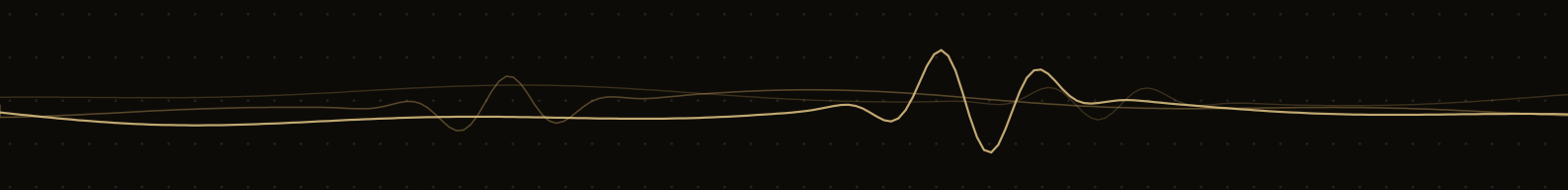


CASE STUDY · REAL-ESTATE MARKET INTELLIGENCE

4,900+ properties a day, structured.

Rent intelligence from the open web — built and operated by a two-to-three-engineer AI-native pod.

CLIENT: NORTHLINE ANALYTICS *



4,900+

PROPERTIES PROCESSED DAILY, IN
PRODUCTION

[97.x]%

7-DAY ROLLING SCRAPE SUCCESS RATE
[CONFIRM]

[9X]%

UNIT-LEVEL FIELD ACCURACY VS HAND-
LABELED EVALS [CONFIRM]

The client & the challenge



ABOUT NORTHLINE ANALYTICS

Northline Analytics provides rent and availability intelligence to institutional owners and operators across the US multifamily market. Its customers reprice thousands of units a week against competitor rents — decisions that are only as good as the freshness and accuracy of the data underneath them.

THE CHALLENGE

Northline needed unit-level rents, availability, and concessions from thousands of property websites, refreshed daily. The sources are hostile: dozens of property-management platforms, aggressive bot protection, layouts that change without notice, and no APIs on offer.

The prior approach — periodic snapshots assembled by a mix of off-the-shelf scrapers and manual checks — delivered stale, partially-verified data. Pure-LLM extraction pilots read every page convincingly but at a cost per record that broke the unit economics, with accuracy nobody could actually state. The business question was blunt: can this be done daily, at platform scale, with accuracy expressed as a measured number?

AT A GLANCE

CLIENT

Northline Analytics *

INDUSTRY

Real-estate market intelligence

SYSTEM

Rent intelligence from the open web

POD

2 engineers

STATUS

In production, daily

ENGAGEMENT

Sprint → Build → Run



Proptech

DATA PLATFORM



2 engineers

SURGEX LABS DELIVERY



In production, daily

SURGEX LABS DELIVERY

* Client name changed. Engagement details anonymized under NDA; detailed numbers and reference calls available on request.

The solution

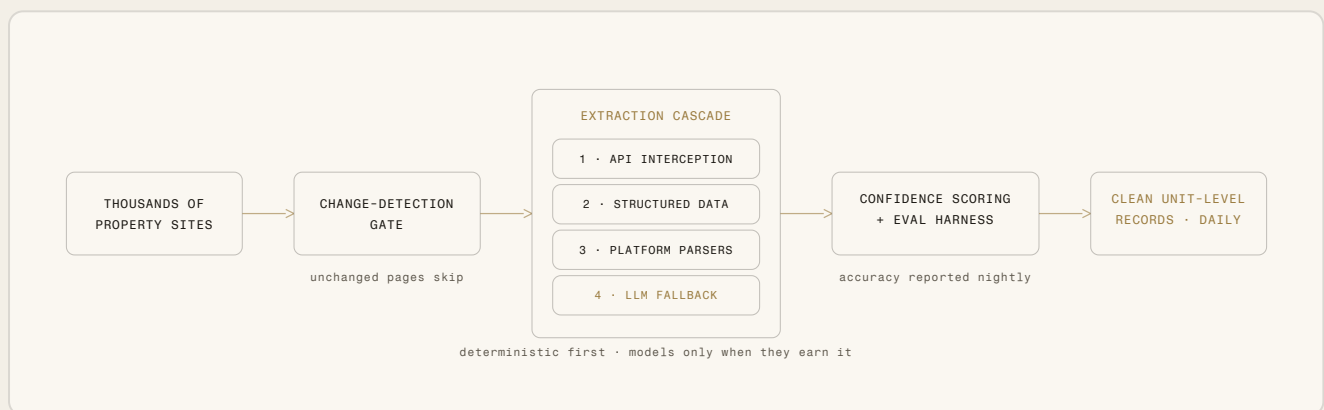


SurgeX Labs built a tiered extraction platform on one principle: deterministic first, model calls only where they earn their cost. A change-detection gate decides whether a page needs scraping at all. Pages that do flow down a cascade — intercepted APIs where they exist, published structured data where present, platform-specific parsers for the major property-management systems, and LLM extraction only as the fallback of last resort.

Every record carries per-field confidence and is traceable to its source page. A hand-labeled evaluation set is replayed continuously, so accuracy is reported nightly rather than asserted. An append-only audit log and day-over-day unit diffing make every data point in the feed explainable after the fact.

The system was built and is operated by a two-engineer pod, with the first production slice live in two weeks and new site platforms absorbed in weekly increments since.

THE TIERED EXTRACTION PIPELINE



HOW IT'S BUILT

- Playwright fleet with change-detection gating (most pages skip a full scrape most days)
- Tiered extraction: API interception → JSON-LD → DOM templates → LLM fallback
- Per-field confidence scoring; eval harness against hand-labeled ground truth
- Append-only audit log; state store diffs units day-over-day

The results



4,900+

PROPERTIES PROCESSED DAILY, IN PRODUCTION

[97.x]%

7-DAY ROLLING SCRAPE SUCCESS RATE [CONFIRM]

[9X]%

UNIT-LEVEL FIELD ACCURACY VS HAND-LABELED EVALS
[CONFIRM]

[7X]%

OF EXTRACTIONS RESOLVED WITHOUT AN LLM CALL
[CONFIRM]

2

ENGINEERS, END TO END — BUILD AND OPERATIONS

“

[Pull quote pending client approval — e.g. "We replaced a weekly snapshot we didn't fully trust with a daily feed we can audit line by line."]

[NAME], [TITLE], NORTHLINE ANALYTICS

Tell us the system you need. We'll tell you the week it ships.

hello@surgexlabs.ai
surgexlabs.ai