



A Season in Review: 2023-2024 Premier League Database

By

Minh Phong Do
Alex Ross



Professor
Dr. Bhaskar,
Vidhyacharan

Introduction

- ▲ An analysis of the 23-24 Premier League Season through SQL database modeling
- ▲ A comprehensive database that contains data from every single match played throughout the season
- ▲ Aimed to uncover trends, outliers, and insights across the entirety of the Premier League
- ▲ Can be utilized by teams and fans alike to predict matches, perform player analysis, and compare your club to the rest of the league
- ▲ Data collected from Football Reference



Entities and Attributes

Club

club(name, founding_year, stadium)

Venue

venue(name, city, capacity, year_opened)

Club Member (*superclass for player and manager*)

club_member(member_id, name, date_of_birth, nationality, height, club FK→club(name))

Player (*subclass of club_member*)

player(member_id FK→club_member(member_id), position, shirt_number, market_value_euros)

Manager (*subclass of club_member*)

manager(member_id FK→club_member(member_id), titles, years_of_service)

Contract

contract(member_id FK→club_member(member_id), start_date, weekly_wage, end_date)

Fixture

fixture(match_id, date, home_club FK→club(name), away_club FK→club(name))

Player Stat

player_stat(stat_id PK, match_id FK→fixture(match_id), member_id FK→club_member(member_id), minutes, goals, assists, pens_made, pens_attempted, shots_attempted, shots_on_target, yellow_cards, red_cards, touches, tackles, interceptions, blocks, xg, x_assist, passes_attempted, passes_completed, dribbles_attempted, dribbles_completed, crosses, throw_ins, corner_kicks, shots_blocked, passes_blocked, clearances, fouls_committed, fouls_drawn, offsides, pens_won, pens_conceded, own_goals, shot_creating_actions, goal_creating_actions, recoveries, ariel_duels_won, ariel_duels_lost, sot_against, goals_against, saves, psxg_against, subbed_in, subbed_off, started)

Club Stat

club_stat(club_stat_id, match_id FK→fixture(match_id), club FK→club(name), goals_scored, possession)

Goal

goal(goal_id, match_id FK→fixture(match_id), scorer_id FK→club_member(member_id), club FK→club(name), minute, half, xg, psxg, yards_out, body_part, method)

Relationships

club – (0,M)-has-(1,1) – club_member:

A club must have 0 or more club members (players and managers).

A club member must belong to exactly one club.

club_member – (1,1)-signs-(1,1) – contract:

Each club member has one active contract.

Each club member must have exactly one contract.

club – (38,38)-plays in-(2, 2) – fixture (as home/away):

A club can play 38 fixtures (19 home, 19 away)

A fixture can only have 2 clubs.

club – (0,M)-has-(1,1) – club_stat:

A club can have multiple match performance records.

The stats of a club in a match can only belong to one club.

club – (1,1)-uses-(1,1) – venue:

A club uses one stadium as its home ground.

fixture – (1,1)-occurs in-(0, 19) – venue:

A fixture is played in a single venue.

A venue can have up to 19 fixtures (number of home matches).

player – (0,M)-has-(1,1) – player_stat:

A player can have multiple fixture performance entries.
player_stats can only belong to one player.

fixture – (22,M)-has-(1,1) – player_stat:

A fixture can have from 22 and above player's performance entries.

player_stats can only belong to one fixture.

fixture – (2,2)-has-(1,1) – club_stat:

Each fixture has two team stats records (one per team).

A club_stat can only belong to one fixture.

fixture – (0,M)-has-(1,1) – goal:

Each fixture can have multiple goals recorded.

A goal can only belong to one fixture.

player – (0,M)-scores-(1,1) – goal:

A player can score multiple goals.

