

A Catalog of Feed-Integrity & Cross-Grain Validation Checks for Retail Visitation Data Methodology, Fail-Closed Verdicts, and a Retrospective Validation Battery

Individ Analytics – Research
identities anonymized (L### labels, Z### zones), volumes relative

September 24, 2026

Abstract

We present a machine-usable catalog of fourteen data-integrity and cross-grain validation checks (FI-01..06, XG-01..02, SV-01..04, MN-01..02) for retail visitation pipelines, together with a fail-closed verdict scheme and a retrospective validation battery run over a frozen fleet snapshot. Each check is grounded in a real incident the platform has already experienced, so the catalog reflects demonstrated failure modes rather than hypothetical risk. We report per-check firing surfaces and rates over the current fleet, and an honest verdict per check — including the checks whose incident surfaces were remediated by the deployed serving fix (a positive outcome, not a round-up).

1 Introduction and framing

Retail visitation data flows through feed, cross-grain, and serving layers before it reaches a report or decision. A defect at any layer — a no-signal feed, a calendar-day hole in one grain, a dropout in dwell classification, a flat fabricated forecast served from an archived artifact — can silently propagate into an analyst’s number. This paper provides a *generalizable* methodology: a fixed catalog of checks, each with an explicit machine-evaluable trigger condition, a fail-closed behavior, an owner, and an evidence pointer; a deterministic replay over a frozen snapshot; and a retrospective validation that checks fire on their recorded incidents.

The motivating incidents are five recorded failures that span three layers (feed integrity, cross-grain consistency, serving/output) plus monitoring hygiene; every check in the catalog is grounded in one of them. The contribution of this draft is the checks themselves, the fail-closed verdict scheme they embody, and the validation battery as the paper’s evidence base.

2 The check-catalog methodology

Each catalog entry (check-catalog/checks.yaml) is a machine-usable record: check-id, layer, severity, trigger condition (with explicit windows, tolerances, thresholds), scope, owner, fail-closed behavior, evidence pointer, and current coverage. The catalog also carries an explicit *negative-results* section so that aspects that admit no additional machine check are recorded as such, rather than over-engineered.

Expressing data-quality expectations as declarative, verifiable checks and automating their evaluation has a strong precedent in large-scale data processing [1]; the catalog in this paper is an instance of that approach, specialized to a retail visitation pipeline and validated retrospectively against real incidents.

Two conventions define the operating semantics:

- **Fail-closed verdicts.** When a check fires, the consumer must not silently continue on the suspect value: it refuses to train/serve/report the suspect grain or surface, or downgrades to a conservative stated value with an explicit flag. This extends the existing forecast-quality guards (which already refuse to serve flat or inactive-activation artifacts) to data feeds.
- **Honest negative results.** A check that cannot be retrospectively validated records NOT-MEANINGFUL /PENDING with the reason; partial detections report the fired subset enumerated, never rounded up.

Verdict class	Checks	n	Meaning
fires	FI-01, FI-02, FI-03, FI-04, XG-01, MN-01	6	the check fires on its recorded incident surface, replayed against the frozen snapshot
fires (partial)	SV-03	1	fires on a subset only; the fired subset is quantified, never rounded up
exposed	FI-05	1	no independent incident replay is possible; the auditable exposed population is identified instead
resolved / deploy-verified	SV-01, SV-02	2	the incident was remediated by the deployed fix; the check’s post-condition now holds (a positive outcome, not a round-up)
not-meaningful	MN-02	1	no numeric retrospective replay is possible; recorded as an honest negative with its reason
pending	FI-06	1	no artifact exists yet to validate against; recorded as pending, not rounded up
eval	XG-02, SV-04	2	candidate concerns surfaced for the platform teams to confirm, not re-derived here

Table 1: Legend of the fail-closed verdict classes used throughout. Class membership and the per-class check count (n) are summed from the checks’ recorded verdicts in the machine-readable results artifact (2); the gloss states each class’s semantics as defined in the catalog.

