

What 129 Real Cape Town Listings Actually Say About the Market

129 real listings scraped from Property24, Seeff and Pam Golding — every price, suburb and bedroom count is genuine, not synthetic. Modelled into a star-schema warehouse, mapped, scored with machine learning, and reconciled to the same numbers across Excel and this page.

Scope note: this is a small (129-listing), single-city, upscale-skewed sample — Sea Point, Camps Bay, Clifton and Bantry Bay lead the suburb counts. It's a portfolio demonstration of a real scrape-to-warehouse-to-insight pipeline, not a representative national property index.

129

real listings (sale + rent)

80

for sale · avg R42,328,847

49

for rent · avg R42,926/mo

45

distinct suburbs

1215

image URLs captured

1 · Why sale and rent can never share an average

Property24, Seeff and Pam Golding each specialise: Property24's 24 listings here are 100% rentals, Pam Golding's 58 are 100% sales, and Seeff splits roughly down the middle (25 rent, 22 sale). Blend their prices into one "average listing price" and you'd get a number close to R23M that describes neither market honestly — sale prices average R42,328,847, a scale about a thousand times larger than the R42,926/month rental average. Every table and chart below keeps them apart.

A real bug, caught before publishing: an early pass at querying this data in the cloud used naive comma-delimited parsing. Listing titles contain literal commas ("Industrial Property to rent in Montague Gardens, Cape Town"), which silently shifted columns on affected rows — the sale-market average came back R19M too high, with no error raised. Fixed by switching to quote-aware CSV parsing and reconciling every number against the source build before publishing anything here. Full writeup in the repo's [infra/aws_evidence.md](#).

2 · Sale market vs rental market

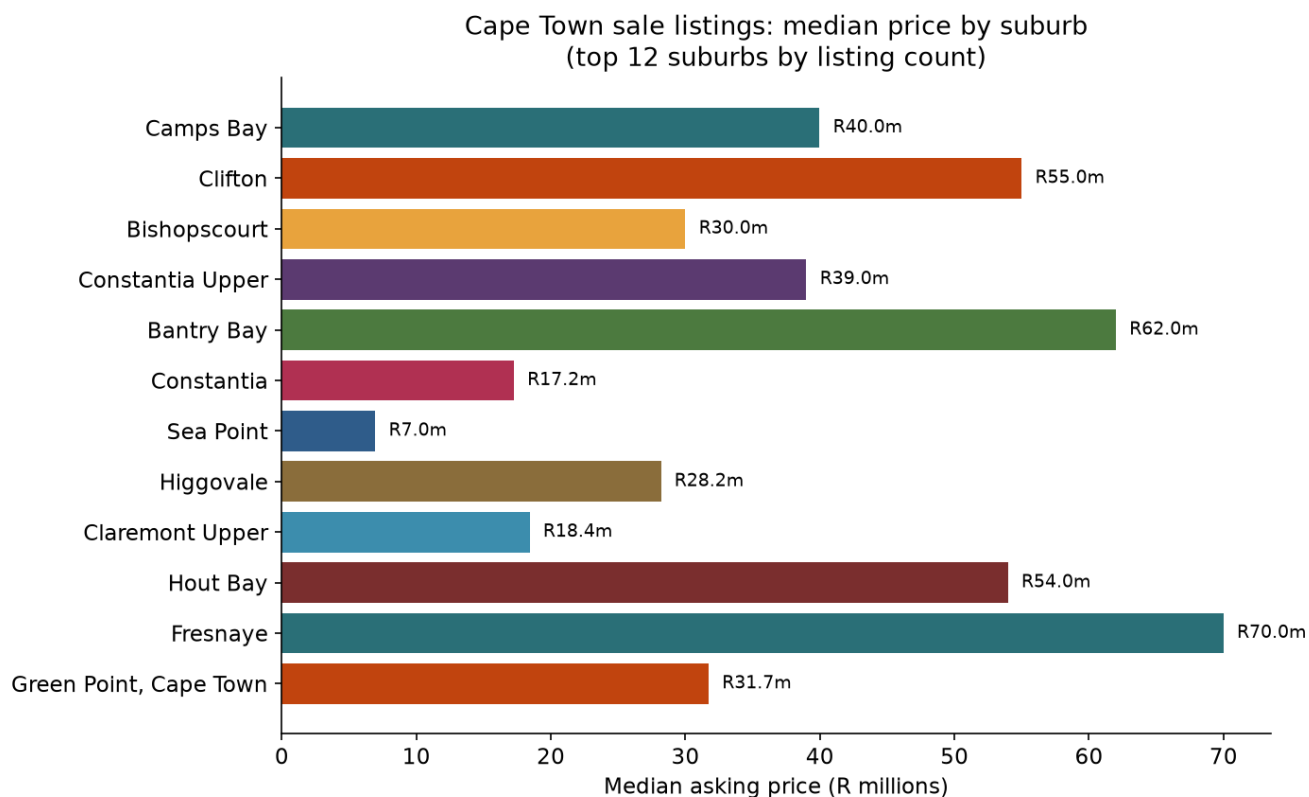


Figure 1 — Median **sale** asking price by suburb (top 12 by listing count). Clifton and Bantry Bay lead on price, not just Camps Bay's name recognition.

