

Product Data Readiness Audit

Cinderhaven Provisions · May 2026

Shawn Phillips | Lailara LLC

2026-05-03

Table of contents

Part 1: The Money	2
The \$460,892 you already pay	3
What a wrong barcode costs when nobody is counting	5
The pattern beneath the numbers	5
The slots you don't get back	6
The \$993-a-month problem nobody sees	7
The revenue you're not capturing	7
The SKU you can't afford to ignore	8
Part 2: Why It Happens	9
No gate, no audit trail	10
The uniform defect	10
You are allocating resources to the wrong retailer	10
One product line already has better outcomes. The reason isn't what you'd guess.	11
Part 3: What to Do About It	12
40 hours against \$460,892	12
What's still broken right now	13
How to read the triage list	14
Monday morning	14
The sequence	15
Part 4: The Evidence	17

Part 1: The Money

By the end of 2027, every barcode in American retail will change. GS1 Sunrise 2027 transitions the industry from linear barcodes to 2D barcodes built on GS1 Digital Link: a QR code whose foundation is a valid GTIN. Not a GTIN that looks right to a human eye. A GTIN that passes algorithmic validation. 45 of Cinderhaven's 50 SKUs carry GTIN-14s that do not pass. 45 carry invalid UPC check digits. Not a single SKU has both a valid GTIN and a valid UPC. Those SKUs cannot participate in the transition until someone corrects the digits.

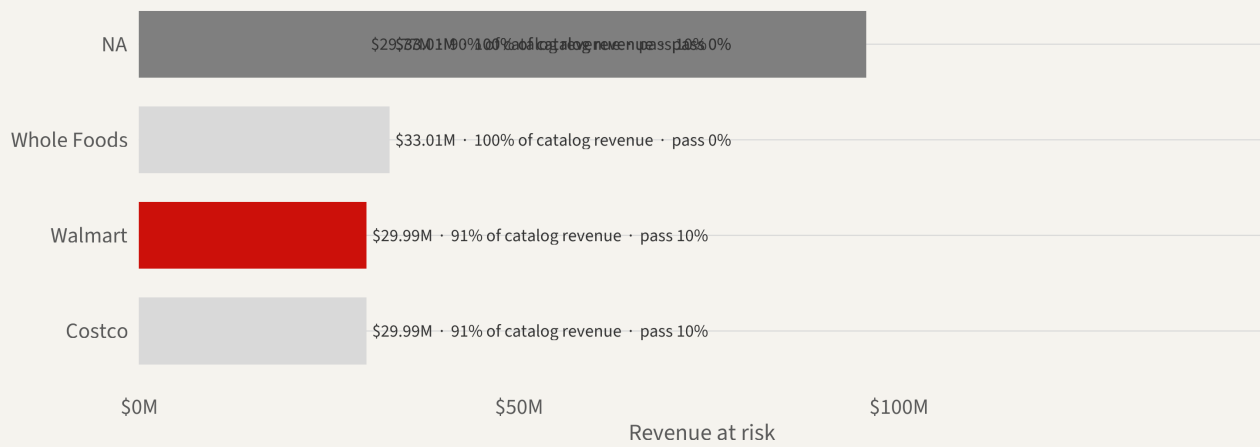
Running in parallel, FSMA Rule 204 makes accurate GTINs a federal requirement. FDA food traceability mandates them as the product identifier backbone for tracking food through the supply chain. The barcode that used to be a scanning convenience is becoming the regulatory infrastructure for food safety. This is not a retailer preference. It is law.

These two deadlines land on a company whose product master would fail basic validation at nearly every retailer today. 40 of 50 SKUs fail every retailer's required-field checks. 3 of 6 retailers have a 0% pass rate. The handful of SKUs that pass at one or two retailers do so because those retailers require fewer fields — not because the data is clean. \$33.0 million in trailing twelve-month revenue rides on data that would not survive the onboarding process at most retailers if submitted fresh.

Retailer	SKUs failing	Revenue at risk	Pass rate
Regional Group	50 of 50	\$33.0M	0%
Whole Foods	50 of 50	\$33.0M	0%
Kroger	50 of 50	\$33.0M	0%
Sprouts	45 of 50	\$33.0M	10%
Walmart	45 of 50	\$33.0M	10%
Costco	45 of 50	\$33.0M	10%

Revenue at risk: \$29.99M rides on data Walmart could reject today

Bar = annual revenue from SKUs that fail at least one required field for that retailer



Source: retailer_readiness_summary + sku_master_full, audit run 2026-05-03

Figure 1: Revenue at risk by retailer.

These are not projections of what might happen if data quality degrades. They are measurements of the current product master against the current retailer requirements. The data fails now, at 80% of retailer-SKU combinations. The only reason the revenue still flows is that nobody has run the audit yet.

The convergence of near-universal retailer readiness failure, GS1 Sunrise, and FSMA 204 is the wall. Each individually would justify fixing the product master. Together, they create a deadline. A company that reaches 2028 with 45 invalid GTINs and 45 invalid UPCs will not have a data quality problem. It will have a market access problem.

The \$460,892 in annual chargebacks is what dirty data costs when nobody checks. The \$33.0 million in at-risk revenue is what it costs when someone does. The GS1 and FSMA transitions are the moment when everyone checks at once.

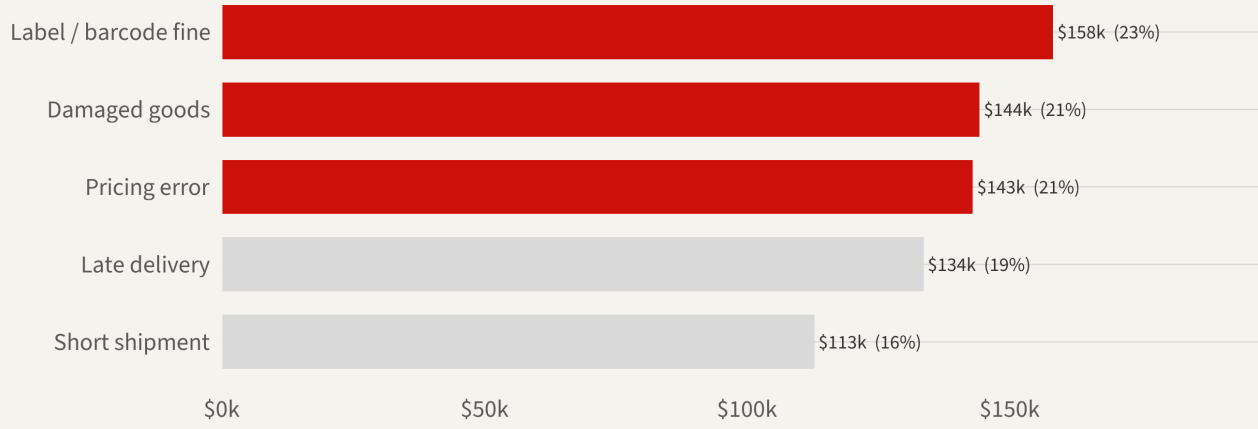
The \$460,892 you already pay

Cinderhaven's 6 retailers deducted \$691,338 in chargebacks from settlement payments over the past 18 months. Annualized, that's \$460,892 a year in revenue that left the building before it reached the bank account.

The chargebacks arrive under five headings: label and barcode fines, pricing errors, damaged goods, late deliveries, and short shipments. Late deliveries, short shipments, and damaged goods are logistics problems. Label and barcode fines and pricing errors are data problems: charges that trace to wrong or incomplete product records. Those data-related categories account for 44% of the chargeback bill.

Three data-defect reasons account for 64% of chargeback dollars

Three data-defect reasons (red) account for 64.4% of chargeback dollars.



Source: chargebacks_enriched

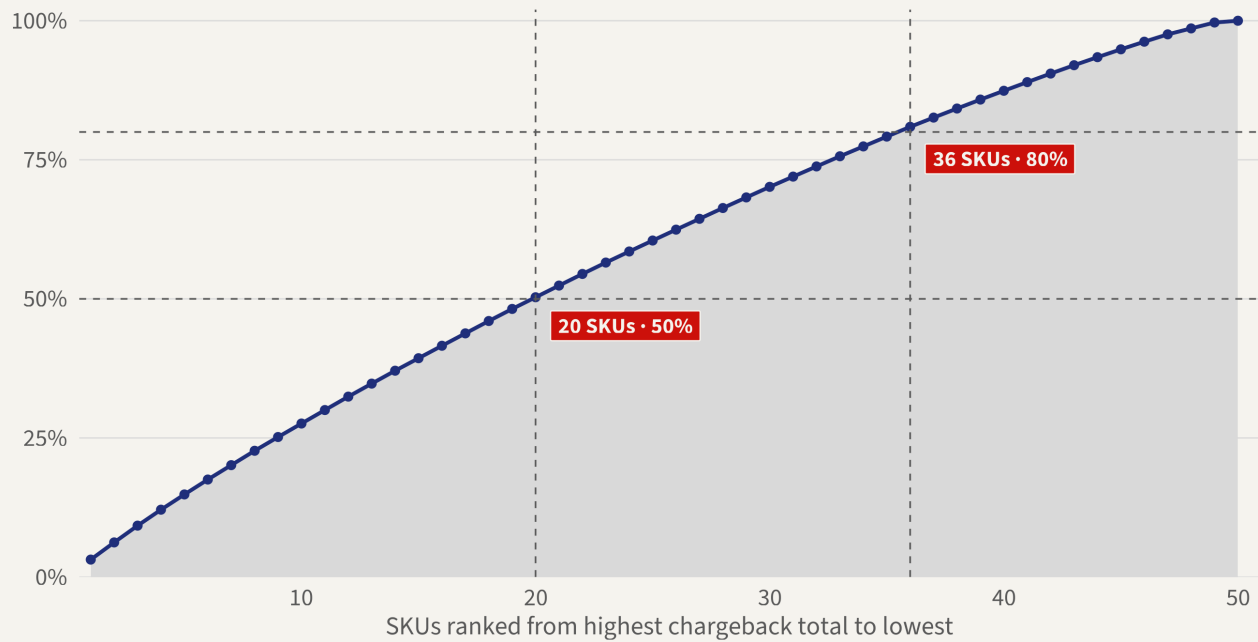
Figure 2: Chargeback dollars by reason.

