Football Player Data Analysis

# Introduction

I started this project to showcase an end-to-end analysis of a small dataset exported from my own personal Football Manager save. I chose this project due to my personal enjoyment for the game itself, as well as the idea of “Moneyball”. I used MySQL and Python to achieve the outcome below. The initial dataset included detailed stats such as player age, nationality, club, goals. expected goals xG), average match rating and a whole load more data points. Key insights were derived from structured queries and then used to present simplistic yet professional data visualizations.

# Dataset Origin

As mentioned, the data set used for this report was exported from a personal Football Manager save. This was achieved by using in-game tools such as custom search views and filters. It includes real simulated stats of players across clubs and leagues. I believe this provides a realistic dataset for performance analysis such as this.

# MySQL Setup & Data Import

The player data was imported into a MySQL schema named `football\_data`. A table called `players\_data` holds all player statistics. All queries were executed in MySQL Workbench and visualized with Python using matplotlib.

# SQL Analysis Queries & Insights

## Top Scorers

***SELECT Name, Club, Position, Gls, `Av Rat` FROM players\_data ORDER BY Gls DESC LIMIT 10;***

This query identifies the top 10 players based on goals scored. It highlights key attackers who are high-impact players, such as Krystian Brzozowski, who leads the list. Useful for scouting goal-scoring efficiency.

## Top Rated Players

***SELECT Name, Club, `Av Rat`, Gls, xG FROM players\_data ORDER BY `Av Rat` DESC LIMIT 10;***

This query ranks players by average match rating. Nathan Gray and Huw Anderson top the list, suggesting consistent match performance. This is useful to assess all-round contribution, not just goals.

## Under 21 Top Performers

***SELECT Name, Age, Club, Position, Gls, `Av Rat` FROM players\_data WHERE Age < 21 ORDER BY `Av Rat` DESC LIMIT 10;***

This query isolates top-performing young players. Huw Anderson and Erik Hadenius are standout U21 players with strong average ratings. A valuable tool for youth scouting and academy focus.

## xG vs Goals

***SELECT Name, Gls, xG, (Gls - xG) AS GxDiff FROM players\_data ORDER BY GxDiff DESC LIMIT 10;***

This query compares expected goals to actual goals. Huw Anderson significantly outperformed his xG, suggesting exceptional finishing ability. It can help identify clutch finishers or players who underperform.

## Goals Per Game

***SELECT Name, Gls, Apps, ROUND(Gls / Apps, 2) AS GoalsPerGame FROM players\_data WHERE Apps > 0 ORDER BY GoalsPerGame DESC LIMIT 10;***

This measures scoring efficiency. Noah Adam leads with a perfect 1.0 ratio, indicating high impact per match despite fewer appearances. Important for assessing impact subs or rotational players.

## Average Age by Position

***SELECT Position, ROUND(AVG(Age), 1) AS AvgAge FROM players\_data GROUP BY Position;***

This analyzes squad age structure. Defensive positions tend to skew older (e.g., RC = 29.5), while midfielders are generally younger. Useful for long-term squad planning.