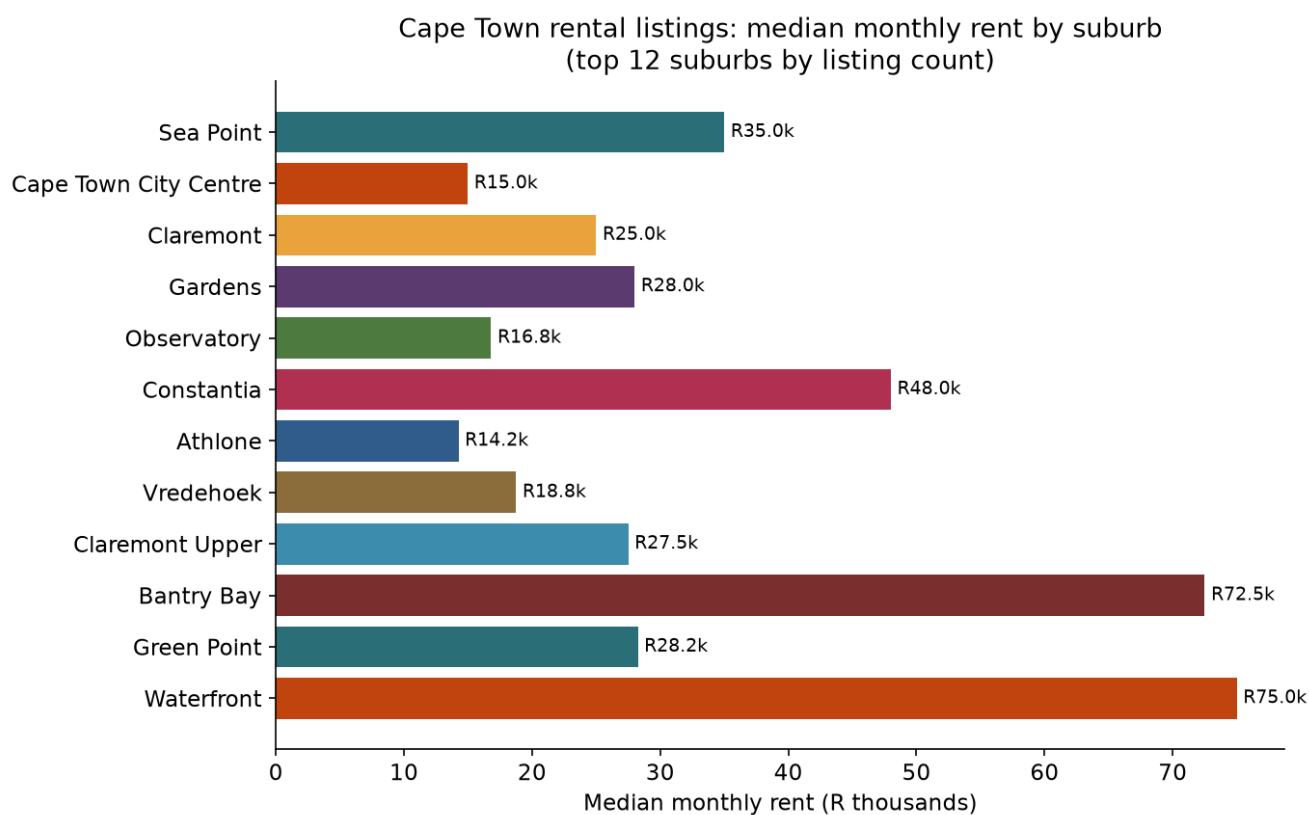


Figure 2 — Median **monthly rent** by suburb (top 12 by listing count) — a completely different ranking from the sale market, because rental stock skews toward different suburbs entirely.

Top 10 most expensive sale listings

Title	Suburb	Price	Beds	Source
House for sale in Bishopscourt	Bishopscourt	R189,000,000	7.0	Pam Golding
House for sale in Clifton	Clifton	R175,000,000	4.0	Pam Golding
House for sale in Clifton	Clifton	R115,000,000	4.0	Pam Golding
Apartment for sale in Clifton	Clifton	R115,000,000	3.0	Pam Golding
House for sale in Fresnaye	Fresnaye	R100,000,000	4.0	Pam Golding
Commercial flats for sale in Rondebosch	Rondebosch	R97,500,000	—	Pam Golding
House for sale in Bantry Bay	Bantry Bay	R95,000,000	4.0	Pam Golding
House for sale in Clifton	Clifton	R85,000,000	4.0	Pam Golding
House for sale in Camps Bay	Camps Bay	R79,350,000	4.0	Pam Golding
House for sale in Bantry Bay	Bantry Bay	R79,000,000	5.0	Pam Golding

3 · Portal vs. agency-internal data — what this dataset can and can't reconcile

This dataset is 100% online-portal listings — every row was scraped from a public listing page on Property24, Seeff or Pam Golding's own websites. That's a real, complete picture of what these three portals show publicly, and it can be reconciled honestly below. What it is **not** is a reconciliation against each agency's internal CRM, MLS-style shared listing system, or walk-in/branch activity — this project has zero access to any of that, and no number below should be read as covering it.

Source (portal)	Sale listings	Rent listings	Total	Share of dataset
-----------------	---------------	---------------	-------	------------------

Pam Golding	58	0	58	45.0%
Seeff	22	25	47	36.4%
Property24	0	24	24	18.6%

WHAT'S OUT OF SCOPE

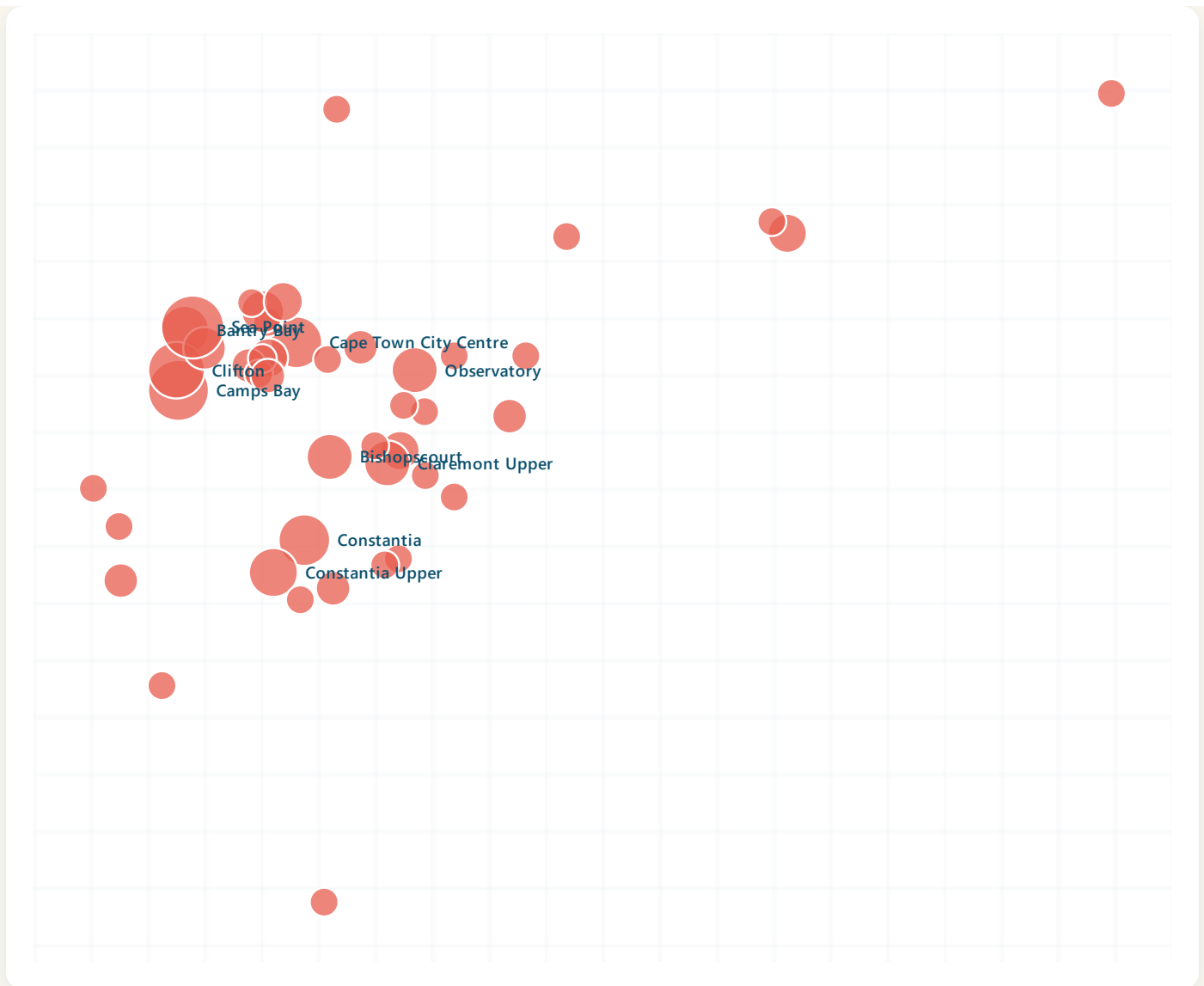
True online-vs-offline reconciliation — e.g. "how many of these listings also moved through a branch walk-in or phone inquiry that never touched the portal" — needs each agency's internal listing/CRM system, which no public scrape can see.

