Final Report

English Premier League 2024/25 Player and Team Performance Analysis

1. Introduction:

Football is the most popular sport in the world, with around 3.5 billion people considering it their favourite. Millions of people work in football, whether as players, coaches, analysts, medical teams, or even content creators.

Football today is very different from how it was 50 years ago. In the past, the game depended mainly on player talent and coaching ideas. Now, football has become more advanced and technical because of the use of data. Every professional team has a data analysis department that studies player performance, team performance, and opponent strategies.

Data analysis now plays a key role in almost every football decision, from choosing the best formation and tactics to selecting starting players. Coaches rely heavily on data to improve performance and achieve better results. This shows how important data analytics has become in modern football and why analysing football data can help teams make smarter decisions.

The English Premier League is one of the most competitive and popular football leagues in the world. Millions of people watch it every week, following their favourite teams and players closely. Each season, the league produces a huge amount of data, from goals and assists to tackles, passes, and saves.

However, this data is not always easy to understand or use. Many fans want to know which players perform best, how their team is improving, and what areas need more work. The best way to answer these questions is through data analysis that turns numbers into clear insights.

This project focuses on the 2024/25 Premier League season, analysing real statistics to understand player and team performance. The goal is to highlight who performed well, how teams compared, and what factors influenced the final league table, including why some teams achieved success and others were relegated.

This analysis will help EPL fans follow their teams and favourite players more closely, using real numbers instead of opinions. It will show which players stand out in attack, defence, and creativity, and how their performances affect team success.

Coaches, analysts, and scouts can also benefit from this data to spot top performers and areas that need improvement. For football content creators, it provides accurate statistics to support discussions, rankings, and season reviews.

Overall, this project makes EPL data easier to understand and helps everyone, from fans to professionals, see the game from a new, data-driven point of view.

2. Data Analysis and Computation:

2.1. Dataset:

Dataset Name: English Premier League - Player Stats - 24/25

Dataset Details: The dataset includes 563 rows representing 562 different players, with 57 columns containing various performance statistics.

Size: The file size is 100.27 kB.

Access: I'll download the dataset from Kaggle.

Source: English Premier League - Player Stats - 24/25

2.2. Data Wrangling and Cleaning:

2.2.1. Data Profiling:

- Apply Excel filters to the dataset.

- Distinct values in some columns:
  - 20 unique clubs.
  - 4 unique positions (GKP, DEF, MID, FWD).

- Errors:
  - The club Brighton & Hove Albion appears twice, once as "Brighton & Hove Albion" and once as "Brighton".
  - Some players have more shots on target than total shots, which isn't possible. Many others are having wrong numbers for shots on target.
  - The conversion rate is not accurate since it depends on incorrect shot data.
  - One goalkeeper is shown with goals scored, which is likely a position or data entry mistake.
  - The positions of a few players are wrong.

- Zeros in data:
  - Many players (especially non-goalkeepers) have zeros in goalkeeper-specific stats.
  - Goalkeepers and players with fewer minutes played have zeros in many statistics because they were less involved in certain aspects of the game.

- Summary of insights:

  The data contains valuable data of player performance in the EPL 2024/25 season, but several statistics need to be corrected before analysis. Errors in shooting and positional data could affect accuracy, so cleaning these issues is essential. Despite that, the dataset remains valuable, allowing us to compare performance across clubs, positions, and individual players.

2.2.2. Data Wrangling:

- “Shots on target” and “Conversion %” columns are removed as they contain inaccurate data.
- Feature engineered “Club” column; substitute any “Brighton” with “Brighton & Hove Albion”.
- Feature engineered “Position” column, substitute “MID” with “GKP” for Alex Palmer and substitute “GKP” with “DEF” for Emerson.
- Featured the engineered “Defensive Contributions” column by adding “Clearances”, “Interceptions”, “Blocks”, and “Tackles” columns.

Data Table Schema:

Field	Type	Description
Player Name	STRING	Name of each player.
Club	STRING	The club that the player is playing for.
Position	STRING	The player’s position on the pitch.
Appearances	INTEGER	Number of matches played by the player.
Minutes	INTEGER	Minutes played by the player.
Goals	INTEGER	Goals scored by the player.
Assists	INTEGER	Assists given by the player.
Shots	INTEGER	Shots by the player.
Touches	INTEGER	Number of times the player touched the ball.
Passes	INTEGER	Number of passes by the player.
Crosses	INTEGER	Number of crosses by the player.
Through Balls	INTEGER	Numbers of through ball passes by the player.
Carries	INTEGER	Number of times ball was carried by the player.
Clearances	INTEGER	Number of times the player cleared the ball.
Interceptions	INTEGER	Numbers of times the player stopped a pass.
Tackles	INTEGER	Number of tackles made by the player.
Fouls	INTEGER	Number of fouls committed by the player.
Yellow Cards	INTEGER	Number of yellow cards for each player.
Saves	INTEGER	Number of saves made by each goalkeeper

2.3. Data Analysis:

The goal of this project is to identify the best players throughout the 2024/25 Premier League season. To achieve this, we analysed player statistics based on their positions and clubs to determine both individual excellence and overall team performance.