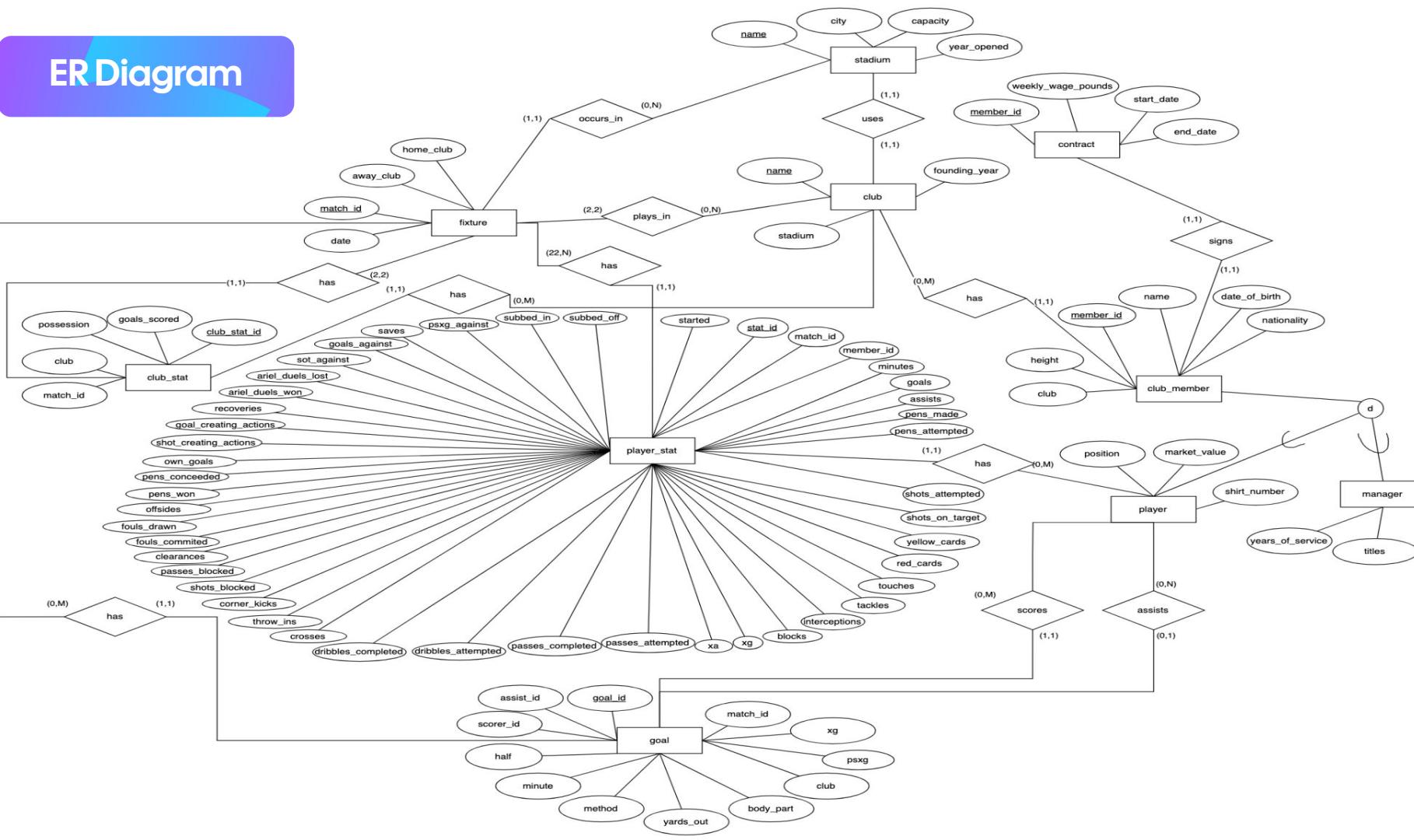
A goal can only belong to one player

player – (0,M)-assists-(0,1) – goal:

A player may assist multiple goals, or none.

An assist can belong to 0 or 1 player.

ER Diagram



1. Oldest Venues



Bramall Lane – Sheffield United

```
SELECT name, city, year_opened  
FROM venue  
ORDER BY year_opened ASC;
```

name	city	year_opened
Bramall Lane	Sheffield	1855
Stamford Bridge	London	1877
Villa Park	Birmingham	1877
Turf Moor	Burnley	1883
Anfield	Liverpool	1884

This query retrieves the five oldest venues in the league by ordering them by the year they opened.

2. Spanish Club Members



Unai Emery – Spanish Manager

```
SELECT COUNT(*) AS spanish_members  
FROM club_member  
WHERE nationality = 'Spain';
```

spanish_members
24

This query counts the number of spanish players and managers in the league.

