Predicting the champions league winner

By: Massour Ouro Bode, Mowafak Seidhamed, Ricardo Carneiro

Project overview

- **Introduction:**
- **This project explores the use of match data to predict the winner of the UEFA Champions League, aiming to identify patterns of success.**

- **Objective:**
- **To develop predictive models that analyze team performance and forecast tournament outcomes.**

Dataset overview

- **Source:**
- **Sofascore, Kaggle**
- **Features:**
- **Team Attributes: Ratings, squad value, coach experience**
- **Match Statistics: Possession, shots, goals, assists**
- **Historical Performance: Win/loss records, previous tournament stages**
- **Collection Period: 2024-2025 season**

Key questions

- **What is a champions league??**
- **What factors most influence a team's success in the Champions League?**
- **Can historical data predict the winner of a tournament?**
- **How do team strategies vary between different stages of the competition?**

Key questions

- What is champions league: The UEFA Champions League is one of the biggest soccer tournament where the best teams from Europe compete to see who's the best. It happens every year and starts with group games, then knockout rounds, and ends with a final match

- What factors most influence a team's success in the champions league: A team's success in the Champions League depends on having great players, a smart coach, and good teamwork. Staying healthy and avoiding injuries is super important too. Teams that work hard, stay focused, and handle pressure well usually go far. Sometimes, a bit of luck and good referee calls can also make a big difference.

- Can Historical data predict the winner of the champions league: Historical data can hint at likely Champions League winners, with strong teams like Real Madrid often doing well. However, soccer is unpredictable, and surprises can always happen.

- How do team strategies vary between different stages of the competition?: Team strategies often change as the Champions League progresses. In the group stage, teams focus on consistency and avoiding losses to secure enough points. In the knockout stages, strategies become more cautious, with teams balancing attack and defense to avoid conceding away goals or falling behind. In the final, teams usually play carefully, knowing one mistake can decide the champion.

Key learning

- Possession is a strong indicator of match outcomes.
- Historical performance trends are valuable predictors.
- Advanced modeling techniques enhance prediction accuracy.

Attackers:

- Goals = 6 points
- Assists = 4 points
- Expected Goals = 2
- Shots on Target = 1.5
- Big chances Missed = -2
- + Goal Conversion

Player Name: Robert Lewandowski

Metric	Value	Points
Goals	7	42.00
Assists	0	0.00
Expected Goals (xG)	5.74	11.48
Shots on Target	8	12.00
Big Chances Missed	2	-4.00
Goal Conversion (%)	46.67	46.67
Total Points		108.15

```python
# Calculating a performance metric for attackers
attackers_data['Performance Metric'] = (
    attackers_data['Goals'] * 6 +
    attackers_data['Assists'] * 4 +
    attackers_data['Expected Goals (xG)'] * 2 +
    attackers_data['Shots on Target'] * 1.5 -
    attackers_data['Big Chances Missed'] * 2 +
    attackers_data['Goal Conversion %']
)
```

Top 10 Attackers:

- The graph ranks the top 10 attackers based on a performance metric that factors in goals, assists, expected goals (XG), shots on target, and goal efficiency while deducting points for missed chances.