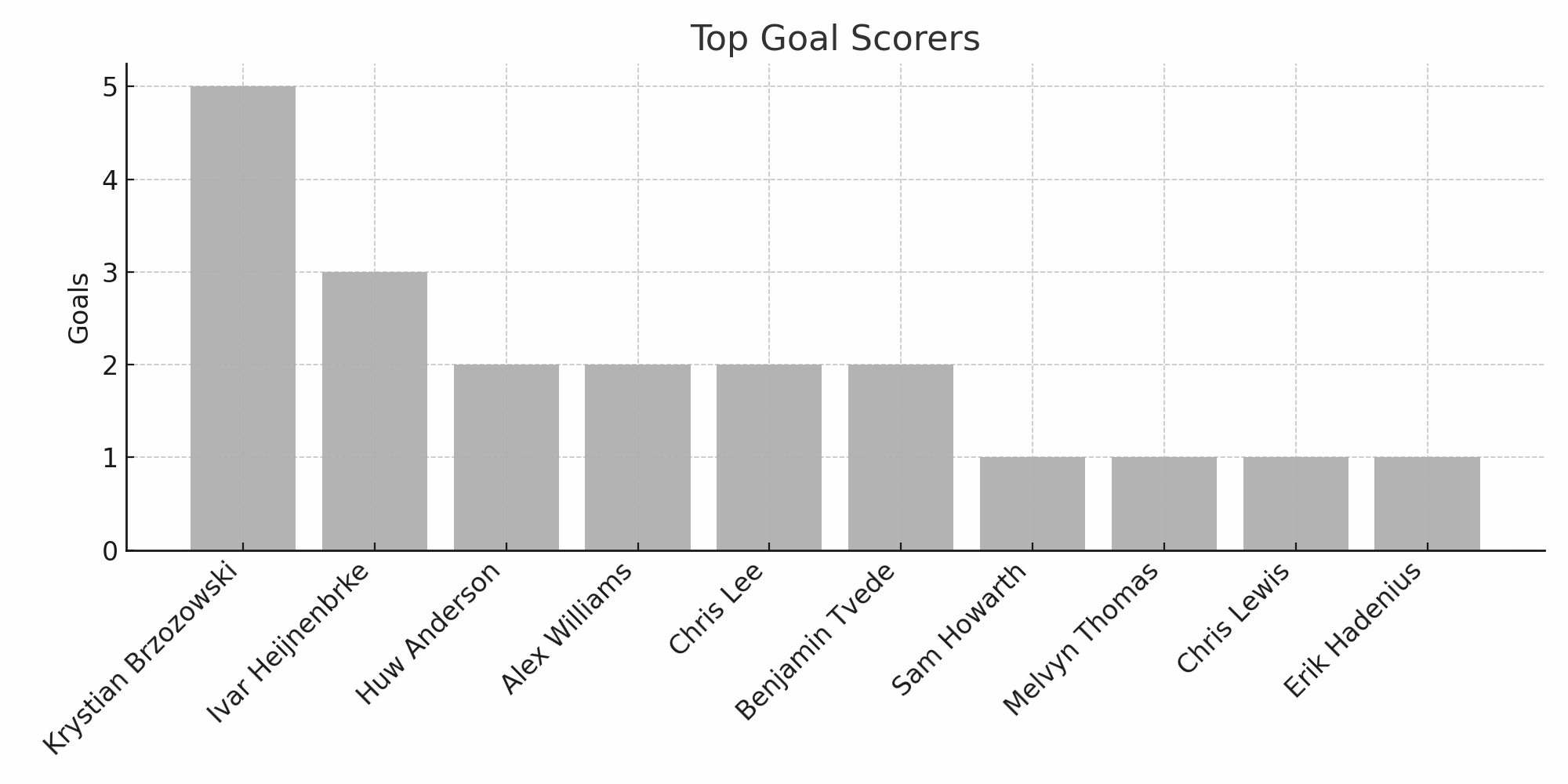
## Top Clubs by Avg Rating

***SELECT Club, ROUND(AVG(`Av Rat`), 2) AS AvgRating, COUNT(\*) AS NumPlayers FROM players\_data GROUP BY Club ORDER BY AvgRating DESC LIMIT 10;***

Clubs like Airbus UK and Colwyn Bay have the highest average-rated players, albeit with fewer data points. Djurgårdens IF appears with more players, indicating broader squad depth. This helps identify high-performance clubs for scouting focus.