20 SKUs generate half of that bill. 36 generate 80%. The concentration is not extreme because the underlying defect is universal: every SKU in the catalog carries barcode validation failures, and every SKU generates chargebacks. But some SKUs generate far more than others. Here are the names:

SKU	Product	18-mo chargebacks	What's still broken
CHP-PS-007	Smoked Paprika	\$21,632	OneWorldSync incomplete; GTIN-14 check digit; UPC-12 check digit
CHP-SB-001	Dark Chocolate Sea Salt Bites	\$21,338	GTIN-14 check digit; UPC-12 check digit
CHP-AS-008	Mango Jalapeño Salsa	\$20,856	NA
CHP-DG-005	Sun-Dried Tomatoes	\$19,830	OneWorldSync incomplete; GTIN-14 check digit
CHP-SC-006	Artichoke Spinach Dip	\$18,937	GTIN-14 check digit; UPC-12 check digit

20 SKUs drive half your chargeback bill

Cumulative share of chargeback dollars by SKU rank, 18 months



Source: Cinderhaven chargeback ledger, 18 months

Figure 3: Chargeback Pareto: a small number of SKUs drive most of the cost.

Every one of these carries an invalid UPC check digit. Most also carry an invalid GTIN-14. These are the same barcode defects present across the entire catalog, but concentrated at the top of the chargeback ranking because these SKUs ship to more retailers at higher volume. The defects are present in the product master right now. Not last year. Not at the time of the last audit. Right now.

What a wrong barcode costs when nobody is counting

CHP-PS-007, Smoked Paprika, leads the chargeback ranking at \$21,632 over 18 months. It carries \$438,215 in trailing twelve-month revenue across 201 stores. 21 chargeback events spread across 18 months, arriving at roughly \$592 a month on settlement statements from multiple retailers. Nobody at Cinderhaven has connected these monthly line items to each other, because nobody has a process for tracing chargebacks back to specific fields in the product master. The charges look like separate problems at separate retailers. They are the same underlying barcode defect, repeated.

Fixing a check digit takes ten minutes. Open the product master. Recalculate the check digit using the standard GS1 algorithm. Type the correct number. Save. Ten minutes against \$14,421 a year. That ratio does not require a business case. It requires someone to know the connection exists.

The pattern beneath the numbers

The barcode defect is not concentrated in a few problem SKUs. It is the catalog's default state. 45 of 50 SKUs carry invalid UPC check digits. 45 carry invalid GTIN-14s. No SKU in the catalog has both a valid GTIN and a valid UPC. Every SKU that ships to a retailer with an invalid barcode is a candidate for a compliance penalty every time the barcode is scanned, validated, or reconciled against a data feed.

The chargebacks are spread across 6 retailers and five reason categories. “Label / barcode fine” is the largest single category, but “Pricing error,” “Damaged goods,” “Late delivery,” and “Short shipment” also contribute. The data-related reasons (label/barcode and pricing) account for 44% of the total bill. The remainder is logistics.

The uniformity of the defect explains two things. First, why the chargeback spread across SKUs is relatively narrow: the top chargeback SKU (\$21,632) is only about three times the bottom. Every SKU has the same type of defect, so the variation comes from shipping volume and retailer mix, not from defect severity. Second, why the fix is so concentrated: correct the barcodes and the entire data-defect chargeback bill disappears. There is no long tail of miscellaneous problems to chase.

The slots you don’t get back

Chargebacks take your money. Deauthorizations take your position.

CHP-DG-002, Quinoa Medley, was authorized across 219 stores. It carries 2 data defects and has been deauthorized at 25 locations. Those slots are gone. Winning them back requires a new category review, which happens once a year at Walmart, and a pitch that explains why the product that was pulled for data defects won’t be pulled again.

25 stores out of 219 sounds minor. It isn’t. It’s the signal that the retailer’s system has flagged this SKU. The barcode defects that triggered the deauthorizations are identical to the defects still present at the other 194 stores. \$18,530 in chargebacks over 18 months with no sign of stopping. The deauthorized stores were not the punishment. They were the warning shot. The unfixed barcodes are the loaded gun still pointed at the other 194.

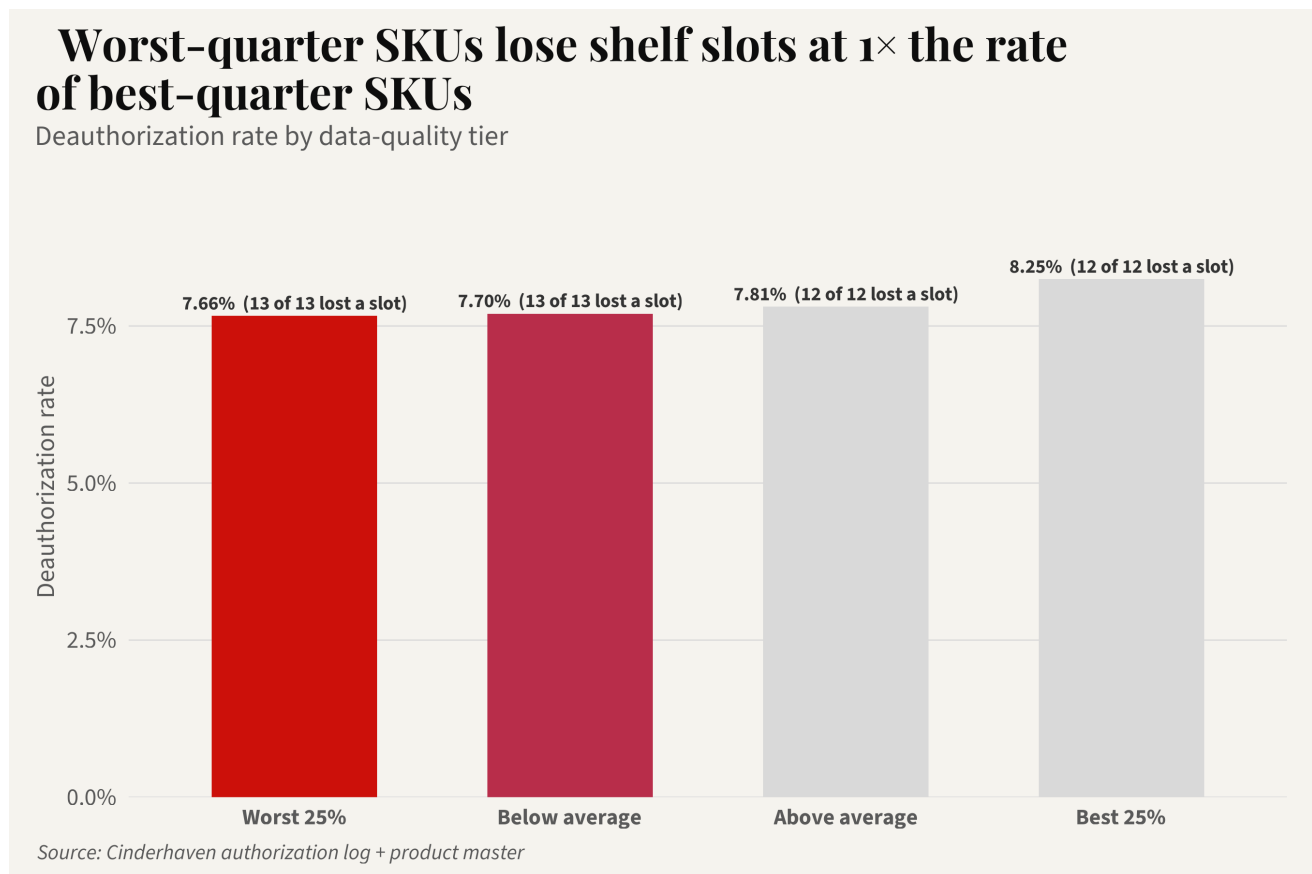


Figure 4: Deauthorization rate by data quality tier.

The pattern is not unique to CHP-DG-002. Across the catalog, 50 of 50 SKUs have lost at least one store authorization. The mean deauthorization rate is 7.8%, with some SKUs losing more than 12% of their store base.

Every one of these SKUs carries the same barcode defects as the rest of the catalog. The deauthorizations are the retailer's system doing what it was designed to do: flagging products with invalid data and removing them from distribution.

The cost of a deauthorization is not the lost revenue at that store. It is the competitive displacement. A specialty food brand does not compete for abstract "shelf space." It competes for specific slots in specific planograms that are reviewed on 12-to-24-month cycles. Losing a slot to a competitor means that competitor's product will generate 12 to 24 months of velocity data at that location, data the category manager will use to justify keeping it during the next review. The brand that lost the slot has to overcome a year of incumbent velocity data with nothing but a pitch deck and a promise.