3 The validation battery: what we ran, and on what

We froze a fleet snapshot (the unchanged data files archived with this paper, whose manifest records window bounds, row counts, sha256 hashes, and the as-of timestamp 2026-09-23) using SELECT-only reads against the production backend, and ran the automatable checks over it. The frozen snapshot guarantees deterministic replay: re-running the runner against the unchanged snapshot reproduces the machine-readable results file byte-identically (its sha256 prefix is `dc2917eb54c00d69...`).¹

¹The run’s machine-readable results artifact — the per-check results file, the frozen snapshot, and its manifest — is archived alongside this paper; every number and figure in this draft is derived programmatically from that artifact, so the text cannot drift from the measured results.

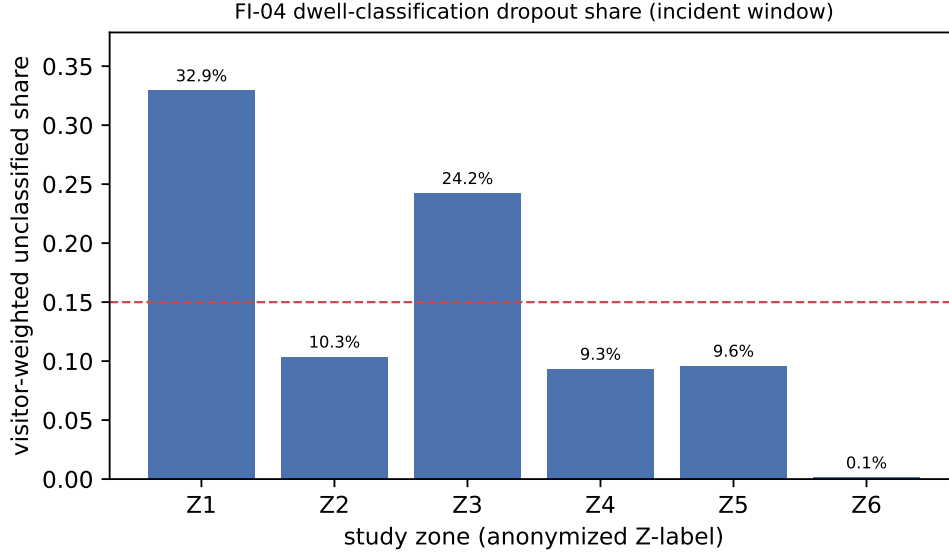


Figure 1: FI-04 visitor-weighted dwell-classification dropout share per study zone over the incident window (values read from the run’s machine-readable results artifact; the dashed line is the 15% threshold). Anonymized Z-labels (Z1, Z2, ..., Z6) denote the six study zones in ascending zone-id order; bar labels repeat the share used in the text, rounded to the same one decimal place.

3.1 Feed-integrity checks

FI-01 (all-zero / no-signal feed). Over a 30-day window the raw condition is a non-archived location whose observed visits are maximally zero. On the frozen snapshot the raw condition fires on 36 non-archived locations, including the recorded all-zero pair (L035, L031). 36 of the fired locations are already expected-no-data-annotated in the model store (the trainer’s fail-closed stamp) and 0 are not — in this snapshot every fired surface is already annotated, so the genuinely unannotated defect case is empty. The 1:1 annotation cross-reference is a positive corroboration of the fail-closed machinery and the precursor to the FI-06 registry. **Validation: fires** (raw condition) on the recorded incident.

FI-02 (feed staleness). A non-archived location whose most recent observed row is older than two days fires. On the snapshot this fires on 33 locations, including the recorded stale-feed pair. **Validation: fires.**

FI-03 (missing calendar day, HOUR grain). Where the day-grain authority reports trade (visitors > 0) on a calendar day but the HOUR grain has zero rows that day, the check fires. This is evaluated fleet-wide over all 6 study zones, which together recorded 88 missing trade days in the incident window; at the recorded defect zone (Z1) it reproduces **13 missing Sundays** (whole calendar days absent from the HOUR grain) in the incident window, matching the recorded 13-day surface, of which 12 fall inside the HOUR coverage window; a distinct in-coverage Mon–Fri gap (of 5 non-Sunday trade days) is an extra finding. 7 pre-coverage boundary days are excluded from the completeness claim. **Validation: fires** on the recorded 13-Sunday surface.

