

The football transfer market: who pays whom, and what age costs

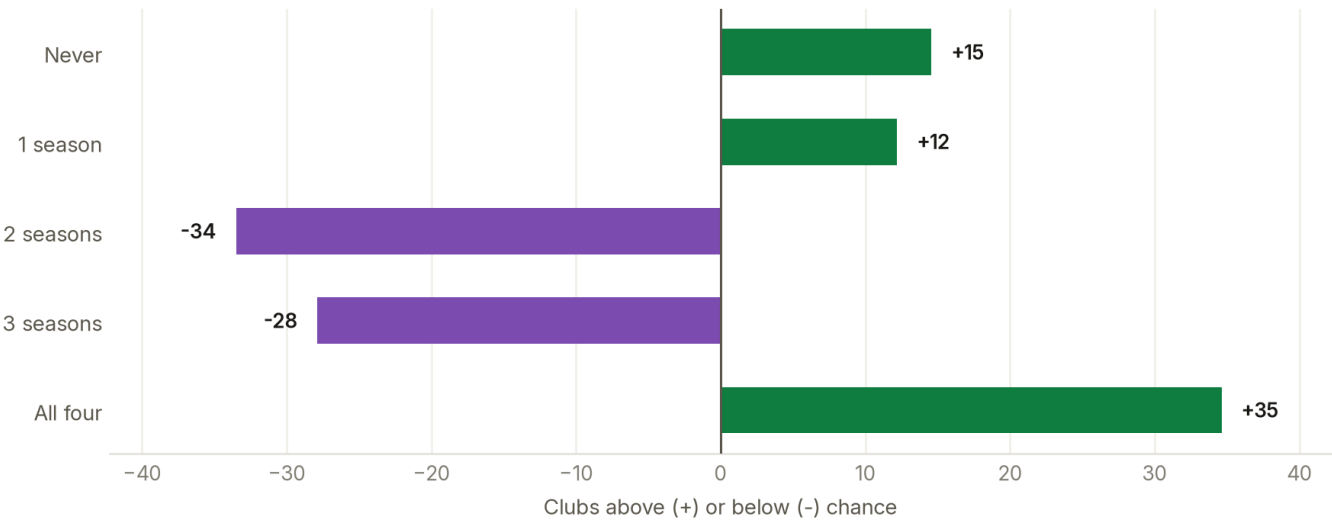
Two things the industry treats as settled turn out to be wrong. Spending is spreading out, not concentrating, and young players carry no price premium. Both were written down as a hypothesis before the analysis — and contradicted by it.

DOMAIN	TOOLS	SCALE
Football / sports finance	SQL · DuckDB · Python · matplotlib	6,716 priced transfers · €37.4bn
SOURCES	WINDOW	PUBLISHED
FIFA TMS census · Transfermarkt (CC0)	2022/23–2025/26 · census 2016–2025	15 August 2026

96 clubs were net sellers in all four seasons; chance alone would have given 61

Clubs above or below what random role assignment would produce, by seasons as a net seller

2022/23–2025/26 · 350 CLUBS PRESENT IN ALL FOUR SEASONS



Source: Transfermarkt via dcaribou/transfermarkt-datasets (CC0), snapshot 2026-08-14 · The baseline is a binomial with the same marginal probability of being a net seller (0.65). Both ends come out over-represented and the middle under-represented: roles persist rather than being redealt each summer.

JA

Clubs above or below what random role assignment would produce. Both tails are over-represented: the role persists rather than rotating each summer.

The eight phases, and what each one settled

Every phase closes on an exit gate checked point by point. When one fails the process goes *back* a phase rather than improvising forward — which is what happened in phase 2.

0 · CHOOSE	A client that can see both sides of the market A multi-club fund, not a selling club: a client on one side of a gap cannot reason about the gap. Source picked only after an integrity test, not before.
1 · ASK	A SMART question, and the hypothesis written down first Three measurable parts: concentration, persistence of role, age. The expected answer was recorded before any analysis — the only way a later contradiction is credible.
2 · PREPARE	Two sources with separate jobs, and a bias caught in time ROCCC on each. The integrity test exposed a survivorship bias that would have inverted the conclusion, forcing the deal-level window down to four seasons.
3 · PROCESS	SQL on DuckDB, with a reconciliation that must hold 175,165 rows down to 8,480 priced deals. Nothing needed correcting; every transformation is a selection, and the checks that came back empty are recorded anyway.
4 · ANALYSE	Five metrics, ten verification blocks Two findings survived unchanged, one was narrowed after failing a robustness check, and the hypothesis fell on both axes.
5 · SHARE	Five figures, five different forms One message per figure, each headline carrying a number. No form repeated from the previous case: matching form to question is part of the craft.
6 · ACT	Three recommendations, each with its risk and its assumption Plus one deliberately withheld: the Saudi channel is a real finding, but four seasons cannot separate a structural change from a spending cycle.
7 · PORTFOLIO	Published against a contract enforced by code The site build fails if a required field is missing. Aggregates cross; raw data stays in the repository.