3. Outside the Box Goals



Alexis Mac Allister – Liverpool

```
SELECT COUNT(*) AS  
goals_outside_box  
FROM goal  
WHERE yards_out > 18;
```

goals_outside_box
154

This query finds the number of goals scored from outside the box (18+ yards) throughout the season.

4.

Most Goals in a Match



Newcastle United – 8 goals

```
SELECT club, goals_scored  
FROM club_stat  
ORDER BY goals_scored DESC  
LIMIT 1;
```

club	goals_scored
Newcastle United	8

This query finds the club that scored the most amount of goals in a single match.

5.

Possession



Manchester City – 65.21%
average possession

```
SELECT club, ROUND(AVG(possession), 2) AS  
avg_possession_percentage  
FROM club_stat  
GROUP BY club  
ORDER BY avg_possession_percentage  
DESC;
```

club	avg_possession_percentage
Manchester City	65.21
Tottenham Hotspur	61.58
Liverpool	61.26
Brighton & Hove Albion	59.84
Chelsea	58.61

This query finds the clubs with the highest average possession per game (the clubs that have control of the ball the most).

1.

Golden Boot



Erling Haaland – 27 goals

```
SELECT cm.name, p.position, cm.club,
COALESCE(SUM(ps.goals), 0) AS total_goals
FROM club_member cm JOIN player p ON
cm.member_id = p.member_id
JOIN player_stat ps ON cm.member_id =
ps.member_id GROUP BY cm.name, p.position,
cm.club ORDER BY total_goals DESC;
```

name	position	club	total_goals
Erling Haaland	Forward	Manchester City	27
Cole Palmer	Forward	Chelsea	22
Alexander Isak	Forward	Newcastle United	21
Dominic Solanke	Forward	AFC Bournemouth	19
Phil Foden	Forward	Manchester City	19

This query finds the players who have scored the most goals throughout the season. The Golden Boot is given to the player with the most goals.

2.

Most Expensive Player



Kevin De Bruyne – £400,000/week

SELECT

cm.name,
c.weekly_wage_pounds

FROM club_member cm

JOIN contract c **ON** cm.member_id = c.member_id

JOIN player p **ON** cm.member_id = p.member_id

ORDER BY c.weekly_wage_pounds **DESC**

LIMIT 1;

name	weekly_wage_pounds
Kevin De Bruyne	400000.00

This query finds the most expensive player in the league by weekly wage (in GBP).

3.

Super Subs



Jhon Durán – 4 goals

```
SELECT cm.name, p.position, p.market_value_euros,  
cm.club, SUM(ps.goals) AS goals_as_sub  
FROM player_stat ps  
JOIN club_member cm ON ps.member_id =  
cm.member_id  
JOIN player p ON ps.member_id = p.member_id  
WHERE ps.subbed_in = TRUE  
GROUP BY cm.name, p.position,  
p.market_value_euros, cm.club  
ORDER BY goals_as_sub DESC;
```

name	position	market_value_euros	club	goals_as_sub
Jáder Durán	Forward	40.00	Aston Villa	4
Leandro Trossard	Forward	35.00	Arsenal	4
Scott McTominay	Midfielder	30.00	Manchester United	4
João Pedro	Forward	50.00	Brighton & Hove Albion	4
Harvey Barnes	Forward	32.00	Newcastle United	4

This query finds the players who have scored the most goals after being subbed into the match.

4.

Goal Contributions



Cole Palmer – 33 goal contributions

```
SELECT cm.name, p.position, p.market_value_euros,  
TIMESTAMPDIFF(YEAR, cm.date_of_birth,  
'2023-08-11') AS age_at_season_start,  
COALESCE(SUM(ps.goals + ps.assists), 0) AS  
goal_contributions FROM club_member cm  
JOIN player p ON cm.member_id = p.member_id  
JOIN player_stat ps ON cm.member_id =  
ps.member_id GROUP BY cm.name, p.position,  
p.market_value_euros, age_at_season_start ORDER  
BY goal_contributions DESC;
```

name	position	market_value_euros	age_at_season_start	goal_contributions
Cole Palmer	Forward	18.00	21	33
Ollie Watkins	Forward	55.00	27	32
Erling Haaland	Forward	180.00	23	32
Mohamed Salah	Forward	55.00	31	28
Phil Foden	Forward	110.00	23	27

This query finds the five players with the most goal contributions (goals + assists) throughout the season.

5.

