M1. Individual Task

Attacking Output: Real vs Expected - Big 5 Leagues 21/22 by Manuel Fernández

Student: Manuel Fernández
Course: PREMSC6 FFA
Date: 30/08/2025

Introduction

In this assignment, an interactive dashboard is developed in Power BI using the dataset *FBREF_Big5_2021_22_Players.csv*, which compiles detailed information on players from the five major European leagues during the 2021/22 season. The main objective is to provide a visual tool to analyze the offensive performance of footballers, combining both traditional and advanced metrics such as goals, assists, expected goals (xG), and expected assists (xAG).

Through a visual and interactive approach, the dashboard aims to facilitate the identification of performance patterns, compare players across different leagues or positions, and highlight those who perform above expectations based on their underlying metrics.

In particular, the dashboard focuses on three key aspects: the relationship between actual and expected production, the detection of players who stand out above their expected performance, and the distribution between finishing and creation roles. As such, it provides a practical tool for individual performance analysis, scouting, and strategic decision-making in the sports domain.
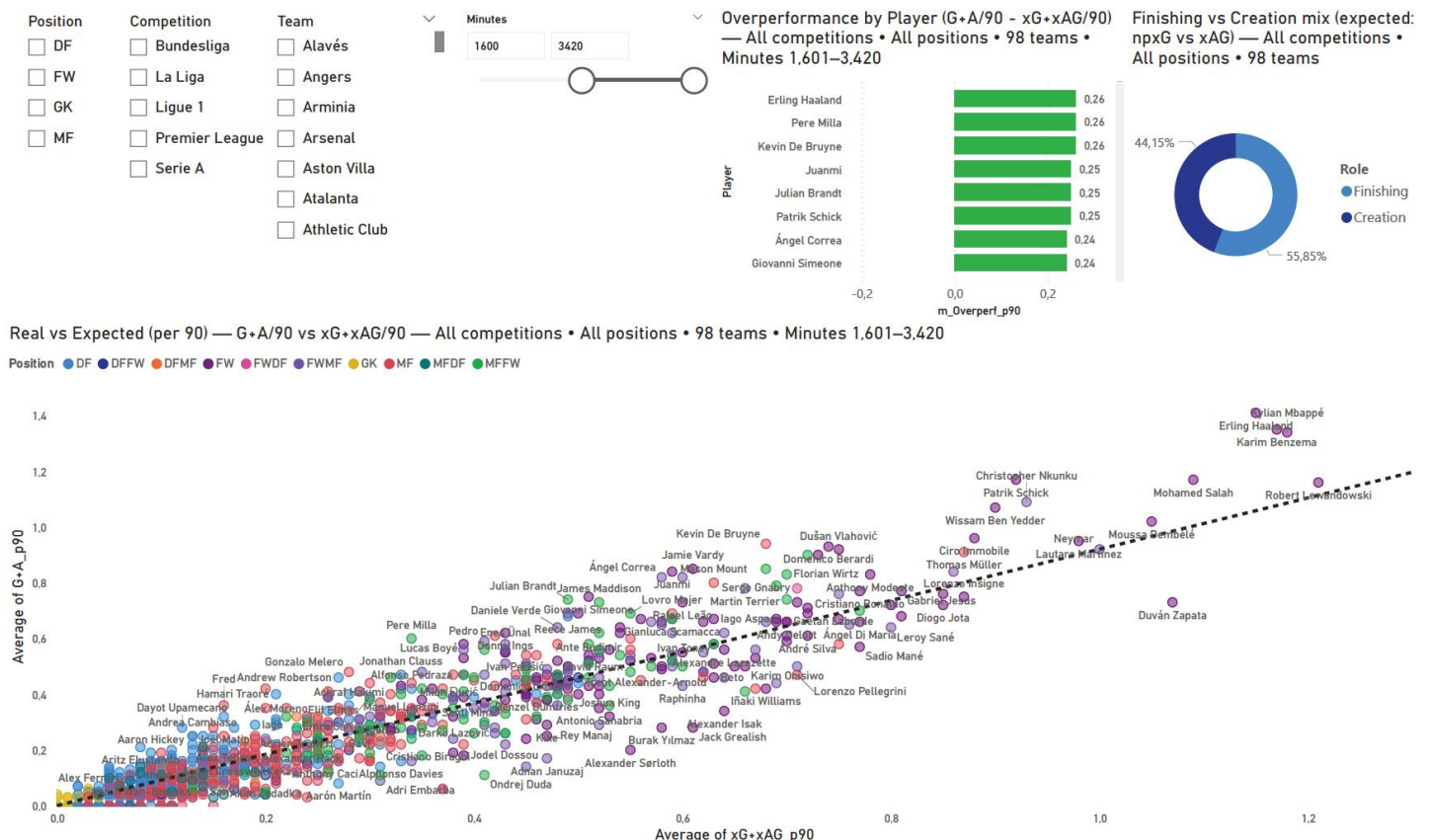
Data Preparation

The main transformations carried out were as follows:

1. Table naming and headers.

The CSV file was imported, and the table was renamed *Players* to facilitate its identification within the model. The first row of the file was promoted as column headers.

