

# FBref / SoccerDarta Metrics Dictionary

This document explains every column in the dataset and what it measures in terms of player performance.

## 1. Context & Identification

### league

- **Type:** Categorical (string)
- **Meaning:** Competition in which the stats were recorded.
- **Examples:** ENG-Premier League , ESP-La Liga
- **Use:** Filter by league, compare players across different competitions.

### season

- **Type:** Numeric / Categorical
- **Meaning:** Season of the data. In your file 2425 = 2024–2025.
- **Use:** Track performance over time, compare players across seasons.

### team

- **Type:** Categorical (string)
- **Meaning:** Club the player represents in that league and season.
- **Example:** Arsenal
- **Use:** Group by team, analyze teammates and team profiles.

### player

- **Type:** Categorical (string)
- **Meaning:** Player’s full name (row identifier).
- **Examples:** Ben White , Bukayo Saka , David Raya
- **Use:** Main identifier of each row (often combined with league/season/team).

### nation

- **Type:** Categorical (string)
- **Meaning:** Player’s nationality (or national team representation).
- **Examples:** ENG , ESP
- **Use:** Nationality-based analysis (talent pipelines, style by country).

### pos

- **Type:** Categorical (string, possibly multi-value)
- **Meaning:** Player’s main playing position(s). In your dataset you have:

Single positions:

- GK → **Goalkeeper**
  - Primary role: shot-stopping, claiming crosses, sweeping behind the defense, distribution.
- DF → **Defender**
  - Covers centre-backs, full-backs, wing-backs. Main tasks: defending the box, winning duels, blocking shots, progressing the ball from the back.
- MF → **Midfielder**
  - Includes defensive, central, and attacking midfielders. Main tasks: ball progression, linking defense and attack, controlling tempo, chance creation.
- FW → **Forward**
  - Includes strikers, center-forwards, wingers/inside forwards. Main tasks: getting on the end of chances, finishing, attacking runs, pressing from the front.

Multiple-position combinations:

- FW, MF
  - Forward **and** midfielder.
  - Typically: attacking midfielders / wide players who can play as wingers or as #10s, or second strikers dropping into midfield (e.g., versatile attacking players).
- MF, FW
  - Midfielder **and** forward (same combination, order reversed).
  - Usually indicates a player whose primary role is midfield but who can also play in advanced attacking positions.
- DF, FW
  - Defender **and** forward.
  - Rare profile; can indicate players used in defense but also as an emergency striker or target man, or very versatile wing-backs that can play as wide forwards.
- FW, DF
  - Forward **and** defender (same combination, order reversed).

- Often similar to DF, FW, but might imply a player primarily used up front who can also fill in at the back.
  - DF, MF
    - Defender **and** midfielder.
    - Typical for full-backs who can play as wide midfielders, or defensive midfielders who can fill in as centre-backs or full-backs.
  - MF, DF
    - Midfielder **and** defender (same combination, order reversed).
    - Usually a midfielder primarily who can step into the back line, or a defensive utility player.
- Use:
  - Position-specific comparisons and modeling (e.g., build different feature sets for GK vs DF vs FW).
  - Multi-position labels are useful for identifying versatile players and for clustering roles beyond classic positions.

#### age

- **Type:** Numeric
- **Meaning:** Age of the player during the season.
- **Use:** Age–performance curves, career stage analysis.

#### born

- **Type:** Numeric (year)
- **Meaning:** Year of birth (e.g. 1997).
- **Use:** Recomputing age across seasons, linking to other datasets.

## 2. Playing Time

These variables measure how much and how often a player appears on the pitch.

#### MP (Matches Played)

- **Type:** Numeric
- **Meaning:** Number of matches where the player appeared (starter or substitute).
- **Use:** Overall involvement; distinguishes between regulars and fringe players.

#### Starts

- **Type:** Numeric
- **Meaning:** Number of matches started in the lineup.
- **Use:** Indicates whether the player is a regular starter or mostly a substitute.

#### Min (Minutes)

- **Type:** Numeric
- **Meaning:** Total minutes played.
- **Use:** Core volume metric; basis for per-90 calculations and for measuring reliability and fitness.

#### 90s

- **Type:** Numeric
- **Meaning:** Minutes divided by 90.
  - Example: 1198 minutes → 13.3 90s.
- **Use:** Standardizes playing time into “full-match equivalents” for per-90 stats.

## 3. Classic Output (Performance)

These are traditional “box score” stats: what directly happened on the pitch.

#### GLs

- **Type:** Numeric
- **Meaning:** Total goals scored.
- **Use:** Core finishing output.

#### Ast

- **Type:** Numeric
- **Meaning:** Total assists (final pass or cross leading to a goal, Opta-style).
- **Use:** Measures playmaking contribution to goals.

#### G+A

- **Type:** Numeric
- **Meaning:** Goals + assists.
- **Use:** Simple total goal contribution (direct involvement in goals).

#### G-PK

- **Type:** Numeric
- **Meaning:** Non-penalty goals (  $Gls - \text{minus penalty goals}$  ).
- **Use:** Measures goal-scoring from open play and non-penalty set pieces.