- The bar chart provides a visual comparison of their performance metrics, clearly showing the standout contributions of the top players, with Gimenez leading the pack
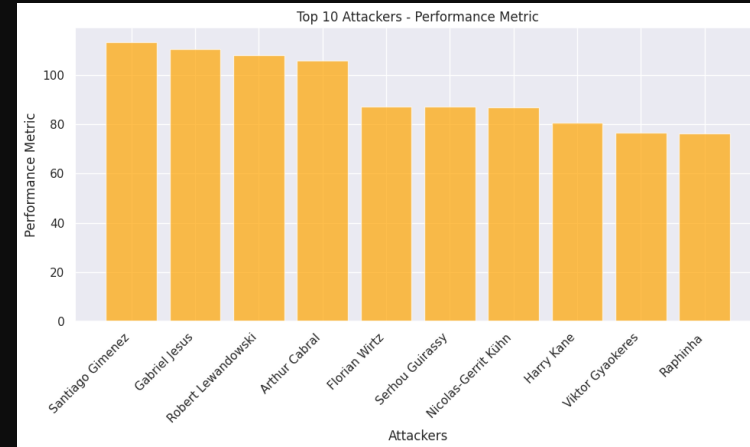

```
Top 10 Attackers:
                   Name    Performance Metric
1       Santiago Gimenez                113.48
2          Gabriel Jesus                110.68
3    Robert Lewandowski                108.15
4          Arthur Cabral                105.80
5          Florian Wirtz                 87.29
6         Serhou Guirassy                87.14
7     Nicolas-Gerrit Kühn                86.79
8             Harry Kane                 80.79
9         Viktor Gyaokeres                76.54
10               Raphinha                 76.38
```
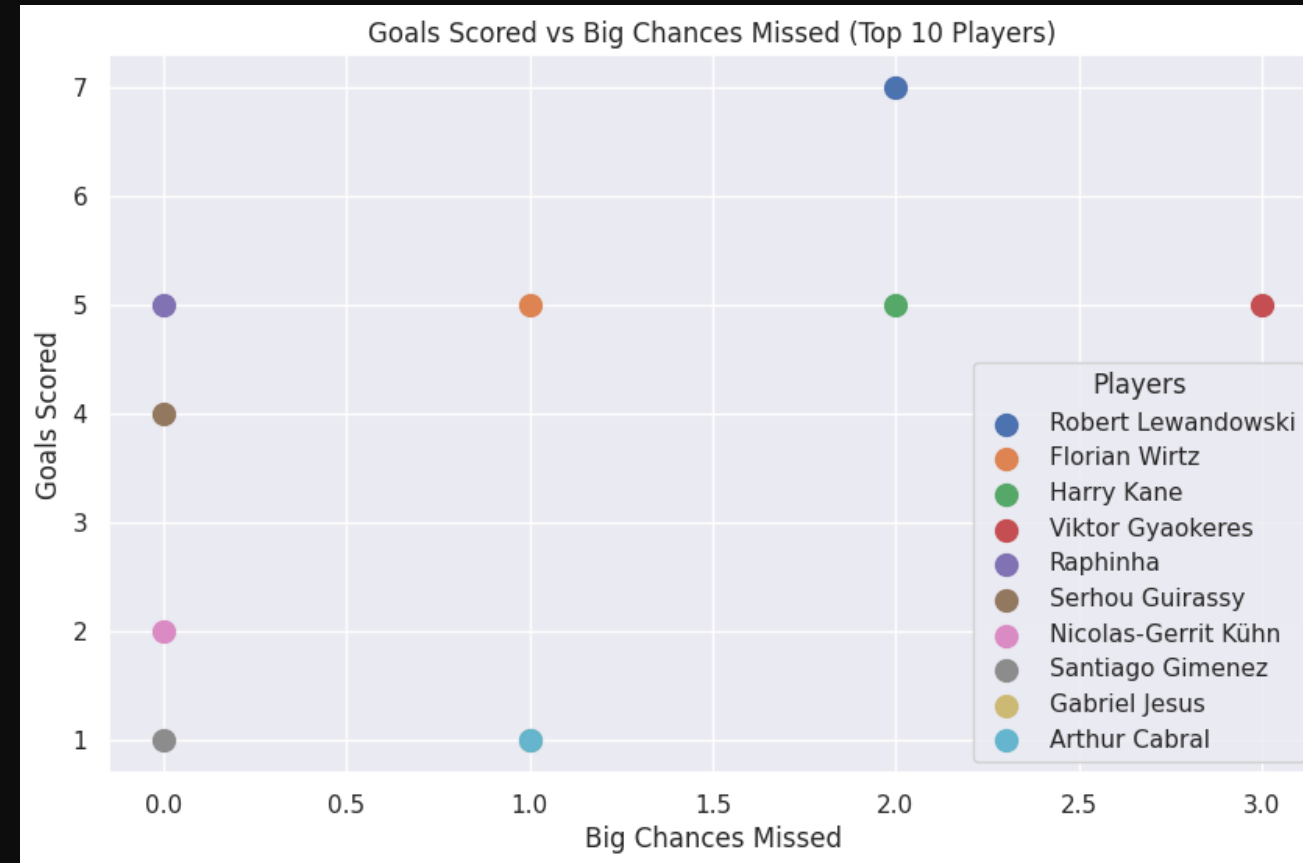
-