# Data Visualizations

## Top Goal Scorers



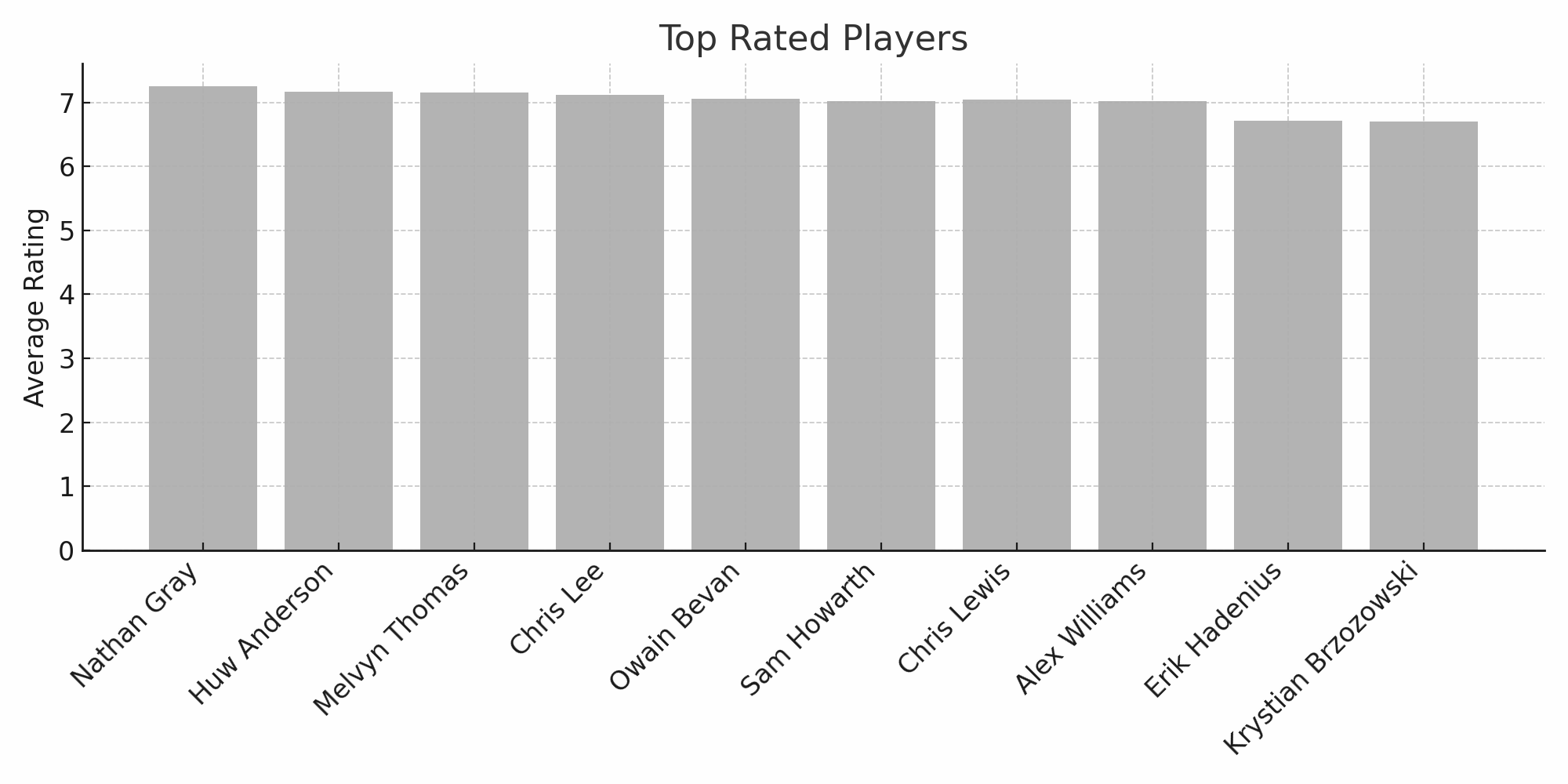
## 📊 Interpretation

This bar chart visualizes the players with the highest number of goals. Krystian Brzozowski leads the list, indicating his role as a primary attacker. The sharp goal count gap between the top player and the rest suggests individual dependency on a few key scorers.

## ✅ Recommendations

Clubs relying heavily on one scorer should consider supporting cast training or recruiting additional attacking depth to balance offensive output.