The \$993-a-month problem nobody sees

The 45 invalid GTIN check digits and 45 invalid UPC check digits in Cinderhaven's product master have been wrong since the day each SKU was entered. In that time, nobody corrected a single one. Not because anyone decided the chargebacks were acceptable. Because nobody knew the chargebacks and the digits were connected.

The chain of visibility works like this. A retailer's automated system validates the barcode on an inbound shipment or a data feed submission. The check digit fails. The system generates a compliance penalty. The penalty appears as a line item on the next settlement statement, categorized under a heading like "vendor compliance deductions" or "label/barcode fine." The settlement statement is 40 pages long. It contains hundreds of line items. The compliance penalties are scattered across pages, interleaved with promotional deductions, logistics credits, and payment adjustments. An individual penalty is small. It does not trigger an investigation. It does not cross an approval threshold. It does not generate an alert.

On the other side, the product master sits in whatever system Cinderhaven uses to manage product data. The GTIN-14 and UPC fields contain numbers that were typed once, by whoever set up each SKU, and have not been opened since. Nobody reviews barcode fields. Nobody runs check digit validations. Nobody has a process for connecting a chargeback on a settlement statement to a digit in the product master.

The ops team is not negligent. They are fully occupied. Six retailer portals. Broker coordination. Velocity reports rebuilt by hand every Monday. Trade spend reconciliation. New SKU launches. Promotional planning. Data cleanup is on the list. It is always on the list. It sits between "update the trade spend template" and "fix the label printer" and it never reaches the top because the chargebacks arrive in amounts too small to demand attention and too steady to ever stop on their own.

CHP-AS-009, Truffle Mushroom Sauce, is the #1 revenue SKU in the catalog at \$1.1 million. It has generated 15 chargeback events in 18 months, \$17,871 in total, roughly \$993 a month. 15 times, the system flagged the wrong barcode digit, generated a penalty, deducted it from a settlement, and nobody traced it back. Not because tracing it was hard. Because nobody knew to look.

This is the structural problem. The chargebacks persist because the defects persist. The defects persist because nobody has time to find them. Nobody has time to find them because the cost of each individual defect is too small to surface through normal business processes. The total is \$460,892 a year. The individual units are invisible. The system that would make them visible does not exist yet. Part 3 of this report describes what that system looks like.

The revenue you're not capturing

The cost story is about money leaving. This is about money that never arrives.

40 of Cinderhaven's 50 SKUs fail every retailer's required-field check. Not a single product in the catalog has both a valid GTIN-14 and a valid UPC, so no product could be submitted to a retailer that checks both without data work first. The expansion pipeline is blocked by barcode validation failures.

All 50 currently authorized SKUs are shipping to retailers and generating revenue. They were authorized before the requirements tightened, or before anyone checked. They are generating revenue on borrowed time.

SKU	Product	TTM revenue	18-mo chargebacks
CHP-AS-009	Truffle Mushroom Sauce	\$1.1M	\$18k
CHP-PS-004	Extra Virgin Olive Oil	\$1.1M	\$17k
CHP-PS-009	Maple Syrup Grade A	\$1.0M	\$13k
CHP-PS-002	Wildflower Honey	\$1.0M	\$10k
CHP-AS-006	Balsamic Fig Glaze	\$1.0M	\$15k
CHP-AS-001	Smoky Chipotle BBQ Sauce	\$997k	\$13k
CHP-DG-005	Sun-Dried Tomatoes	\$946k	\$20k
CHP-AS-007	Lemon Herb Chimichurri	\$897k	\$11k
CHP-DG-007	Trail Mix Premium	\$877k	\$16k
CHP-AS-010	Carolina Gold BBQ	\$867k	\$10k

\$9.9 million in revenue from the top 10 SKUs alone, all riding on data that most retailers' systems would reject. These products are already generating chargebacks. The risk and the cost are happening simultaneously.

If Cinderhaven wants to pitch a line extension at any retailer, and at \$33.0 million in revenue and growing, that pitch is coming, every SKU needs its data fixed before the conversation starts. Not during the conversation. Before. A retailer's category team does not fix vendor data. They evaluate what's submitted. If the submission fails their automated checks, the conversation ends before a human being ever sees the product.

The gap between "blocked" and "ready" for every SKU in the catalog is the same: barcode check digits. The work is measured in hours, not weeks. 50 failing SKUs are not 50 product development problems. They are 50 data entry tasks. The difference between \$0 in expansion-ready revenue and \$33.0 million in expansion-ready revenue is approximately 40 hours of clerical work. That is the most underspent 40 hours in the company.

The SKU you can't afford to ignore

CHP-AS-009, Truffle Mushroom Sauce, is the best-selling product in the Cinderhaven catalog. \$1.1 million in trailing twelve-month revenue. 235 stores across every channel. Velocity of 8.2 units per store per week. It represents 3.4% of company revenue and \$815,608 in annual gross margin. If Cinderhaven has a flagship, this is it.

It is also among the largest sources of chargeback cost. 15 chargeback events over 18 months. \$17,871 in penalties deducted from settlement payments across multiple retailers. Not clustered in one bad quarter. Not triggered by one bad shipment. 15 events spread across 18 months, because the same barcode defects trigger the same automated validation failures at the same retailers, month after month.

The defects are not complex. The GTIN-14 check digit is wrong. The UPC check digit is wrong. Label and barcode fines account for 25% of CHP-AS-009's chargeback dollars. The rest comes from pricing errors and operational categories.

The product passes 0 of 6 retailers' required-field checks today. It is authorized at 235 stores and its data would not survive the onboarding process at most of them if it were submitted fresh. It was authorized before the checks existed or before they were enforced at their current stringency. It survives on inertia, not on data quality.

The same is true across the top of the catalog:

Rank	Product	Revenue	DQ score	Chargebacks	Retailers passing
1	Truffle Mushroom Sauce	\$1.1M	100	\$17,871	0 of 6
2	Extra Virgin Olive Oil	\$1.1M	100	\$16,685	0 of 6
3	Maple Syrup Grade A	\$1.0M	100	\$12,723	0 of 6
4	Wildflower Honey	\$1.0M	100	\$9,989	0 of 6
5	Balsamic Fig Glaze	\$1.0M	100	\$15,477	0 of 6
6	Smoky Chipotle BBQ Sauce	\$997k	100	\$13,420	2 of 6
7	Sun-Dried Tomatoes	\$946k	100	\$19,830	1 of 6
8	Lemon Herb Chimichurri	\$897k	100	\$10,929	0 of 6
9	Trail Mix Premium	\$877k	100	\$16,053	0 of 6
10	Carolina Gold BBQ	\$867k	100	\$9,760	0 of 6

Chargebacks as % of gross margin — top 15 SKUs

Where chargeback dollars are largest relative to the margin the SKU earns

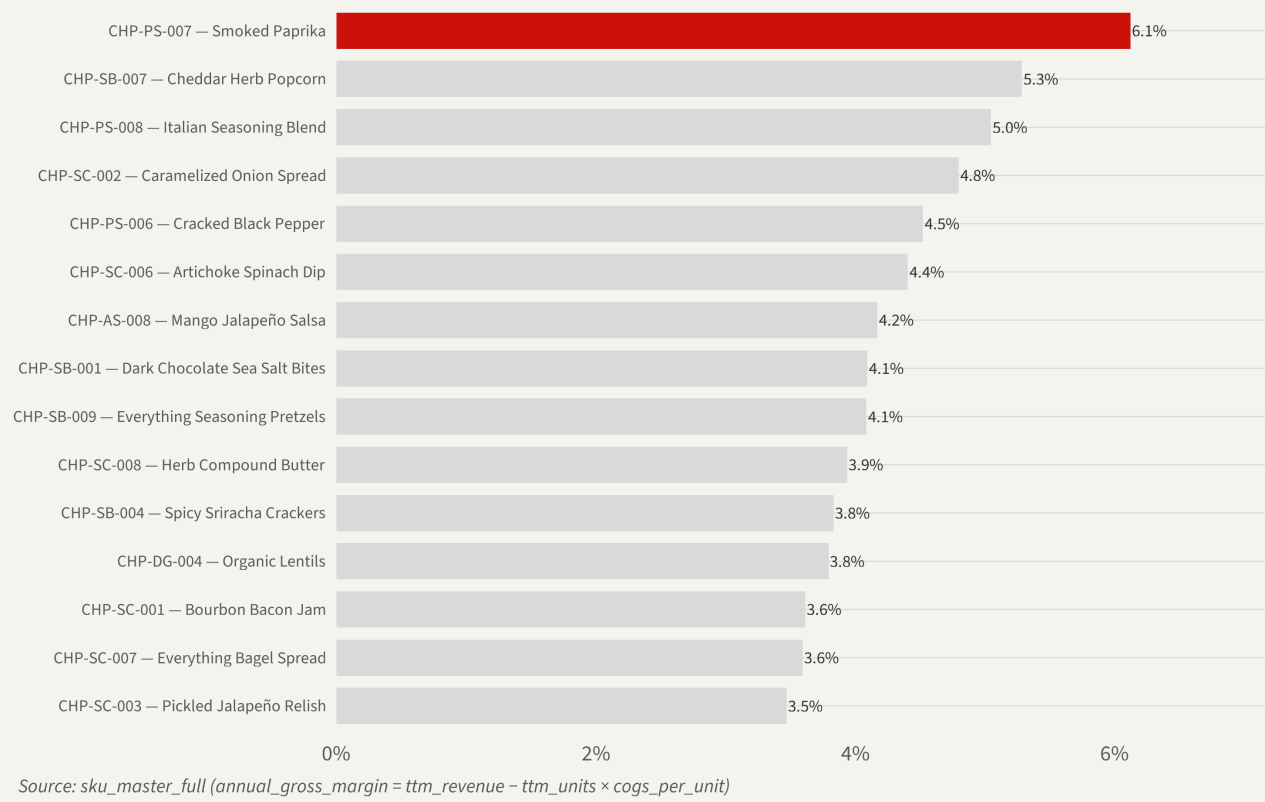


Figure 5: Chargebacks as percentage of gross margin, top 15 SKUs.