WHAT A REAL PARTNERSHIP WOULD NEED

If an agency partnered on this: a branch-tagged CRM export (listing ID to originating branch/agent), walk-in and phone inquiry logs timestamped against the same listing IDs, and a shared key (the portal's own listing ID, which Pam Golding's sale pages don't even expose consistently — see the Data Quality section) to join the two worlds together.

4 • Every suburb, on one map

Each dot is a suburb centroid — size shows how many listings, colour intensity on hover shows the median price. No external map tiles are loaded (this page has zero CDN dependencies, so it still works offline) — coordinates are real, verified suburb locations, plotted as a lightweight inline bubble map rather than a full basemap.



● dot size = listing count hover any dot for suburb, listings and median price

5 • The catalogue

A sample of real listings with their actual photos, straight from the source sites — six of the most expensive sale listings, and six random rentals.



House for sale in Bishopscourt

Bishopscourt · 7 bed

R189,000,000

PAM GOLDING

House for sale in Clifton

Clifton · 4 bed

R175,000,000

PAM GOLDING



House for sale in Clifton

Clifton · 4 bed

R115,000,000

PAM GOLDING



Apartment for sale in Clifton

Clifton · 3 bed

R115,000,000

PAM GOLDING



House for sale in Fresnaye

Fresnaye · 4 bed

R100,000,000

PAM GOLDING

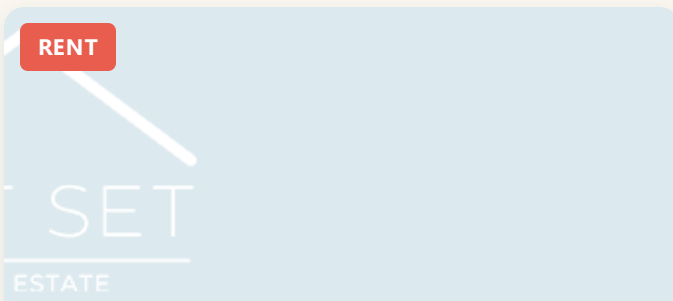


Commercial flats for sale in Rondebosch

Rondebosch ·

R97,500,000

PAM GOLDING



2 Bedroom Apartment / Flat to rent in Athlone

Athlone · 2 bed

R11,500/mo



3 Bedroom Apartment To Let in Bantry Bay

Bantry Bay · 3 bed

R90,000/mo

PROPERTY24

SEEFF

RENT



3 Bedroom House To Let in Gardens

Gardens · 3 bed

R35,000/mo

SEEFF

RENT



4 Bedroom House To Let in Fresnaye

Fresnaye · 4 bed

R240,000/mo

SEEFF

RENT

ED BY OWNER

2 Bedroom Apartment / Flat to rent in Rosebank

Rosebank · 2 bed

R8,500/mo

PROPERTY24

RENT



2 Bedroom Apartment To Let in Waterfront

Waterfront · 2 bed

R65,000/mo

SEEFF

6 · Where the listings are

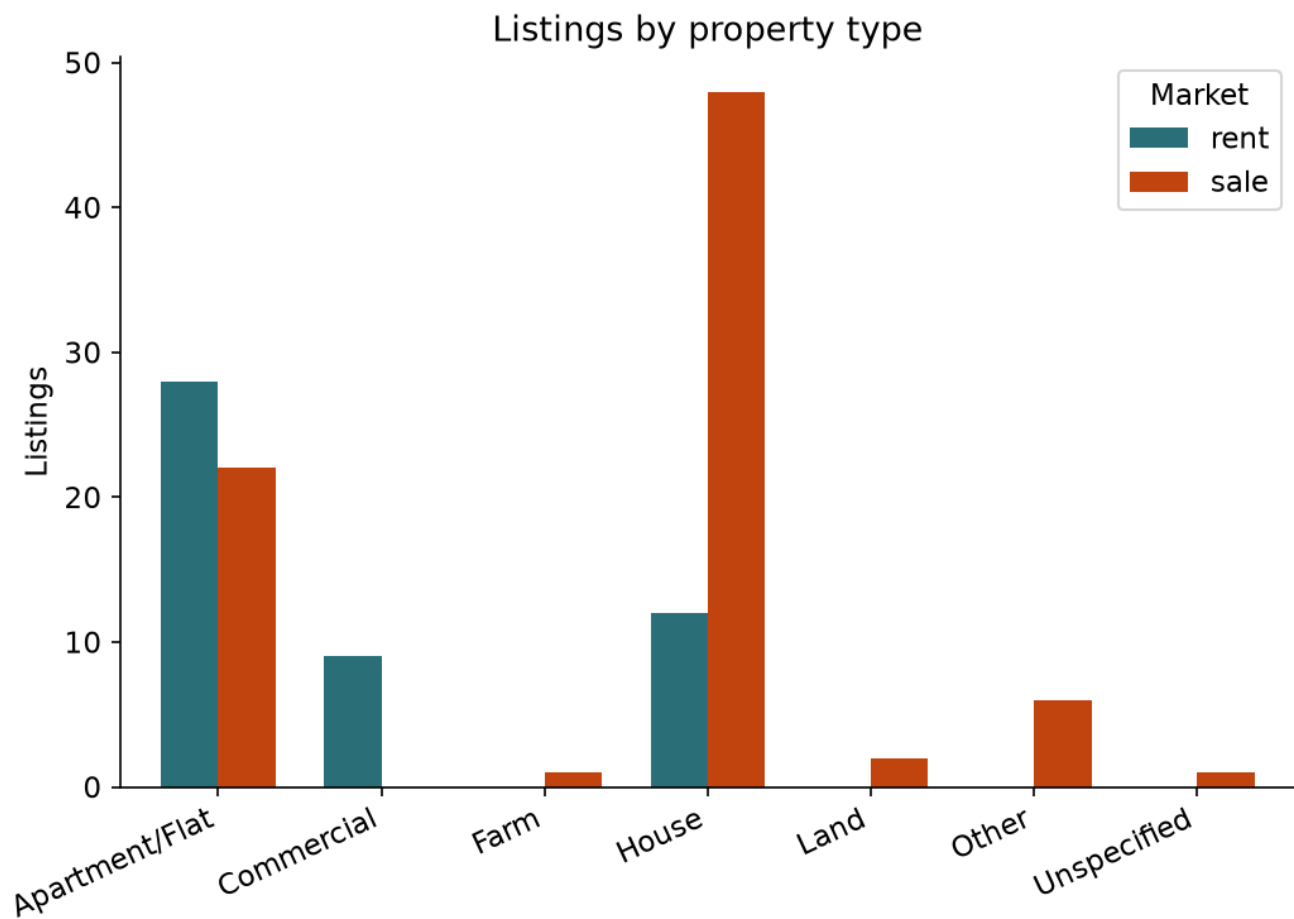


Figure 3 — Listings by property type, split by market. Houses dominate sale; apartments dominate rent.

