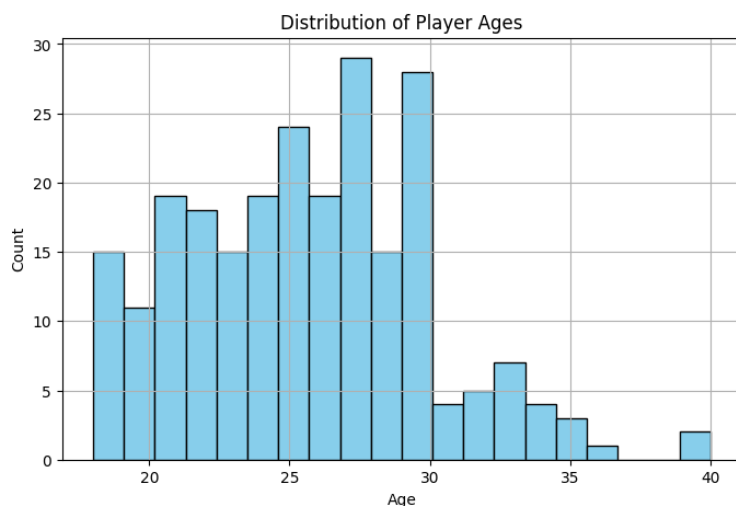


Sheffield United Technical Assessment

Section 1- Exploratory Data Analysis

What trends exist in the data with regards to player age and performance metrics?

This section explores how player age relates to key right-back (RB) performance outputs across the provided dataset. To keep comparisons fair, I focused on RB designations (DR/DMR) and applied a minute's floor ($\geq \sim 8$ full games) to reduce small-sample noise. Counting stats were converted to per-90 rates (e.g., assists/90, progressive carries/90) so age differences aren't confounded by playing time.



Defensive volume declines with age. The strongest negatives are Successful Tackles/90 ($r=-0.30$) and Interceptions/90 ($r=-0.21$), with medians stepping down from ≤ 21 through 31+. Physical defensive actions taper as players age.

Ball carrying eases with age.

Progressive Carries/90 ($r=-0.17$) and Ball Progression by Carrying/90 ($r=-0.16$) trend down after the mid-20s. Older RBs push forward with the ball slightly less.

Creative peak in mid-20s.

Open-play Key Passes/90 shows a mild decline with age ($r=-0.08$), while xA/90 peaks in the 25–27 band (0.068) compared with ≤ 21 (0.048), 22–24 (0.052), 28–30 (0.049) and 31+ (0.048). the most productive blend of athleticism and decision-making sits around 25–27.

Crossing output edges up with age.

Completed Crosses/90 has a small positive trend ($r=+0.14$) and is highest in the 31+ band. experience/selection effects-older RBs may attempt fewer but better crosses.

Possession security is broadly stable.

% Passing is essentially flat across ages ($r=-0.03$; medians ~ 79 –80%). passing accuracy doesn't materially deteriorate with age.

Threat shifts from carrying to passing.

xT/90 ($r=+0.18$) and xT Passing/90 ($r=+0.19$) rise with age despite fewer carries. older RBs may create comparable danger via positioning and pass selection.

Scoring contribution fades with age (as expected for RBs).

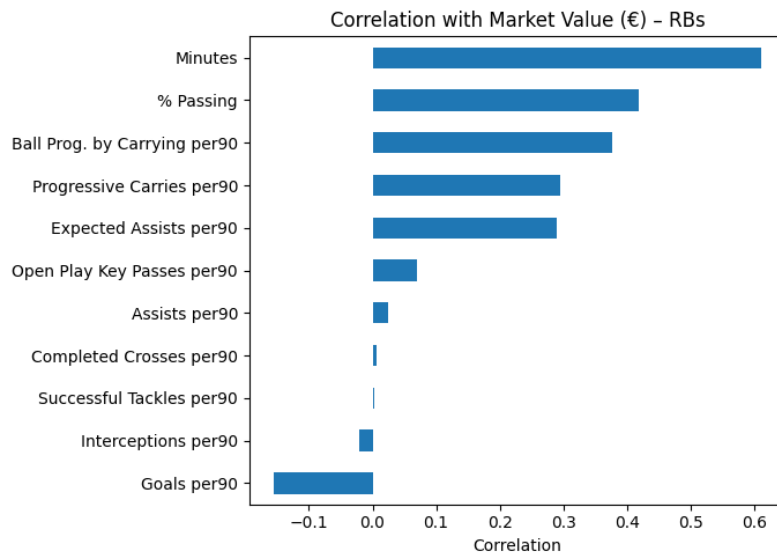
Goals/90 ($r=-0.18$) and Assists/90 ($r=-0.09$) trend down with age.