## Top Rated Players



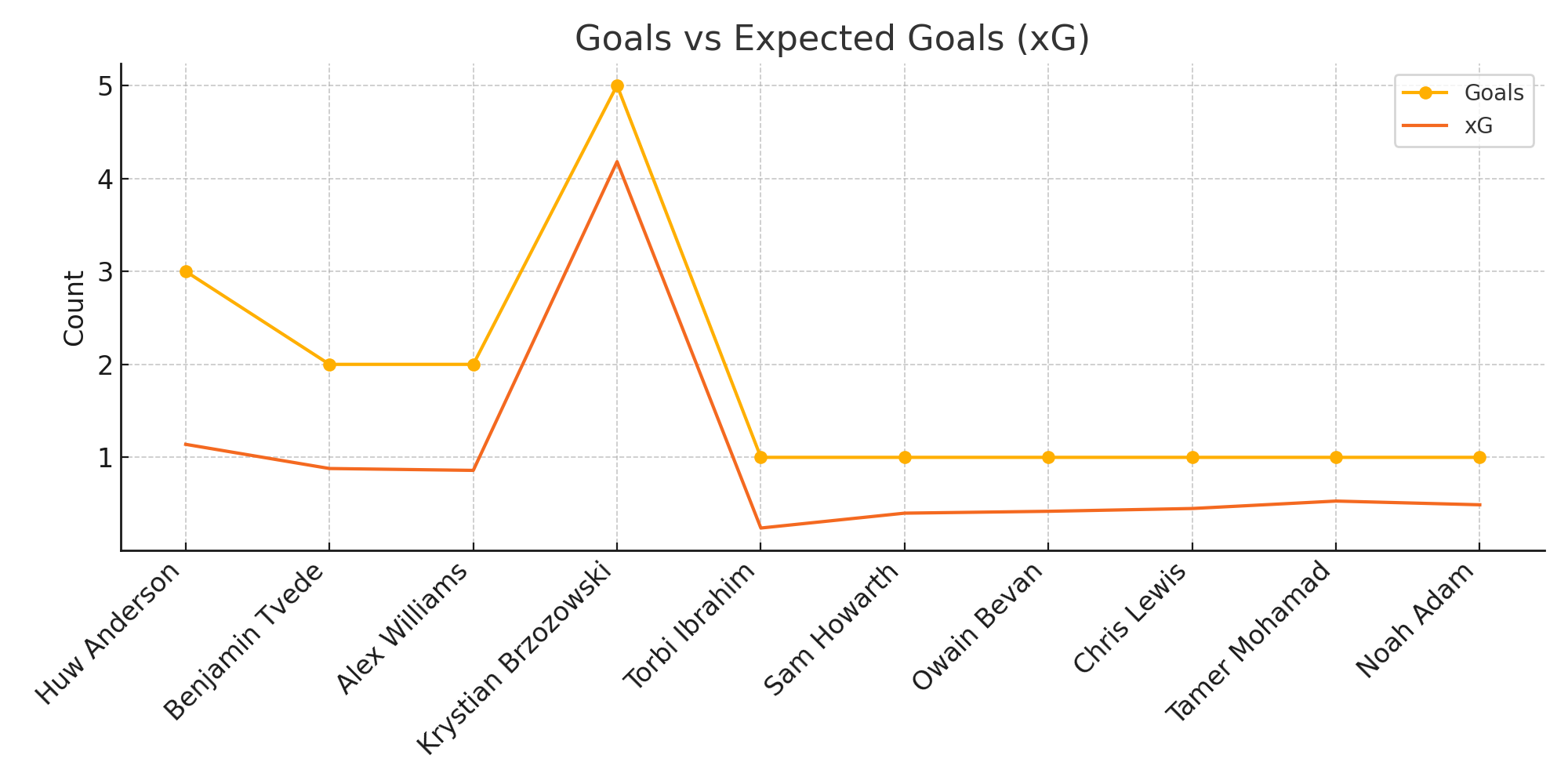
## 📊 Interpretation

This chart shows players with the highest average match ratings. These ratings reflect all-around performance, not just goal scoring. Players like Nathan Gray and Huw Anderson consistently perform across all areas.

## ✅ Recommendations

High-rated players should be retained and considered for leadership roles. Clubs may also want to analyze what contributes to their consistency for replication across the squad.