The analysis is divided into six main parts: goalkeepers' performance, defenders' performance, midfielders' defensive performance, midfielders' attacking performance, forwards' performance, and team performance.

1. Goalkeepers' performance:

This section shows how the goalkeepers performed throughout the season. The analysis highlights who stood out as the best goalkeeper based on key metrics.

During this analysis, we used the saves, punches, high claims, and goals prevented columns to create meaningful charts that show which goalkeepers performed best in each category. These columns were selected because they contain zeros for all other positions, confirming that they are specific to goalkeepers.

Comparing and analysing these features for each goalkeeper allows us to create charts in the dashboard, showing the audience which goalkeeper had performed the best that season.

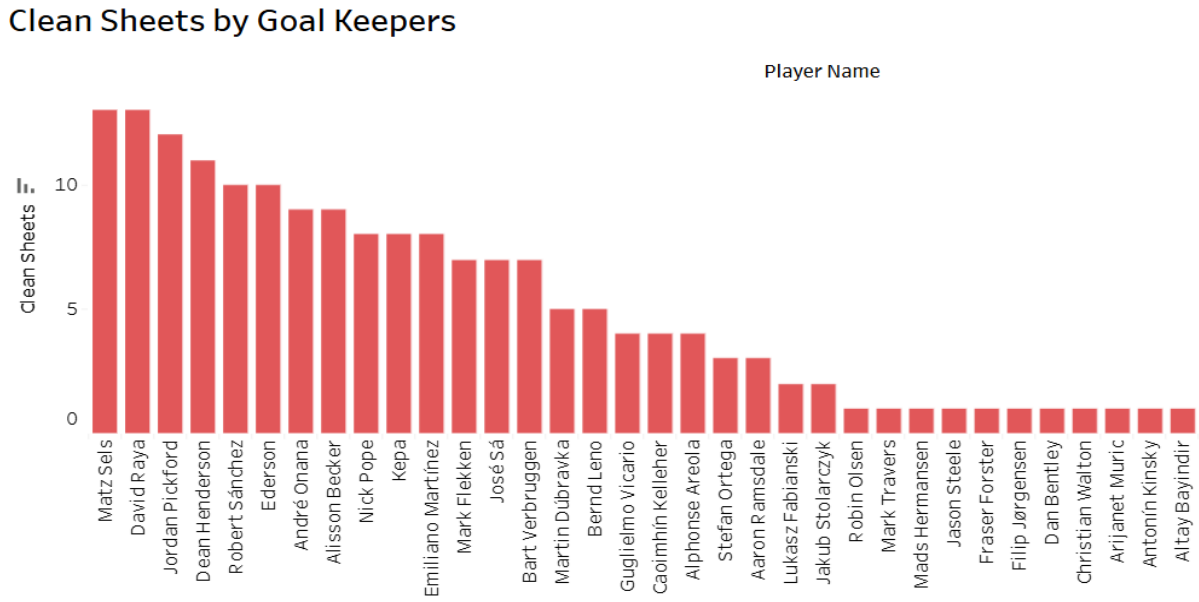

Figure 1. Clean Sheets by Goal Keepers.

Figure 1 shows the number of clean sheets achieved by each goalkeeper, an important metric for identifying the best-performing goalkeepers this season.

2. Defenders' performance:

After analysing the goalkeepers, the next step focuses on the defenders to understand the strongest backline performers.

We used the feature-engineered column (Defensive Contributions), which is the sum of Clearances, Interceptions, Blocks, and Tackles, to show which defenders made the most contributions throughout the season across all defensive metrics. This approach is more inclusive and helps save time and space in the dashboard by combining multiple features into one.

We also used the Clean Sheets column to show which defenders helped their team keep a clean sheet during matches. In addition, we included the Successful Ground Duels % and Successful Aerial Duels % columns instead of total duels. Using percentages gives a clearer picture of player effectiveness, since a high number of duels does not always mean strong performance if many were lost.

We also considered the goals and assists for each defender to compare their attacking contributions, which is especially important for full-backs.

Comparing and analysing these features for each defender allows us to create clear charts in the dashboard, helping the audience see which defenders had performed the best that season.