How do statistics such as pass completion percentage, goals and assists affect a player's market value?

This section examines which performance metrics most strongly relate to a player's market value (€) among right-backs (RBs).

I again restricted the sample to DR/DMR designations with a minute's floor ($\geq \sim 8$ full games) to reduce small-sample effects.

Counting stats were converted to per-90 rates so that playing time doesn't inflate correlations.



Availability dominates valuation.

Minutes played shows by far the strongest relationship with market value ($r = +0.61$).

Clubs value reliability and consistent selection-players trusted to play regularly command higher prices.

Technical security matters.

Pass completion (% Passing, $r = +0.42$) is the next-strongest correlate.

Ball-secure RBs are valued more highly in modern possession-based systems.

Progression is rewarded.

Ball Progression by Carrying/90 ($r = +0.38$) and Progressive Carries/90 ($r = +0.30$) both show notable positive links.

Full-backs who advance play are priced above purely defensive profiles.

Expected creativity signals quality better than realised assists.

Expected Assists/90 ($r = +0.29$) outperforms raw Assists/90 ($r = +0.02$).

Underlying chance creation carries more weight than one-off outputs.

Final-third passing adds some signal.

Open-Play Key Passes/90 has a mild positive correlation ($r = +0.07$), suggesting that regular involvement in attacking build-up modestly boosts value.

Defensive volume has little effect.

Successful Tackles/90 ($r = 0.00$) and Interceptions/90 ($r = -0.02$) are neutral or weakly negative.

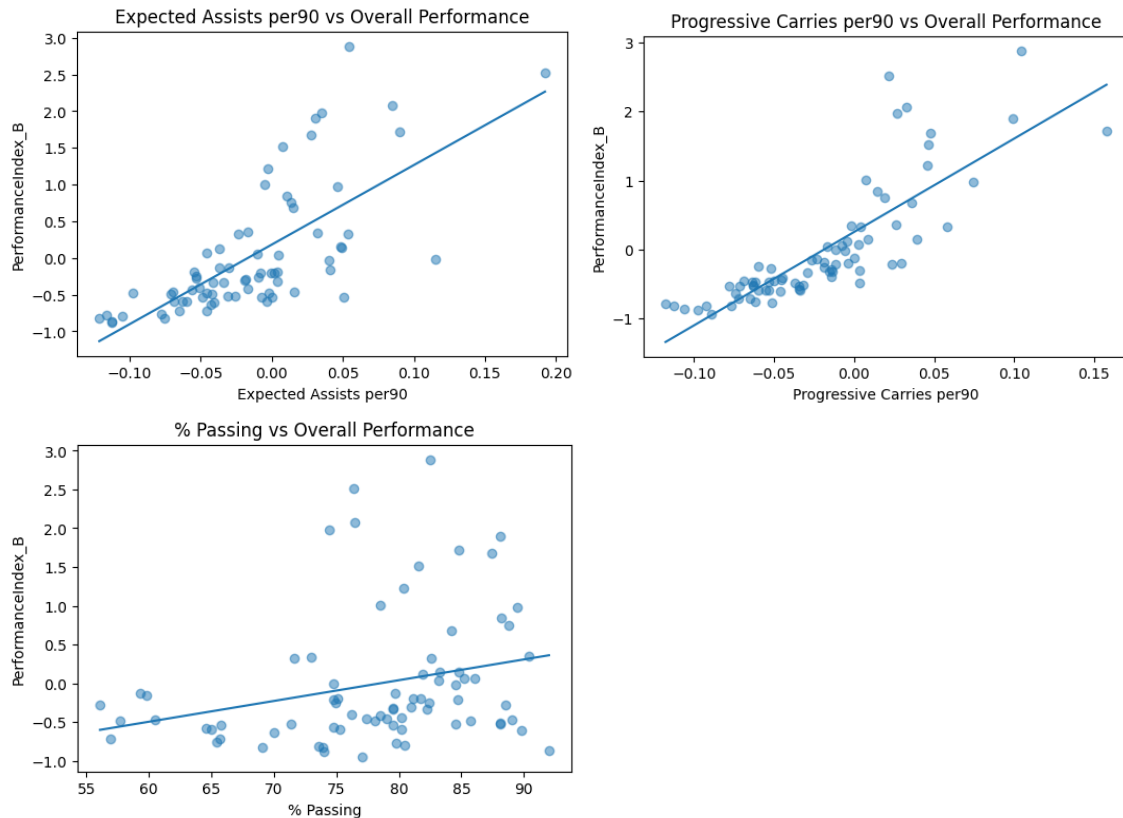
High-value RBs often play in teams that defend less frequently, so defensive counts don't drive price.