## Goals vs Expected Goals



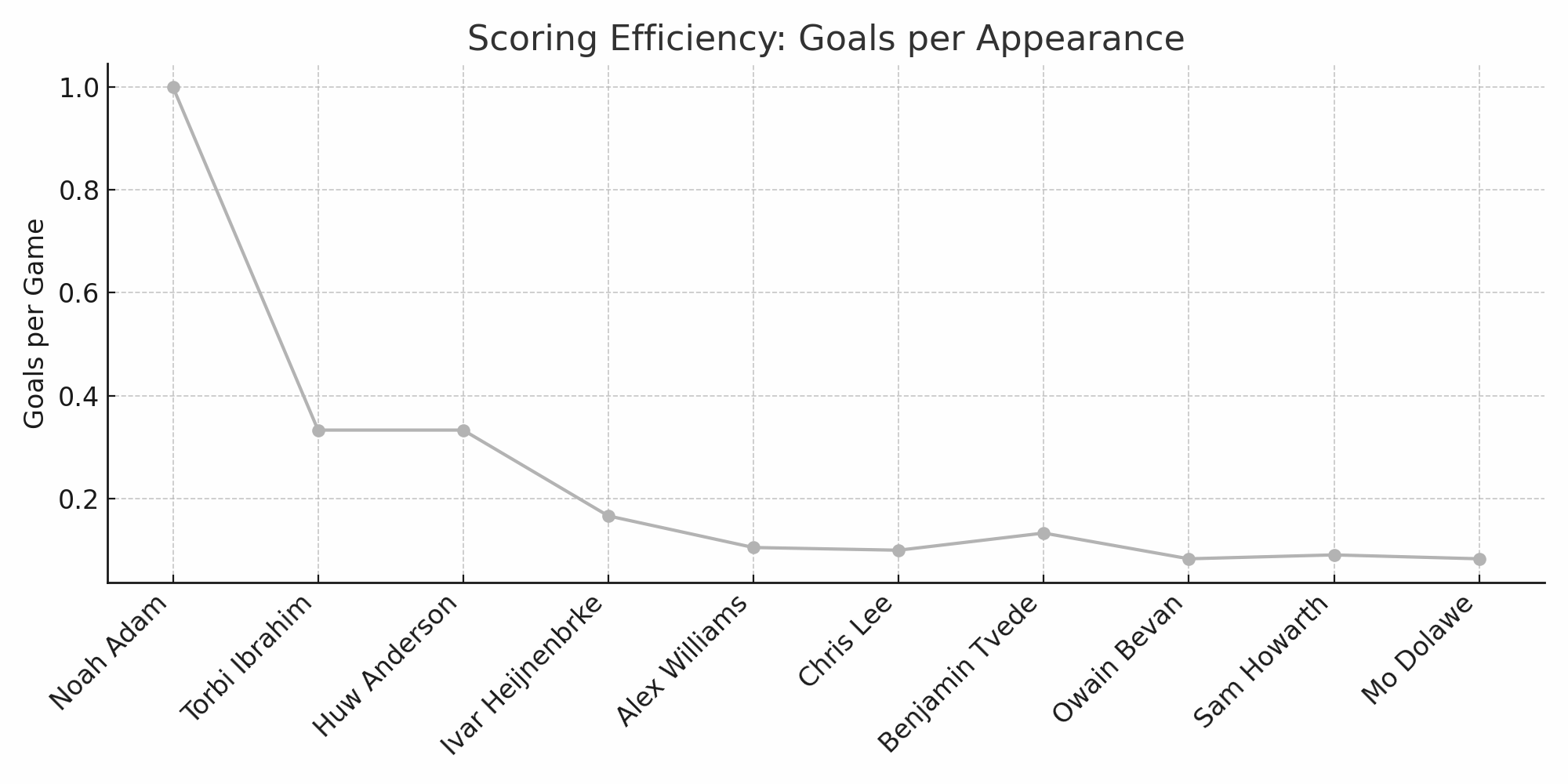
## 📊 Interpretation

This line chart compares actual goals to expected goals (xG). Huw Anderson's performance well above xG suggests elite finishing skill. Conversely, players underperforming their xG may indicate inefficiency or poor shot selection.

## ✅ Recommendations

Focus on training underperformers or reevaluating their roles. Overperformers might deserve more playtime or responsibility in high-stakes matches.

## Goals per Game Efficiency



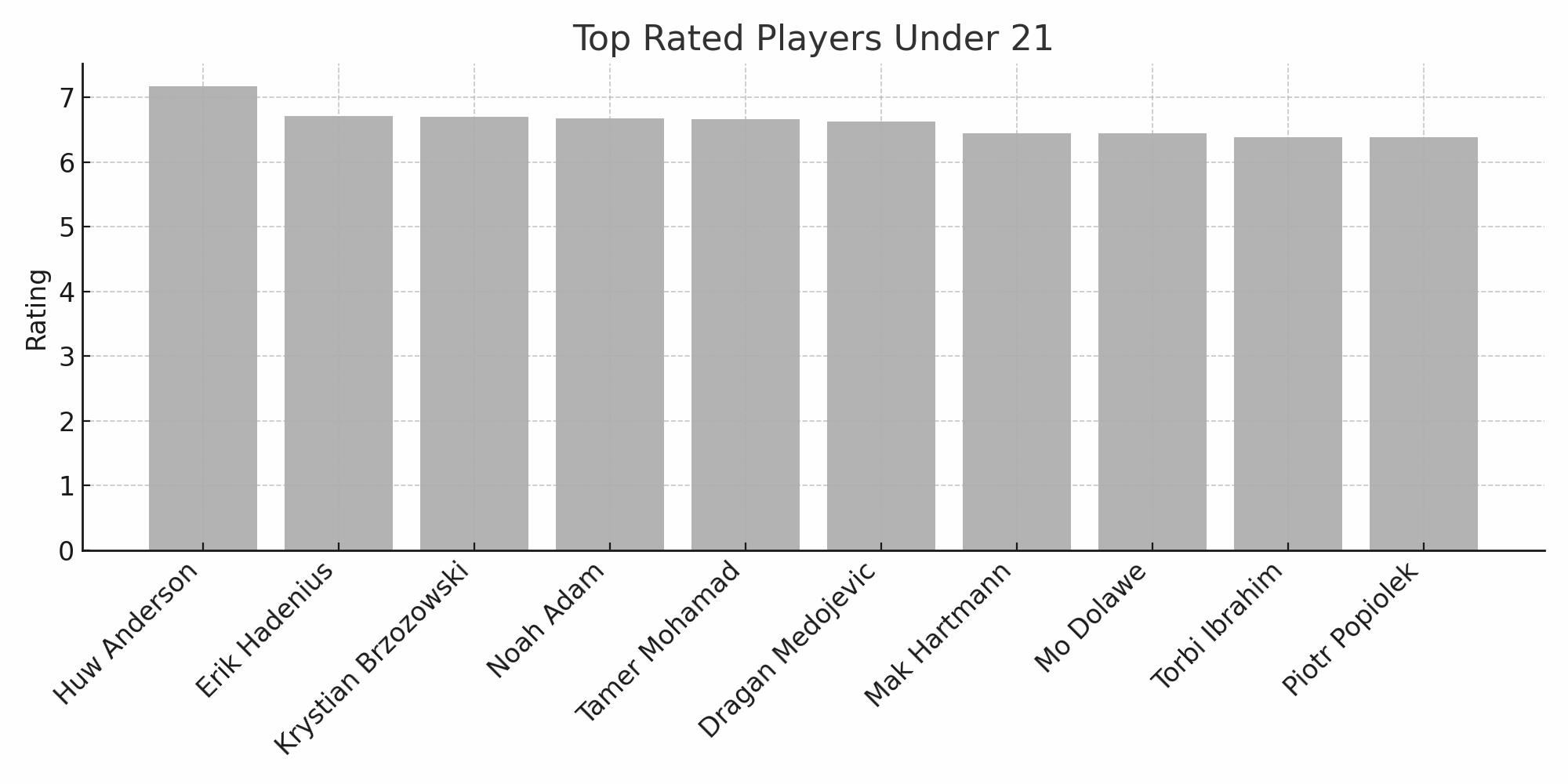
## 📊 Interpretation

The graph highlights players who score efficiently relative to appearances. High goals-per-game ratios spotlight impactful players who make the most of limited minutes.