#### PK

- **Type:** Numeric
- **Meaning:** Penalty kicks scored (converted).
- **Use:** Shows who takes penalties and how many they convert.

#### PKatt

- **Type:** Numeric
- **Meaning:** Penalty kicks attempted (scored + missed).
- **Use:** Used with PK to calculate penalty conversion rates.

#### CrdY

- **Type:** Numeric
- **Meaning:** Yellow cards received.
- **Use:** Discipline, risk of suspension.

#### CrdR

- **Type:** Numeric
- **Meaning:** Red cards received.
- **Use:** Serious disciplinary issues and impact on team (suspensions, playing with 10 men).

---

## 4. Expected Metrics (xG / xAG)

These are model-based stats measuring chance quality rather than just outcomes.

#### xG (Expected Goals)

- **Type:** Numeric
- **Meaning:** Sum of goal probabilities of all shots taken by the player.
- **Use:** Indicates the quality and volume of chances the player gets.

#### npG (Non-Penalty xG)

- **Type:** Numeric
- **Meaning:** xG excluding penalties.
- **Use:** Focuses on open-play and non-penalty set-piece chance quality.

#### xAG (Expected Assisted Goals)

- **Type:** Numeric
- **Meaning:** Total xG of shots that result from the player's passes.
- **Use:** Measures chance creation quality, independent of whether teammates score.

#### npG+xAG

- **Type:** Numeric
- **Meaning:**  $npG + xAG$  .
- **Use:** Combined open-play attacking contribution (shooting + creative).

---

## 5. Progression Metrics

These stats capture how a player helps move the ball up the pitch into more dangerous areas.

PrgC (Progressive Carries)

- **Type:** Numeric
- **Meaning:** Number of times the player carries the ball significantly closer to the opponent’s goal.
- **Use:** Identifies strong ball carriers and line-breaking dribblers.

PrgP (Progressive Passes)

- **Type:** Numeric
- **Meaning:** Completed passes that move the ball significantly closer to the opponent’s goal.
- **Use:** Identifies line-breaking passers and progression from deeper roles.

PrgR (Progressive Passes Received)

- **Type:** Numeric
- **Meaning:** Times the player **receives** a progressive pass.
- **Use:** Captures off-ball movement and ability to find space in advanced zones.

## 6. Per 90 Minutes Metrics

All of the following stats are rate versions:  
[  $\text{Stat per 90} = \frac{\text{Total Stat}}{\text{90s}}$  ]

They allow fair comparison between players with different amounts of playing time.

Gls per 90min

- **Type:** Numeric
- **Meaning:** Goals per 90 minutes.
- **Use:** Scoring rate normalized for playing time.

Ast per 90min

- **Type:** Numeric
- **Meaning:** Assists per 90 minutes.
- **Use:** Creative output per full game equivalent.

G+A per 90min

- **Type:** Numeric
- **Meaning:** (Goals + Assists) per 90 minutes.
- **Use:** Total direct goal contributions per full game.

G-PK per 90min

- **Type:** Numeric
- **Meaning:** Non-penalty goals per 90 minutes.
- **Use:** Open-play scoring rate per 90.

G+A-PK per 90min

- **Type:** Numeric
- **Meaning:** Non-penalty goal contributions per 90 minutes.
- **Note:** Essentially  $(G-PK + Ast) / 90s$  .
- **Use:** Total attacking output per 90, excluding penalty goals.

xG per 90min

- **Type:** Numeric
- **Meaning:** Expected Goals per 90 minutes.
- **Use:** Quality and volume of chances the player gets each full game.

xAG per 90min

- **Type:** Numeric
- **Meaning:** Expected Assisted Goals per 90 minutes.
- **Use:** Chance creation quality per game.

xG+xAG per 90min

- **Type:** Numeric
- **Meaning:**  $xG + xAG$  per 90 minutes.
- **Use:** Total expected attacking contribution per game.

npxG per 90min

- **Type:** Numeric
- **Meaning:** Non-penalty xG per 90 minutes.
- **Use:** Open-play expected scoring per game.

npxG+xAG per 90min

- **Type:** Numeric
- **Meaning:** npxG + xAG per 90 minutes.
- **Use:** Best single measure of non-penalty attacking contribution per full match (shooting + creating).

---

## 7. Category Overview

---

- **Context/ID:**  
league, season, team, player, nation, pos, age, born
- **Playing Time:**  
MP, Starts, Min, 90s
- **Classic Output:**  
Gls, Ast, G+A, G-PK, PK, PKatt, CrdY, CrdR
- **Expected Metrics:**  
xG, npxG, xAG, npxG+xAG
- **Progression:**  
PrgC, PrgP, PrgR
- **Per 90:**  
Gls per 90min, Ast per 90min, G+A per 90min,  
G-PK per 90min, G+A-PK per 90min,  
xG per 90min, xAG per 90min, xG+xAG per 90min,  
npxG per 90min, npxG+xAG per 90min