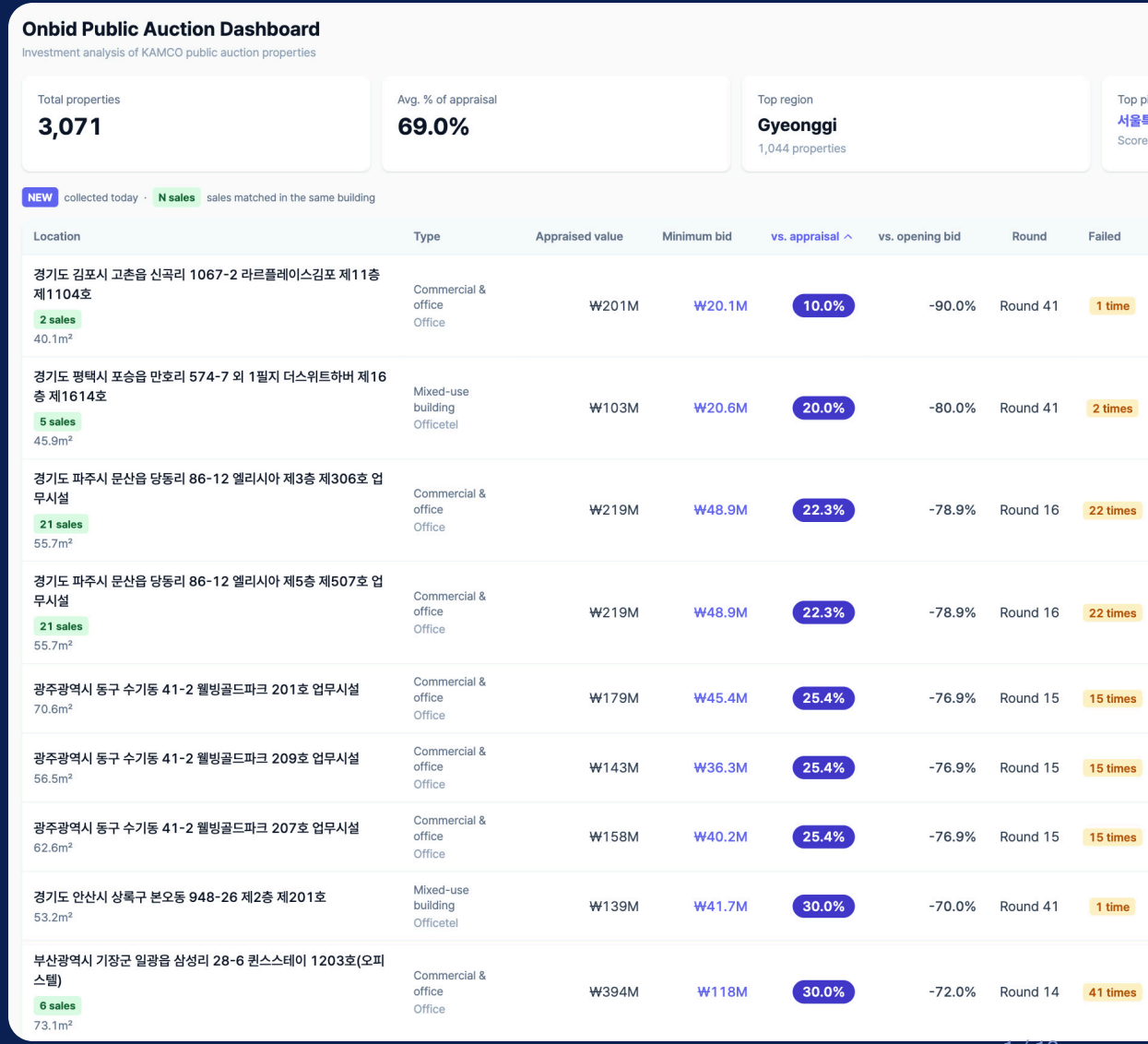
DATA ENGINEERING CASE STUDY

Public Auction Data Pipeline

Every morning it finds the government auction listings that are cheap compared with real sale prices.