The decisions that defined the case

Eight of the twenty-one recorded in **CASO.md**. The discarded alternative is the line that matters: it shows the choice was reasoned rather than reflexive.

The client is a multi-club fund, not a selling club

The question is about a gap between the two sides of the market, so the client has to be able to see both.

Discarded: An academy club that sells talent — it would only ever have seen its own side.

DuckDB as the engine, not BigQuery

The result has to rebuild from the repository alone. DuckDB reads the compressed CSVs directly and needs no server and no account.

Discarded: BigQuery — it would have forced a cloud account on anyone verifying the work.

Two sources, never joined at record level

FIFA is a census with no survivorship bias but never names a club; Transfermarkt names everyone but its coverage is uneven. Each covers the other's blind spot.

Discarded: A single source — Transfermarkt alone would have produced a confident, wrong decade trend.

Deal-level analysis restricted to four seasons

Coverage before 22/23 thins backwards and differentially by age, which is the very axis under study.

Discarded: The full 2010–2026 range, or a statistical correction — neither could be validated, and a decade of wrong-but-confident trend is worse than four honest seasons.

Two declared universes instead of one

Deals with only the buyer identified are 60% aged 18–23 at a quarter of the median price — the cheap purchases of young talent that are the phenomenon itself.

Discarded: Keeping only fully-identified deals: 93% of the money, and biased against exactly what the case set out to measure.

The youth-premium claim narrowed after verification

The reading "the ratio fell from 1.25 to 0.955" appeared only when cutting by season and hung on one starting observation.

Discarded: Publishing the fall — a better headline, and an artefact of specification.

Level claims from FIFA, "who bought from whom" from Transfermarkt

Our youth share runs 4–8 points below the census every year, because FIFA counts only international transfers.

Discarded: Trusting one source for both — it would have understated the youth share by up to 8 points.

Outliers investigated and kept

Bellingham at €127m, Gvardiol at €90m and Højlund at €79.8m are real deals and the subject of the case. The skew is handled with medians.

Discarded: Trimming the top percentile — statistically tidier, and it would delete the phenomenon.

The moment the case nearly went the other way

Survivorship bias, caught in phase 2 rather than discovered in phase 5

Transfermarkt rebuilds each player's history from the players present in its base *today*. Coverage therefore thins going backwards — and, critically, it thins **differentially by age**.

Taken at face value the source says the youth share of spending collapsed since 2010. That is not football: it is who survives in the database. The naive reading would have answered the **opposite** of the hypothesis, for reasons unrelated to the sport.

SEASON	PRICED DEALS	MEAN AGE	SHARE AGED 18–21
2010/11	143	20.8	53.8%
2013/14	372	21.5	44.9%
2017/18	936	23.2	27.5%
2021/22	1,343	23.9	25.5%
2022/23	1,746	24.4	19.9%
2025/26	2,367	24.4	21.2%

The real 2010/11 market had thousands of priced transfers, not 143. From 22/23 the series stabilises — mean age flat at 24.4 — so deal-level analysis uses those four seasons and every claim about a decade comes from FIFA's census instead. **Cross-checking two sources is what caught it**; a single source would have looked perfectly consistent.

