

PREMIER LEAGUE ANALYSIS

Season 2021/2022



OVERVIEW

- Web Scraping
- Database Schema
- Data cleaning
- Data visualization



TOOLS

- Python
- MySQL
- Excel
- Power bi
- Tableau



WEB SCRAPING PREMIER LEAGUE DATA

What is web scraping?

Why use web scraping for
Premier League data?

Overview of the scraping
process.

```
url = "https://fbref.com/en/comps/9/2021-2022/schedule/2021-2022-Premier-League-Scores-and-Fixtures"
```

Page 11

```
headers = {  
    "User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko)"  
}
```

```
response = requests.get(url, headers=headers)
```

```
soup = BeautifulSoup(response.text, 'html.parser')
```

```
table = soup.find("table", {"id": "sched_2021-2022_9_1"})
```

```
headers = [th.text.strip() for th in table.find("thead").find_all("th")]
```

```
rows = []
```

```
for row in table.find("tbody").find_all("tr"):  
    cols = [td.text.strip() for td in row.find_all(["th", "td"])]  
    if cols: # Ensure it's not an empty row  
        rows.append(cols)
```

```
df = pd.DataFrame(rows, columns=headers)
```

```
df.to_csv("Premier_League_2021-2022_Scores_Fixtures.csv", index=False, encoding="utf-8")
```

DATA SOURCE & TARGET WEBSITE

Dribbling is more than just moving the ball; it's about outmaneuvering opponents and creating scoring opportunities. This skill forms the foundation of an attacking mindset in soccer.

Website used: FBref.com

What kind of data was
extracted?
(Match schedules, results,
teams, etc.

Challenges of extracting
data from sports websites

01

Teams&Dates

- Information about each team (name, stadium, league) .
Match dates, weeks, months, and years.
-

02

wages

- Salary details of teams.
-

03

Stats

- Performance details like wins, goals, and points.