Seonmin Kim
Sole developer, Global Vision Co., Ltd.
Jan to Jul 2026



Onbid Public Auction Dashboard

Investment analysis of KAMCO public auction properties

Total properties	Avg. % of appraisal	Top region	Top pi
3,071	69.0%	Gyeonggi 1,044 properties	서울특 Score

NEW collected today · N sales sales matched in the same building

Location	Type	Appraised value	Minimum bid	vs. appraisal	vs. opening bid	Round	Failed
경기도 김포시 고촌읍 신곡리 1067-2 라르플레이스김포 제11층 제1104호 2 sales 40.1m²	Commercial & office Office	₩201M	₩20.1M	10.0%	-90.0%	Round 41	1 time
경기도 평택시 포승읍 만호리 574-7 외 1필지 더스위트하버 제16층 제1614호 5 sales 45.9m²	Mixed-use building Officetel	₩103M	₩20.6M	20.0%	-80.0%	Round 41	2 times
경기도 파주시 문산읍 당동리 86-12 엘리시아 제3층 제306호 업무시설 21 sales 55.7m²	Commercial & office Office	₩219M	₩48.9M	22.3%	-78.9%	Round 16	22 times
경기도 파주시 문산읍 당동리 86-12 엘리시아 제5층 제507호 업무시설 21 sales 55.7m²	Commercial & office Office	₩219M	₩48.9M	22.3%	-78.9%	Round 16	22 times
광주광역시 동구 수기동 41-2 웰빙골드파크 201호 업무시설 70.6m²	Commercial & office Office	₩179M	₩45.4M	25.4%	-76.9%	Round 15	15 times
광주광역시 동구 수기동 41-2 웰빙골드파크 209호 업무시설 56.5m²	Commercial & office Office	₩143M	₩36.3M	25.4%	-76.9%	Round 15	15 times
광주광역시 동구 수기동 41-2 웰빙골드파크 207호 업무시설 62.6m²	Commercial & office Office	₩158M	₩40.2M	25.4%	-76.9%	Round 15	15 times
경기도 안산시 상록구 본오동 948-26 제2층 제201호 53.2m²	Mixed-use building Officetel	₩139M	₩41.7M	30.0%	-70.0%	Round 41	1 time
부산광역시 기장군 일광읍 삼성리 28-6 퀸스스테이 1203호(오피스텔) 6 sales 73.1m²	Commercial & office Office	₩394M	₩118M	30.0%	-72.0%	Round 14	41 times

THE PROBLEM

It started with a 3-hour manual search, every day

BEFORE

3 hours

a day. One person went through the auction site by hand to find listings that met the company's conditions: discount, failed rounds, region, private sale.

AFTER

0 hours

The pipeline runs overnight. The same conditions are now filters, and every listing is ranked against real sale prices.

3,000+
listings refreshed daily

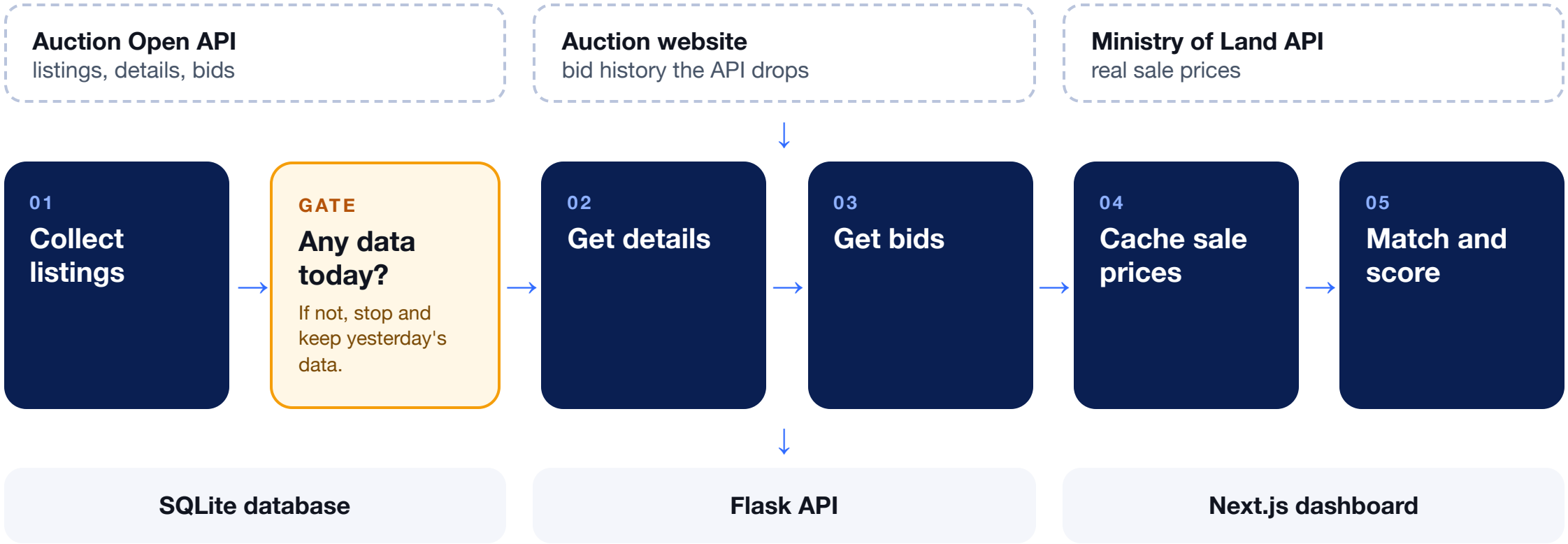
490K+
real sale prices to compare against

