



Automating the Closed File Audit

A deterministic, rules-driven pipeline that replaces a manual insurance-claim audit selection

Gunek Singh CHADDHA (21012537)

Advised by Dr. Desmond TSOI

Industrial Host: HSBC Life

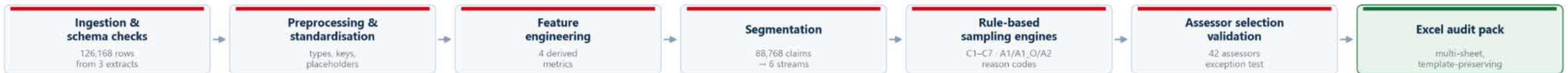


SYSTEM ARCHITECTURE & DATA PIPELINE

Python pandas NumPy openpyxl Streamlit Jupyter Excel

PRESENTATION LAYER · Streamlit

upload 5 inputs → single Run action → status panel & previews → download audit pack | core logic in a Jupyter notebook



THE PROBLEM

- Selection applied by hand in spreadsheets across three schemes and two claim types.
- About 1.5 business days of analyst effort every month.
- Not reproducible — the same extract could yield a different sample.
- No record of why any individual claim was selected.

OBJECTIVES

- ✓ A reproducible pipeline, run monthly on real production extracts
- ✓ An explicit reason code attached to every selected claim
- ✓ Automatic detection of assessor-coverage exceptions
- ✓ An interface the audit team operates without touching code

VERIFICATION & EVIDENCE

31 / 31

unit tests pass

5 / 5

reconciliation checks

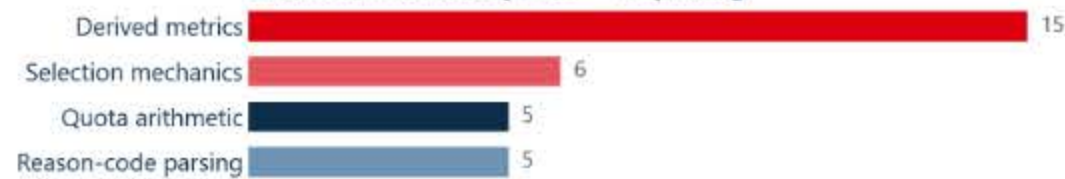
0

coverage exceptions

88,747 + 21,500 + 15,921 = 126,168

scheme rows reconcile exactly

Executed unit tests by area — all passing



- Deterministic by construction: one explicit rounding rule and a stable sort remove every source of run-to-run variation. The three assessors with no selected claim each met no criterion, so no exception arose.

KEY TECHNICAL CONTRIBUTIONS

Business-day metrics with defensive sentinels

Turnaround measured on a working-day calendar. A malformed date returns a sentinel rather than aborting a 126k-row run.

Window marking over a stable sort

One claim spans several rows; a stable sort restores order so every row of a selected claim is swept as one unit.

Quota allocation with roll-over

A 2% budget gives 15% to C1 and splits 80/10/10 across C5–C7; unfilled specialised quota rolls into materiality.

Tie-inclusive materiality cut-off

All claims at or above the Nth-largest paid amount are taken, so the sample cannot depend on the order of ties.

IMPACT & CONCLUSION

BEFORE

~1.5 business days

manual, rationale unrecorded



AFTER

minutes

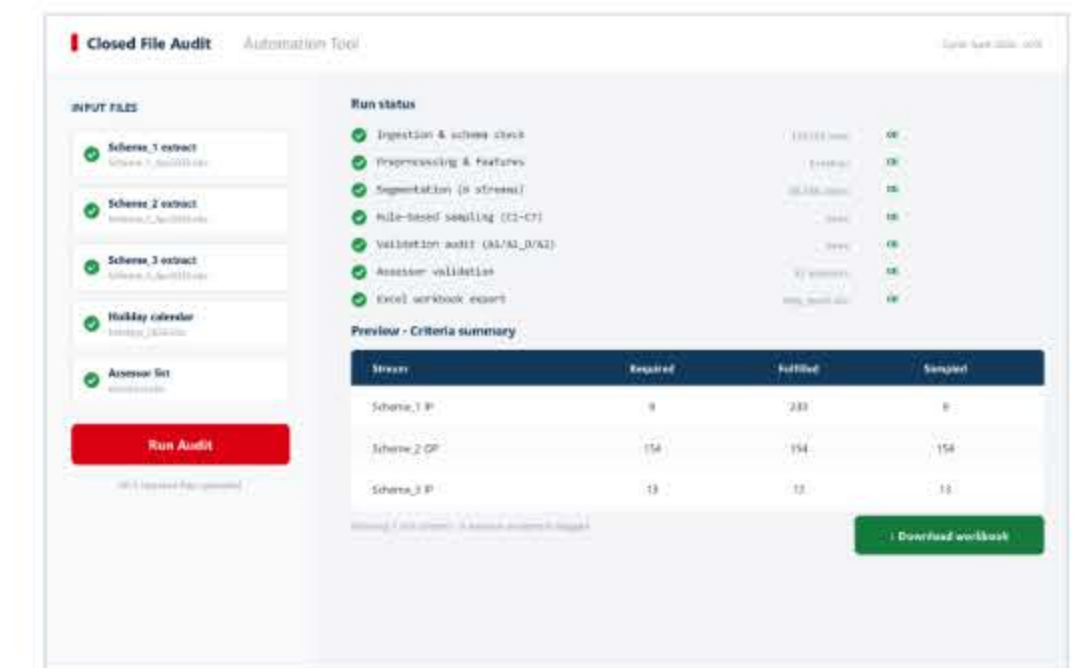
reason code on every claim

The Closed File Audit selection is now an automated control rather than a manual exercise. It runs in minutes on a real production population, reconciles exactly to the ingested extract, and attaches a defensible reason code to every sampled claim. No assessor met a selection criterion without being sampled, so the cycle raised zero coverage exceptions and the effort saved is defensible rather than merely faster.

FUTURE WORK

- Expose thresholds through the interface with change controls and an audit log.
- Retain results across cycles to review trends and repeat selections.
- Compare every cycle against the retained baseline automatically.

THE DELIVERED SYSTEM & APRIL 2026 RESULTS



The operational front end. Five monthly inputs are uploaded, one action runs the whole pipeline, a status panel confirms every stage, preview tables show the criteria summary and the assessor check, and the finished audit workbook is downloaded.

126,168

claim rows ingested

88,768

unique claims

6

scheme × type streams

42

assessors validated

39

with selected claims

1.5 d → min

monthly effort

Unique claims per scheme × claim-type stream