DATABASE SCHEMA EXPLANATION



The code

```
-- creating star schema

) CREATE TABLE teams (Team_ID INT PRIMARY KEY AUTO_INCREMENT, Team_Name TEXT NOT NULL, Stadium TEXT, League TEXT);
) CREATE TABLE dates (Date_ID INT PRIMARY KEY AUTO_INCREMENT, Match_Date DATE NOT NULL, Week INT, Month INT, Year INT);
) CREATE TABLE wages (Wages_ID INT PRIMARY KEY AUTO_INCREMENT, Team_ID INT, Number_of_Players INT, Weekly_Wages INT, Annual_Wages INT,
FOREIGN KEY (Team_ID) REFERENCES teams(Team_ID));
) CREATE TABLE stats (Stats_ID INT PRIMARY KEY AUTO_INCREMENT, Team_ID INT, Wins INT, Losses INT, Draws INT, Goals_For INT, Goals_Against INT,
Points INT, FOREIGN KEY (Team_ID) REFERENCES teams(Team_ID));
) CREATE TABLE scores_fixtures (
    Fixture_ID INT PRIMARY KEY AUTO_INCREMENT, Team_ID INT, Opponent_ID INT, Date_ID INT, Stats_ID INT, Wages_ID INT, Score TEXT,
    Attendance INT,
    FOREIGN KEY (Team_ID) REFERENCES teams(Team_ID),