5
people on the real estate team used it

208 / 210
days the daily run completed

HOW IT WORKS

Five stages and one gate



Safe to rerun. Every write is an upsert, so a failed run can simply start again. Scheduled daily with GitHub Actions, packaged with Docker Compose.

WHAT WAS HARD

Three problems I had to solve

0

bad days overwrote good data

Government APIs sometimes return an empty answer that looks like success. The gate stops the run before anything is replaced.

17K

deleted bid records recovered

The API drops a listing's bid history once it moves to private sale. A second collector reads it from the public web page. That is 57% of all history.

4

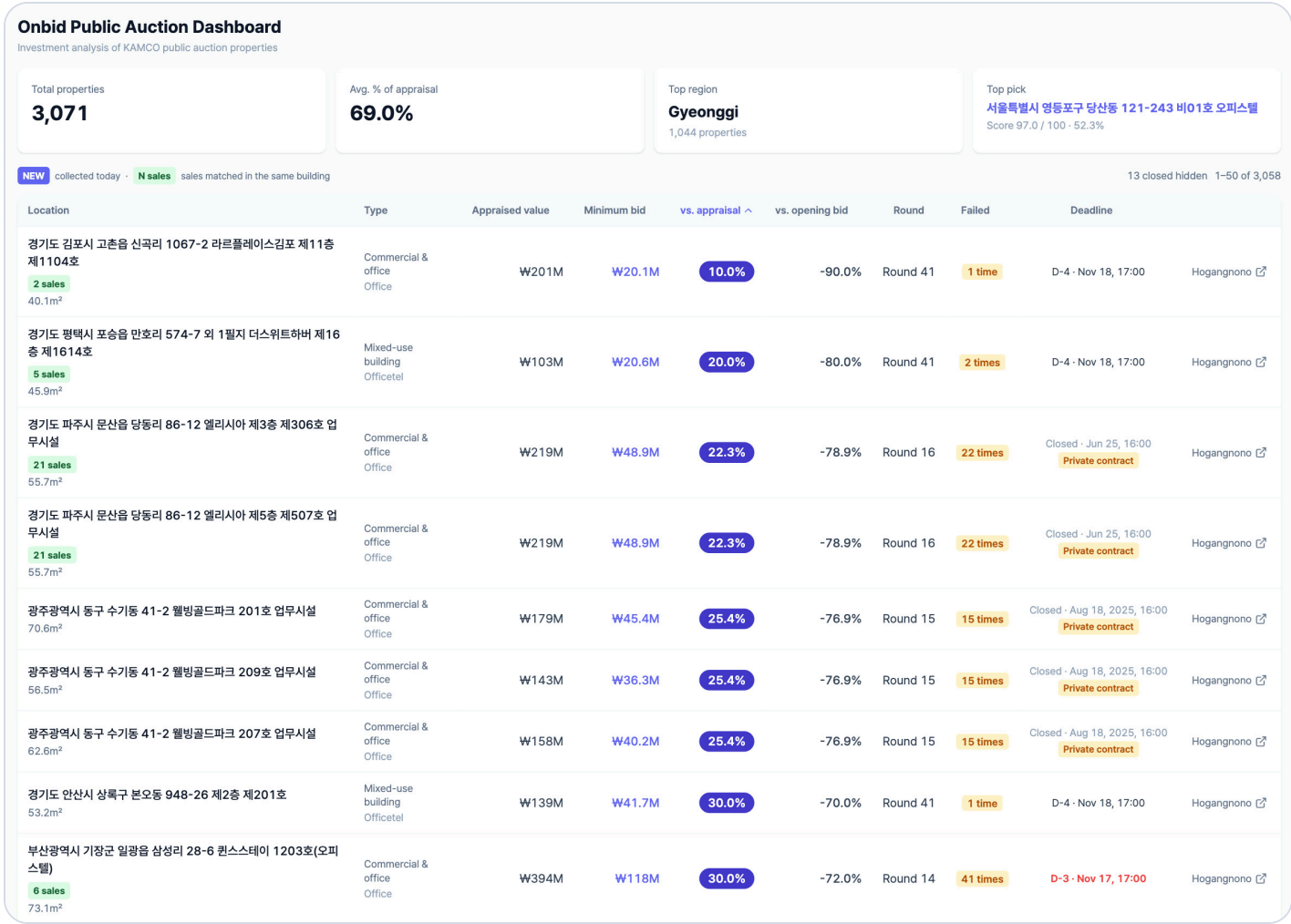
levels of price matching

Each listing is matched to real sales, from the same lot number down to the same district. The level is saved, so users see how sure the price is.