## ✅ Recommendations

Such players should be prioritized for starting roles or used strategically in tight matches. This also helps in squad rotation planning.

## Top Rated Players Under 21



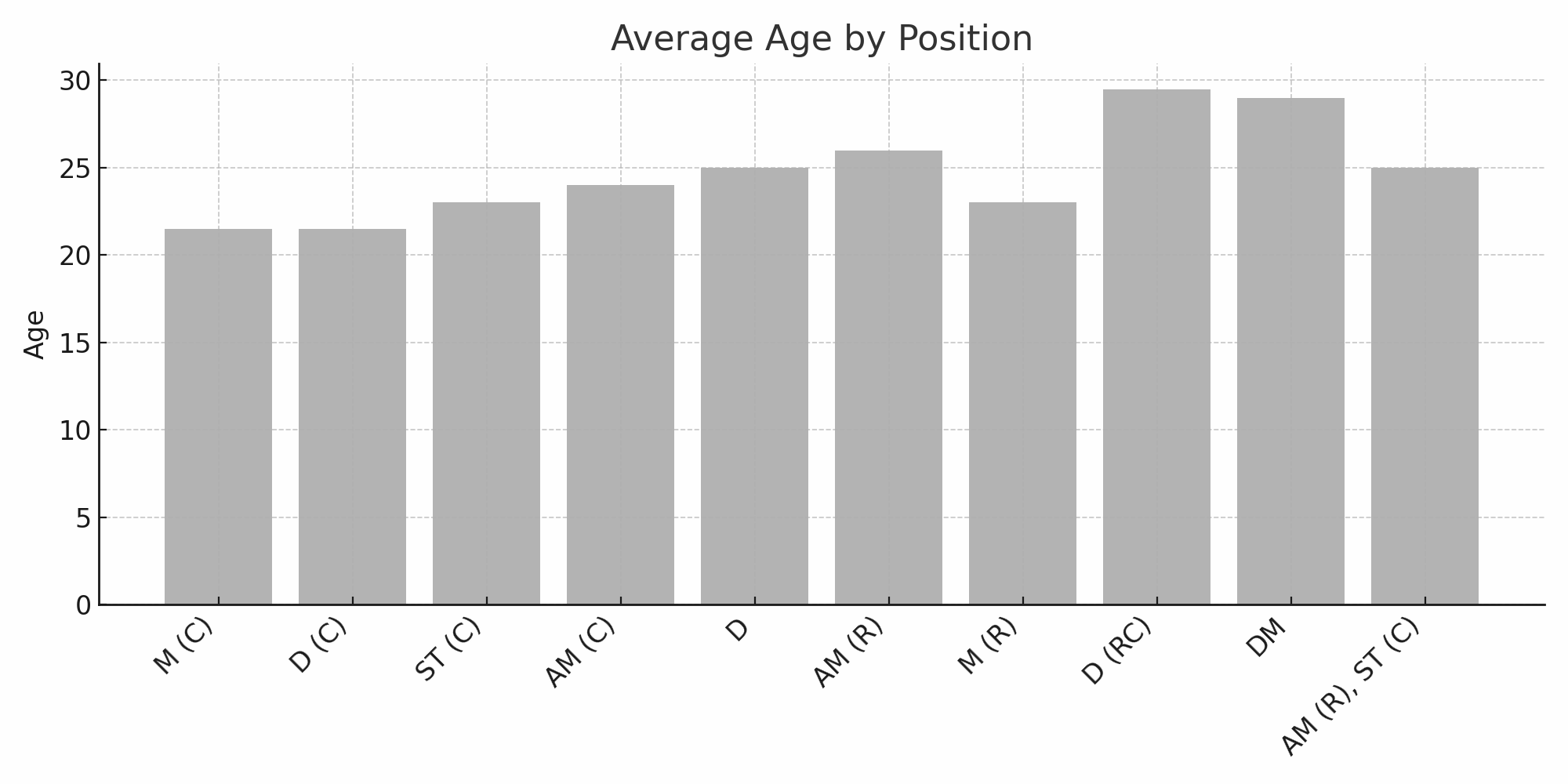
## 📊 Interpretation

This bar chart presents the top young talents based on rating. It shows the presence of U21 players who can perform at senior levels. Identifying future stars like Huw Anderson ensures proactive talent development.

## ✅ Recommendations

These players should receive focused training, mentorship, and gradual exposure to tougher matches to groom future team pillars.