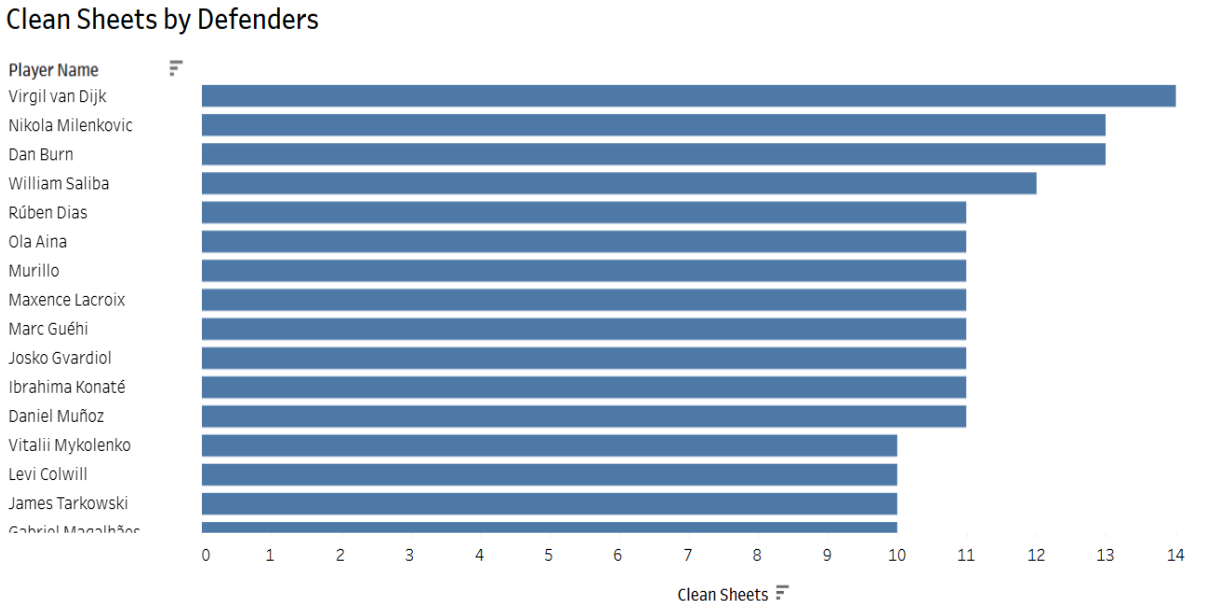

Figure 2. Clean Sheets by Defenders.

Figure 2 shows the number of clean sheets achieved by each defender, an important metric for identifying the best-performing defenders this season.

3. Midfielders' Performance:

Midfielders play a crucial role in both attacking and defending, making them the most important players on the pitch in my opinion. This section will analyse their overall impact using metrics that reflect their creativity, passing effectiveness, and defensive contributions.

The reason for analysing both attacking and defensive performance is that the top-performing midfielders in attack are often completely different from the top performers in defence. Therefore, to accurately determine who the best midfielders were throughout the season, it is necessary to evaluate both categories. This allows us to highlight not only specialist attacking or defensive midfielders but also those who make balanced contributions in both areas.

For the attacking analysis, we will use goals and assists to measure direct contributions, along with passing and possession metrics such as through balls, successful passes %, successful final third passes %, touches, carries, and progressive carries to show how effectively midfielders create chances and help move their team from defensive zones into attacking positions.

For the defensive analysis, we will use defensive contributions (the sum of tackles, interceptions, blocks, and clearances) to show how midfielders supported their defence. We will also use successful ground duels % and successful aerial duels % to reflect their effectiveness in winning the ball back and competing physically.

Comparing and analysing these features for each midfielder will allow us to create clear bar and line charts in the dashboard, helping the audience identify the best-performing midfielders in both attacking and defensive roles throughout the season.