Scoring is largely irrelevant for RBs.

Goals/90 correlates slightly negatively ($r = -0.16$), reflecting role context - RBs aren't valued for goal output.

Explore which metrics correlate to high performance

This section builds a simple, role-aware overall performance index for RBs by combining key attacking, progression and defensive metrics (z-scored and averaged). I restricted to DR/DMR with a minutes floor (~8 full games) and converted counting stats to per-90 to avoid time-on-pitch bias. I then correlated individual metrics against this overall index.



Progression is the strongest signal of quality.

Ball Progression by Carrying/90 ($r = +0.87$) and Progressive Carries/90 ($r = +0.82$) show the tightest relationships with overall performance.

The top RBs consistently advance play with the ball.

Both sides of the ball matter.

Successful Tackles/90 ($r = +0.75$) and Interceptions/90 ($r = +0.67$) correlate strongly, indicating high-performing RBs pair progression with defensive output.

Creative involvement is a key differentiator.

Open-Play Key Passes/90 ($r = +0.72$) and Expected Assists/90 ($r = +0.70$) rank highly.

Sustained chance creation aligns with overall impact better than raw assists.

Technical security helps but is secondary.

% Passing shows a moderate positive link ($r = +0.27$).

Possession reliability supports performance but doesn't define it.

Raw box outputs are weak or noisy for RBs.

Assists/90 ($r = -0.11$) and Goals/90 ($r = -0.15$) are poor standalone indicators once broader contributions are considered.

Section 2 - Recruitment Scenario

Sheffield United is preparing to recruit a new right-back for the 2025/26 season. Your task is to design a clear data-driven methodology for identifying the most suitable players from the dataset for scouts to prioritize for further evaluation.

Before building a data-driven shortlist, it's important to frame the analysis around Sheffield United's tactical context under Chris Wilder, who returned to the club in 2023. Wilder typically deploys a 3-5-2 / 5-3-2 system featuring overlapping centre-backs and wing-backs providing attacking width. His sides prioritise direct, vertical play, rapid transitions, and cross-based attacks rather than sustained possession. The wide defenders are therefore required to carry the ball forward, combine in the final third, and recover quickly when possession is lost.

For a right-back (or right-sided wing-back), this means a balanced blend of progressive carrying, crossing accuracy, and defensive reliability in wide channels. While raw defensive volume may appear lower due to the structure, Wilder's system rewards players who can progress play under pressure, deliver quality balls into the box, and maintain tactical discipline when the team drops into a compact block.

With that tactical context established, the recruitment framework was designed to identify right-backs who best align with Wilder's approach.

The model translates these tactical demands into measurable data pillars - Build-up, Creation, and Defending - each derived from objective performance metrics within the dataset.

Every player is standardised on a per-90 and league-adjusted basis to ensure fair comparison across leagues and roles.

The final output ranks players not only on technical suitability but also on practical feasibility, giving scouts a clear, data-backed shortlist to prioritise for further evaluation.

The recruitment model translates tactical fit into a structured, transparent scoring framework. Each player is evaluated through a sequence of quantitative steps designed to balance football relevance, data robustness, and scout usability.

The process follows five main stages:

Eligibility filtering – restrict the dataset to realistic right-back or right wing-back candidates (DR/DMR) with sufficient playing time (≥ 1000 minutes) to ensure data reliability.