League Standings



Manchester City – 91 points

```
SELECT c.name AS club_name, COUNT(cs.match_id) AS games_played, SUM(CASE WHEN cs.goals_scored > opp.goals_scored THEN 1 ELSE 0 END) AS wins, SUM(CASE WHEN cs.goals_scored = opp.goals_scored THEN 1 ELSE 0 END) AS draws, SUM(CASE WHEN cs.goals_scored < opp.goals_scored THEN 1 ELSE 0 END) AS losses, SUM(CASE WHEN cs.goals_scored > opp.goals_scored THEN 3 WHEN cs.goals_scored = opp.goals_scored THEN 1 ELSE 0 END) AS points, SUM(cs.goals_scored - opp.goals_scored) AS goal_difference FROM club c JOIN club_stat cs ON c.name = cs.club JOIN club_stat opp ON cs.match_id = opp.match_id AND cs.club != opp.club GROUP BY c.name ORDER BY points DESC, goal_difference DESC;
```

club_name	games_played	wins	draws	losses	points	goal_differen...
Manchester City	38	28	7	3	91	62
Arsenal	38	28	5	5	89	62
Liverpool	38	24	10	4	82	45
Aston Villa	38	20	8	10	68	15
Tottenham Hotspur	38	20	6	12	66	13
Chelsea	38	18	9	11	63	14

This query orders the clubs by points where 3 is a win, 1 is a draw, and 0 is a loss. If clubs are tied in points, goal difference is the tiebreaker.

1.

Worst Finishing Club



Everton – 16.3 less goals than expected

```
WITH club_goals AS (SELECT cs.match_id, cs.club,
SUM(cs.goals_scored) AS match_goals FROM
club_stat cs GROUP BY cs.match_id, cs.club),
club_xg AS (SELECT cm.club, ps.match_id,
SUM(ps.xg) AS match_xg FROM player_stat ps
JOIN club_member cm ON ps.member_id =
cm.member_id GROUP BY cm.club, ps.match_id)
SELECT cx.club, ROUND(SUM(cx.match_xg), 2) AS
total_xg, SUM(cg.match_goals) AS total_goals,
ROUND(SUM(cx.match_xg) - SUM(cg.match_goals),
2) AS xg_minus_goals FROM club_xg cx JOIN
club_goals cg ON cx.club = cg.club AND cx.match_id =
cg.match_id GROUP BY cx.club ORDER BY
xg_minus_goals DESC LIMIT 1;
```

club	total_xg	total_goals	xg_minus_goals
Everton	56.30	40	16.30

This query finds the club with the largest positive difference between expected goals (xG) and actual goals scored, meaning the club that underperformed their xG the most, implying poor finishing ability.

2.

Best Finishers



Phil Foden – 8.60 goals more than expected

```
SELECT finishers.name, finishers.position,
finishers.club, c.founding_year, finishers.total_goals,
finishers.total_xg, (finishers.total_goals -
finishers.total_xg) AS overperformance FROM
(SELECT cm.name, p.position, cm.club, SUM(ps.goals)
AS total_goals, SUM(ps.xg) AS total_xg
FROM player_stat ps JOIN club_member cm ON
ps.member_id = cm.member_id JOIN player p ON
ps.member_id = p.member_id GROUP BY cm.name,
p.position, cm.club) AS finishers JOIN club c ON
finishers.club = c.name WHERE finishers.total_xg >= 5
ORDER BY overperformance DESC;
```

name	position	club	total_goals	total_xg	overperformance
Phil Foden	Forward	Manchester City	19	10.40	8.60
Son Heung-min	Forward	Tottenham Hotspur	17	11.70	5.30
Jean-Philippe Mateta	Forward	Crystal Palace	16	11.00	5.00
Michael Olise	Forward	Crystal Palace	10	5.30	4.70
Diogo Jota	Forward	Liverpool	10	5.40	4.60

This query finds the best finishers in the league by finding the difference between their total goals and total expected goals (xG).

3.