FEATURE 1 OF 8

A ranked table of every open listing

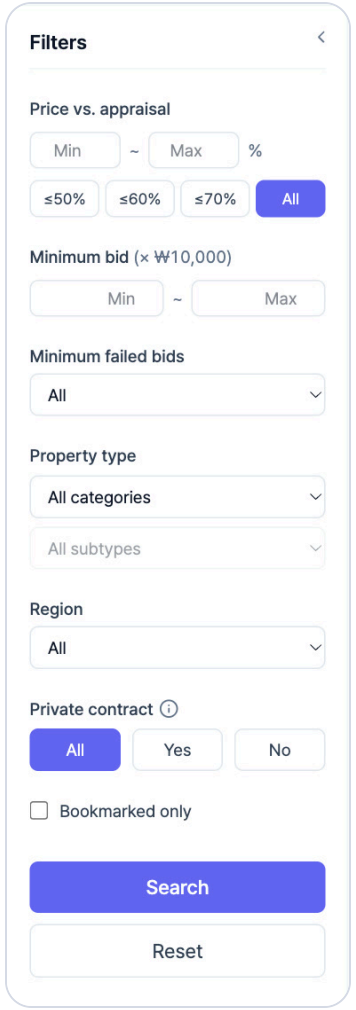
- **Summary strip:** total properties, average % of appraisal, top region, top pick
- **Sort** by price, discount, failed rounds, or deadline
- **Color** marks how far below appraisal each listing is
- **Badges** show matched real sales and days left
- **Columns** can be dragged into any order, and the order is remembered



FEATURE 2 OF 8

The manual checklist became filters

- Every condition the team used to check by hand is one control
- Filters combine, so a 3-hour search is a few clicks
- Reset returns to the full list