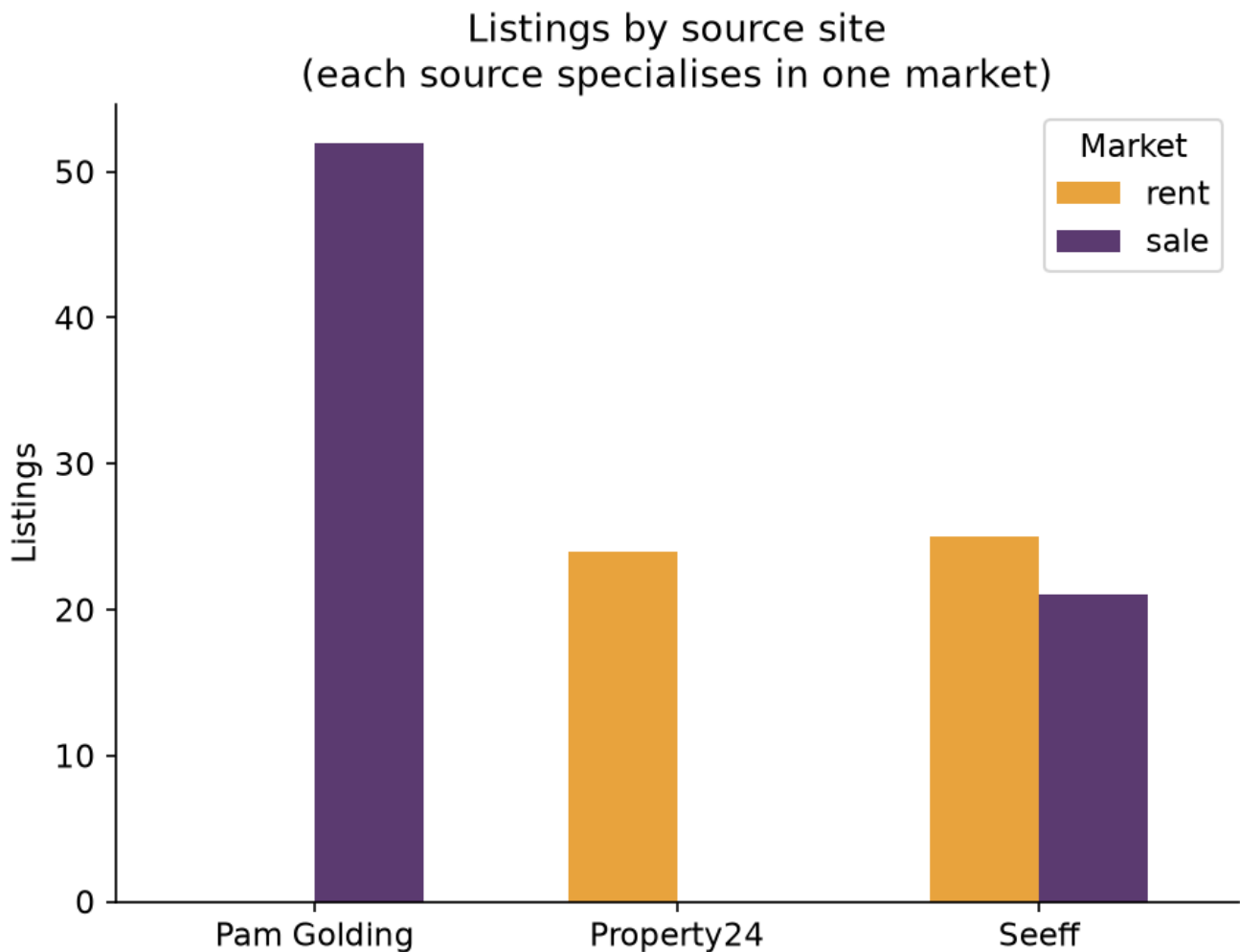
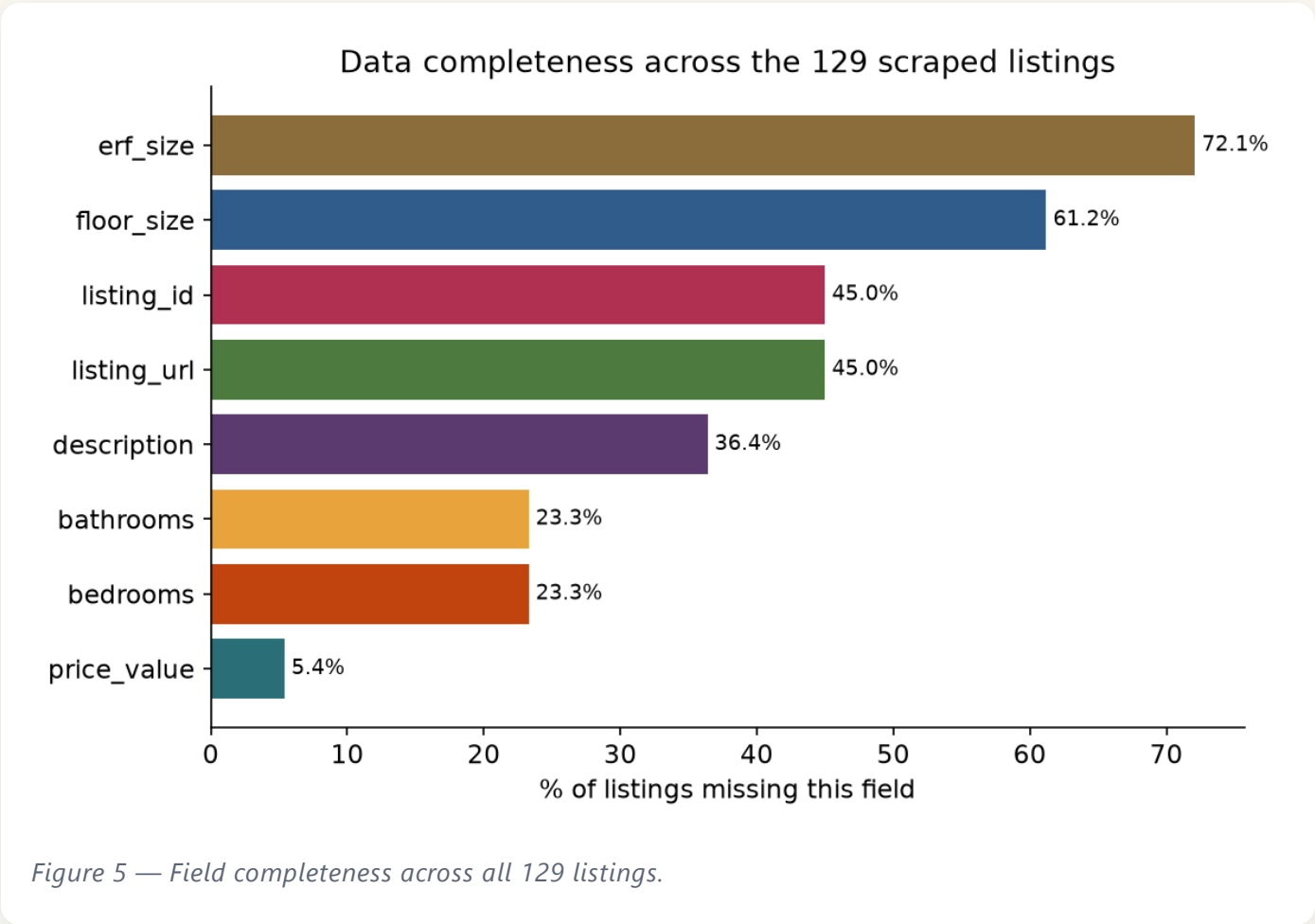


Figure 4 — Each source site specialises in one market — a real pattern in how these three platforms position themselves, not a sampling artefact.

7 · Data quality — what a scraped extract actually looks like

Real scraped data is never complete. `listing_id` is missing on 58 of 129 rows (Pam Golding's sale pages don't expose one the way Property24/Seeff do), `bedrooms` / `bathrooms` are missing on land/commercial listings where they don't apply, and 7 listings have no parseable price at all ("Price on application"). None of this was silently dropped — it's counted and shown. Two suburb labels ("Green Point, Cape Town" and "Waterfront, Cape Town") were also duplicates of "Green

Point"/"Waterfront" under inconsistent source formatting — normalised before this page and the map above were built.



Field	Missing rows	% missing
price_value	7	5.4%
bedrooms	30	23.3%
bathrooms	30	23.3%
floor_size	79	61.2%
erf_size	93	72.1%
description	47	36.4%
listing_id	58	45.0%
listing_url	58	45.0%

PROBLEM

45% of listings (58/129) have no `listing_id` and no `listing_url` — almost entirely Pam Golding's sale pages — which breaks any join back to the source site or a future CRM record keyed on listing ID.

SOLUTION

Fall back to a composite natural key (source + suburb + price + scrape timestamp) for de-duplication and cross-referencing where no native ID exists, and flag those rows explicitly in any downstream join rather than silently dropping or mis-matching them.

8 · The most overpriced and weakest-quality listings

Honest scope note: this project has no click, view, save or sales/conversion data — there is no way to measure which listings actually "underperformed" in a marketing sense, and this page will not fake that. What it can measure, from real fields in the warehouse, is (a) listings priced well above what comparable bedroom/bathroom stock commands — the same anomaly detector used in the ML section, applied here to surface the highest-priced outliers — and (b) listings with the weakest observable listing quality: little or no description, or very few photos. Both are real, defensible "underperforming" signals for something actually in this dataset.