Best Creators



Martin Ødegaard – 220 shot creating actions

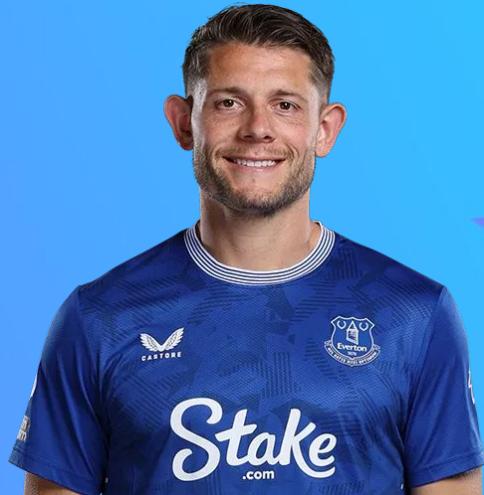
```
SELECT cm.name, cm.club, p.position,  
sca_top.total_sca FROM (SELECT ps.member_id,  
SUM(ps.shot_creating_actions) AS total_sca  
FROM player_stat ps JOIN player p2 ON  
ps.member_id = p2.member_id GROUP BY  
ps.member_id) AS sca_top JOIN club_member cm  
ON sca_top.member_id = cm.member_id JOIN player  
p ON sca_top.member_id = p.member_id ORDER BY  
sca_top.total_sca DESC;
```

name	club	position	total_sca
Martin Ødegaard	Arsenal	Midfielder	220
Bruno Fernandes	Manchester United	Midfielder	210
Bukayo Saka	Arsenal	Forward	188
Pascal Groß	Brighton & Hove Albion	Midfielder	185
Luis Díaz	Liverpool	Forward	163

This query finds the players with the most shot creating actions (dribbles, passes, drawing fouls) throughout the entire season.

4.

Best Defenders



James Tarkowski – defensive score: 612

```
SELECT cm.name, cm.club, p.position,
defender_stats.defensive_score
FROM club_member cm JOIN (SELECT
ps.member_id, SUM(ps.tackles + ps.interceptions +
ps.blocks + ps.clearances + ps.shots_blocked +
ps.passes_blocked + ps.ariel_duels_won) AS
defensive_score FROM player_stat ps JOIN player p2
ON ps.member_id = p2.member_id
WHERE p2.position IN ('Defender', 'Midfielder')
GROUP BY ps.member_id) AS defender_stats ON
cm.member_id = defender_stats.member_id
JOIN player p ON cm.member_id = p.member_id
ORDER BY defender_stats.defensive_score DESC;
```

name	position	club	defensive_score
James Tarkowski	Defender	Everton	612
Max Kilman	Defender	Wolverhampton Wanderers	482
Jarrad Branthwaite	Defender	Everton	475
Joachim Andersen	Defender	Crystal Palace	459
Virgil van Dijk	Defender	Liverpool	457

This query finds the best defenders in the league by creating a defensive score metric, which is the sum of defensive-minded statistics. It also only considers midfielders and defenders.

5.

Best Keepers



José Sá – 7.70 less goals allowed than expected

```
SELECT cm.name, cm.club, p.position, gk_stats.total_saves,
gk_stats.total_goals_against, gk_stats.total_psxg_against,
(gk_stats.total_psxg_against - gk_stats.total_goals_against)
AS overperformance FROM club_member cm
JOIN (SELECT ps.member_id, SUM(ps.saves) AS total_saves,
SUM(ps.goals_against) AS total_goals_against,
SUM(ps.psxg_against) AS total_psxg_against FROM
player_stat ps JOIN player p2 ON ps.member_id =
p2.member_id WHERE p2.position = 'Goalkeeper' GROUP BY
ps.member_id HAVING SUM(ps.saves) >= 50
) AS gk_stats ON cm.member_id = gk_stats.member_id
JOIN player p ON p.member_id = cm.member_id
ORDER BY overperformance DESC;
```

name	club	position	total_saves	total_goals_agai...	total_psxg_agai...	overperformance
José Sá	Wolverhampton Wanderers	Goalkeeper	131	58	65.70	7.70
Arijanet Muric	Burnley	Goalkeeper	62	16	21.70	5.70
Emiliano Martínez	Aston Villa	Goalkeeper	95	48	50.90	2.90
André Onana	Manchester United	Goalkeeper	146	58	59.80	1.80
Jordan Pickford	Everton	Goalkeeper	117	51	52.30	1.30

This query finds the best keepers in the league based on the difference between expected goals scored against post shot (psxg) and goals allowed.