- The graph compares the top 10 attackers' goals scored against their big chances missed. Each goal adds 6 points to their performance metric, while every big chance missed deducts 2 points. The plot shows how players like Robert Lewandowski maximize their scoring impact despite missing some chances, while others balance scoring with efficiency.

Goals Scored vs Big Chances Missed (Top 10 Players)

Players
- Robert Lewandowski
- Florian Wirtz
- Harry Kane
- Viktor Gyaokeres
- Raphinha
- Serhou Guirassy
- Nicolas-Gerrit Kühn
- Santiago Gimenez
- Gabriel Jesus
- Arthur Cabral

Midfielders:

- Assist = 5 points
- Key passes = 3 points
- Accurate Passes % = 2 points
- Accurate long Balls = 1.5 points
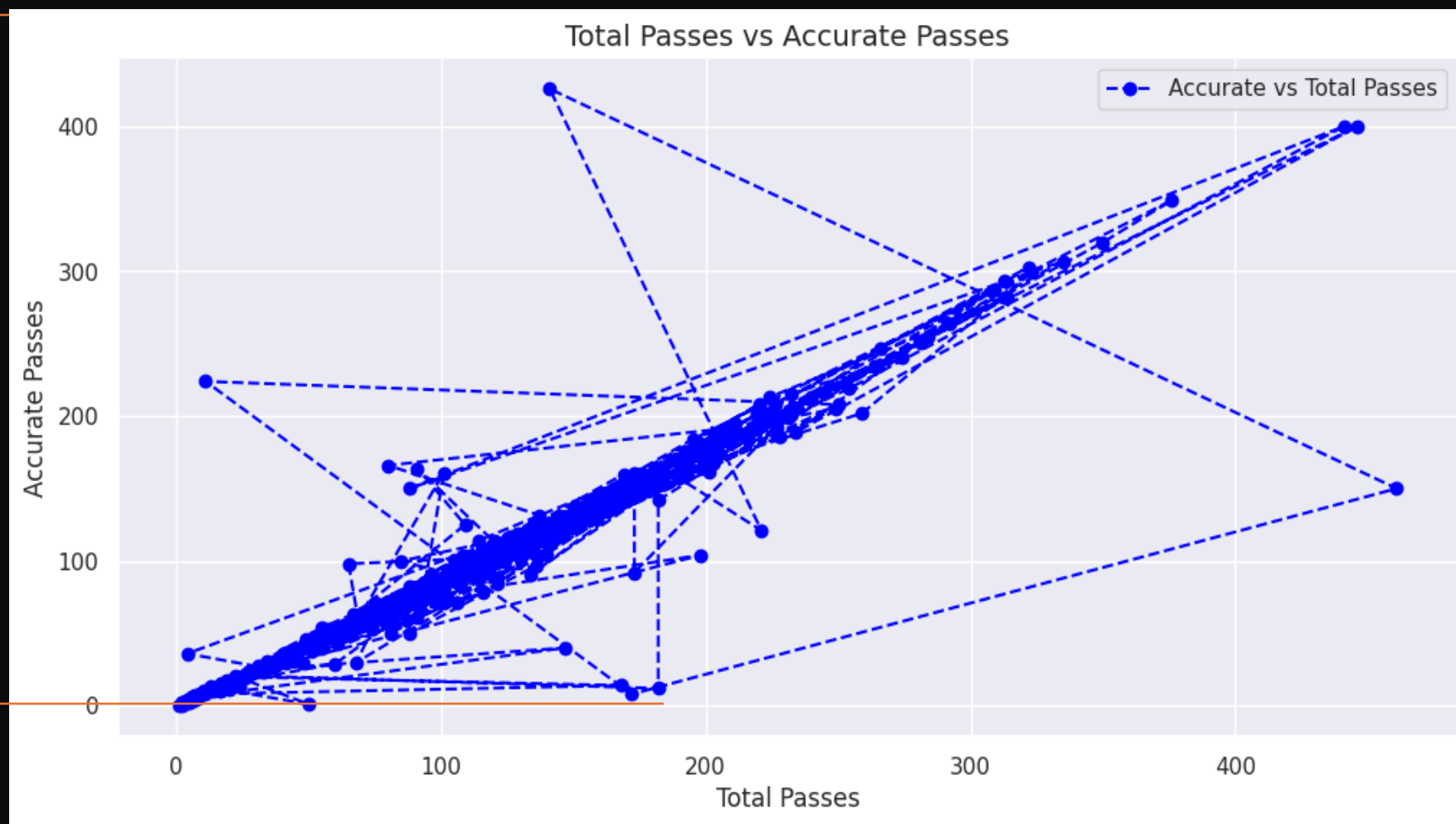- Accurate Passes = 0.5 points

```python
# Calculating a performance metric
midfielders_data['Performance Metric'] = (
    midfielders_data['Assists'] * 5 +
    midfielders_data['Key Passes'] * 3 +
    midfielders_data['Accurate Passes %'] * 2 +
    midfielders_data['Accurate Long Balls %'] * 1.5 +
    midfielders_data['Accurate Passes'] * 0.5
)
```