What the data says

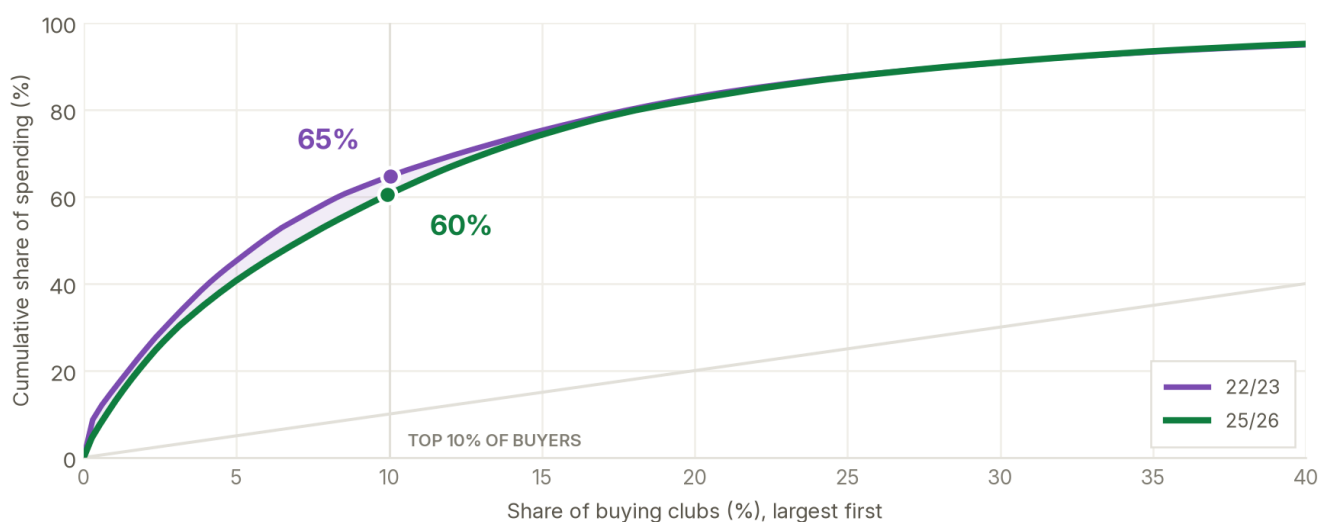
Each finding was written as a sentence carrying a number before any chart was drawn — the last check that it was closed rather than a topic.

1 · The money is spreading out, not concentrating

The money spread out instead of concentrating, the opposite of what was expected

Cumulative share of European transfer spending, clubs ordered from biggest buyer down. X axis trimmed to the first 40% of buyers, where the difference lives

2022/23 VS 2025/26 · 5,109 PRICED DEALS · BOTH CLUBS IDENTIFIED



Source: Transfermarkt via dcaribou/transfermarkt-datasets (CC0), snapshot 2026-08-14 · The straight line is a perfectly even split. The further a curve bulges above it, the fewer clubs hold the money. The hypothesis, written before the analysis, expected the bulge to grow.

JA

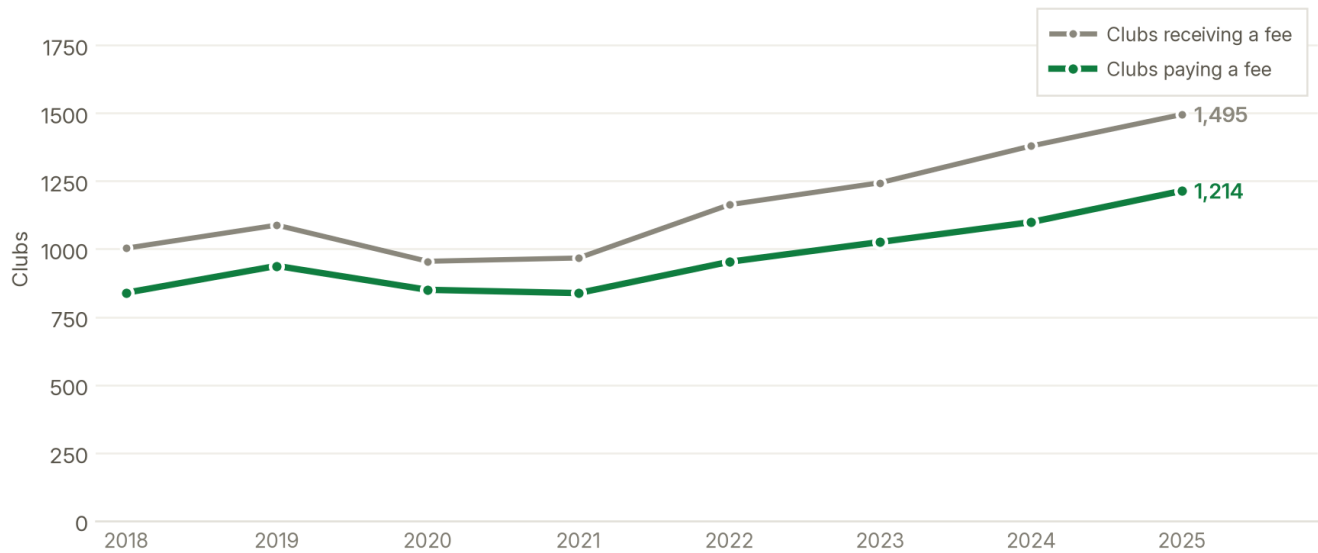
Top 10% of buyers: 65% of spending in 2022/23, 60% in 2025/26.