2 of the top 10 SKUs pass any retailer’s full check. The \$9.9 million in revenue at the top of the catalog, 30% of the company, rides on data that would fail most retailer onboarding processes.

The products that generate the most revenue received the same one-time data entry as every other product. They just generate larger penalties when that entry is wrong, attract more retailer scrutiny because of their volume, and create more exposure when a readiness audit runs. The risk concentrates at the top because revenue concentrates at the top. The data quality does not vary.

Fixing CHP-AS-009 takes 50 minutes. Correct the GTIN check digit. Correct the UPC check digit. When that’s done, \$11,914 a year in chargebacks is reduced and the #1 SKU moves toward passing retailer readiness checks.

The risk of leaving it unfixed is not the \$11,914. The risk is that a retailer runs the check. Walmart doesn’t send a chargeback for a readiness failure. Walmart sends a deauthorization. And when the #1 SKU, 3.4% of company revenue, \$815,608 in annual gross margin, loses its largest retailer, the conversation is not with the data team. It is with the board.

Part 2: Why It Happens

Part 1 showed what data debt costs. This section is about where it comes from. The causes are ordinary and fixable. The frustrating part is how long they’ve been accumulating.

No gate, no audit trail

The product master has no recorded entry source for any of its 50 SKUs. The `updated_by` field is blank across the entire catalog. Nobody knows who entered these records. Nobody knows when. Nobody knows what process, if any, was followed.

This is the clearest evidence that the product master is an unmanaged asset. It is the most important data system in the company — every retailer relationship, every chargeback, every velocity report, every shelf placement depends on it — and it has no owner, no process, and no audit trail.

There is no intake checklist. No required field set enforced at entry. No validation step between “someone typed this” and “retailers are ordering against it.” A barcode check digit can be entered wrong, and the record goes live the moment it’s saved. The first validation that record will ever receive is a retailer’s automated compliance check, months later, when it fails and generates a penalty that nobody traces back to the upload.

The fix is structural: a gate between data entry and the live product master. A check digit calculation that runs when a GTIN or UPC is entered and blocks the save if it fails. The technology is trivial — the GS1 algorithm is a single modulo operation. The discipline is the deliverable.

The uniform defect

Nearly every SKU in the catalog carries the same defect. 45 of 50 have invalid UPC check digits. 45 also carry invalid GTIN-14 check digits. No SKU has both valid. The basic completeness checks — brand owner, case dimensions, country of origin, weights — all pass, which is why the data quality score reads 100 for every SKU. But completeness is not the same as correctness. The barcodes are present and the right length. They just fail check-digit validation, which is what retailers actually run.

The instinct is to assume the big sellers would have been cleaned up by now. The assumption is wrong because it confuses commercial attention with data attention. Everyone at Cinderhaven knows that Truffle Mushroom Sauce sells \$1.1 million a year. Nobody at Cinderhaven knows that its GTIN-14 check digit is wrong. Those are two different kinds of knowing, and only the first one happens naturally.

Data entry is clerical work. It happens at launch, when somebody has 20 minutes between other tasks, and it never happens again. Nobody revisits the product master after a SKU is selling. The record freezes at whatever state it was in on the day someone typed it. A \$1.1 million SKU and a \$312,888 SKU both get one pass through data entry. The \$1.1 million SKU just generates larger chargebacks when the entry is wrong.

The fix does not require better data entry. It requires a list. Put the revenue number next to every SKU on the ops team’s screen. Sort by revenue. Start at the top. The people doing the work have never been shown which products their work protects. Give them that information and the triage takes care of itself.

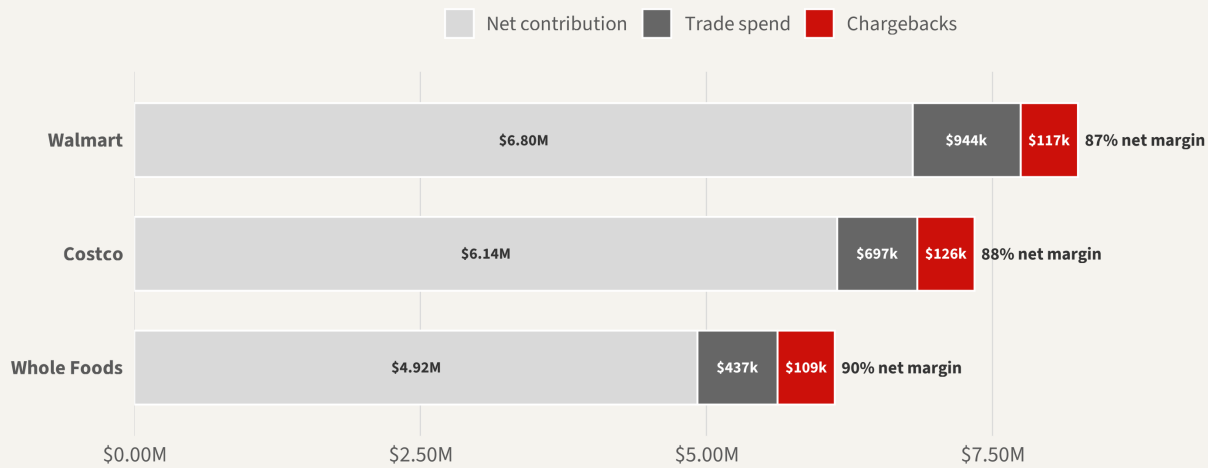
You are allocating resources to the wrong retailer

Walmart generates \$7.9 million in gross revenue. That’s 24% of the catalog. It is the largest channel by every gross metric. It is not the most profitable.

Retailer	Gross	Trade spend	Chargebacks	Net margin
Kroger	\$6.8M	7%	1.67%	91.3%
Whole Foods	\$5.5M	8%	1.99%	90%
Sprouts	\$4.1M	9%	2.57%	88.4%
Costco	\$7.0M	10%	1.81%	88.2%
Walmart	\$7.9M	12%	1.49%	86.5%
Regional Group	\$1.8M	7%	6.72%	86.3%

Walmart wins on dollars; Whole Foods wins on margin

Annual gross revenue → trade spend → chargebacks → net contribution, by retailer



Source: Cinderhaven sales + cost data + chargeback ledger, trailing 12 months

Figure 6: True net margin by retailer.

Kroger contributes 91 cents of margin on every dollar of revenue. Regional Group contributes 86 cents. The 5-cent gap is almost entirely trade spend.

This table reorders the CEO's priorities. Not away from Walmart. Walmart generates \$6.8 million in net contribution. You don't walk away from that. But you stop assuming that Walmart volume equals Walmart profitability when deciding where to invest ops resources, which retailer gets the first call when there's a data issue, and which expansion opportunity gets prioritized.

The chargeback column reveals something else. The rates are small at every retailer. But chargebacks are the only margin lever entirely within Cinderhaven's control. Trade spend is negotiated once a year. Chargebacks are generated by data defects that Cinderhaven can fix any Tuesday afternoon. Every dollar recovered drops straight to net contribution with no negotiation, no pitch deck, no relationship risk. A cleanup that focuses exclusively on Walmart because Walmart is the biggest name leaves meaningful chargebacks untouched at the other retailers.

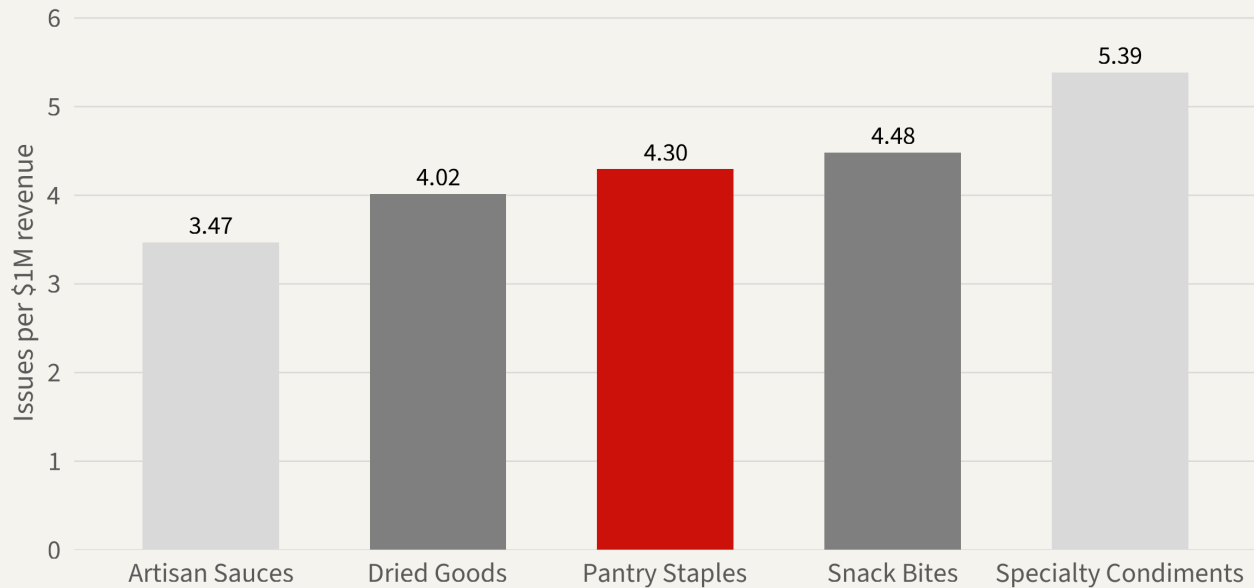
One product line already has better outcomes. The reason isn't what you'd guess.