Passes completed:

- Out of 307 Champions League midfielders, 11 players completed over 300 passes, and 3 managed more than 400 in just 5 games, averaging roughly one pass per minute. This highlights their exceptional ability to control the game and maintain effective ball distribution.


Created using SoccerTutor.com Tactics Manager

Total Passes VS Accurate passes

Top 10 Midfielders:

	Name	Performance Metric
1	Savinho	535.745
2	Pascal Groe	517.780
3	Angelo Stiller	505.415
4	Atakan Karazor	488.285
5	Ulkay Gundogan	486.390
6	Granit Xhaka	483.065
7	Marc Casade	457.610
8	Tijjani Reijnders	455.565
9	Vitinha	452.320
10	Hans Vanaken	449.395

Top 10 Midfielders - Performance Metric
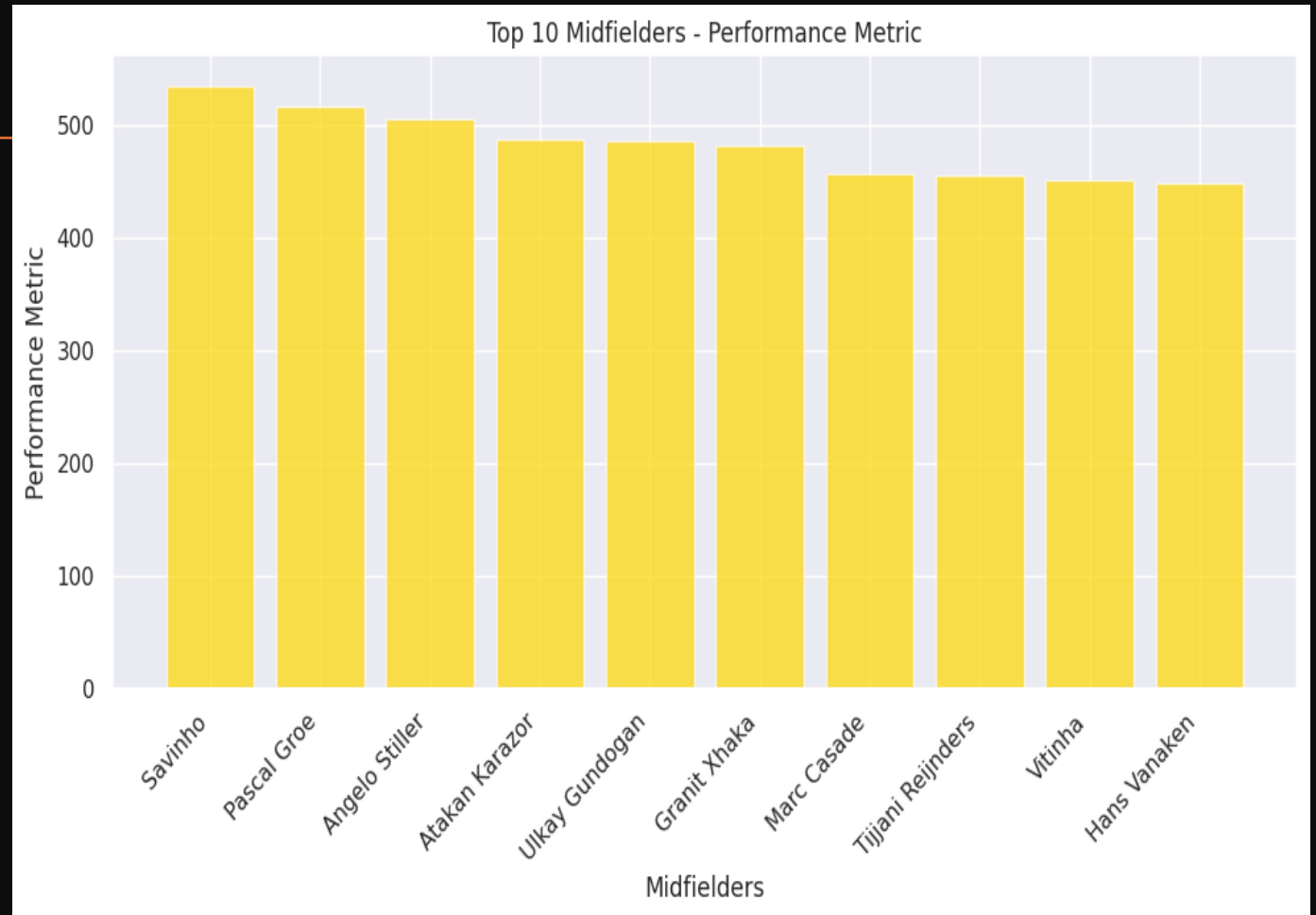
Defenders:

- **Positive Contributions:**

- **Tackles:** Effective in stopping the opposition; weighted at 2.

- **Interceptions:** High value for disrupting plays; weighted at 3.

- **Clean Sheets:** Strong defensive outcomes; high impact with a weight of 5.

- **Clearances:** Clearing danger contributes moderately; weighted at 1.5.

- **Negative Contributions:**

- **Errors Leading to Goals:** Costly mistakes, penalized with a weight of 2.