FI-04 (dwell-classification dropout). The visitor-weighted share of rows where visitors > 0 but no dwell classification was captured. On the incident window the study zones show dropout shares of 32.9% (the defect zone, Z1), 24.2% (the milder defect zone, Z3), and 10.3% (a healthy control, Z2) — the elevated share reproduced at the defect zones. The firing set is exactly the

Retrospective-validation verdict per check

Check	Retrospective verdict class
FI-01	fires (replay on incident)
FI-02	fires (replay on incident)
FI-03	fires (replay on incident)
FI-04	fires (replay on incident)
FI-05	exposed (no replay)
FI-06	pending
XG-01	fires (replay on incident)
XG-02	eval
SV-01	resolved / deploy-verified
SV-02	resolved / deploy-verified
SV-03	fires (partial)
SV-04	eval
MN-01	fires (replay on incident)
MN-02	not-meaningful



Figure 2: Per-check retrospective-validation verdict class (fires / partial / resolved / pending / not-meaningful / eval), read from the run’s machine-readable results artifact.

two defect zones (Z1 and Z3), which clear the 15% threshold; the healthy controls sit at or just below it, the highest one measuring 10.3%. *Basis disclosure:* these shares are measured on the day-grain unclassified-visitor column over the full incident window; the recorded incident figures were measured on a different basis (the API dwell classification over the dropout hours). The battery therefore validates *elevation* at the defect zones (vs. healthy controls at most $\sim 10.4\%$), not the exact recorded percentages. **Validation:** fires on the defect zones.

FI-05 (training-sufficiency gate exposure). We expose the below-gate (no-active-model) population as an auditable skip result: 18 entities have no active model, reproducing the trainer’s documented minimum-data gate (50 non-zero days). **Validation:** exposed-skip population identified; the full recorded incident surface falls within it, but no independent incident replay is possible, so this is reported as exposure rather than a firing.

FI-06 (expected-no-data registry). PENDING: the registry enrichment that would materialize a machine-readable expected-no-data table has not yet been delivered, so no such table exists to validate against. Recorded as pending, not rounded up.

3.2 Cross-grain checks

XG-01 (DAY-vs-HOUR reconciliation). The same surface as FI-03 seen as reconciliation: at the defect zone (Z1) the DAY authority reports trade on days where the HOUR grain sums to zero. **Validation:** fires on the recorded incident surface.

XG-02 (zone-sum vs location-total invariant). Comparing the same window (both grains), 94 multi-zone locations were checked against the documented invariant that zone-sum exceeds the location total; 5 locations showed the location total exceeding its zone sum (an `eval` verdict — candidate invariant violations surfaced for the platform teams to confirm, not re-derived here).

3.3 Serving/output checks

SV-01 (serve-only-active-model). Resolved / deploy-verified. The deployed serving-worker image, verified as of this snapshot to contain the active-only lookup fix in its lineage, remediated the incident surface: the five flat-serving locations have **0** served rows, and the recorded all-zero pair has no served rows in the post-fix era (the 264 rows still present in the snapshot were all generated 2026-09-05..2026-09-07, before the fix deployed; fail-closed “no model”). Recorded as a positive resolution (deploy-verified: zero post-fix served rows), not a numeric firing.

SV-02 (flat-serving guard). Resolved / deploy-verified. No location in the current served surface exhibits the recorded flat-serving defect, whose constant sat far above a typical store’s daily output; the current served-surface scan returns 9 flat-signature candidates (low-volume stores), none at the incident scale. The recorded flat-serving class is gone from the served fleet.

SV-03 (served-version attribution). Fires (partial). Across 168120 served rows, 60.6% carry `model_version` = NULL; version-based guards and rollback attribution cannot operate on those rows. The NULL class persists fleet-wide (was ~all-NULl at the recorded serving incident), so the check fires, though coverage has improved post-fix.