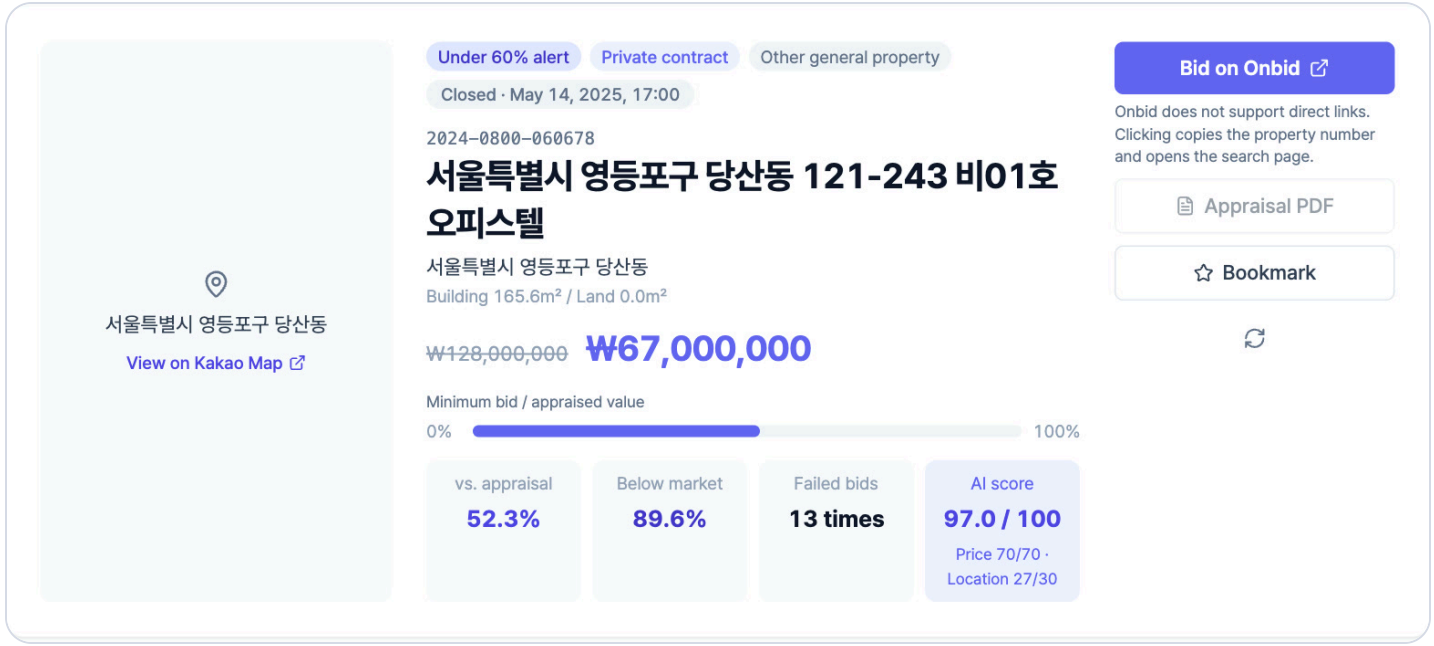


WHAT THE TEAM LOOKED FOR	FILTER
Cheap compared with appraisal	**Price-to-appraisal range**
Within budget	**Starting bid range**
Nobody else wanted it	**Minimum failed rounds**
The right kind of property	**Type, two levels**
The right place	**Region**
Can be bought by negotiation	**Private contract: yes or no**
Already on the shortlist	**Bookmarked only**

FEATURE 3 OF 8

One listing at a glance

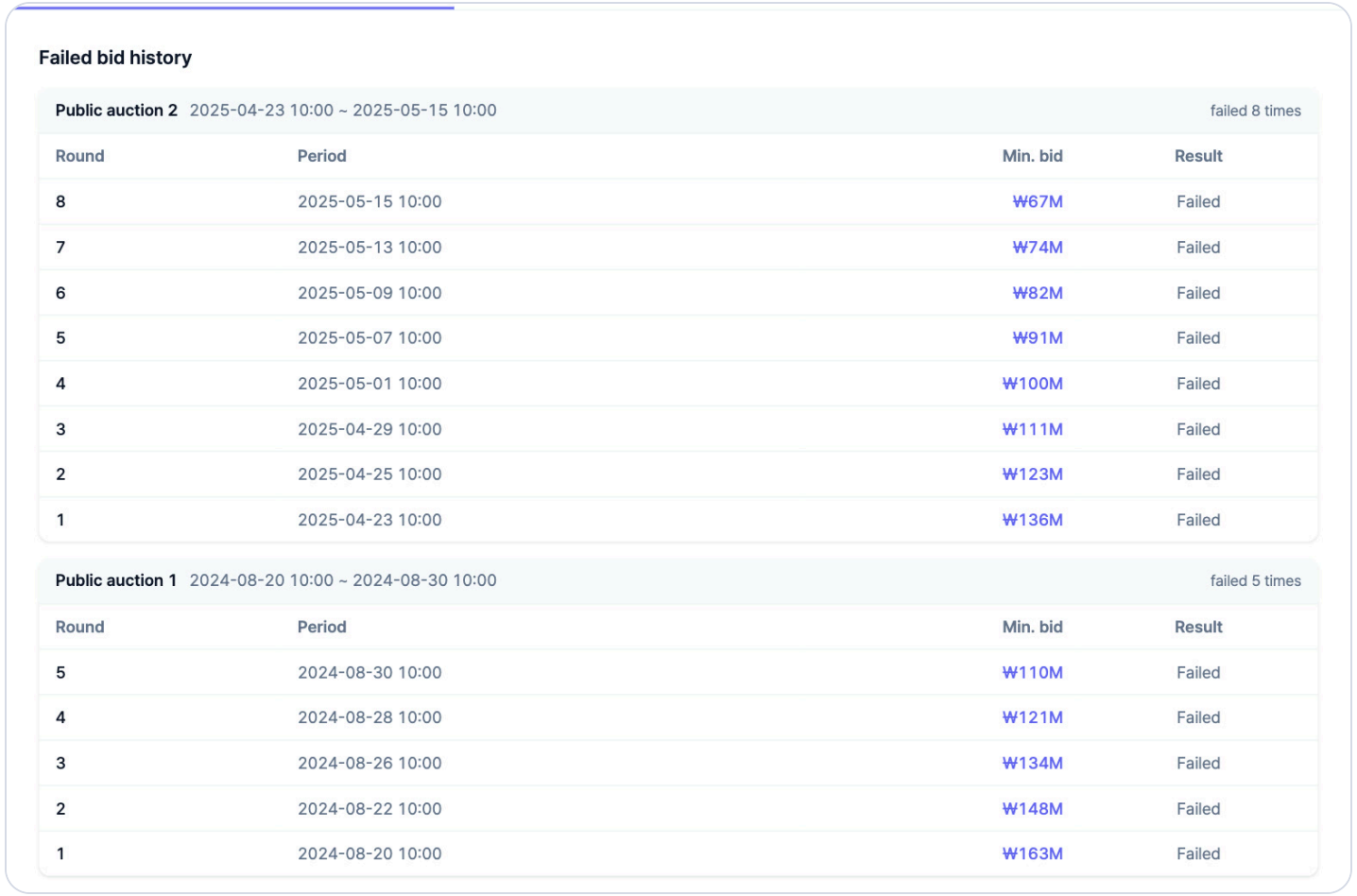
- **Alert tag** when the price falls under 60% of appraisal
- **Price bar:** current bid against the appraised value
- **Four numbers:** price vs appraisal, below market, failed bids, score
- **Score breakdown:** price points and location points shown separately
- **Buttons:** open the official page, appraisal report, bookmark, refresh this listing now