- **Dribbled Past:** Direct indicator of defensive lapses, subtracted outright.

```python
# Calculating a performance metric
defenders_data['Performance Metric'] = (
    defenders_data['Tackles'] * 2 +
    defenders_data['Interceptions'] * 3 +
    defenders_data['Clean sheet'] * 5 +
    defenders_data['Clearances'] * 1.5 -
    defenders_data['Errors lead to goal'] * 2 -
    defenders_data['Dribbled past']
)
```
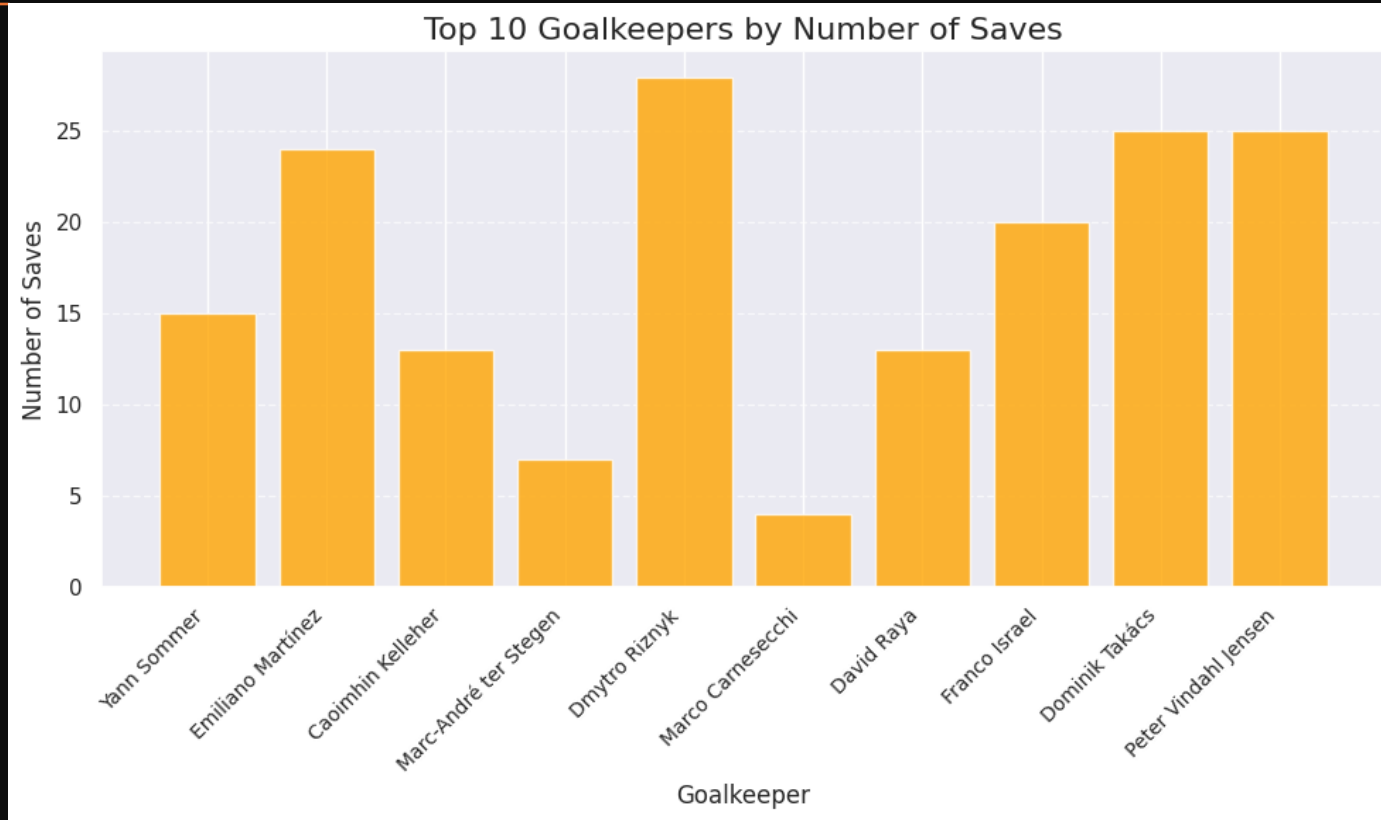
Top 10 defenders:

Top 10 Defenders:

|	Name	Performance Metric
1	Brendan Chardonnet	88.0
2	Jhon Lucumi	78.5
3	Tomas Araujo	76.5
4	Rafa Torrente	67.5
5	William Saliba	66.0
6	Stefan Posch	65.0
7	Yann Bisseck	63.0
8	Nico Schlotterbeck	61.0
9	Gabriel Magalhaes	60.0
10	Nicolas Otamendi	60.0