## Average Age by Position



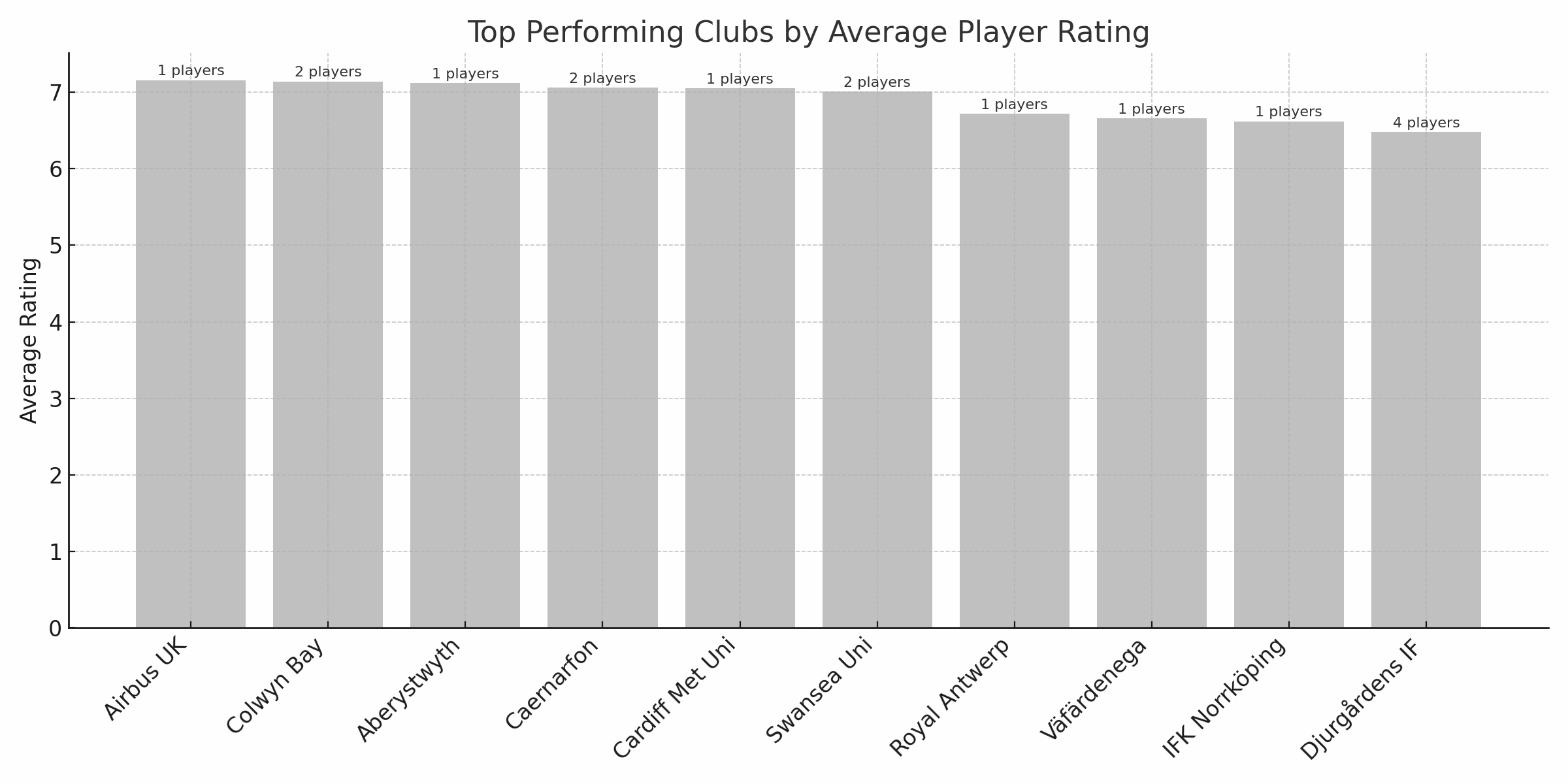
## 📊 Interpretation

This chart reveals the age profile by position. Defensive roles generally have older averages, which may reflect experience dependence. Midfield and forward roles skew younger.

## ✅ Recommendations

Helps in planning succession. Older defensive lines might need replacements in the next transfer window.