FEATURE 4 OF 8

Every failed round, kept

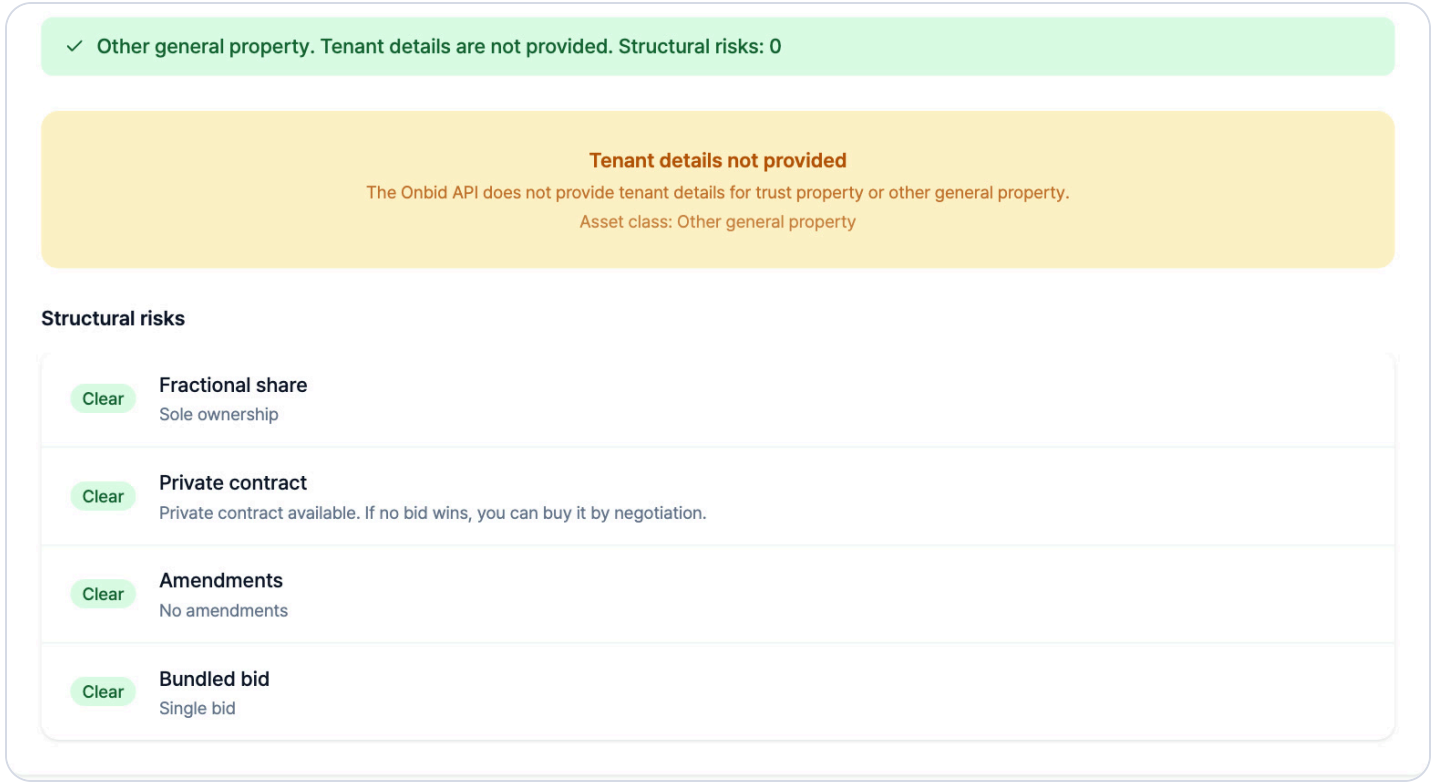
- **Full bid history** grouped by auction, newest first
- **Each round:** date, starting bid, result
- **The price falling** round by round is the signal the team looked for
- This history is what the official API deletes. The pipeline **recovered 17K records** so it stays visible