Top 10 Defenders - Performance Metric
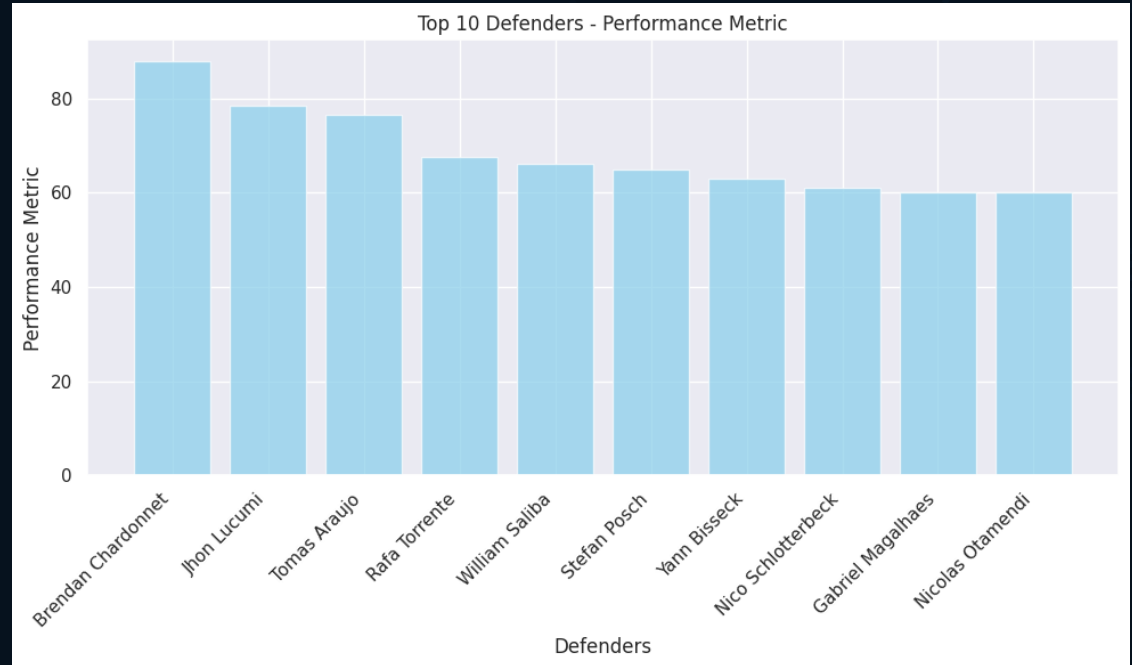
Goalkeepers:

- The graph highlight the top 10 goalkeepers in the Champions League. While saves are essential for a goalkeeper's performance, as shown by players like Dmytro Riznyk leading in total saves, clean sheets hold more importance. Clean sheets, achieved most notably by Yann Sommer, directly impact the team's table placement by preventing goals and securing crucial points.