Normalisation – convert counting stats to per-90 values and apply league-aware z-scores, standardising each metric relative to its competition to offset league strength bias.

Pillar scoring – summarise player performance into three weighted pillars:
Build-up (40%) – progression, composure under pressure, and passing quality.
Creation (35%) – chance generation, delivery, and attacking output.
Defending (25%) – 1v1 reliability, interceptions, and aerial effectiveness.

Bonuses and feasibility adjustments – apply small, transparent bonuses for ideal age (20–27) and high minutes (≥ 1800), then combine football performance with feasibility factors such as market value and contract expiry.

Ranking and sensitivity testing – produce final Recruitment and Priority scores, then re-rank under varied weights to test robustness and highlight consistent top performers.

Results:

The model outputs two complementary measures:

Recruitment Score — pure football suitability based on tactical fit and on-pitch performance.

Priority Score — the same value adjusted for feasibility (market value, contract, and age), showing which players are most practical for Sheffield United to pursue.

Top-Ranked Candidates

The shortlist highlights a mix of Premier League benchmarks and high-value lower-league options, reflecting the balance between performance excellence and recruitment realism:

Name	Team	Age	Minutes	(£) Market Value	BuildUp	Creation	Defending	Overall	Bonuses	Recruitment_Score	Feasibility_Score	Priority_Score	Flags
Fraser Murray	Wigan Athletic	26	1010	375	-0.34	2.11	0.08	0.62	0.1	0.72	0.81	0.58	Low minutes
Pedro Porro	Tottenham	26	2950	38000	0.28	1.71	0.09	0.73	0.15	0.88	0.25	0.22	
Matheus Nunes	Manchester City	27	1881	35000	1.13	0.1	0.31	0.57	0.15	0.72	0.26	0.19	
Amad Diallo	Manchester United	23	2127	45000	1.39	0.89	-0.38	0.77	0.15	0.92	0.18	0.16	
Remeao Hutton	Gillingham	27	1032	275	-1.47	1.89	0.18	0.12	0.1	0.22	0.72	0.16	Low minutes
Tom Barkhuizen	Barrow	32	1032	300	0.04	0.83	-0.33	0.22	0	0.22	0.71	0.16	Low minutes
Kieran Trippier	Newcastle	35	1465	4000	0.23	0.74	-0.16	0.31	0	0.31	0.47	0.14	
Mats Wieffer	Brighton	25	1233	25000	0.29	-0.37	1.45	0.35	0.1	0.45	0.29	0.13	
Aaron Wan Bissaka	West Ham United	27	3551	24000	1.12	0.36	-0.03	0.56	0.15	0.71	0.18	0.13	
Kyle Walker Peters	Southampton	28	3265	15000	1.49	-0.07	-1.2	0.27	0.05	0.32	0.34	0.11	Progressive; 1v1 risk
Malo Gusto	Chelsea	22	2164	35000	0.51	0.05	0.59	0.37	0.15	0.52	0.2	0.1	
Valentino Livramento	Newcastle	22	3187	40000	0.79	-0.13	-0.25	0.21	0.15	0.36	0.25	0.09	

Recruitment Score = football fit in Wilder's model (Build-up 40%, Creation 35%, Defending 25%) + small age/minutes bonuses.

Feasibility Score = practicality (log market value, contract, GBE).

Priority Score = Recruitment × Feasibility → who we should look at first.

Fraser Murray (Wigan) — Priority 0.58 (Recruitment 0.72 × Feas 0.81). Creation pillar is a genuine outlier; very affordable. Flag: low minutes → verify availability and repeatability of crossing output.