SV-04 (forecast-quality guard suite extension). eval: the flat-serving and related quality gates run over the served surface as the feed-extension of the existing guard family.

3.4 Monitoring / alert hygiene

Alerting on an unbounded error metric rather than on the characteristic its rule claims to control is a statistical-process-control concern; the classical remedy is a control chart on the intended characteristic [2].

MN-01 (alert-rule metric alignment). Fires. The live *SMAPE*-named alert rules threshold the unbounded `model_quality_mape` column (the recorded metric mis-aim), still present in the snapshot’s rule copy: a named-SMAPE rule must not fire on MAPE values.

MN-02 (alert provenance). NOT-MEANINGFUL: the recorded 99.94%/100% figures are unrecoverable from any live surface; the durable constructive product is the no-gauge-fails-closed behavior and the provenance guard, not a numeric replay.

4 Negative results / honesty

The catalog records that (i) attributing a detected no-signal store (expected-no-data vs defect) is a human decision, not a machine check; (ii) which grain to quote when HOUR is defective is policy (the check is XG-01, the policy is documentation); (iii) the exact 99.94/100 alert numbers are unrecoverable; (iv) no incident here motivates a “correlated fleet retrain” data-integrity check. These are stated so the catalog does not over-engineer.

5 Threats to validity and scope of the validation

The battery is *retrospective and in-sample*: every check was designed from the same recorded incidents it is then validated against, so the results demonstrate that the checks *replay/cover* the

known incidents, not that they can predict unseen ones. There is no held-out incident set, and the false-positive rate is unmeasured by design. The closest thing to a negative control is the FI-01 case, where the fired surfaces are already expected-no-data-annotated (no unannotated false fires); that is reported as such and should not be read as a blind-detection result. Accordingly the validation is offered as operational coverage and replay evidence — sufficient to catch a regression of the recorded failure modes — not as a blind predictive benchmark. Blind/hold-out validation and false-positive-rate measurement are explicitly out of scope for this study and are candidates for follow-up work, not claims made here.

6 Discussion and adoption path

6.1 Adoption path

These checks are intended to be consumed in the analysis layer rather than run as a one-off study. In the MCP analysis layer (the platform component that fields report and answer requests), the checks can run at pre-flight, before an answer is computed, so an answer built on suspect data is refused or explicitly flagged rather than silently served; and at post-flight, so the fired-check verdicts are attached to the produced result and the reader can see which data-integrity conditions were assessed and how each resolved. This matches the catalog’s fail-closed semantics: when a check fires, the consumer must not silently continue on the suspect value. Consistent with the catalog’s honesty rule, attribution of a detected no-signal store (expected-no-data intent versus an unknown defect) remains a human decision at the registry; the automated checks flag a surface and require an annotation, they never auto-classify.

7 Validation summary

Validated: FI-01, FI-02, FI-03, FI-04, XG-01 (replay on recorded incident), MN-01 (live mis-aim reproduced). Exposed, no incident replay: FI-05. Remediated (positive resolution; checks’ post-conditions now hold): SV-01, SV-02. Partial: SV-03 (NULL class persists). Pending: FI-06. Not-meaningful: MN-02. Eval: XG-02, SV-04. This is the evidence base: each automated check either fires on its recorded incident, reconciles honestly why it cannot, or is recorded as exposure/remediation with its basis stated.

References

- [1] S. Schelter, D. Lange, P. Schmidt, M. Celikel, F. Biessmann, and A. Grafberger, “Automating large-scale data quality verification,” *Proceedings of the VLDB Endowment*, vol. 11, no. 12, pp. 1781–1794, 2018. Verified via Crossref DOI 10.14778/3229863.3229867.
- [2] W. H. Woodall and D. C. Montgomery, “Research issues and ideas in statistical process control,” *Journal of Quality Technology*, vol. 31, no. 4, pp. 376–386, 1999. Verified via Crossref DOI 10.1080/00224065.1999.11979944.