Top-10 share fell from 32.1% to 26.7% while buying clubs rose from 339 to 382. It holds without English buyers (30.6% → 24.0%), and FIFA's census backs it over eight years: **45% more clubs paying fees** than in 2018.

45% more clubs pay transfer fees than eight years ago

Number of clubs worldwide paying and receiving at least one transfer fee per year

2018-2025 · FIFA TMS CENSUS OF EVERY INTERNATIONAL TRANSFER



Source: FIFA Global Transfer Report (FIFA TMS), editions 2018-2025 · Our four seasons cannot carry a claim about a decade.
This series can: it is a census, not a sample, and carries no survivorship bias.

JA

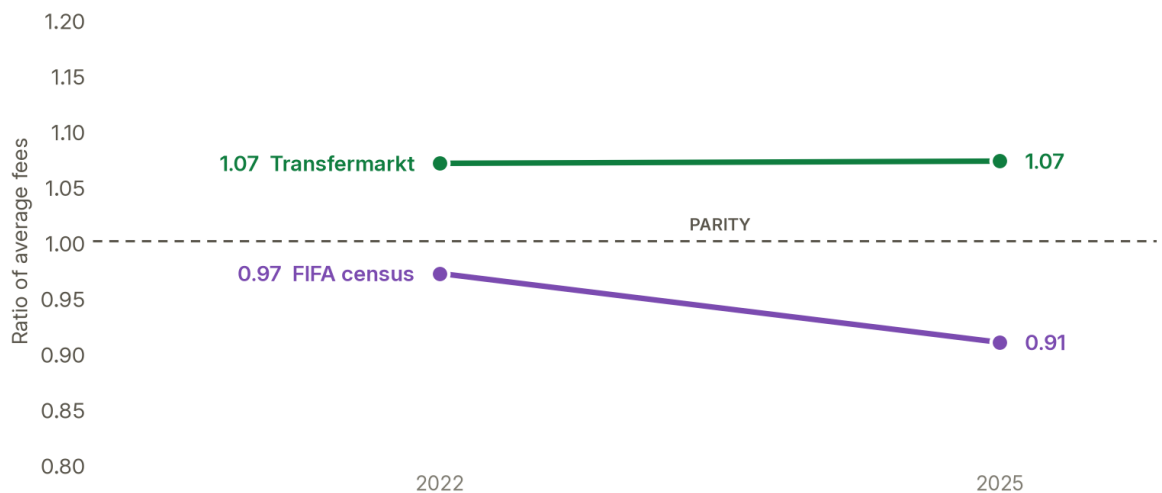
FIFA TMS census: clubs paying a fee up 45% since 2018, clubs receiving one up 49%.

2 · Youth carries no premium

A young player costs about what a peak-age one costs, and still does

Average fee for an 18-23 player divided by the average fee for a 24-29 player

2022 VS 2025 · CALENDAR YEARS · BOTH SOURCES, AVERAGES FOR COMPARABILITY



Source: FIFA Global Transfer Report (FIFA TMS), editions 2018-2025; Transfermarkt via dcaribou/transfermarkt-datasets (CC0), snapshot 2026-08-14 · Above the parity line the young player is dearer; below, cheaper. Levels differ because the universes do; neither moves far, and neither escalates.

JA

Ratio of average fees, 18-23 against 24-29, in both sources.