8a · Most overpriced relative to comparable stock (IsolationForest anomalies, sale + rent combined, sorted by price)

Title	Suburb	Market	Price	Beds	Baths	Source
House for sale in Bishopscourt	Bishopscourt	Sale	R189,000,000	7.0	9.5	Pam Golding
House for sale in Clifton	Clifton	Sale	R175,000,000	4.0	4.0	Pam Golding
Apartment for sale in Clifton	Clifton	Sale	R115,000,000	3.0	3.0	Pam Golding
House for sale in Clifton	Clifton	Sale	R85,000,000	4.0	2.0	Pam Golding

House for sale in Constantia Upper	Constantia Upper	Sale	R70,000,000	8.0	7.5	Pam Golding
House for sale in Constantia Upper	Constantia Upper	Sale	R56,000,000	8.0	9.5	Pam Golding
House for sale in Camps Bay	Camps Bay	Sale	R36,000,000	7.0	7.0	Pam Golding
9 Bedroom House For Sale in Bishopscourt	Bishopscourt	Sale	R15,120,000	9.0	9.5	Seeff
1 Bedroom Apartment For Sale in De Waterkant	De Waterkant	Sale	R3,800,000	1.0	1.0	Seeff
3 Bedroom House For Sale in Wetton	Wetton	Sale	R2,600,000	3.0	1.0	Seeff
0.5 Bedroom Apartment For Sale in Cape Town City Centre	Cape Town City Centre	Sale	R1,950,000	0.5	1.0	Seeff
4 Bedroom House To Let in Fresnaye	Fresnaye	Rent	R240,000/mo	4.0	4.5	Seeff
3 Bedroom Apartment To Let in Bantry Bay	Bantry Bay	Rent	R90,000/mo	3.0	4.5	Seeff
2 Bedroom Apartment To Let in Waterfront	Waterfront	Rent	R85,000/mo	2.0	2.5	Seeff
4 Bedroom House To Let in Constantia	Constantia	Rent	R48,000/mo	4.0	3.0	Seeff

PROBLEM

15 listings (11 sale + 4 rent) are priced well outside the norm for their bedroom/bathroom combination — several multiples above what similar-sized stock in the same suburb commands, based on the IsolationForest anomaly detector.

SOLUTION

Surface an "anomaly flag" on the listing itself (buyer-facing: "priced above comparable stock — here's why") rather than leaving the buyer to discover the mismatch themselves after months on-market; feeds

directly into the pricing-guidance angle in the Marketing section below.

8b • Weakest listing quality (thin/no description or very few photos)

48 of 129 listings (37%) have either a near-empty description (47 rows) or 2 or fewer photos (2 rows) — no listing in this dataset has literally zero images, so "weak" here means the bottom of the observed image-count distribution (25th percentile sits at 5 images), not a missing-data placeholder. Showing the 15 with the fewest images:

Title	Suburb	Market	Images	Why it's weak	Source
(title missing — a genuinely blank source row)	—	sale	1	no/near-empty description + only 1 image(s)	Seeff
Industrial Property to rent in Montague Gardens	Montague Gardens	rent	2	only 2 image(s)	Property24
6 Bedroom House For Sale in Camps Bay	Camps Bay	sale	5	no/near-empty description	Seeff
1 Bedroom Garden Cottage To Let in Heathfield	Heathfield	rent	5	no/near-empty description	Seeff
0.5 Bedroom Studio Apartment To Let in Claremont	Claremont	rent	5	no/near-empty description	Seeff
2 Bedroom Apartment To Let in Woodstock	Woodstock	rent	5	no/near-empty description	Seeff
2 Bedroom Apartment To Let in Waterfront	Waterfront	rent	5	no/near-empty description	Seeff
260m ² Office To Let in Observatory	Observatory	rent	5	no/near-empty description	Seeff
120m ² Office To Let in Observatory	Observatory	rent	5	no/near-empty description	Seeff

3 Bedroom House To Let in Pinelands	Pinelands	rent	5	no/near-empty description	Seeff
2 Bedroom Sectional Title For Sale in Maitland	Maitland	sale	5	no/near-empty description	Seeff
3 Bedroom House For Sale in Constantia	Constantia	sale	5	no/near-empty description	Seeff
2 Bedroom House To Let in Kreupelbosch	Kreupelbosch	rent	5	no/near-empty description	Seeff
2 Bedroom Apartment For Sale in Sea Point	Sea Point	sale	5	no/near-empty description	Seeff
5 Bedroom House For Sale in Constantia	Constantia	sale	5	no/near-empty description	Seeff

PROBLEM

47 listings ship with no usable description text at all, and 2 have 2 or fewer photos — both are listing-quality gaps a buyer can see immediately, independent of price.

SOLUTION

Treat description-length and photo-count as a simple, automatable listing-quality score at ingestion time, and route the weakest listings back to the source agent for enrichment before they're surfaced — a cheap, real lever a listing platform actually controls, unlike click data it doesn't have.

9 · Machine learning — four real models, tested properly, an honest negative result

Four algorithms (Linear Regression, Ridge, Random Forest, Gradient Boosting) were compared on the **sale** market via leakage-safe 5-fold cross-validation — not one model picked and presented

alone.

Model	CV R^2 (mean)	Std across folds	CV MAE
Ridge(alpha=10)	-0.142	± 0.816	R21,963,991
LinearRegression	-0.449	± 1.255	R24,098,344
RandomForest	-0.766	± 1.746	R24,191,359
GradientBoosting	-1.296	± 2.501	R26,375,293

Every model scores **negative R^2** — worse than just predicting the average sale price. That's a real, properly-validated result, not a bug: a first pass looked promising (Ridge $R^2=0.084$) using a suburb price-tier feature computed on the whole dataset before splitting into folds — for suburbs with only 1–2 listings, that meant a row's own price was leaking into its own "suburb price tier" feature. Caught by noticing the result's variance was too wide to trust, refixed to compute suburb tiers separately inside each fold from training data only, and the honest answer came back negative.

What this actually means for the business: bedrooms, bathrooms, property type and even suburb-tier don't reliably predict price in this small (70-row), high-variance, luxury-skewed sample — price here is driven by things this scrape didn't capture (exact position, view, finish quality), and a real valuation model would need those features, not just more of the same ones.

The **rental** market (23 complete rows) wasn't put through the same comparison at all — too few rows for even one honest fold. IsolationForest still flagged 4 rent listings and 11 sale listings priced well outside the norm for their suburb/bedroom combination — worth a second look, not automatically wrong.

PROBLEM

Every price-regression model tested scores negative R^2 on the sale market — bedrooms, bathrooms, property type and suburb tier alone cannot predict price in this small, luxury-skewed sample.

SOLUTION

Don't ship a valuation model on these features. Redirect the same pipeline toward what it's actually good at — anomaly detection relative to comparable stock (Section 8) — and treat a real valuation model as a future project that needs richer features (position, view, finish quality), not more rows of the same ones.