FEATURE 5 OF 8

Rights and risk check

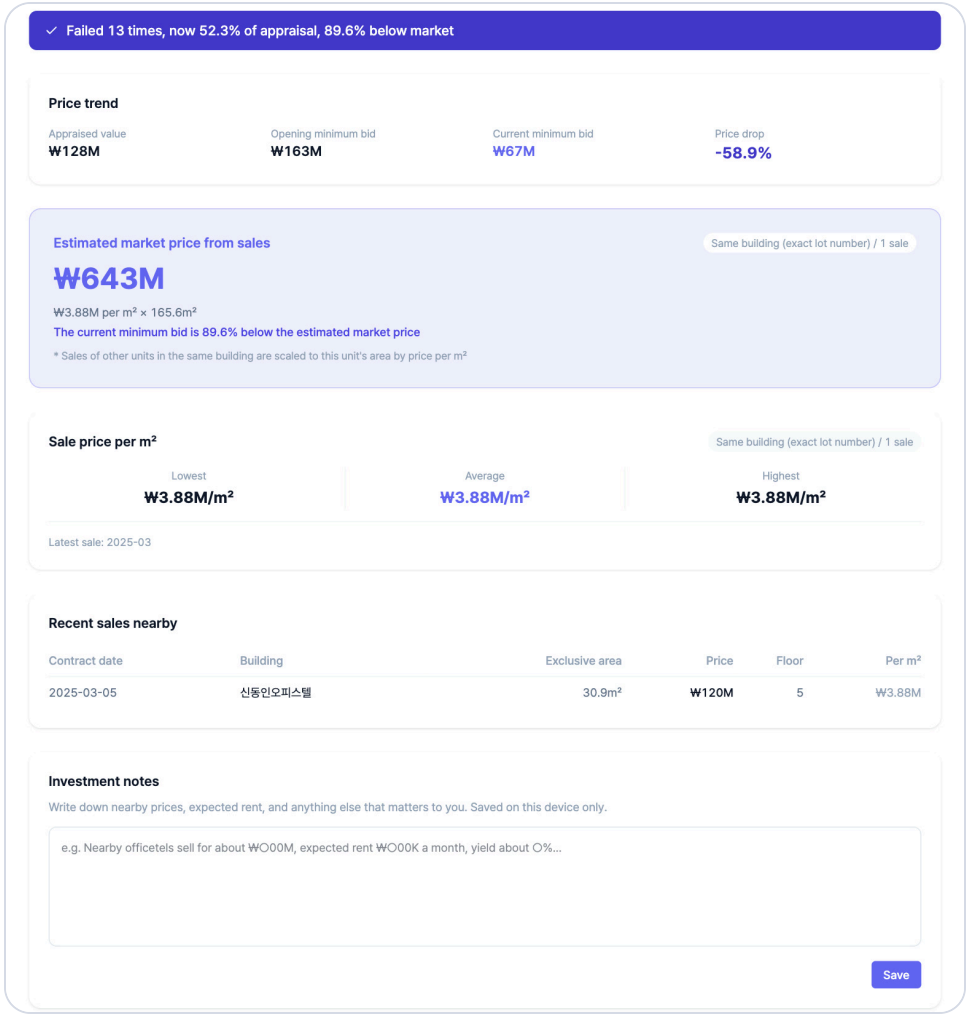
- **Tenants:** listed when the source provides them, and clearly marked when it does not
- **Fractional share:** whole property or only a share
- **Private contract:** whether it can be bought by negotiation after a failed auction
- **Amendments:** whether the notice was changed
- **Bundled bidding:** sold alone or as a package



FEATURE 6 OF 8

What is it really worth?

- **Estimated market value** from real sales of the same building
- **Price history:** appraisal, first bid, current bid, total drop
- **Unit price** low, average, and high per square meter
- **Nearby sales** listed with date, size, and price
- **Match level** shown on the card, so users know how sure it is
- **Notes** field for the reviewer's own judgment



