DATA DIAGNOSTIC | LAND ACQUISITION SERIES

Six Plots, One Outlier, Too Many Gaps: A Data-Quality Audit

An unvarnished analytical review of the six-plot options dataset — what the numbers actually say, where the dataset fails basic standards, and what must be fixed before any acquisition decision is underwritten on it.

- Six plots spanning AED 20–203M; portfolio ask totals AED 336M
- All FAR values reconcile exactly to GFA ÷ plot size — the arithmetic is clean
- Maritime City is an outlier on every metric (FAR 6.50, AED 750/sq ft, AED 203M) — verify before trusting
- Critical fields are missing: no plot IDs, no sale status, no owner, no dates — the table cannot support decisions as-is

PREPARED BY STRATEGIC ADVISORY • JULY 2026 • CONFIDENTIAL — FOR INTERNAL USE

Executive Summary

The dataset under review lists just **six plot options** — Dubai Island, Maritime City, Meydan, Majan, DLRC (Verdania) and Dubai South — with high-level attributes (plot size, GFA, FAR, approval height, pricing, usage). Our blunt assessment: **this is a marketing summary, not analytical data**. It lacks plot IDs, ownership, sale status, phasing and dates; two entries carry "Details awaited" placeholders instead of locations. Within those limits, the numeric content is internally consistent — every stated FAR reconciles exactly to GFA ÷ plot size — but one record, **Maritime City, is an outlier on every single metric** and must be independently verified before it anchors any underwriting.

Bottom line: fix the data first, then trust the analysis. The recommendations in Section 5 specify exactly which fields to acquire and which values to verify.

6	AED 336M	3.72	AED 452	0
plot options in dataset	combined asking price	mean FAR (range 1.80–6.50)	mean price / sq ft (range 267–751)	plot IDs, statuses or sale dates provided

1. Data Extraction & Standardisation

The fields present under "Plot Options" were extracted and standardised: Project, Plot_Size_SqFt, GFA_SqFt, FAR, Construction Approval (height), PricePerSqFt_AED and TotalPrice_AED_M, plus Usage and Comments. Currency is unlabelled but is evidently AED; all areas are in square feet. Numeric types were verified, and FAR was confirmed against the industry definition (gross floor area ÷ land area) for every row — **no calculation errors were found**. Illustratively, Maritime City: 270,410 ÷ 41,613 ≈ 6.498, matching the stated 6.50.

2. The Dataset at a Glance

Project	Plot size (sq ft)	GFA (sq ft)	FAR	Price/sq ft (AED)	Total price (AED M)	Usage
Dubai Island	25,833	80,857	3.13	482.33	39	Residential
Maritime City	41,613	270,410	6.50	750.71	203	Mixed use
Meydan (DL)	15,069	51,066	3.39	567.89	29	Mixed use
Majan (DLRC)	16,522	66,091	4.00	317.74	21	Mixed use
DLRC (Majan/Verdania)	20,800	72,827	3.50	329.55	24	Residential
Dubai South	41,555	74,800	1.80	267.38	20	Residential
Mean	**26,899**	**102,675**	**3.72**	**452**	**56**	—

Table 1 — Summary statistics per project. Maritime City (highlighted) drives the means: it alone represents 60% of total asking value.