2. Text cleaning in categorical variables.

In the *Nation* and *Comp* columns, extra text with abbreviations appeared. To avoid noise in the analysis, the *Extract Text Before Delimiter* transformation was applied, removing this redundant information and leaving only the clean country or competition name.

3. Removal of irrelevant columns.

The variables *Matches* and *-9999* were deleted, as they did not provide useful information for the analysis.

4. Review of data formats.

Special attention was given to the configuration of decimal separators, ensuring that the relevant variables were being read correctly with the appropriate decimal format.

Treatment of Multiple Positions

An important feature of the dataset was that many players appeared in the *Pos* column with more than one position encoded (for example, "DFMF" or "MFFW"), reflecting the versatility of those footballers who can play in different areas of the pitch. While this characteristic is valuable, it complicates analyses filtered by a single position.

To address this issue, the following steps were taken:

1. Creation of an initial bridge table.

From the *Players* table, an auxiliary table called *PositionsBridge* was generated, retaining only the key variables for this analysis: *Player*, *Pos*, *Squad*, and *Comp*.

2. Construction of a unique key.

To ensure record uniqueness and avoid ambiguities in cases of players with the same name, a PlayerKey identifier was created by concatenating player, team, and competition.

3. Splitting positions.

The *Pos* column was split into two (*Pos1* and *Pos2*), so that when a player had more than one assigned position, both were reflected. An *unpivot* operation was then applied on these columns, generating a single field *PosSingle*. In this way, each player was represented in as many rows as positions they could perform, allowing for more precise position-based analysis.

4. Relationship with the main table.

Finally, the *PositionsBridge* table was related to the *Players* table through the PlayerKey field in a one-to-many (1: *) relationship. This ensured that each unique record in *Players* could be linked to one or more rows in *PositionsBridge*, depending on the number of positions played.

Data Model and DAX Measures

After the initial preparation and cleaning of the dataset, the next step was to structure the data model in Power BI and create the necessary measures for the analysis.

The final model was composed of four main tables:

- **Players:** the base table containing all performance metrics at player level (goals, assists, xG, xAG, minutes, etc.).
- **PositionsBridge:** a bridge table generated from *Players*, with one row per player and position (*PosSingle*), enabling filters by positional role without losing information about versatile players.
- **Roles:** a disconnected table with two rows (*Finishing*, *Creation*), used in the donut plot to compare the relative weight of finishing (npxG) versus creation (xAG).
- **Logos:** an auxiliary table with the team's name and the logo URL, used in customized tooltips.
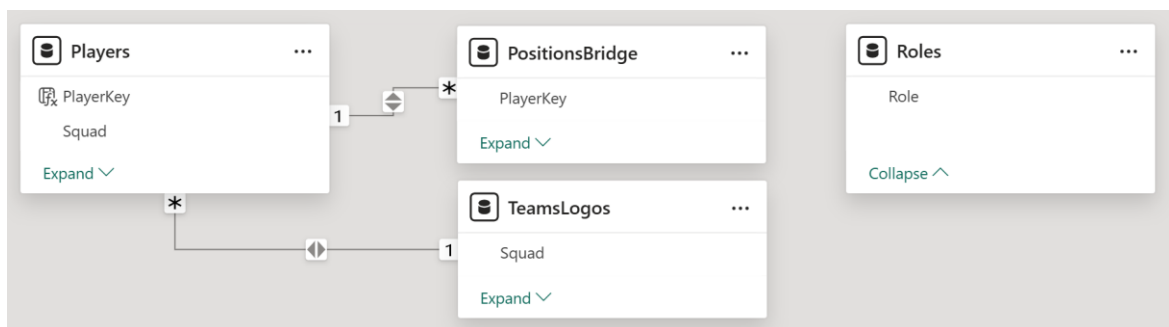
The key measures used in the visuals are summarized below. For clarity, they are grouped according to each of the main charts in the dashboard.