    FOREIGN KEY (Opponent_ID) REFERENCES teams(Team_ID),
    FOREIGN KEY (Date_ID) REFERENCES dates(Date_ID),
    FOREIGN KEY (Stats_ID) REFERENCES stats(Stats_ID),
    FOREIGN KEY (Wages_ID) REFERENCES wages(Wages_ID)
);

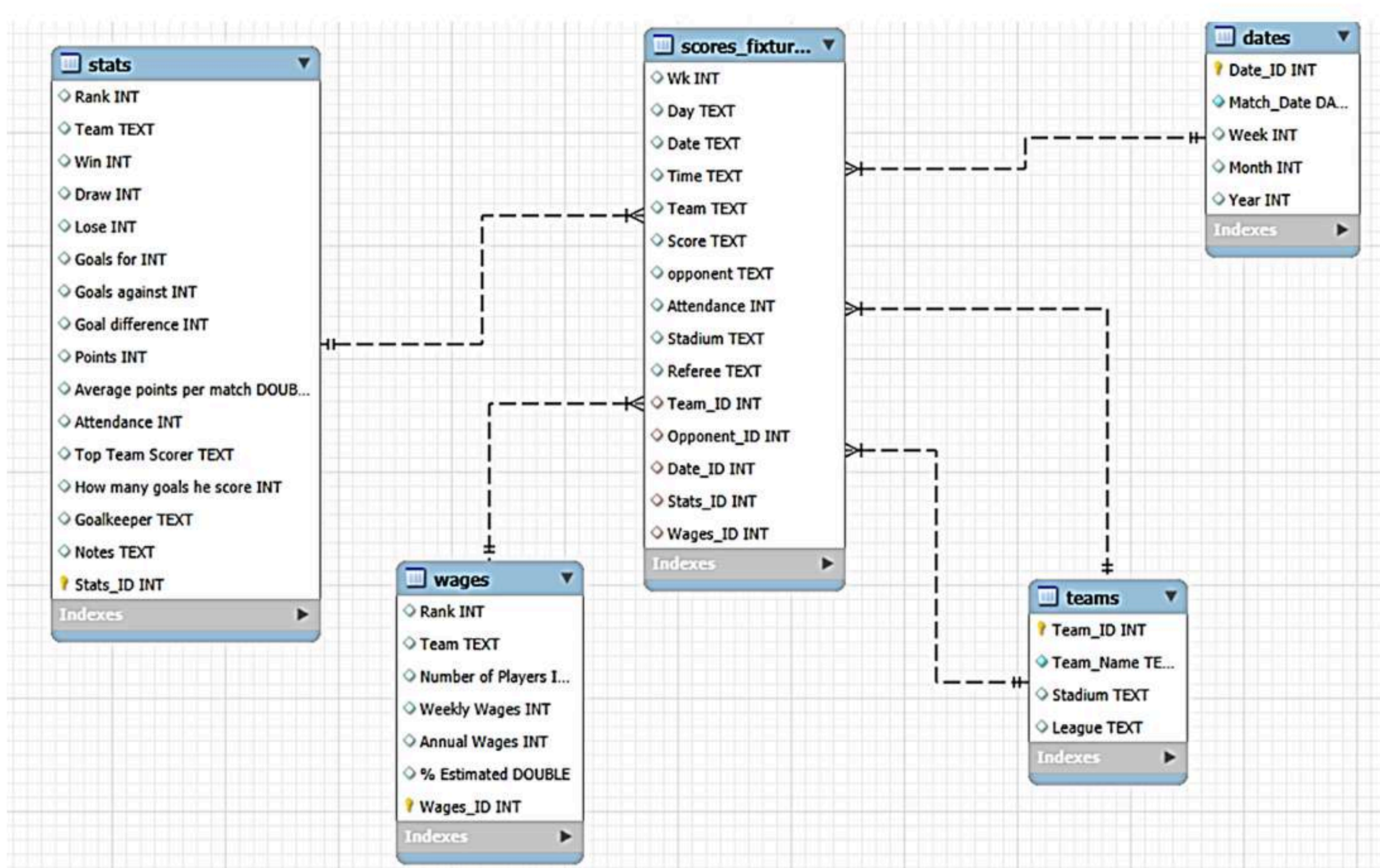
-- .....

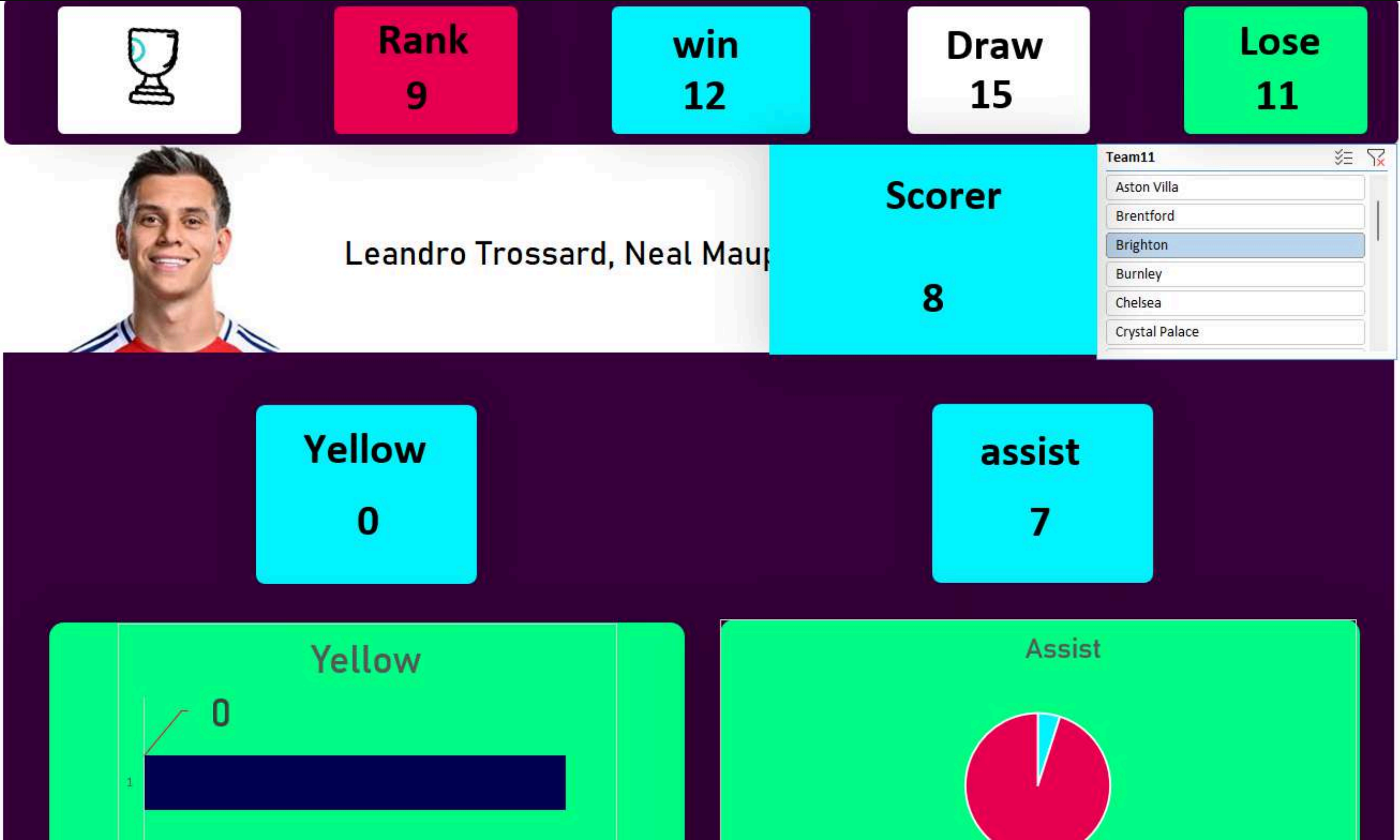
) ALTER TABLE scores_fixtures ADD COLUMN Team_ID INT, ADD COLUMN Opponent_ID INT, ADD COLUMN Date_ID INT,
ADD COLUMN Stats_ID INT, ADD COLUMN Wages_ID INT;

-- .....

) ALTER TABLE scores_fixtures ADD CONSTRAINT fk_scores_team FOREIGN KEY (Team_ID) REFERENCES teams(Team_ID),
ADD CONSTRAINT fk_scores_opponent FOREIGN KEY (Opponent_ID) REFERENCES teams(Team_ID),
ADD CONSTRAINT fk_scores_date FOREIGN KEY (Date_ID) REFERENCES dates(Date_ID),
ADD CONSTRAINT fk_scores_stats FOREIGN KEY (Stats_ID) REFERENCES stats(Stats_ID),
ADD CONSTRAINT fk_scores_wages FOREIGN KEY (Wages_ID) REFERENCES wages(Wages_ID);
```