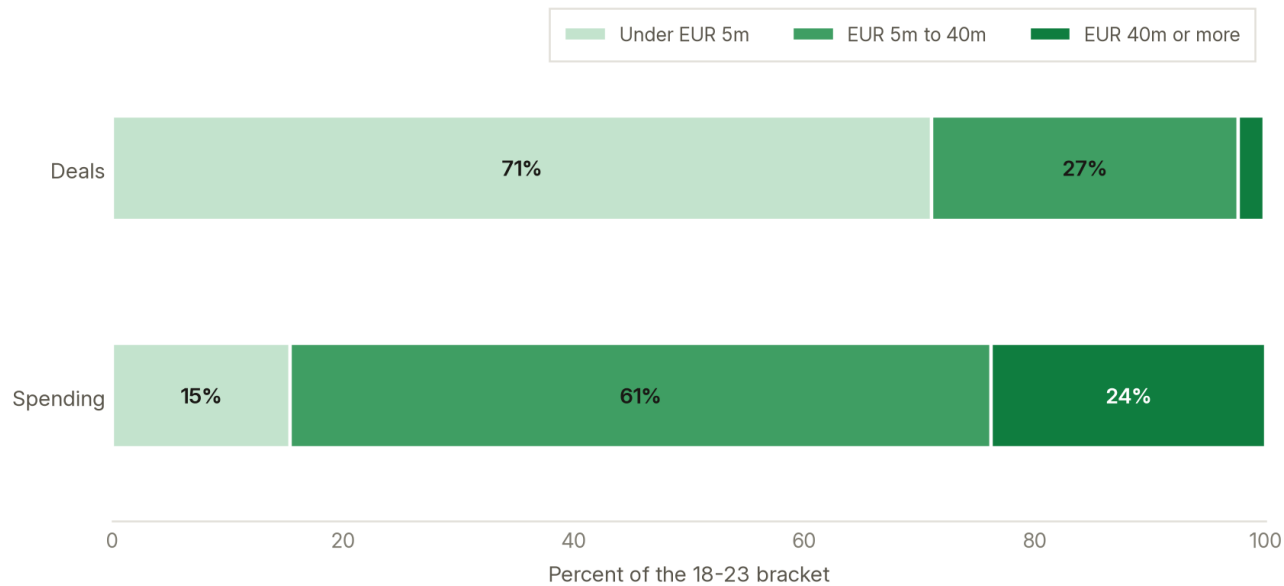
An 18-23 player costs about what a 24-29 player costs, and the ratio does not escalate in either source. Levels differ because the universes do; the direction is what matters, and neither rises.

3 · But the young market runs at two speeds

2% of young-player deals take a quarter of the money spent on them

Deals for players aged 18-23 by fee size: share of deals against share of spending

2022/23-2025/26 · 3,029 PRICED DEALS FOR 18-23 YEAR OLDS



Source: Transfermarkt via dcaribou/transfermarkt-datasets (CC0), snapshot 2026-08-14 · The headline signings are real, but they sit on top of a long cheap tail, and it is the tail that moves the aggregate.

JA

Share of deals against share of spending, players aged 18-23.

Deals of €40m or more are **2.3% of transactions and 23.8% of the money**; the 71% under €5m account for 15%. Seven of the ten largest buyers are English.

What each check ruled out

A finding that survives only one way of measuring it is a choice, not a result.

CHECK	WHAT IT RULED OUT	OUTCOME
V1 · Census sanity	That our youth share tracks FIFA	Gap widens to 7.1 pts in 2025 — no claim made on that axis
V1b · Both sources	That the finding depends on our data	Neither source shows escalation
V2 · Recomputation	An aggregation error	Identical from deal level
V3 · Denominators	A wrong base	Every share sums to 100 within its season
V4 · Without England	A Premier League artefact	Holds: 30.6% → 24.0%
V4b · Census decade	A four-season blip	+44.7% buyers over eight years
V5 · Effect size	Differences too small to act on	–5.4 points of concentration
V5b · Specification	A result that depends on how time is cut	It did — the claim was narrowed
V6a · Null model	That persistence is arithmetic	Both tails over-represented
V6b · Largest buyer	That one club drives the ratio	Barely moves without it

Conclusions, limits and what this case demonstrates

Recommendations

- Buy into a persistent net seller in a mid-tier league, not an elite club. Ajax banked €267m net across four seasons, Salzburg €234m; Chelsea (–€800m) and Manchester United (–€672m) never sold.
- Cap prices against the market band and write into the mandate that age carries no premium. A policy line, not a project.
- Make each club pick its lane — cheap volume or elite few.

What these data cannot answer

- Four seasons are not a trend. Every claim about direction over time rests on the census.
- The youth spending share is not claimable in level or direction: the two sources disagree and the 2025 gap doubles without a defensible explanation.
- Nothing here is causal, and net transfer income is not profit — without club accounts, "sells well" is not "makes money".
- Under-18s are not analysable at 2–7 priced deals a season.

What this case demonstrates

A hypothesis written down before the analysis and published unchanged when the data contradicted it, on both axes. Underneath it, the reason the answer is trustworthy: a data-quality decision taken in phase 2 rather than discovered in phase 5, by reconciling a scraped dataset against an official census instead of trusting either alone.

Full phase log, cleaning log with every discarded alternative, ROCCC source records and the ten verification blocks: [CASO.md](#) in the case repository — github.com/BobbyTarantino099/football-transfer-market