FEATURE 7 OF 8

What the site looks like

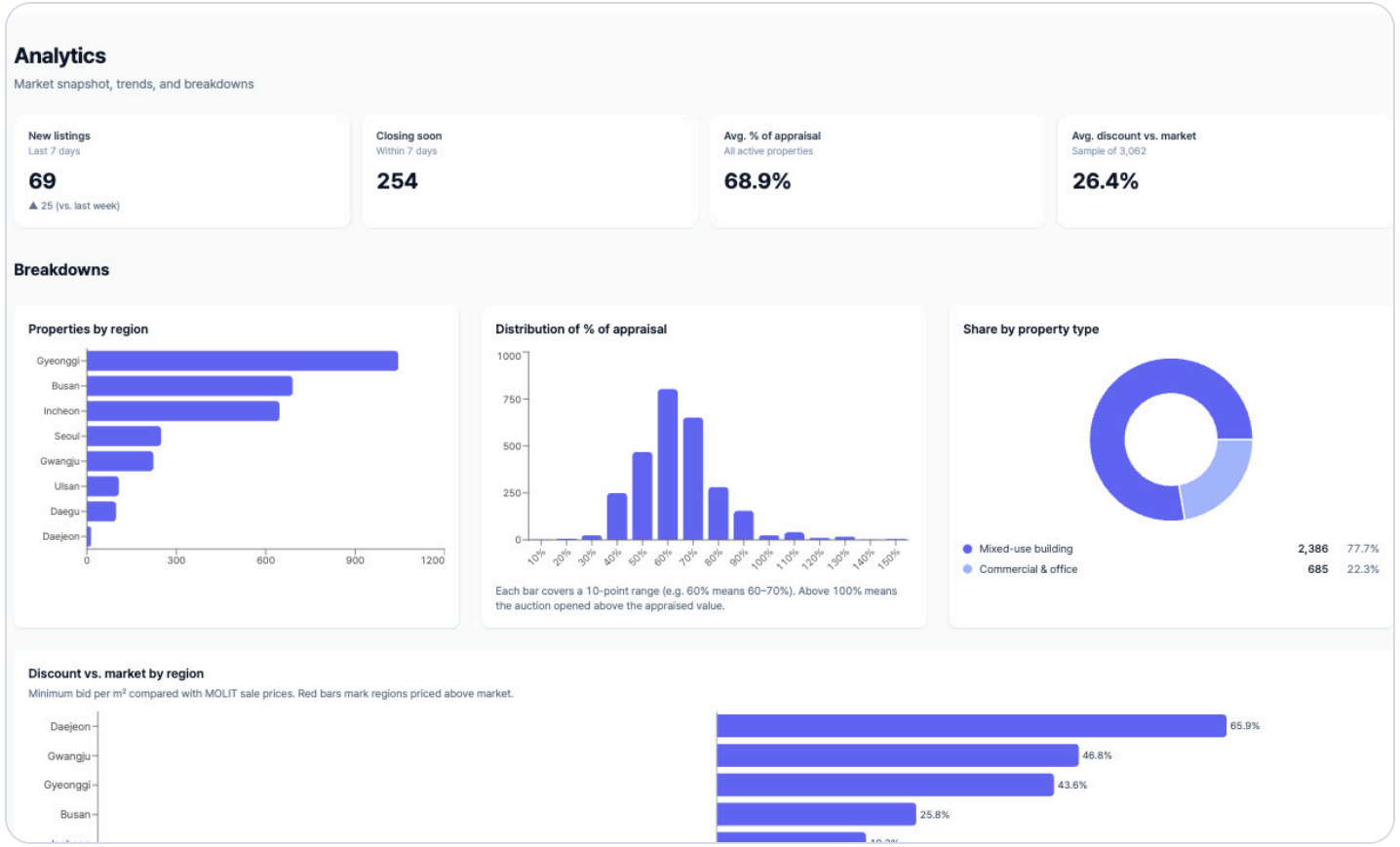
- **Photo** from the official notice
- **Location and use** as described by the seller
- **Key facts:** address, type, building and land area, bid period
- A reminder to confirm against the original notice



FEATURE 8 OF 8

The whole market on one page

- **New listings** this week against last week, and how many close within 7 days
- **Average discount** against appraisal and against market
- **By region:** where the listings are
- **Distribution** of price-to-appraisal ratios
- **Discount by region:** where auctions run cheapest against real prices
- Region and type filters apply to every chart



BEHIND THE SCREENS

What keeps it running

Daily schedule

GitHub Actions runs the five stages every morning and redeploys the demo with fresh data.

API

A Flask API with 15 endpoints serves listings, detail, bid history, market prices, and analytics. One endpoint re-pulls a single listing on demand.

Packaging

Docker Compose starts the pipeline, API, dashboard, and web server together.

Tests and measurement

Lot-number parsing has unit tests. A script measures how often listings match real sales, so matching quality is a number.

Python · SQL · SQLite · Flask · Next.js · Docker Compose · GitHub Actions

Live demo: **onbid-demo.onrender.com**

Code: **github.com/nimnusmik/onbid-auction-pipeline**