## Top Performing Clubs by Avg Rating



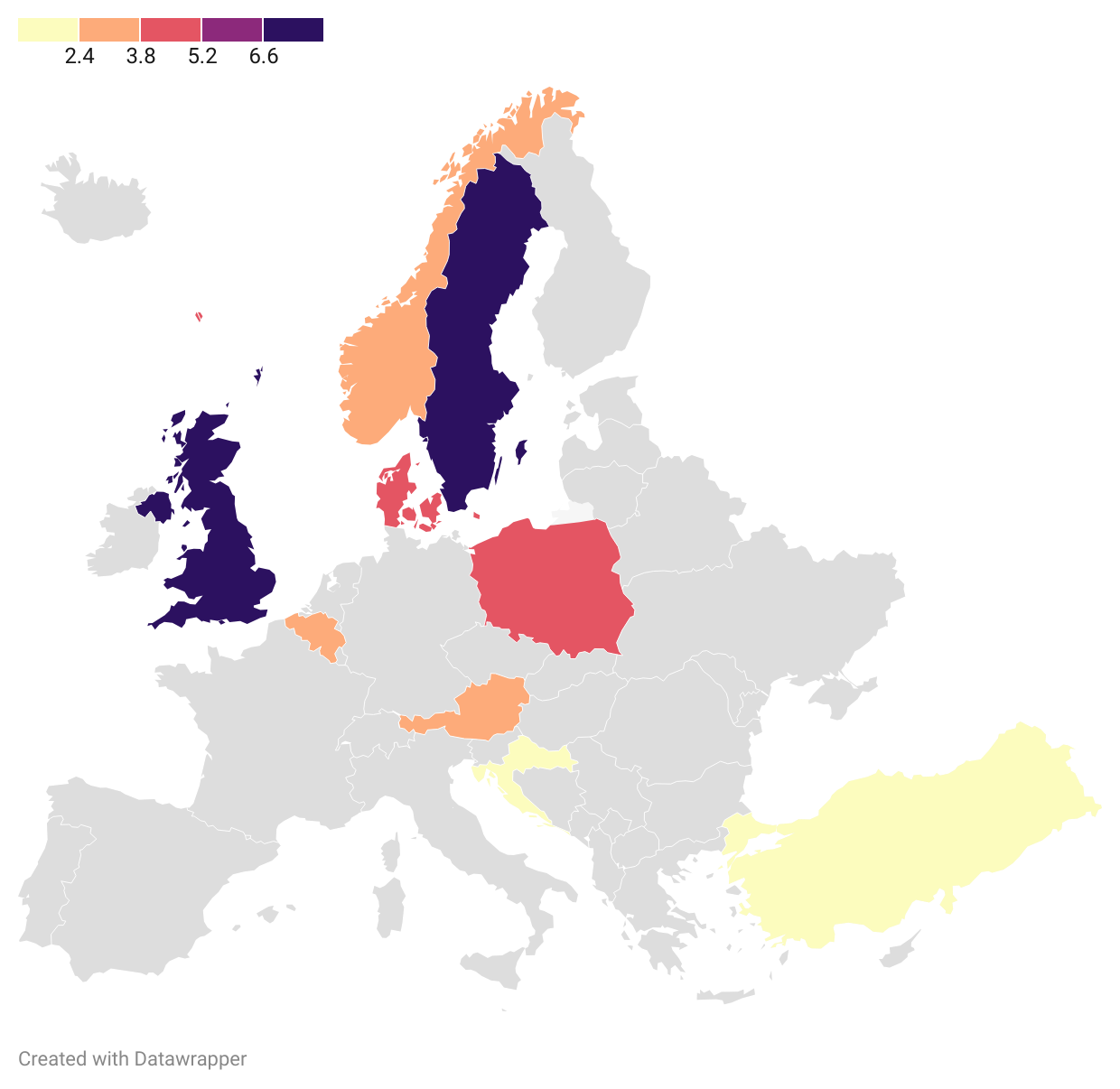
## 📊 Interpretation

This chart highlights clubs with top average player ratings. Airbus UK, Colwyn Bay, and Aberystwyth rank high—suggesting star individuals. Djurgårdens IF appears with depth (4 players), indicating a balanced, capable squad.

## ✅ Recommendations

Clubs with high-rated individuals should build systems around them. Those with overall high averages (like Djurgårdens IF) might explore more ambitious competitive goals.

# Nationality Distribution Map

Due to technical rendering limitations, the nationality map was created using Datawrapper. The map illustrates the concentration of players across European countries, highlighting strong national representation in the UK and Scandinavia. This provides insight into international player sourcing and national player pool trends.

# Practical Use Case: Striker Replacement Analysis

A real-world simulation was run to replace a key striker. His attributes were extracted from a Football Manager save, emphasizing strength, heading, jumping reach, and finishing. SQL was used to rank candidates using a MatchScore calculation based on how closely each player's attributes matched the original striker.

## SQL MatchScore Query

***SELECT Name, Club, Position, Age, Gls, `Av Rat`, Hea, Fin, Str, Jum,  
 (Fin + Hea + Str + Jum + (`Av Rat` \* 10)) AS MatchScore  
FROM players\_data  
WHERE Position LIKE '%ST%'  
ORDER BY MatchScore DESC  
LIMIT 10;***

## Interpretation & Recommendations

Tamer Mohamad was the highest scoring candidate based on the scoring algorithm. He exhibited strong physical and technical stats. Other players showed promise and were flagged for developmental potential. This approach simulates real-world recruitment scenarios using in-game data