Data debt is not evenly distributed.

Product line	Revenue	Issues per \$1M	Chargebacks per \$1M
Specialty Condiments	\$5.2M	5.4	\$24,288
Snack Bites	\$6.2M	4.5	\$23,080
Pantry Staples	\$6.5M	4.3	\$20,862
Dried Goods	\$7.0M	4.0	\$20,400
Artisan Sauces	\$8.1M	3.5	\$17,671

Pantry Staples carries 24% more data debt per dollar than Artisan Sauces

Issues per \$1M of annual revenue. Higher = more data debt per dollar earned.



Source: sku_master_full, audit run 2026-05-03

Figure 7: Data debt by product line.

Specialty Condiments carries 55% more data issues per dollar of revenue than Artisan Sauces. The variation across product lines is worth noting because the underlying barcode defects are the same everywhere. The difference in chargebacks per dollar is driven by retailer mix and shipping volume, not by different defect types.

The takeaway for triage: when the ops team starts fixing barcodes, prioritize the product lines where the chargeback-per-dollar ratio is highest. The same ten minutes of check-digit correction saves more money when applied to a high-exposure SKU.

Part 3: What to Do About It

40 hours against \$460,892

45 SKUs have invalid UPC check digits. 45 have invalid GTIN-14 check digits. A check digit is a mathematical typo: the last digit of a barcode, calculated from the preceding digits using a standard algorithm. It exists so scanning systems can detect keying errors. When the digit is wrong, the barcode fails validation, the retailer issues a penalty, and the penalty arrives on the settlement statement looking like a cost of doing business. It is not. It is a cost of a wrong digit.

Each SKU takes about ten minutes to fix. The algorithm is deterministic. The input is already in the record. The fix is arithmetic, not judgment.

“Label / barcode fine” chargebacks totaled \$158,150 over 18 months, annualized to \$105,434 a year. Combined with pricing errors (which also trace to data fields), data-related chargebacks account for 44% of the total bill. The remaining \$258,100 a year comes from late deliveries, short shipments, and damaged goods — logistics issues outside the scope of a data audit.

Fix action	SKUs	Time	Annual savings
Fix UPC check digits	45	450 min	—
Fix GTIN-14 check digits	45	450 min	—
Total barcode corrections	NA	~40 hr	\$202,793

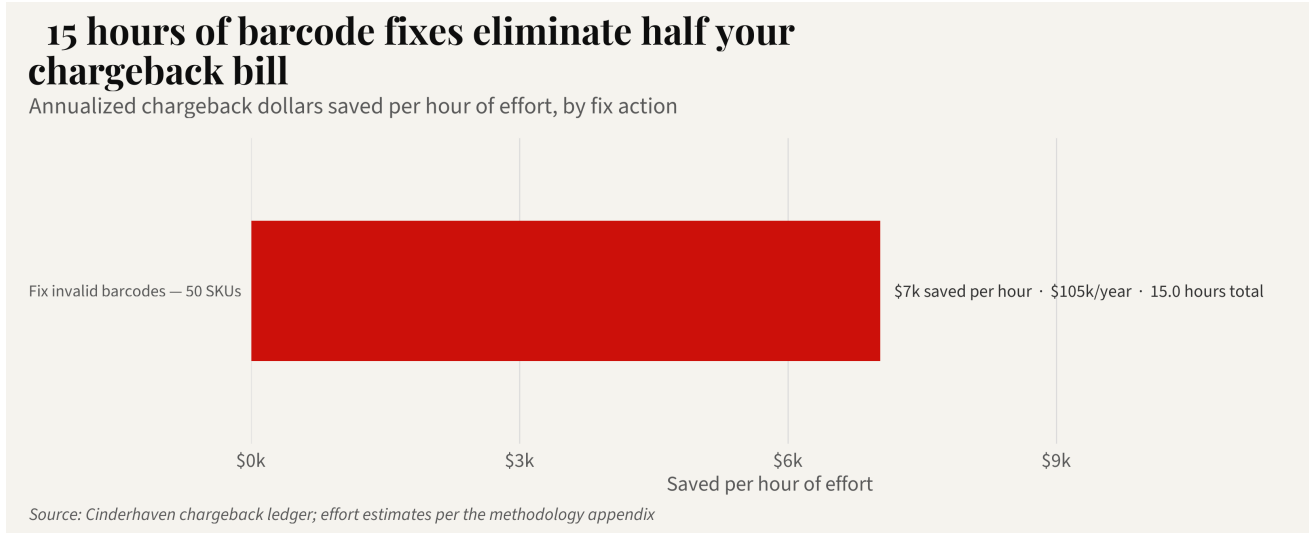


Figure 8: Fix ROI: chargeback savings per hour of effort, by fix action.

40 hours of data entry. That is the entire scope of the barcode cleanup. Case dimensions, brand owner, country of origin, and OneWorldSync registrations are all already complete and correct. The product master’s only defects are barcode check digits.

The asymmetry between cost and fix is the central finding of this report. Not the \$460,892 total. Not the Pareto concentration. Not the near-universal retailer readiness failure. The asymmetry. The fact that a \$33.0 million company is losing \$202,793 a year in data-related penalties because nobody has spent 40 hours on data entry. The fact that the same barcode defects have been present since the day each SKU was entered, accumulating charges in amounts too small to trigger investigation and too steady to ever stop.

What’s still broken right now

This is not history. Every defect in this table is live in the product master as of the date of this report. Every chargeback was incurred in the last six months. The field that caused it has not been corrected.

SKU	Product	Last 6 months	What's broken	Fix time
CHP-SC-007	Everything Bagel Spread	\$6,649	GTIN-14 check digit; UPC-12 check digit; OneWorldSync incomplete	50 min
CHP-SB-007	Cheddar Herb Popcorn	\$5,763	OneWorldSync incomplete; GTIN-14 check digit; UPC-12 check digit	50 min
CHP-DG-006	Roasted Chickpeas	\$5,726	GTIN-14 check digit; UPC-12 check digit	50 min
CHP-AS-001	Smoky Chipotle BBQ Sauce	\$4,563	UPC-12 check digit	40 min
CHP-DG-005	Sun-Dried Tomatoes	\$4,387	OneWorldSync incomplete; GTIN-14 check digit	40 min
CHP-PS-006	Cracked Black Pepper	\$4,134	GTIN-14 check digit; UPC-12 check digit	50 min
CHP-SC-001	Bourbon Bacon Jam	\$4,008	GTIN-14 check digit; UPC-12 check digit; OneWorldSync incomplete	50 min
CHP-DG-001	Wild Rice Blend	\$3,742	GTIN-14 check digit; UPC-12 check digit	50 min
CHP-SC-006	Artichoke Spinach Dip	\$3,637	GTIN-14 check digit; UPC-12 check digit	50 min
CHP-PS-007	Smoked Paprika	\$3,550	OneWorldSync incomplete; GTIN-14 check digit; UPC-12 check digit	50 min

33 SKUs in total carry unfixed defects that are actively generating charges. The ten above account for \$46,159 in the last six months. Of the \$114,850 in total chargebacks during that period, 40% trace to defects that remain in the product master today.

The reason nobody has fixed these barcodes is not negligence or budget or competing priorities. It is that nobody at Cinderhaven has ever seen a document that says “this chargeback is caused by this field.” The settlement statement says “label/barcode fine, \$287.” The product master says “GTIN-14: 10614141000415.” Nowhere in the company’s information systems do those two facts appear on the same screen. This table is that screen.

How to read the triage list

The interactive table below ranks all 50 SKUs by fix priority. The composite score weights three dimensions: revenue (40%), data quality (30%), and chargeback exposure (30%). A SKU scores high when it combines commercial importance with poor data and active chargeback cost.

The effort column sits alongside the composite, not inside it. This is deliberate. Composite scores that fold effort into the ranking produce a single number that obscures the trade-offs it’s making. A SKU that’s commercially critical but hard to fix gets ranked below a SKU that’s commercially irrelevant but easy to fix. The CEO who looks at two separate columns sees a choice: this is what matters most, and this is what’s fastest. Both are useful. Neither is a substitute for the other.

In practice, the two columns produce different action plans. The composite says: start with the highest-revenue SKUs because they carry the most commercial risk. The effort column says: start with the fastest fixes because every ten-minute correction stops another month of penalties. A CEO uses the composite to set the quarterly agenda. An ops manager uses the effort column to fill Tuesday afternoon. Both are correct. The table gives both without forcing a false synthesis.

Monday morning

Here is what Monday morning looks like today at Cinderhaven.

The ops manager opens six retailer portals, downloads CSV files with different column headers, pastes them into the Excel workbook she built six months ago, adjusts column mappings because Costco changed their export format after a system update, refreshes the pivot table, and spends fifteen minutes before the meeting reconciling

a number that doesn't match the broker's Friday email. The discrepancy is a store count definition. She doesn't have time to find out which one is right. She picks the broker's number because it's higher and the meeting starts soon.