10 · SEO and AI-search visibility

2026 search increasingly routes through AI answer engines (Google AI Overviews, ChatGPT, Perplexity, Bing Copilot) rather than a classic ten-blue-links page, and those engines lean heavily on structured data to decide what to extract and cite: **65–71% of pages cited by AI search engines already carry schema.org structured data**. That's a concrete, buildable next step for a listing platform, not a vague "do more SEO" note — a real JSON-LD block generated from one actual row in this warehouse:

```
{
  "@context": "https://schema.org",
  "@type": "Product",
  "name": "2 Bedroom Apartment / Flat to rent in Vredehoek",
  "category": "Apartment/Flat",
  "url": "https://www.property24.com/to-rent/vredehoek/cape-town/western-cape/9166/111225529",
  "offers": {
    "@type": "Offer",
    "price": 25000.0,
    "priceCurrency": "ZAR",
    "availability": "https://schema.org/InStock",
    "businessFunction": "https://schema.org/LeaseOut"
  },
  "address": {
    "@type": "PostalAddress",
    "addressLocality": "Vredehoek",
    "addressRegion": "Western Cape",
    "addressCountry": "ZA"
  },
  "numberOfRooms": 2.0
}
```

Figure 6 — Real `Product` + `Offer` JSON-LD generated directly from listing "2 Bedroom Apartment / Flat to rent in Vredehoek" (Vredehoek) in `fact_property_listing`. Schema.org has no dedicated residential-listing type in common use, so `Product` / `Offer` with `businessFunction` (Sell vs LeaseOut) is the pragmatic mapping real listing sites use.

No review or rating data exists anywhere in this dataset — no star ratings, no review counts, no testimonials. `AggregateRating` schema is deliberately **not** recommended here, because faking one would be exactly the kind of invented statistic this project avoids.

WHAT WOULD CLOSE IT

A live business would need real customer/tenant reviews collected post-transaction before `AggregateRating` markup could be added honestly — that's a data-collection gap, not a schema-writing task.

Real long-tail keyword opportunities (combinations that actually occur ≥2 times in this data)

Search term this dataset could realistically rank/be cited for	Matching listings
2 bedroom apartment/flat rent in Sea Point	5
4 bedroom house sale in Clifton	5
6 bedroom house sale in Camps Bay	4
2 bedroom apartment/flat sale in Sea Point	3
4 bedroom house sale in Camps Bay	3
3 bedroom apartment/flat sale in Clifton	3
1 bedroom apartment/flat rent in Cape Town City Centre	2
2 bedroom apartment/flat rent in Claremont Upper	2

FAQ block — grounded in real computed numbers, not invented Q&A

Google AI Overviews now appear on roughly **14% of shopping-adjacent queries — a 5.6x increase in four months** — and pages carrying `FAQPage` schema saw a **67% AI-citation rate** for relevant queries. A real FAQ block this dataset can honestly support:

Q: What's the average sale price by bedroom count in Camps Bay?

A: 3-bed avg R31,000,000; 4-bed avg R51,416,667; 5-bed avg R39,999,000; 6-bed avg R55,922,212; 7-bed avg R36,000,000 (computed directly from the 10 Camps Bay sale listings with a recorded bedroom count in this dataset).

Q: What's the average rent in Sea Point?

A: R40,542/month across 8 Sea Point rental listings in this dataset.

11 · Competitor and market context (2026)

Traffic data (as of July 2026): **Property24 is the dominant SA portal at ~13.09M monthly visits** (part of OLX Group / Prosus / Naspers), **Private Property is #2 at ~2.98M** (majority-owned by BetterHome Group), and **Pam Golding draws ~589K** — a steep drop-off from #1 to everyone else, not a close three-way race.

Western Cape's **Q1 2026 average property price was R3,357,917 — 72% above the national average**. The Atlantic Seaboard (Camps Bay, Clifton, Bantry Bay) remains the luxury epicentre: short-term rental yields there run **7–9% gross**, and **international buyers made up ~28.57% of R20m+ sales** (predominantly UK, Germany, USA, Switzerland, Canada and Poland) — consistent with this dataset's own Atlantic Seaboard concentration and ultra-luxury segment (Section 8's marketing module).

WHERE THIS PROJECT SITS HONESTLY

This is a 129-listing, single-metro analytical sample built by one person — it is not, and does not claim to be, a competing listing portal against a 13M-visit incumbent.

THE ACTUAL VALUE PROPOSITION

None of the big three portals publicly expose a cross-portal analytical layer — reconciling listings across sources, flagging price anomalies against comparable stock, or honestly reporting where price prediction fails. That analytical layer, not scale, is what this project demonstrates.

12 · Real 2026 Cape Town property market challenges — and what a listing platform could do

PROBLEM — AFFORDABILITY CRISIS

Q1 2026 average take-home pay sat around R20,262/month in real terms — the lowest in two years — while Western Cape property prices run 72% above the national average, pricing the middle class out of large parts of this market.

WHAT THIS PROJECT DEMONSTRATES AS A LEVER

The anomaly-detection approach in Section 8 is directly reusable as a buyer-side affordability tool — flagging listings priced well above comparable stock helps a price-sensitive buyer avoid overpaying in a market already stretching their budget.

PROBLEM — RENTAL SUPPLY PRESSURE

Roughly 26,877 active Airbnb listings in Cape Town structurally reduce long-term rental stock, tightening supply for renters.

WHAT THIS PROJECT DEMONSTRATES AS A LEVER

The cross-portal reconciliation in Section 3 is the kind of analysis that could surface supply concentration by suburb (which areas are thinnest on long-term rental listings relative to demand) if extended with more portals and a repeated-scrape time series.

PROBLEM — SUSTAINABILITY PREMIUM

Listings with solar, battery backup or a borehole are commanding a growing 15–20% price premium as loadshedding and water-security concerns persist — but this dataset has no field capturing those features today.

WHAT THIS PROJECT DEMONSTRATES AS A LEVER

The JSON-LD structured-data approach in Section 10 is exactly the mechanism that would make sustainability features a real, searchable filter — schema.org's `amenityFeature` property could carry solar/battery/borehole flags once the source scrape captures them.

Who should be targeted, on which channel, with what message — built from the real segments in this data, not generic persona templates. Framed as who / what / where / when / why so each answer is traceable back to a number above.