1. Positional Market Value



Virgil Van Dijk, Defender

$\gamma_{position, \text{SUM}(\text{market_value_euros}) \rightarrow \text{TotalValue}}(\text{PLAYER})$

Legend:

- γ - Aggregation Operation

position	TotalValue
Defender	4176.30
Forward	4124.00
Midfielder	4397.00
Goalkeeper	630.50

This query groups all records in the player table by position, calculate the sum of market_value_euros for each position group, and label this sum as TotalValue in the result.

2.

Liverpool Players



Liverpool F.C.

$$\rho_{\text{LiverpoolPlayers}}(\pi_{\text{name}, \text{nationality}, \text{club}}(\sigma_{\text{club}=\text{'Liverpool'}}(\text{CLUB_MEMBER})))$$

Legend:

- γ - Aggregation Operation
- ρ - Rename Operation
- π - Projection Operation
- σ - Selection Operation

name	nationality	club
Virgil van Dijk	Netherlands	Liverpool
Luis Díaz	Colombia	Liverpool
Alexis Mac Allister	Argentina	Liverpool
Mohamed Salah	Egypt	Liverpool
Alisson	Brazil	Liverpool
Trent Alexander-Arnold	England	Liverpool
Dominik Szoboszlai	Hungary	Liverpool
Darwin Núñez	Uruguay	Liverpool
Wataru Endo	Japan	Liverpool
Andrew Robertson	Scotland	Liverpool
Joe Gomez	England	Liverpool

This query filters the club_member table to find all members associated with Liverpool. Keep only their name, nationality, and club information, and name this resulting set of data as LiverpoolPlayers.

3.

Venues By City



Villa Park, Birmingham

$\gamma_{\text{city}, \text{COUNT}(* \rightarrow \text{StadiumCount})}(\text{VENUE})$

Legend:

- γ - Aggregation Operation

city	StadiumCount
Liverpool	2
Sheffield	1
London	6
Manchester	2
Brentford	1
Luton	1
Wolverhampton	1
Newcastle	1
Falmer	1
Nottingham	1
Burnley	1
Birmingham	1
Bournemouth	1

This query groups the venue table by city and count the number of venues (rows) in each group, naming the count column as StadiumCount.

4.

Largest Venues



Anfield, Liverpool

$$\pi_{\text{name}, \text{city}, \text{capacity}}(\sigma_{\text{capacity} > 40000}(\text{VENUE}))$$

Legend:

- π - Projection Operation
- σ - Selection Operation

name	city	capacity
Anfield	Liverpool	61276
Emirates Stadium	London	60704
Etihad Stadium	Manchester	53400
London Stadium	London	62500
Old Trafford	Manchester	74310
St. James' Park	Newcastle	52305
Stamford Bridge	London	40343
Villa Park	Birmingham	42918
Vitality Stadium	Bournemouth	42918

This query returns just the stadium name, city, and capacity for all venues with capacity over 40,000.

5.

Possession (notation)



Aston Villa – 53.10% average possession

$$\begin{array}{l} \tau_{\text{AvgPossession}} \text{ DESC}(\gamma_{\text{club}, \text{AVG}(\text{possession})} \\ \rightarrow \text{AvgPossession} (\text{CLUB_STAT})) \end{array}$$

Legend:

- τ - Sort Operation
- γ - Aggregation Operation

club	AvgPossession
Manchester City	65.205128
Tottenham Hotspur	61.578947
Liverpool	61.263158
Brighton & Hove Albion	59.842105
Arsenal	58.692308
Chelsea	58.605263
Aston Villa	53.105263
Newcastle United	52.210526
Fulham	50.605263
Manchester United	50.578947
Wolverhampton Wa...	48.868421
Burnley	47.051282
Brentford	45.157895
AFC Bournemouth	44.394737
Luton Town	42.368421
Crystal Palace	42.289474
West Ham United	41.105263

This query groups the records in the club_stat table by club, calculates the average possession for each club, and then sorts the results in descending order based on the AvgPossession value.

