	Walmart	PARTIAL	$165.54	$10,165.10	0 days

STUB	RETAILER	MATCH STATUS	MATCHED	UNMATCHED	DAYS REMAINING
`walmart_stub_02.pdf`	Walmart	UNMATCHED	$0.00	$3,758.06	0 days
`walmart_stub_03.pdf`	Walmart	UNMATCHED	$0.00	$11,805.97	0 days
`walmart_stub_broken_arithmetic.pdf`	Walmart	UNMATCHED	$0.00	$7,806.70	0 days

Labor Estimate

Manual keying of a single remittance stub — opening the PDF, reading each line item, typing invoice numbers and amounts into a spreadsheet, checking arithmetic — takes an experienced analyst 8 to 15 minutes per stub. At 15 stubs, that is 150 to 225 minutes of manual work. The automated pipeline processes the same set in under 15 seconds.

150–225 min
Manual keying estimate

< 15 sec
Automated extraction

The cost is not the keying time. It is the delay. Every day a deduction sits unreconciled is a day closer to the dispute window closing. Manual processes create a bottleneck that turns recoverable dollars into write-offs.

Reconciliation against Cinderhaven reference invoices (data/cinderhaven_reference.json). Dispute window: 90 days from deduction date. Labor estimates based on industry benchmarks for manual remittance keying.

The Validation Loop

Every extracted stub passes two deterministic checks before it reaches the ledger. First, arithmetic: does net cash plus the sum of deductions equal the gross invoice? Second, code mapping: does every reason code resolve to a known category in the retailer's configuration? A stub that fails either check is flagged for human review.

13
Stubs verified

2
Stubs flagged for review

STUB	RETAILER	STATUS	ARITHMETIC	CODE MAPPING	ISSUES
`costco_stub_01.pdf`	Costco	VERIFIED	Balanced	All mapped	0
`costco_stub_02.pdf`	Costco	VERIFIED	Balanced	All mapped	0
`costco_stub_03.pdf`	Costco	VERIFIED	Balanced	All mapped	0
`costco_stub_broken_unmapped_code.pdf`	Costco	FLAGGED	Balanced	Unmapped codes	1
`keHE_stub_01.pdf`	KeHE	VERIFIED	Balanced	All mapped	0
`keHE_stub_02.pdf`	KeHE	VERIFIED	Balanced	All mapped	0
`keHE_stub_03.pdf`	KeHE	VERIFIED	Balanced	All mapped	0
`unfi_stub_01.pdf`	UNFI	VERIFIED	Balanced	All mapped	0
`unfi_stub_02.pdf`	UNFI	VERIFIED	Balanced	All mapped	0
`unfi_stub_03.pdf`	UNFI	VERIFIED	Balanced	All mapped	0
`unfi_stub_multipage.pdf`	UNFI	VERIFIED	Balanced	All mapped	0
`walmart_stub_01.pdf`	Walmart	VERIFIED	Balanced	All mapped	0
`walmart_stub_02.pdf`	Walmart	VERIFIED	Balanced	All mapped	0
`walmart_stub_03.pdf`	Walmart	VERIFIED	Balanced	All mapped	0
`walmart_stub_broken_arithmetic.pdf`	Walmart	FLAGGED	Discrepancy ($42.50)	All mapped	1

Of 15 stubs processed, 13 balanced to the penny and mapped to known reason codes. 2 did not. Flagged stubs enter the review queue, where an analyst can inspect the original PDF alongside the extracted data, correct errors, and revalidate.

The flags are not failures of the system — they are the system working. Every flagged stub is a deduction that would otherwise have been keyed wrong or missed entirely.

Validation is deterministic: no model confidence scores, no probability thresholds. If the arithmetic balances and the codes map, the stub is verified. If not, it is flagged. The trust signal is the data, not the model.

What This Means for a $25M Brand

A specialty food company running four retail channels and two distributors generates roughly 50 to 80 remittance stubs a month. At 10 minutes per stub, that is 8 to 13 hours of manual keying — work that typically falls to a controller or AP lead who has other responsibilities, which is why the spreadsheet runs three weeks behind.

Three weeks is half the dispute window. By the time a deduction is keyed, categorized, and matched against the invoice, the remaining days are often in single digits or already zero. The result is not a process problem. It is a write-off schedule that runs on autopilot.

A configured parser eliminates the keying lag entirely. Stubs are extracted, validated, and reconciled the day they arrive. Deductions that do not balance or do not map to a known reason code route to a review queue instead of sitting in a folder. The dispute window starts full instead of half-spent.

For Cinderhaven, the automated pipeline flagged $207,338.44 in unreconciled deductions across 15 stubs. Of that, $181,164.64 had already passed the 90-day dispute window — forfeit. Another $26,173.80 remained inside the window but narrowing. The deductions were there. The data to dispute them was there. What was missing was the time between the PDF arriving and someone reading it.

Lailara sets up a parser against your actual retailer and distributor formats in two to three weeks. The deliverable is a working pipeline tuned to your stubs, a back-populated deduction ledger, and a standing flag list of what to dispute before it expires.

The alternative is the spreadsheet. It will always be behind.

Built by Lailara LLC — data hygiene and analytics consulting for specialty food brands scaling into national retail.