Database Schema Diagram





EXCEL DASHBOARD

01

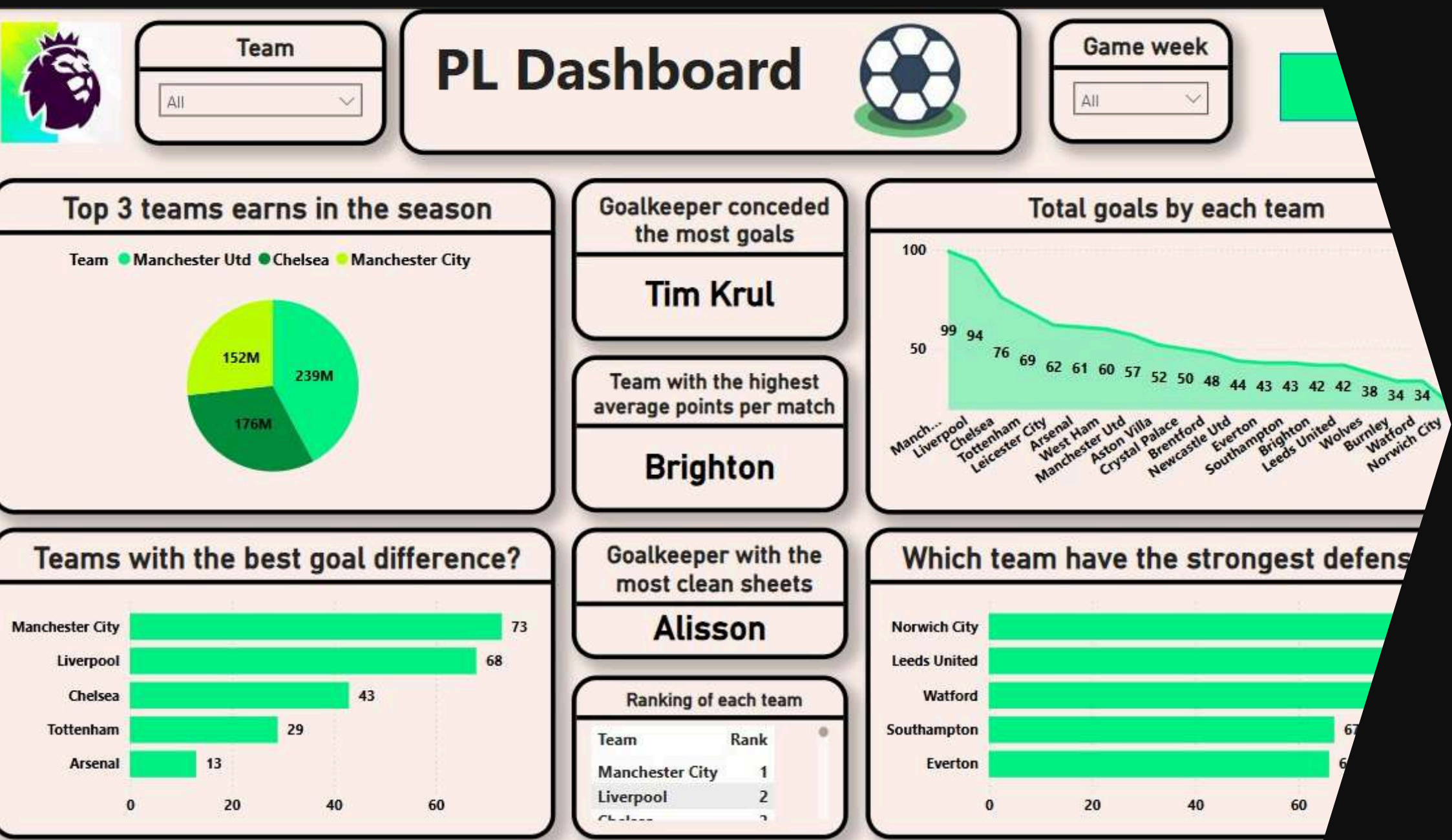
Top Performing Teams & Players

- Manchester City dominated the league in multiple aspects:
- Most goals scored.
- Best goal difference (+73).
- Ranked #1 overall.
- Liverpool & Chelsea also performed well in terms of goal difference and earnings.
- Mohamed Salah & Son Heung-min were the top goal scorers with 23 goals each.

02

Financial Insights

- Manchester United had the highest annual wages, spending £239M on salaries.
- Chelsea & Manchester City followed closely with £176M and £152M, respectively.
- The top 3 earning teams were:
Manchester United - £239M
Chelsea - £176M
Manchester City - £152M