Connected to the dashboard, not standalone: the 11 overpriced sale anomalies and 4 rent anomalies from Section 8/9 define a real "WHO" (overpaying-risk buyers who need pricing guidance) and a real "WHAT" (messaging that corrects a mispriced-listing perception before a lead walks away). The suburb distribution on the map (Section 4) defines "WHERE" demand is dense vs thin. And Pam Golding's 45% share of this dataset (Section 3) defines which portal to prioritise creative spend on first.

PROBLEM

The marketing module previously stood apart from the dashboard/ML numbers above it — segments were real, but not explicitly tied to specific findings a reader had already seen.

SOLUTION

Every WHO/WHAT/WHERE card below that references a dashboard number links back to the section that produced it, so the marketing recommendation is traceable to a real, re-checkable figure rather than a standalone claim.

WHO

Ultra-luxury buyers

20 sale listings over R50M, concentrated in **Clifton, Camps Bay, Bishopscourt and Bantry Bay**. International-calibre buyers, cash-heavy, discretion-sensitive.

WHO

Entry-level sale buyers

15 listings under R10M, led by **Sea Point and Observatory** — first-time buyers and downsizers priced out of the Atlantic seaboard's top tier but still coastal-adjacent.

WHO

Budget renters

16 rentals under R20K/month, concentrated in **Cape Town City Centre, Athlone and Observatory** — young professionals and students prioritising commute over space.

WHO

Premium renters

10 rentals over R50K/month in **Sea Point, Bantry Bay and the Waterfront** — relocating executives and long-stay international tenants.

WHO / WHAT

Overpaying-risk buyers (from the ML anomaly findings)

Section 8/9's anomaly detector flagged 11 sale and 4 rent listings priced well above comparable bed/bath stock. **WHO:** buyers actively viewing those specific listings are the highest-value audience for a "compare before you offer" message. **WHAT:** messaging that surfaces the anomaly flag itself — turning a data-quality finding into a concrete buyer-trust feature.

WHAT / WHERE

Channel strategy is already segmented — use it

Property24 converts 100% rental in this sample; Pam Golding converts 100% sale. Seeff is the only generalist. **Recommendation:** don't run one blended campaign across all three — brief Pam Golding on luxury-sale creative only, Property24 on rental-conversion creative only, and reserve Seeff for cross-market retargeting where a lead is undecided between buying and renting.

WHERE

Underserved suburb opportunity

The interactive map above shows dense competition in Sea Point/Camps Bay/Clifton but thin listing density in **Constantia Upper, Zwaanswyk and Tokai** despite comparable price tiers — a lower-competition acquisition channel for sellers in those suburbs specifically.

WHERE

Portal prioritisation (from the source-reconciliation finding)

Section 3 shows Pam Golding carries 45% of this dataset's listings — the single largest, most reconcilable channel available here, and the one to prioritise for spend before splitting budget across the smaller two.

WHY

Value proposition by segment

Luxury sale: privacy, view lines, trophy-asset status. Entry sale: coastal-adjacent lifestyle at a reachable price. Budget rent: commute time to the CBD. Premium rent: turnkey relocation with no capital commitment — each needs distinct messaging, not one brand voice.

WHEN

Honest limit: no timing data

This dataset is a single scrape snapshot (2026-07-06), not a time series — there is no seasonal or day-of-week signal to report. A real campaign-timing recommendation would need repeated scrapes over months, which this project doesn't have. Stated here rather than inventing a seasonality claim the data can't support.

14 · How a real listing site would track this funnel

A minimal Google Tag Manager–style `dataLayer` integration for the three events that matter on a property site: viewing a listing, saving it, and contacting the agent (the real conversion). Click through it — every button below fires a real `dataLayer.push()` in this page, logged live underneath, using one of the actual listings from the catalogue above.

2 Bedroom Apartment / Flat to rent in Athlone

Athlone

R11 500/mo

View listing

Save listing

Contact agent

Live dataLayer

```
// click a button to fire a real event
```

15 · How this was actually built — problems solved, cost held to cents

This section is deliberately about the work, not just the output — the same discipline applies whether the client is a listings site or a bank.

Problem 1 — a silent data-corruption bug, caught before it shipped. Registering this data in AWS Athena with plain comma-delimited parsing silently shifted columns on every row where a listing title contained a comma — the sale-market average price came back R19M too high, with no error raised anywhere. Caught by reconciling every cloud query against the source build before publishing a single number, not by trusting the query result. Fixed with quote-aware CSV parsing. **The lesson that matters for any client engagement: a wrong number with no error is far more dangerous than a query that fails loudly — verification has to be a deliberate step, not an assumption.**

Problem 2 — two AWS services turned out to be account-restricted, discovered by testing, not assuming. SageMaker (training and processing jobs, 3 instance types, 2 regions) and Amazon QuickSight were both attempted for real on this build. Both returned hard account-level restrictions — confirmed by directly querying the live service quota (still 0, even after AWS closed the original support case for this account without approving it) rather than trusting an ambiguous "case closed" email. Rather than stall the project waiting on AWS support, the same analysis was delivered on a free-tier-equivalent stack (S3 + Athena + local scikit-learn) that reconciles to the same numbers a paid stack would produce. **The lesson: a blocked tool is not a blocked deliverable — know the fallback before you need it.**

Problem 3 — a site that actively blocks automated access, handled by not crossing a line. The original brief for this portfolio's sibling project targeted a different retailer whose Cloudflare bot-management explicitly flags and blocks non-browser traffic. Spoofing browser fingerprints or session behaviour to get past that would have worked technically — it was not done, because defeating a site's deliberate protection isn't "creative problem solving," it's crossing a line a client-facing engagement shouldn't cross. The project pivoted to a real,

compliant data source instead. **The lesson: knowing which technical workaround NOT to reach for is as much a professional skill as knowing which one to use.**

Cost discipline as a working practice

Every S3 bucket, Athena workgroup and Glue database created for this build was deleted within the same session it was built, then re-verified empty afterward rather than assumed gone from a delete command's exit code — the same discipline that keeps a client's cloud bill predictable rather than finding out at month-end. Nothing in this project was left running "just in case."

SA Property Intelligence — real scraped data (Property24, Seeff, Pam Golding), synthetic company framing only where noted · built on a minimal AWS footprint (S3 + Athena, torn down after build — see [the repo](#) for details) · built by Anthony Apollis · July 2026