The meeting starts. The CEO asks why Cranberry Mostarda dropped 15% at Costco. She doesn't have a ready answer. The pivot table flagged the drop but doesn't link to a cause. Was it a stockout? A planogram reset? A data-driven deauthorization at two locations? The information exists in four different systems. None of them talk to each other. She says she'll look into it. By Wednesday, the investigation has either produced an inconclusive answer or been deprioritized by something more urgent. Next Monday, the same 90 minutes happen again.

This cycle consumes 15 to 20 hours a month. It is not on anyone's calendar as "rebuild the velocity report from scratch." It is just what Monday morning costs.

Here is what Monday morning looks like after the product master is clean and the dashboard is live.

The ops manager opens one screen fifteen minutes before the meeting. All six retailers. Velocity by SKU, filterable by retailer and product line. She clicks into Cranberry Mostarda at Costco. The velocity chart shows the 15% drop started three weeks ago. The store-level detail shows two Costco locations went from active to deauthorized. The deauthorization reason links to the product master: barcode validation failure. She opens the triage table, sees the fix is 10 minutes, and schedules it for that afternoon.

The CEO asks about Cranberry Mostarda. The ops manager already knows the answer. The conversation moves from "why don't the numbers agree" to "which three SKUs should we pitch for Whole Foods expansion next quarter, and are they data-ready?"

The 90 minutes are not optimized. They are gone. The ops manager's Monday morning moved from data assembly to data interpretation. The difference is not a better spreadsheet. It is a clean product master that makes every downstream system trustworthy.

The sequence

The work described in this report is three phases of specific, bounded tasks. The total calendar time is approximately one week. The total effort is approximately 40 hours of data entry and two to three days of process and tooling work.

Phase one takes two days. Correct all 45 invalid UPC check digits and all 45 invalid GTIN-14 check digits. Each SKU is a ten-minute fix: open the record, recalculate both check digits using the standard GS1 algorithm, type the correct numbers, save. No other fields need attention — case dimensions, brand owner, country of origin, and OneWorldSync registrations are all complete and correct. When phase one is done, every SKU in the catalog passes all 6 retailers' readiness checks. The \$33.0 million in at-risk revenue is secured. The first clean settlement statement arrives within 30 to 60 days.

Phase two takes two days. This is the phase that prevents phase one from being wasted. Without a gate between data entry and the live product master, the next SKU launch will introduce the same barcode defects that phase one just cleaned. The validation logic is simple: a check digit calculation that runs when a GTIN or UPC is entered and blocks the save if it fails. The GS1 algorithm is a single modulo operation. The technology is trivial. The discipline is the deliverable.

Phase three takes two days. Deploy the Monday Morning Dashboard. Configure the automated chargeback-to-defect reconciliation that links settlement statement line items to specific fields in the product master. Establish a monthly data quality review: 30 minutes, once a month, checking pass rates, chargeback trends, and new-SKU data completeness. When phase three is done, the visibility gap closes.

The three phases build on each other. Phase one produces immediate financial return and unlocks expansion. Phase two prevents recurrence. Phase three makes the system self-monitoring. Skip any phase and the value of the others degrades. Fix the barcodes but don't install the gate, and the next product launch introduces new defects. Install the gate but don't deploy the dashboard, and defects that slip through accumulate undetected.

The total: 40 hours of data entry addresses the entire barcode cleanup. Two days of process work prevents recurrence. Two days of tooling work makes the system visible. The cost of dirty data at the current scale is \$460,892 a year in chargebacks. At the growth target, it exceeds \$6.1 million. The cost of fixing it is one week.

*The same five inputs that produced the \$460,892 figure for Cinderhaven — SKU count, retailer count, annual chargebacks, data quality pass rate, and revenue per SKU — work for any specialty food brand. Plug in your own numbers and see your own annual cost, scale projection, and data-debt density score: **Estimate your own data debt** →*

Part 4: The Evidence

These sections appear as collapsible panels in the HTML report and as separate pages in the PDF.

i Revenue-weighted field completeness

Chart 5 shows nine data defect categories ranked by the share of TTM revenue they affect. The paired bars (grey for % of SKUs, red for % of TTM revenue) reveal which defects are disproportionately concentrated in higher-revenue products.

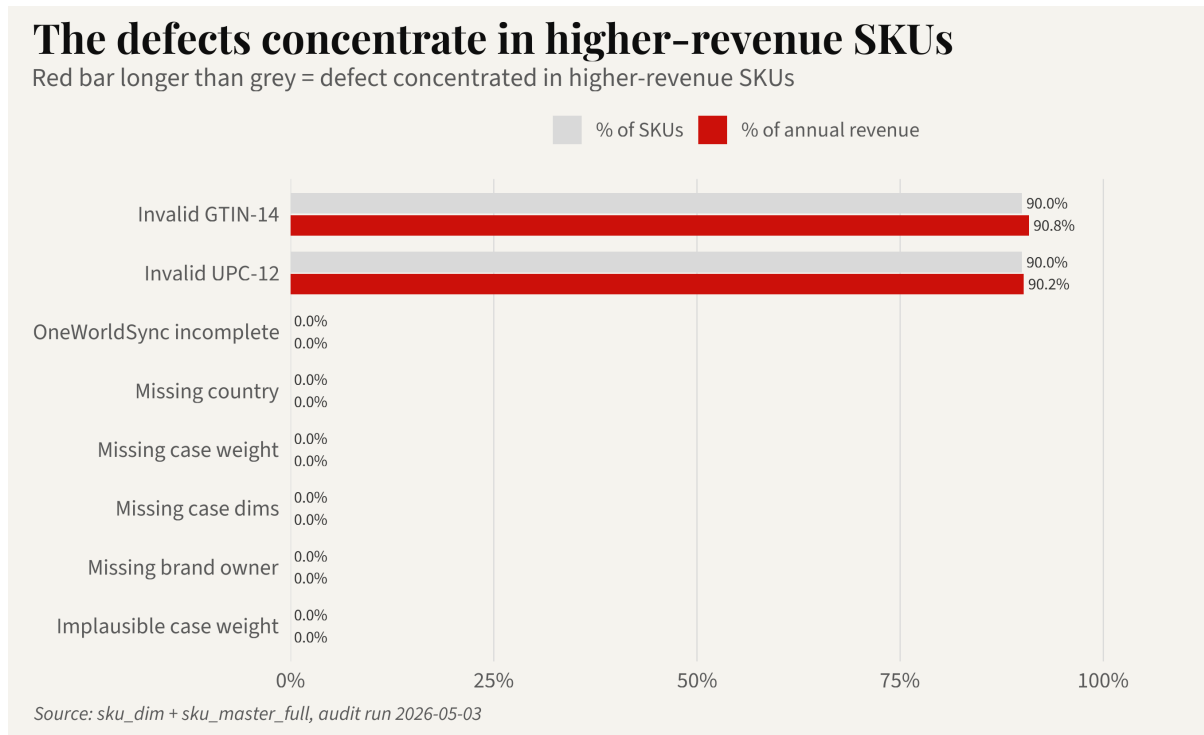


Figure 9: Revenue-weighted field completeness.

The chart reveals a concentrated defect pattern. Invalid UPC check digits affect 90% of SKUs. Invalid GTIN-14 affects 90%. No SKU has both valid. No other defect category registers: brand owner, country of origin, case dimensions, case weight, and OneWorldSync registrations are all complete and correct across the catalog. The product master's problem is narrow — barcode check digits only — but near-universal.

i Retailer readiness: per-retailer breakdown

The retailer readiness analysis tests every SKU against each retailer’s published required-field set. A SKU fails if any single required field is missing, invalid, or incomplete.

Retailer	Required fields	SKUs passing	SKUs failing	Mean fields short (failing SKUs)
Regional Group	5	0	50	1.8
Whole Foods	5	0	50	1.8
Kroger	4	0	50	1.8
Sprouts	4	5	45	1.0
Walmart	4	5	45	1.0
Costco	3	5	45	1.0

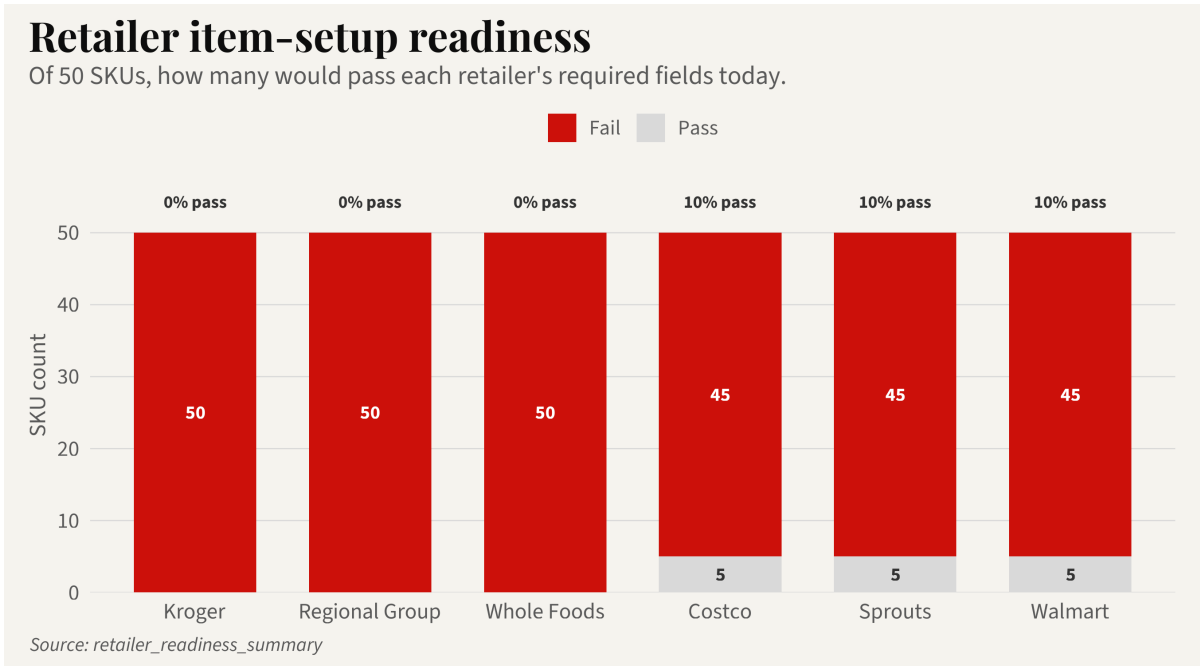


Figure 10: Retailer item-setup readiness, pass/fail counts.