POWER BI DASHBOARD

- 01

Key Metrics Shown

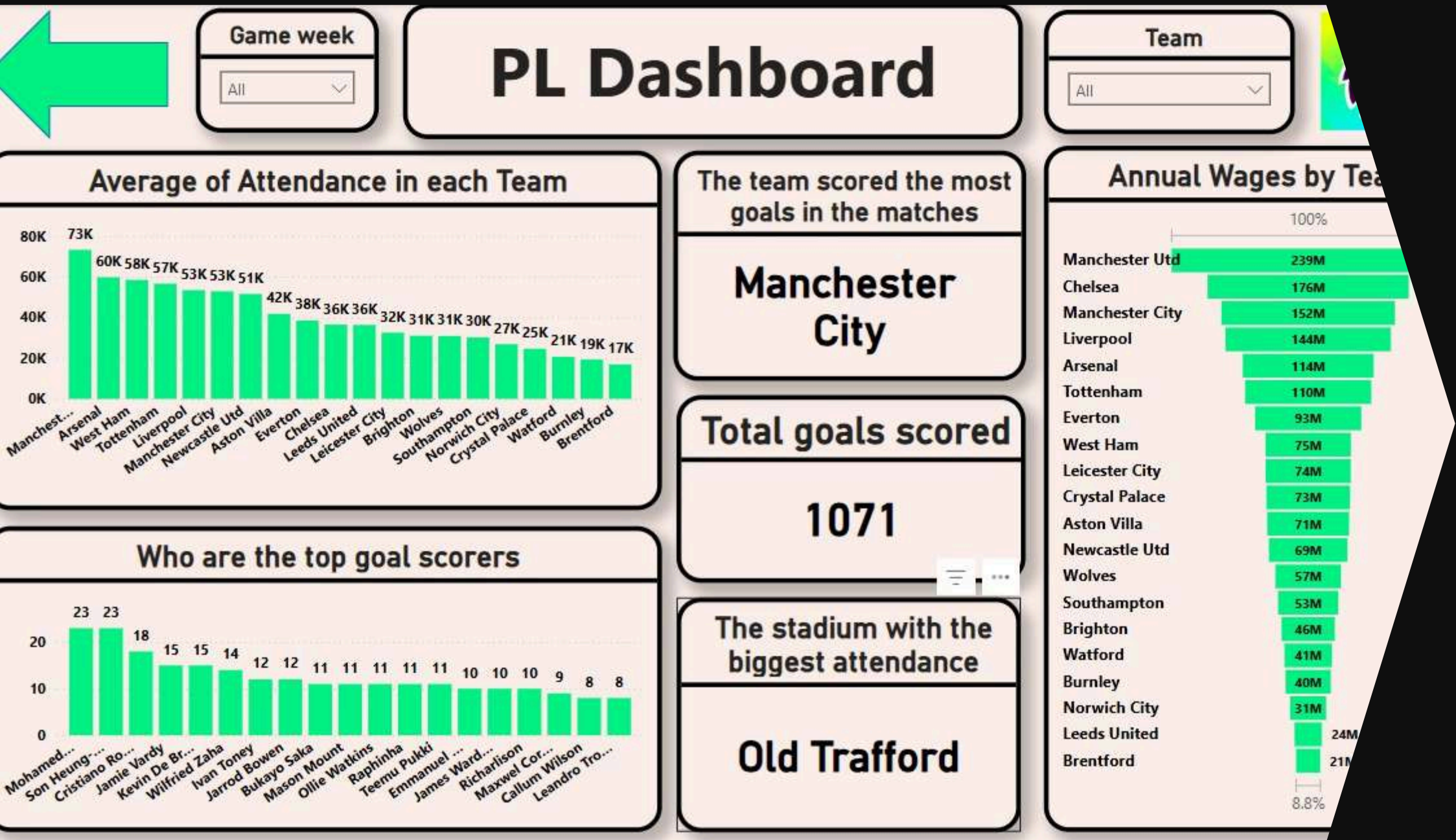
 - Teams with the best goal difference.
 - Total goals by each team.
 - Top teams by earnings.
 - Defensive strength of team
- 02

Interactive Features

 - Dropdowns for team and game week selection.
- 03

Insights Drawn

 - Manchester City achieved the best goal difference.
 - Tim Krul conceded the most goals, while Alisson had the most clean sheets.
 - Norwich City had the weakest defense.



POWER BI DASHBOARD

01

Key Metrics Shown

- Average attendance per team.
- Top goal scorers.
- Annual wages by team.
- Team with the most goals and the stadium with the biggest attendance.

02

Interactive Features

- Filters by game week and team.

03

Insights Drawn

- Manchester City had the highest goals scored.
- Old Trafford hosted the biggest average attendance.
- Manchester United led in annual wages.



INSIGHTS FROM VISUALIZATIONS

01

Teams with strong offensive performance (Manchester City, Liverpool).

02

Impact of attendance on team performance.

03

Wage spending vs. performance (e.g., Manchester United vs. Manchester City).