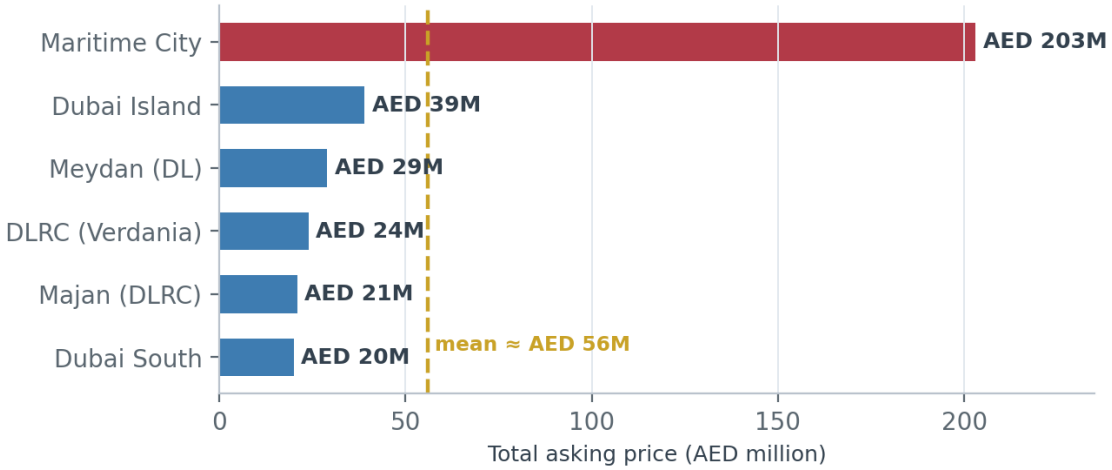

Exhibit 1 — Total asking price by plot. Maritime City (AED 203M) sits >5× the next-largest ticket and single-handedly pulls the mean to AED 56M against a median of AED 26.5M.

3. Distribution & Outlier Analysis

Plotting density against unit price makes the structure of the dataset obvious. Five plots cluster in a FAR band of 1.8–4.0 and a price band of AED 267–568 per sq ft. Maritime City sits alone at FAR 6.50 and AED 750.71 per sq ft — roughly double the cluster on density and 66% above the next-highest unit price.

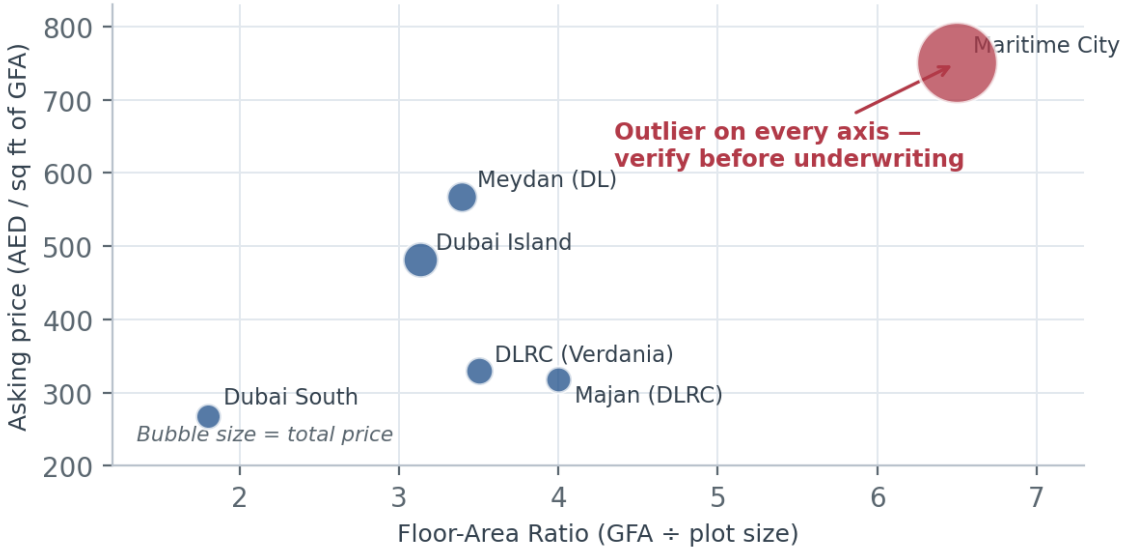

Exhibit 2 — Density vs unit price; bubble size = total asking price. Five plots form a coherent mid-market cluster; Maritime City stands apart on all three dimensions.