Top 10 Goalkeepers by Number of Saves
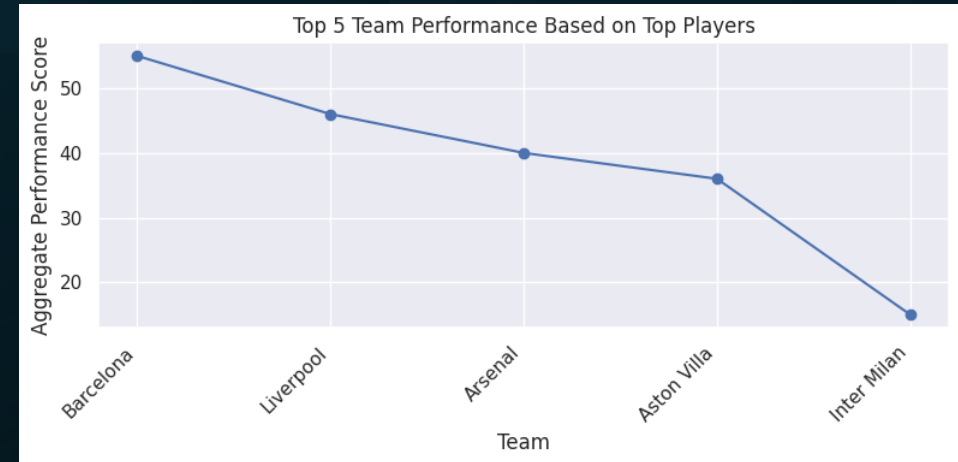
Top 10 Goalkeepers by Clean Sheets
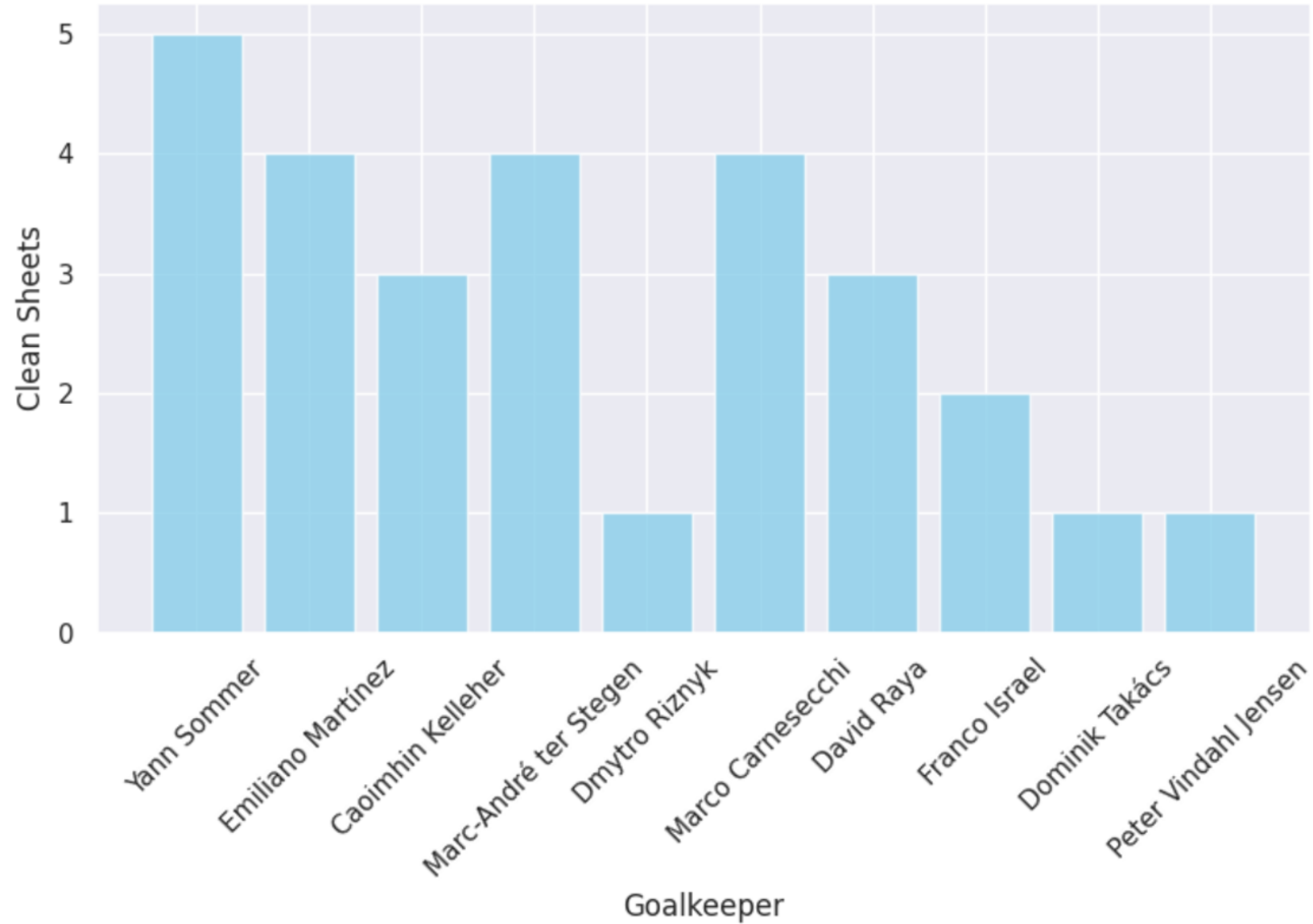
Final score calculation:

- Points from Goalkeepers=$4\times(10-$Index in Top 10$)$
- Points from Defenders=$3\times(10-$Index in Top 10$)$
- Points from Midfielders=$2\times(10-$Index in Top 10$)$
- Points from Attackers=$3\times(10-$Index in Top 10$)$

Winners

Predicted League Standings:

	Team	Score
1	Barcelona	55.0
2	Liverpool	46.0
3	Arsenal	40.0
4	Aston Villa	36.0
5	Inter Milan	15.0

Top 5 Team Performance Based on Top Players

Thank You