04

Players contributing most to team success (top scorers)

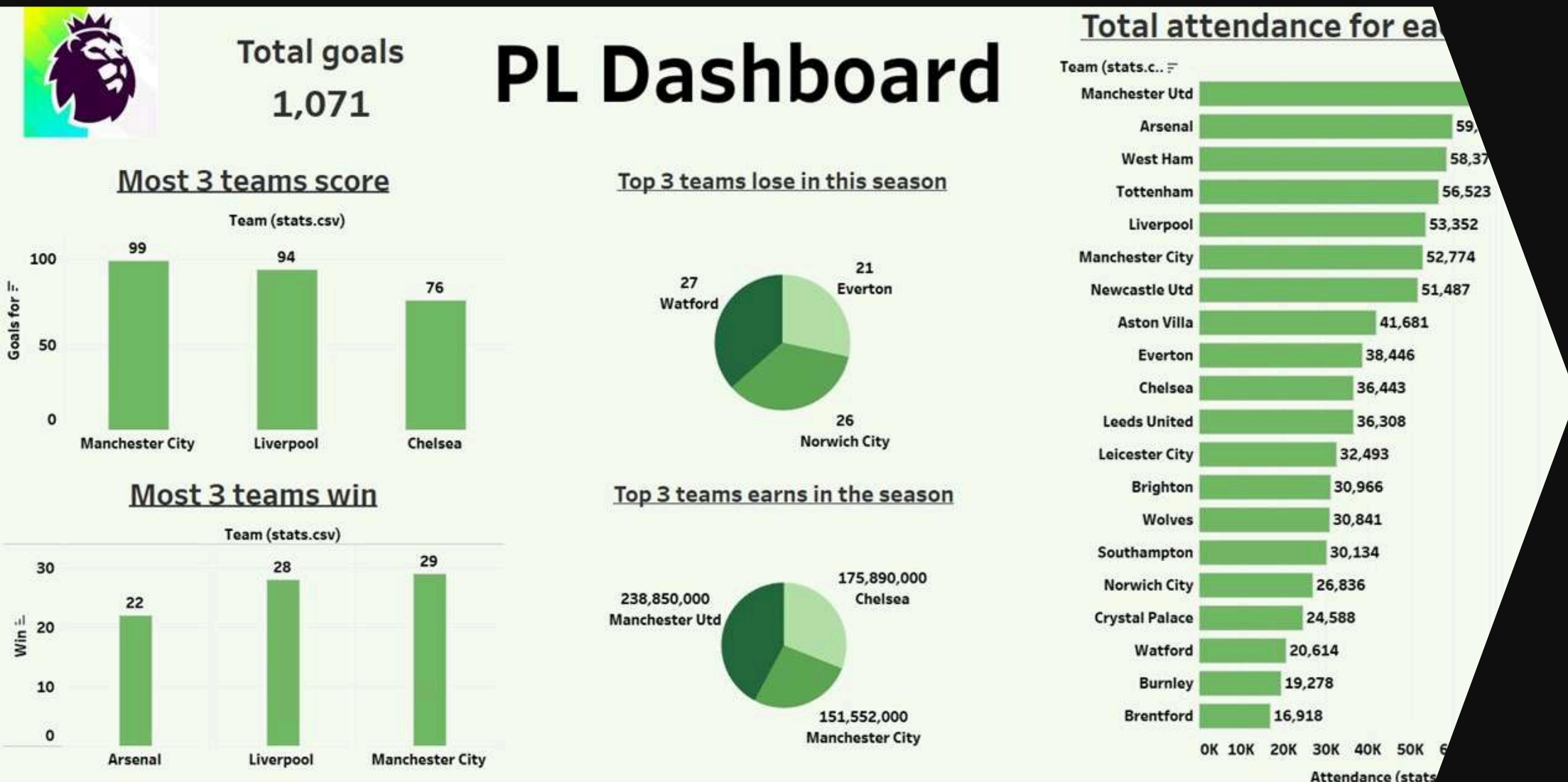


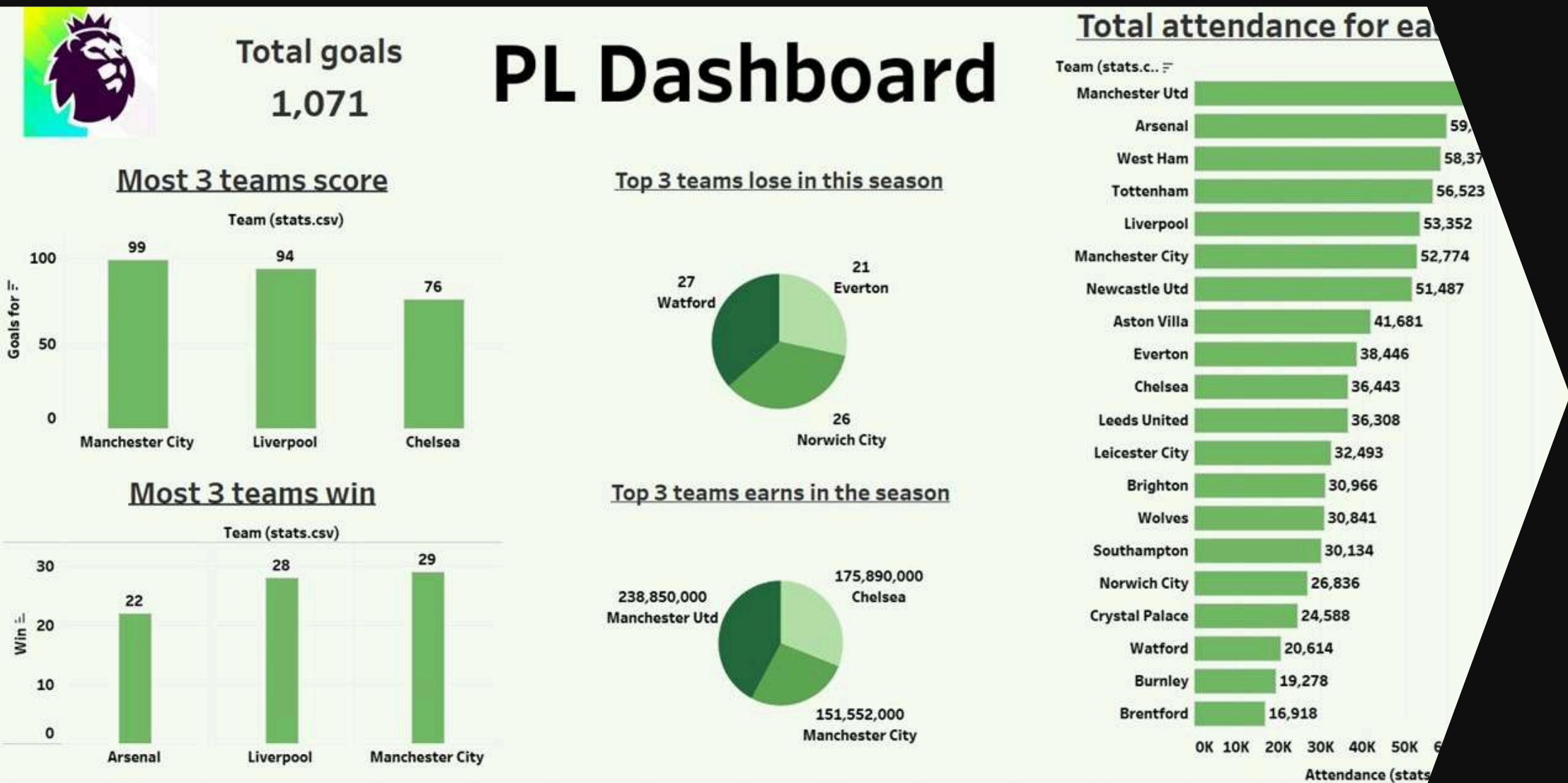
TABLEAU DASHBOARD

01

Defensive Strength

- Norwich City had the strongest defense with 84 defensive actions.
- Leeds United & Watford also showed strong defensive performances.
- Alisson (Liverpool) had the most clean sheets (most matches without conceding a goal).

TABLEAU DASHBOARD



02

Attendance & Stadium Insights:

- Manchester United (Old Trafford) had the highest stadium attendance.
- The average attendance for Arsenal & Tottenham was above 60K.
- Key Takeaways
- Manchester City was the most dominant team in terms of goals, rankings, and goal difference.
- Manchester United led in wages, showing a strong financial investment in players.
- Brighton overperformed based on their average points per match.
- Defensive strength varied, with Norwich City leading in defensive actions.

THANK YOU

Supervised by:
Eng: Huda Hemdan

Seif Wael Hassan

Mohab Emad Morrris

Mohanad Mohamed Saleh