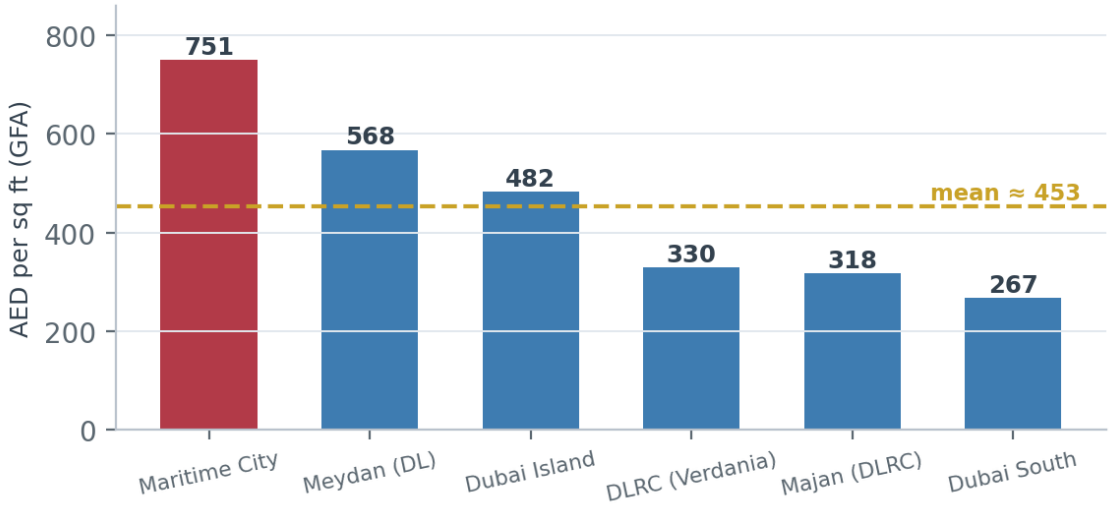

Exhibit 3 — Asking price per sq ft of GFA. The spread from AED 267 (Dubai South) to AED 751 (Maritime City) implies fundamentally different land-use and planning contexts.

How to treat the Maritime City outlier

Statistical best practice (per NIST) is clear: outliers come from a distribution different from the bulk of the data and **should not be auto-removed — they are trying to tell us something**. The values are mathematically valid; a FAR of 6.5 with mixed use could reflect a legitimate high-density waterfront scheme. But it could equally be a transcription error. Cross-check against official project documents before this record informs any decision.

Secondary patterns: (i) usage splits three mixed-use versus three residential plots, which plausibly explains part of the unit-price dispersion; and (ii) Dubai South combines the lowest FAR (1.80) with the lowest unit price (AED 267), consistent with a low-density, early-stage planning context very different from the urban-core plots.

4. Data-Quality Audit

No blank cells appear among the core numeric fields, and no contradictory entries were found. The serious problems are structural:

Issue	Severity	Detail	Required correction
Missing key columns	Critical	No Plot ID, phase/block, availability or sales status, owner/developer, or sale dates. The table cannot link to transactions or track which plots remain available.	Acquire or create these fields from project documents and integrate before analysis.
Placeholder locations	High	DLRC and Dubai South show "Details awaited" instead of a location — a missing value disguised as text.	Fill actual coordinates; where unknown, mark NULL/N-A consistently.
Unverified outlier	High	Maritime City's FAR (6.50) and price (AED 203M) are valid arithmetic but far outside the cluster.	Verify against official project plans; do not drop or accept without investigation.
Units & currency implicit	Medium	Areas in sq ft, prices presumed AED — neither labelled. In multi-region data this would be disastrous.	Encode units in column names (e.g. PricePerSqFt_AED) or add a currency field.
Extraneous text	Medium	Footnotes mixed into the data ("prices indicative", "5% VAT on commission", MOP note).	Move disclaimers to a separate notes field or appendix.
Formatting noise	Low	Excess decimal precision (FAR to 12 digits); risk of numerics stored as text.	Round to 2–3 decimals; enforce numeric dtypes throughout.

5. Conclusions & Next Steps

The workbook offers a snapshot, not a foundation. The arithmetic is trustworthy; the structure is not. Until identifiers, statuses and dates exist, any conclusion drawn from this table stands on shaky ground. Recommended sequence:

1. **Complete the schema:** add Plot ID, phase, availability status, owner and date fields sourced from project documentation.
2. **Verify Maritime City** against official plans before it influences any ranking, average or shortlist.
3. **Clean the table:** isolate disclaimers, resolve "Details awaited" placeholders, standardise units and precision.
4. **Recompute all metrics** on the completed dataset, and re-run the distribution checks (histograms, box plots) to confirm the outlier picture.
5. **Only then underwrite:** use the cleaned data and exhibits above to support shortlist and acquisition decisions.

> **A mentor's blunt advice**
>
> Fix the data first, then trust the analysis. Six rows are enough to reveal an outlier; they are nowhere near enough to hide one.

Methodology: fields extracted and standardised programmatically (pandas); FAR validated as GFA ÷ plot size per industry definition; outlier treatment follows NIST/SEMATECH e-Handbook guidance on exploratory data analysis.