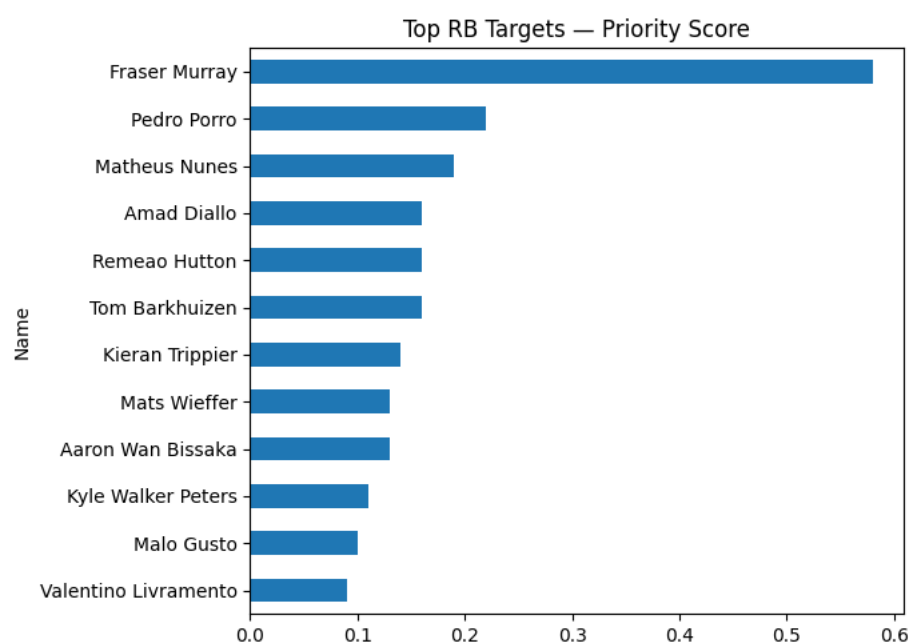
Pedro Porro (Spurs) — Priority 0.22 with elite Recruitment 0.88; benchmark for the desired style (ball progression + delivery). Feasibility moderates ranking.

Matheus Nunes (Man City) — Priority 0.19; strong carrier, balanced profile; feasibility limits.

Amad Diallo (Man Utd) — Priority 0.16; high Build-up/Creation, younger profile; costly.

Remeao Hutton (Gillingham) / Tom Barkhuizen (Barrow) — value creators who rise on feasibility; both carry low-minutes flags → verify physical robustness and defensive transition clips.

Kyle Walker-Peters — flagged “Progressive; 1v1 risk” → targeted video on wide 1v1s and recovery runs.



Weight shifts didn't overturn the core ordering

Creation ↑ (+10%): Murray remains #1; Hutton and Barkhuizen move up (creation-led).

Defending ↑ (+10%): Murray still #1; Wieffer rises (defensive pillar), while creation-only profiles slip slightly. This indicates the shortlist is stable, not overly sensitive to small preference changes.

What the model is rewarding (and why it fits SUFC)

Build-up/progression (e.g., progressive carries, ball progression by carrying) – aligns with Wilder's direct, vertical wide play.

Creation/delivery (xA/90, key passes/90, completed crosses/90, cross efficiency) – reflects cross-based final actions.

Defending the flank (tackles/90, interceptions/90, 1v1 proxy) – ensures recovery and wide compactness.

Feasibility then converts quality into actionability, pushing attainable EFL options above elite, expensive benchmarks without hiding the latter's value as style references.

Scout plan (who to watch, and for what)

Tier A – Immediate priority (high Priority_Score)

Fraser Murray — confirm repeatability of creation vs stronger opposition; clips: early crosses under pressure, underlaps/cut-backs, defensive recovery after overlap.

Remeao Hutton — cross quality vs set defence, decision-making in transition, defensive body shape 1v1.

Tom Barkhuizen — similar checklist; add stamina/90 and late-game actions due to minutes flag.

Tier B – Stretch/benchmark quality (high Recruitment_Score, moderate feasibility)

Pedro Porro, Matheus Nunes, Amad Diallo — use as style comps; if market shifts (loan/contract dynamics), re-assess feasibility quickly.

Tier C – Role-specific fits / monitor

Mats Wieffer — defending weight boosts him; good if we bias to a more conservative RWB.

Kyle Walker-Peters — excellent build-up; verify 1v1 and back-post defending before escalation.

Malo Gusto / Livramento — younger PL profiles; keep on watchlist for loans or late-window opportunities.

Conclusion

The methodology converts tactical requirements into measurable pillars, ranks on football value, and then prioritises by feasibility. The resulting shortlist surfaces actionable targets (Murray, Hutton, Barkhuizen) alongside elite style benchmarks (Porro, Nunes, Diallo). This gives scouts a clear, data-backed order of work, plus targeted video checks to validate fit before live scouting or engagement.