1. Most Expensive Clubs



Arsenal – £130,769 average weekly player wage

$$\tau_{\text{AverageWage}} \text{ DESC } (\gamma_{\text{club}, \text{AVG}(\text{weekly_wage_pounds}) \rightarrow \text{AverageWage}} (\text{CLUB_MEMBER} \bowtie \text{CONTRACT}))$$

Legend:

- τ - Sort Operation
- γ - Aggregation Operation
- \bowtie - Natural Join Operation

club	AverageWage
Manchester City	150000.000000
Manchester United	132375.000000
Arsenal	130769.230769
Liverpool	89870.967742
Chelsea	83000.000000
Tottenham Hotspur	74200.000000
West Ham United	73629.629630
Aston Villa	62375.000000
Newcastle United	61575.757576
Fulham	50769.230769
Crystal Palace	46185.185185
Everton	44259.259259
Brighton & Hove Al...	39516.129032
Nottingham Forest	38088.235294
Brentford	37241.379310
Burnley	36406.250000
Wolverhampton W...	35000.000000

This query finds the average weekly wage for players at each club by joining the club_member and contract tables, then groups the results by club. Finally, it sorts the results in descending order by average wage so that the highest-paying clubs appear first.

2.

Man. United Members



Arsenal – £130,769 average weekly player wage

$$\begin{aligned}
 & \tau_{\text{total_minutes}} \text{ DESC}(\gamma_{\text{cm.name} \rightarrow \text{club_member}}, \text{SUM}(\text{ps.minutes})) \\
 & \rightarrow \text{total_minutes} \\
 & \sigma_{\text{cm.club} = \text{'Manchester United'}}(\text{CLUB_MEMBER CM}) \\
 & \bowtie_{\text{CM.member_id} = \text{PS.member_id}} (\text{PLAYER_STAT PS}))
 \end{aligned}$$

Legend:

- τ - Sort Operation
- γ - Aggregation Operation
- \bowtie - Left Outer Join Operation

club_member	total_minutes
Erik ten Hag	NULL
Daniel Gore	1
Donny van de Beek	3
Ethan Wheatley	16
Omari Forson	70
Jadon Sancho	79
Hannibal Mejbri	133
Facundo Pellistri	172
Willy Kambwala	330
Amad Diallo	390
Sergio Reguilón	412
Anthony Martial	448
Mason Mount	516

This query retrieves all Manchester United club members and calculates their total playing time in minutes by using left outer join to include all members in the result.

1.

Top First Half Scorers



Dominic Solanke – 12 goals in 2023

$$\begin{aligned}
 & \pi_{\text{club.name}, \text{cm.name}, \text{cm.nationality}, \text{ps.goals}} \\
 & \sigma_{\text{goals} \geq 10} (\text{CLUB} \bowtie_{\text{CLUB.name} = \text{CM.club}} \text{CLUB_MEMBER} \text{ AS } \\
 & \text{CM} \bowtie_{\text{CM.member_id} = \text{PS.member_id}} (\gamma_{\text{member_id}, \text{SUM(goals)} \rightarrow \text{goals}} \\
 & \sigma_{\text{match_id} \text{ IN } (\pi_{\text{match_id}} (\sigma_{\text{date} \geq '2023-01-01' \wedge \\
 & \text{date} \leq '2023-12-31')} (\text{Fixture}))} (\text{PLAYER_STAT})) \text{ AS PS})
 \end{aligned}$$

Legend:

- π - Projection Operation
- γ - Aggregation Operation
- ρ - Rename Operation
- σ - Selection Operation
- \bowtie - Natural Join Operation

dub_name	player_name	nationality	goals
Manchester City	Erling Haaland	Norway	16
AFC Bournemouth	Dominic Solanke	England	12
Tottenham Hotspur	Son Heung-min	South Korea	12
Liverpool	Mohamed Salah	Egypt	12
West Ham United	Jarrod Bowen	England	11
Wolverhampton Wanderers	Hwang Hee-chan	South Korea	10

This query shows all players who scored more than 10 goals in 2023, along with their club, nationality, and the exact number of goals they scored, sorted from highest goal scorers to lowest.

Conclusion

- ▲ Our system transforms complex football data into actionable insights for clubs, analysts, and fans
- ▲ Provides essential tools for extracting key statistics about players, clubs, and matches
- ▲ Enables clubs to make data-driven decisions by revealing performance patterns and trends
- ▲ The database delivers quick lookups and complex analytics in one scalable platform
- ▲ Creates a foundation for future enhancements including predictive modeling



THANK YOU

By

Minh Phong Do
Alex Ross



Professor
Dr. Bhaskar,
Vidhyacharan