1. Scatter Plot: Expected vs Real

Axes:

- *m_GA_p90 = AVERAGE(Players[G+A_p90])*
- *m_xGA_p90 = AVERAGE(Players[xG+xAG_p90])*
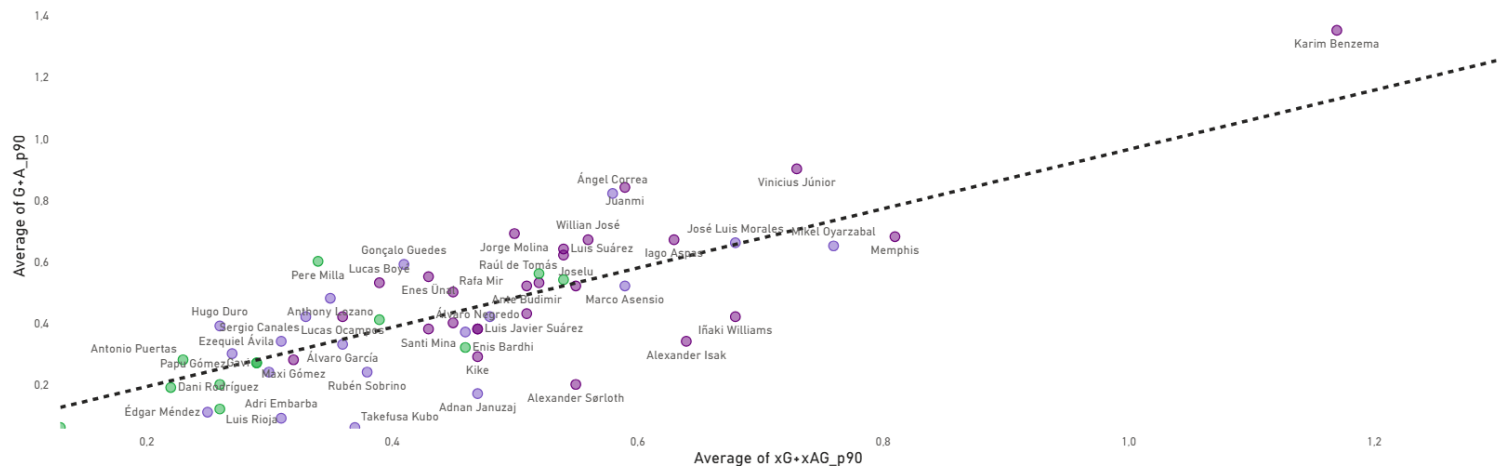
Derived metric for tooltips:

- *m_Overperf_p90 = [m_GA_p90] - [m_xGA_p90]*

Dynamic title:

- *m_Title_Scatter = "Expected vs Real (per 90) — G+A/90 vs xG+xAG/90 — " & [m_Sel_Comp] & " • " & [m_Sel_Pos] & " • " & [m_Sel_Team] & " • " & [m_Sel_MinRange]*

Real vs Expected (per 90) — G+A/90 vs xG+xAG/90 — La Liga • FW • 20 teams • Minutes 1,604–3,420

Position ● FW ● FWMF ● MFFW
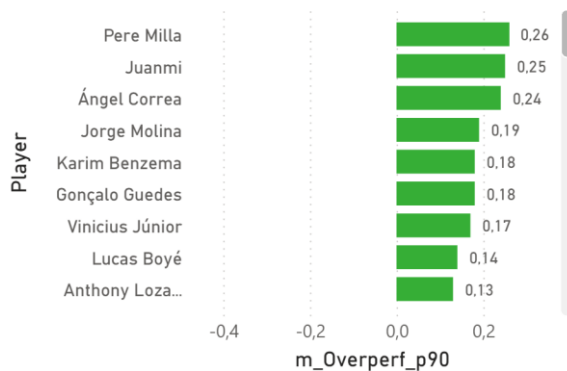

2. Bar Plot: Overperformance by Player

Metric:

- *m_Overperf_p90 = [m_GA_p90] - [m_xGA_p90]*

Dynamic title:

- *m_Title_Bar = "Overperformance by Player (G+A/90 – xG+xAG/90) — " & [m_Sel_Comp] & " • " & [m_Sel_Pos] & " • " & [m_Sel_Team] & " • " & [m_Sel_MinRange]*

Overperformance by Player (G+A/90 – xG+xAG/90)
— La Liga • FW • 20 teams • Minutes 1,604–3,420