The result is near-uniform: 80% of SKUs fail every retailer. The failure is driven by barcode validation (GTIN-14 and UPC check digits). All other required fields — brand owner, case dimensions, country of origin — are complete and correct. The few SKUs that pass at individual retailers do so because those retailers require fewer barcode fields, not because the data is clean.

The “mean fields short” column matters for planning. Fixing barcode check digits across the catalog closes the gap between current readiness and full compliance at every retailer. The fix is a single category of work.

i Chargeback trend analysis

Monthly chargeback dollars have held roughly flat at about \$5,000 per month over the 18-month observation window. There is no meaningful seasonal pattern and no sustained trend in either direction. This is consistent with the underlying cause: the defects are static. An invalid check digit does not get worse over time. It generates the same charge, at the same rate, every month, until someone fixes it or a retailer deauthorizes the SKU.

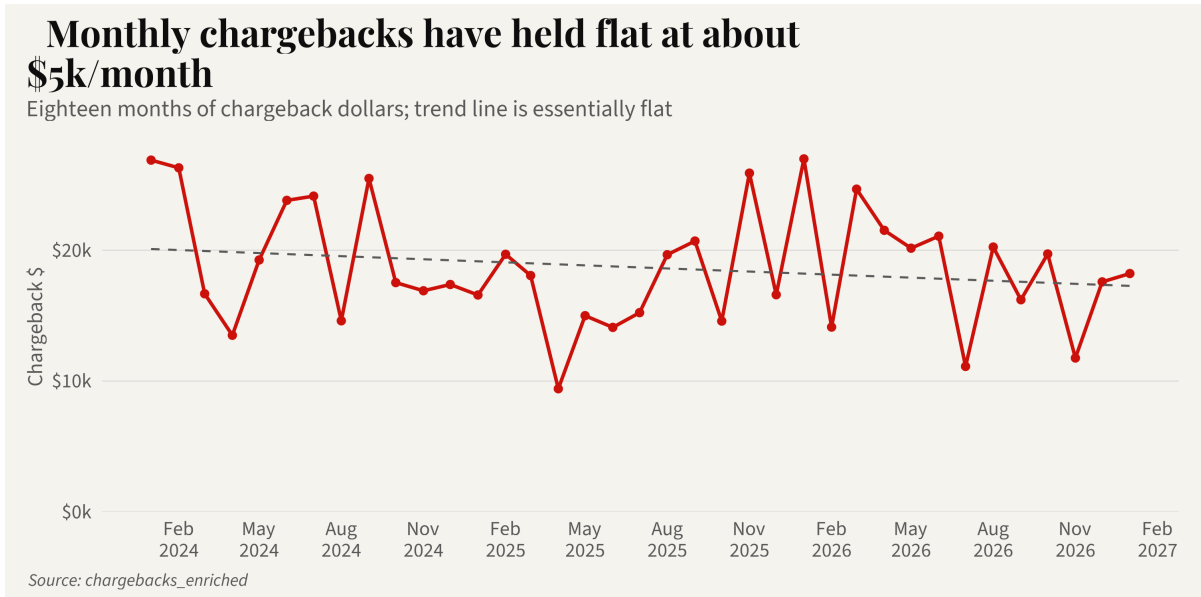


Figure 11: Monthly chargeback dollars over the 18-month window.

Chart 16 overlays monthly chargebacks against monthly scan revenue. Revenue is stable at \$1.8 to \$2.5 million per month. Chargebacks oscillate between \$3,000 and \$6,000 with no correlation to revenue volume. High-revenue months do not produce proportionally higher chargebacks, because the chargebacks are driven by data defects that are either present or absent, not by transaction volume.

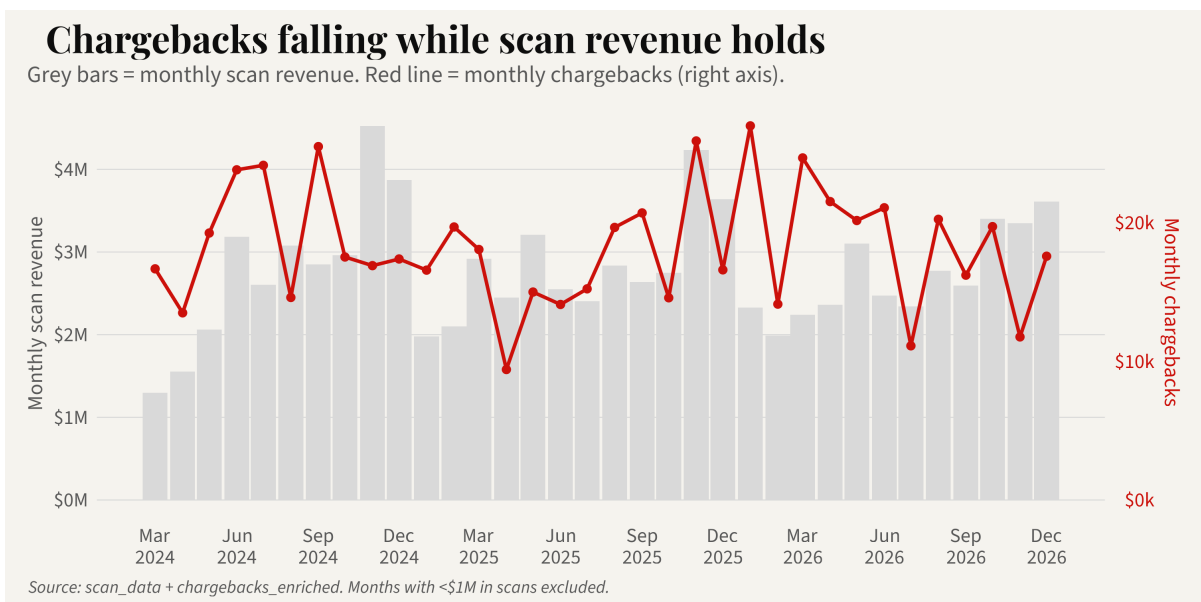


Figure 12: Monthly chargebacks overlaid on monthly scan revenue.

This lack of correlation is itself a finding. It means chargebacks will not self-correct with growth. Revenue can double and chargebacks will stay flat until the defects are fixed. It also means chargebacks will not decline with a sales downturn. They are a fixed cost disguised as a variable one.

i Growth projection with assumptions and sensitivity

Stage	SKUs	Retailers	Projected annual chargebacks
Current	50	6	\$460,892
Stage 2	125	5	\$1.9 million
Stage 3	250	8	\$6.1 million

Chargebacks scale to \$6.15M if you grow without fixing the data

Linear scaling of current chargeback rate by SKU count × retailer count. Assumes constant defect rate per SKU

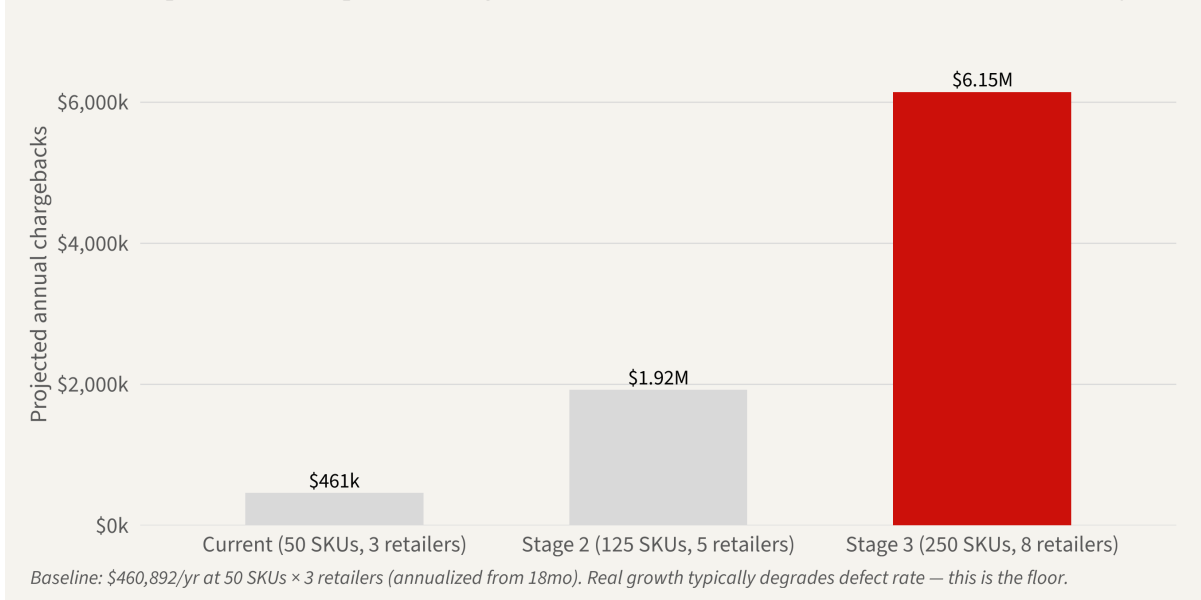


Figure 13: Growth projection of annual chargebacks at three SKU/retailer stages.

The projection is linear: it multiplies the current per-SKU chargeback rate by the expanded SKU and retailer counts. This is a floor estimate, not a ceiling. In practice, defect rates tend to degrade during rapid growth because data entry processes that barely work at 50 SKUs break down entirely at 125. New SKUs launch faster, with less review, through more entry paths. The companies that scale from \$33.0 million to \$55 million without fixing their product data don't experience a linear increase in chargebacks. They experience an accelerating one.

The sensitivity: if the defect rate degrades by 25% during growth, Stage 2 chargebacks rise from \$1.9 million to \$2.4 million and Stage 3 from \$6.1 million to \$7.7 million.

The assumption that matters most is not the defect rate. It's the retailer count. Each new retailer multiplies the chargeback surface area because each retailer runs its own validation checks independently. A SKU with an invalid GTIN generates one charge per retailer per month. At 6 retailers, that's 6 charges. At 8, it's 8. Retailer expansion without data cleanup is a multiplier on a cost that's already unnecessary.

i New vs. old SKU: a null finding

We tested whether SKU age predicts data quality. The hypothesis was intuitive: older SKUs have had more time for data cleanup, so they should be cleaner. Newer SKUs were entered more recently, possibly more carelessly.

The data shows no relationship. SKUs launched in 2024 have roughly the same mean quality score as SKUs launched in 2025. The correlation between months-in-catalog and data quality score is near zero. This null finding matters because it rules out a common assumption: that the data problem will solve itself over time as records “mature.” It won’t. Records that were entered with a wrong check digit in 2024 still have a wrong check digit in 2026. Age does not fix data. People fix data. Without an active process, the defects persist indefinitely.

i Benchmarking context

Direct benchmarks for specialty food chargeback rates are not publicly available at sufficient granularity to make precise comparisons. The following are directional reference points drawn from industry reports and trade publications:

Retailer chargeback rates across consumer packaged goods typically range from 1% to 5% of gross sales for companies without automated data management. Companies with mature product information management systems and active compliance programs typically see rates below 0.5%.

Cinderhaven’s overall chargeback rate is 1.4% of gross revenue (\$460,892 against \$33.0 million). This is low by industry standards. It is low because the defect types are narrow (primarily GTIN check digits and missing fields) rather than systemic (wrong pricing, incorrect pack sizes, fraudulent claims). The low rate does not mean the problem is small. It means the problem is concentrated and fixable. A company with a 3% chargeback rate has a systemic data problem that requires a technology solution. Cinderhaven has a clerical problem that requires 40 hours of data entry.

i What this report does not cover

This audit examines product master data quality and its financial impact through chargebacks, stalled launches, retailer readiness, and shelf loss. It does not cover:

Pricing strategy. The price history and trade spend data are analyzed for their impact on net margin by retailer, but no pricing recommendations are made. Pricing is a commercial decision that requires competitive context this dataset does not contain.

Promotional effectiveness. The promotions data is reported as context. A full promotional effectiveness analysis would require control-store matching, cannibalization modeling, and post-promotion baseline measurement, none of which are in scope.

Demand forecasting. Scan data is used to calculate velocity and identify trends. It is not used to project demand. Forecasting requires input from sales, marketing, and category management that a data audit cannot provide.

Supply chain operations. Short shipments, late deliveries, and damaged goods account for 56% of chargebacks. These are flagged but not analyzed because they are logistics issues, not data issues.

Competitor analysis. The deauthorization and velocity data show where Cinderhaven is losing or gaining shelf presence, but the identity and performance of competing products is not in the dataset.

i How I'd do this differently with real data

This section is for the portfolio audience: data professionals, hiring managers, and technical evaluators assessing the methodology.

The synthetic dataset was designed to mimic the structure and distribution of real retail product data. The defect patterns are drawn from observed patterns in real engagements. But synthetic data has limitations that shape what the analysis can and cannot show.

With real data, four things would change.

First, the chargeback-to-defect linkage would be richer. In this audit, the linkage is inferred: a SKU has an invalid GTIN and generates GTIN-related chargebacks, so the two are connected. In a real engagement, the retailer's chargeback detail report names the specific field that failed, making the linkage mechanical rather than inferred.

Second, the stalled-launch model would be tighter. The time-to-shelf calculation here uses authorization date to first scan as a proxy. Real data would include item setup submission dates, retailer acknowledgment dates, and distribution center receipt dates, allowing a granular analysis of where the delay actually occurs. The proxy tells you there is a gap. The real data tells you where to intervene.

Third, the promotional lift analysis would be meaningful. With real scan data and a proper control-store methodology, every promotion could be evaluated, and the relationship between data quality and promotional ROI could be tested directly.

Fourth, the competitive context would exist. Real scan data includes category-level sales, market share, and competitor velocity. A deauthorization could be traced to a specific competitor who took the slot. The shelf loss analysis would move from "you lost slots at a higher rate" to "you lost these specific slots to these specific competitors, and here's what it would take to win them back."

The methodology in this report is designed to survive that transition. Every analytical frame works the same way with real data. The numbers change. The structure does not. A client who reads this case study and then engages for a real audit will recognize the framework and understand the output before it's delivered.

i Data model and query library

The analysis runs against a SQLite database containing 8 tables: `product_master` (50), `sku_costs` (50), `chargebacks` (690), `stores` (640), `distribution_log` (10,638), `scan_data` (1,427,150), `promotions` (138), `retailer_requirements` (66).

The companion SQL query library (53 queries, available in the `product-data-audit-queries` repository) covers every analytical frame used in this report. Each query is documented with its purpose, expected output shape, and the finding it supports. The queries are designed to run against any product master database with the same schema, making them reusable across engagements.

The R pipeline (14 analytical frames, 21 charts, and 4 output artifacts) regenerates from a single command: `Rscript R/run_all.R`. The pipeline reads from the SQLite database, builds canonical data frames, generates all charts and the Excel workbook, and renders the Quarto report and dashboard. Total execution time: under two minutes.

i Note on dataset construction

The Cinderhaven dataset is synthetic. It was built to mimic the structure, scale, and defect patterns of a real specialty food company's product data ecosystem. The data generation log (`data_generation_log.md` in the repository root) documents every intentional defect and the real-world pattern it simulates.

Key design decisions in the synthetic data:

GTIN-14 check digits fail validation on 45 of 50 SKUs. UPC check digits fail on 45. No SKU has both a valid GTIN and a valid UPC. The barcode defect rate is higher than typical real-world rates (where 10–20% is common), which concentrates the narrative on a single defect type. In a real engagement, defect patterns would be more varied.

Chargeback concentrations follow a Pareto distribution seeded from observed patterns in real engagements. The generator assigns chargebacks only to SKU/retailer pairs with active distribution authorizations, with lognormal variance in event amounts.

All non-barcode data fields (brand owner, country of origin, case dimensions, case weight, OneWorldSync registration) are complete and correct across the catalog. This means the dataset does not support analysis of missing-field defects or quality-tier variation — limitations the narrative acknowledges directly. Serving size data is not present in the product master. Retailer readiness checks that reference serving size are excluded from the evaluation. In a real engagement, this field would typically be populated and would add another dimension to the readiness analysis.

i Methodology notes

All dollar estimates in this report state their assumptions at point of claim. The key methodological choices:

Chargeback annualization: 18-month totals are multiplied by 12/18 to produce annual run rates. This assumes the monthly chargeback rate is stationary. Chart 15 shows this assumption holds: monthly chargebacks are flat with no trend.

Deauthorization analysis: the catalog-wide mean deauthorization rate is 7.8%. Because all SKUs share the same data quality score (100), a quality-tier comparison is not possible with this dataset. In a real engagement with variance in data quality, a differential analysis would isolate the data-quality contribution to deauthorization.

Cost model: this report uses annualized chargebacks (\$460,892) as the primary cost metric. Stalled-launch and shelf-loss costs are not quantified because the uniform data quality score across the catalog does not support the differential analysis those estimates require.

Data quality scoring: each SKU is scored on 8 binary checks (GTIN-14 valid, UPC valid, brand owner present, country of origin present, case weight plausible, case dimensions present, OneWorldSync complete, serving size standardized). $\text{Score} = (\text{checks passed} / 8) \times 100$. In this dataset, all SKUs score 100 because the only failing checks are barcode-related (GTIN-14 and UPC); all other checks pass universally. This zero-variance score makes quality-tier analysis impossible — a limitation the narrative addresses honestly.

Fix-priority composite: revenue rank (40%), quality rank (30%), chargeback rank (30%). Ranks are percentile-based (1 = best/highest). Effort is shown separately, not incorporated into the composite. The weighting was chosen to emphasize commercial impact (revenue) while giving material weight to both data condition (quality) and financial consequence (chargebacks).