Player	m_Overperf_p90
Pere Milla	0,26
Juanmi	0,25
Ángel Correa	0,24
Jorge Molina	0,19
Karim Benzema	0,18
Gonçalo Guedes	0,18
Vinicius Júnior	0,17
Lucas Boyé	0,14
Anthony Loza...	0,13

Vinicius Júnior

Team:	Real Madrid
Comp:	La Liga
Min:	2690
90s:	29.90
Expected (xG+xAG/90):	0,73
Real (G+A/90):	0,90
Difference (Overperf/90):	0,17

3. Donut Chart: Finishing vs Creation Mix

Base measures:

- m_RoleValue = VAR r = SELECTEDVALUE(Roles[Role] RETURN SWITCH(r, "Finishing", [m_npxG_total], "Creation", [m_xAG_total], BLANK())
- *m_npxG_total = SUM(Players[npxG])*
- *m_xAG_total = SUM(Players[xAG])*

Supporting measures for tooltips:

- m_ShareFinisher = DIVIDE([m_npxG_total], [m_Exp_total], 0)
- m_ShareCreator = DIVIDE([m_xAG_total], [m_Exp_total], 0)
- *m_Exp_total = [m_npxG_total] + [m_xAG_total]*

Dynamic title:

- *m_Title_Donut = "Finishing vs Creation mix (expected: npxG vs xAG) — " & [m_Sel_Comp] & " • " & [m_Sel_Pos] & " • " & [m_Sel_Team]*

Finishing vs Creation mix (expected: npxG vs xAG) — La Liga • FW • 20 teams

34,65%
65,35%

Role
● Finishing
● Creation
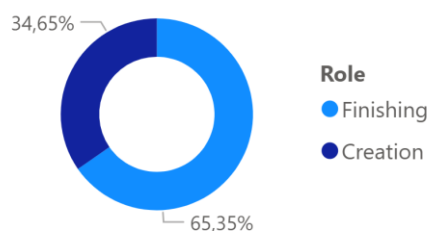
Creation	
Share (of expected):	34.7 %
This total (expected):	196
Total expected (npxG + xAG):	565

Finishing	
Share (of expected):	65.3 %
This total (expected):	369
Total expected (npxG + xAG):	565

4. Context and filter measures (used across all charts)

- *m_Sel_Comp = IF(HASONEVALUE(Players[Comp]), SELECTEDVALUE(Players[Comp]), "All competitions")*
- *m_Sel_Pos = IF(HASONEVALUE('PositionsBridge'[PosSingle]), SELECTEDVALUE('PositionsBridge'[PosSingle]), "All positions")*
- *m_Sel_Team = VAR n = COUNTROWS(VALUES(Players[Squad])) RETURN IF(HASONEVALUE(Players[Squad]), SELECTEDVALUE(Players[Squad]), IF(n = 0, "All teams", FORMAT(n, "0") & " teams"))*
- *m_Sel_MinRange = VAR minSel = MINX(ALLSELECTED(Players), Players[Min]) VAR maxSel = MAXX(ALLSELECTED(Players), Players[Min]) RETURN "Minutes " & FORMAT(minSel,"#,0") & "–" & FORMAT(maxSel,"#,0")*

Slicers

The dashboard includes several interactive filters (slicers) that allow the visualization to be adapted to the desired context and enable fair comparisons:

- **Position:** created from the *PositionsBridge* table, it assigns a single primary position to each player, allowing for clearer filtering by role.
- **Competition:** filters by one of the five major European leagues or allows for an aggregated analysis.
- **Team:** restricts the data to a specific club.
- **Minutes:** a numeric slider that sets a minimum range of accumulated minutes, avoiding misleading readings from players with very limited playing time.

Position	Competition	Team	Minutes
☐ DF	☐ Bundesliga	☐ Alavés	1600 — 3122
■ FW	■ La Liga	☐ Athletic Club	
☐ GK	☐ Ligue 1	☐ Atlético Madrid	
☐ MF	☐ Premier League	☐ Barcelona	
	☐ Serie A	☐ Betis	
		☐ Cádiz	
		☐ Celta Vigo	

Conclusions

The developed dashboard allows for an intuitive exploration of the relationship between expected performance (xG, xAG) and actual performance (G+A) in the five major European leagues. Thanks to the combination of visualizations (scatter, bar, and donut charts) and interactive slicers, it enables a flexible analysis that ranges from an overall view to the detailed performance of a specific player or role.

Overall, it serves as a useful tool and starting point for scouting, performance analysis, and profile comparison, which could be further expanded with contextual metrics or measures of impact in specific phases of play.