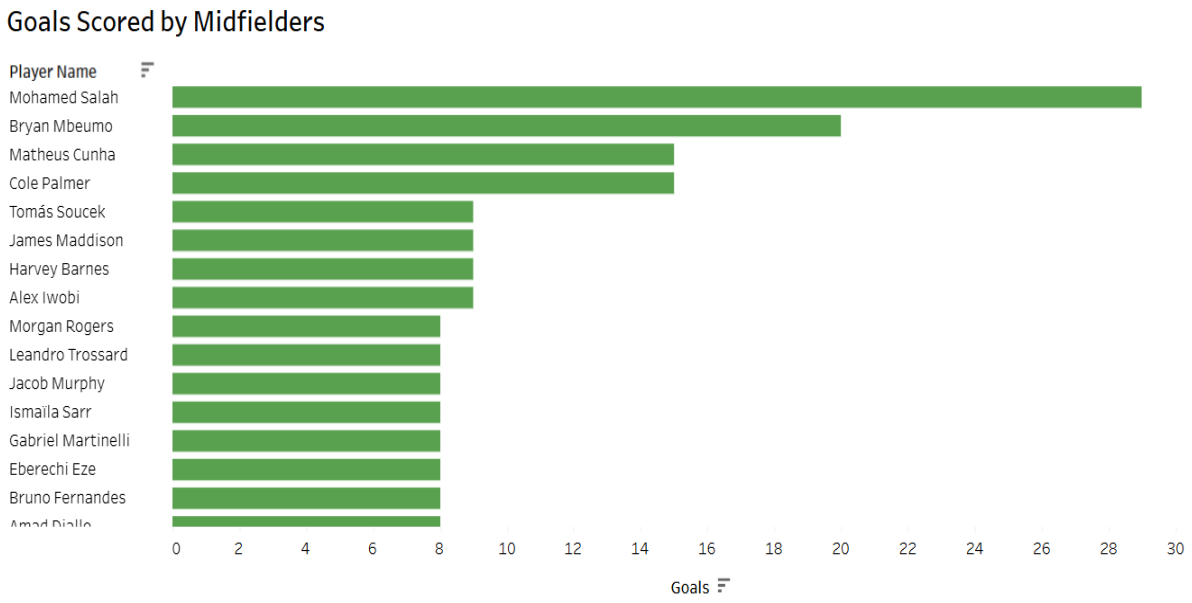

Figure 3. Goals Scored by Midfielders.

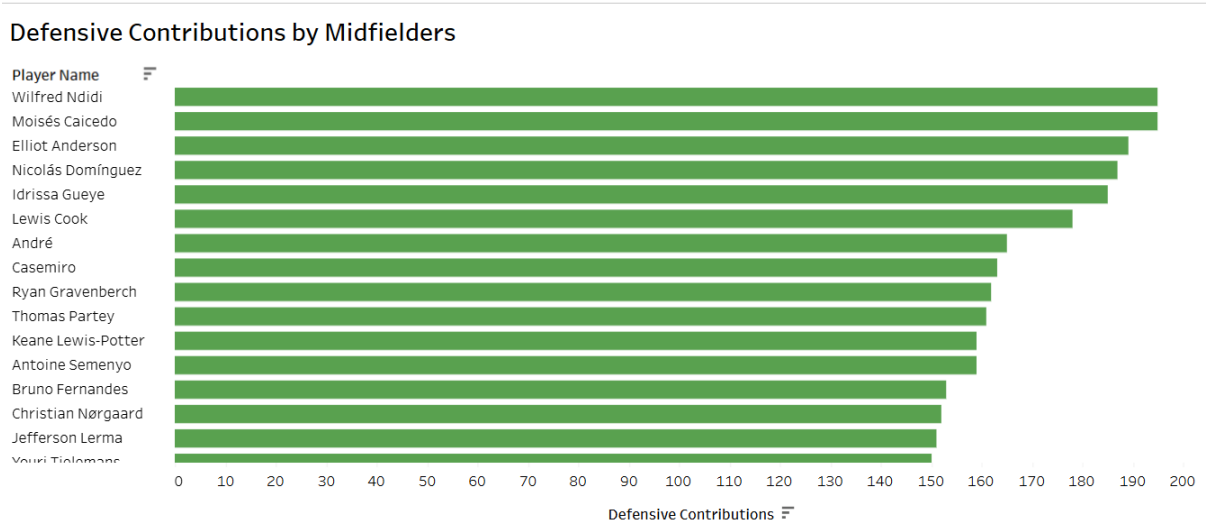

Figure 4. Defensive Contributions by Midfielders.

Figure 3 shows the number of goals scored by each midfielder, an important metric for identifying the best-performing attacking midfielders this season, and Figure 4 shows the number of defensive contributions achieved by each midfielder, an important metric for identifying the best-performing defending midfielders this season.

4. Forwards' Performance:

Forwards are primarily responsible for scoring goals, making them the most decisive players in the final third. This section analyses their attacking effectiveness using key metrics that show both productivity and efficiency in front of goal.

The main aim of this analysis is to identify which forwards had the highest impact on their team's attacking output throughout the season. Unlike other positions, the performance of forwards is mainly judged on their ability to score goals, convert chances, create assists, and apply pressure on the opposition's defence.

To measure goal-scoring performance, we will use goals, shots, and big chances missed. Comparing shots to goals will help us understand which forwards are more clinical and efficient in front of goal. This will be clearly shown in the charts created for this section.

To measure creative involvement, we will also use the Assists column to show which forwards contributed to the build-up play, not just finishing chances.

Analysing and comparing these metrics for each forward will allow us to create clear charts in the dashboard, helping the audience determine which forwards were the most dangerous, efficient, and valuable to their teams throughout the season.

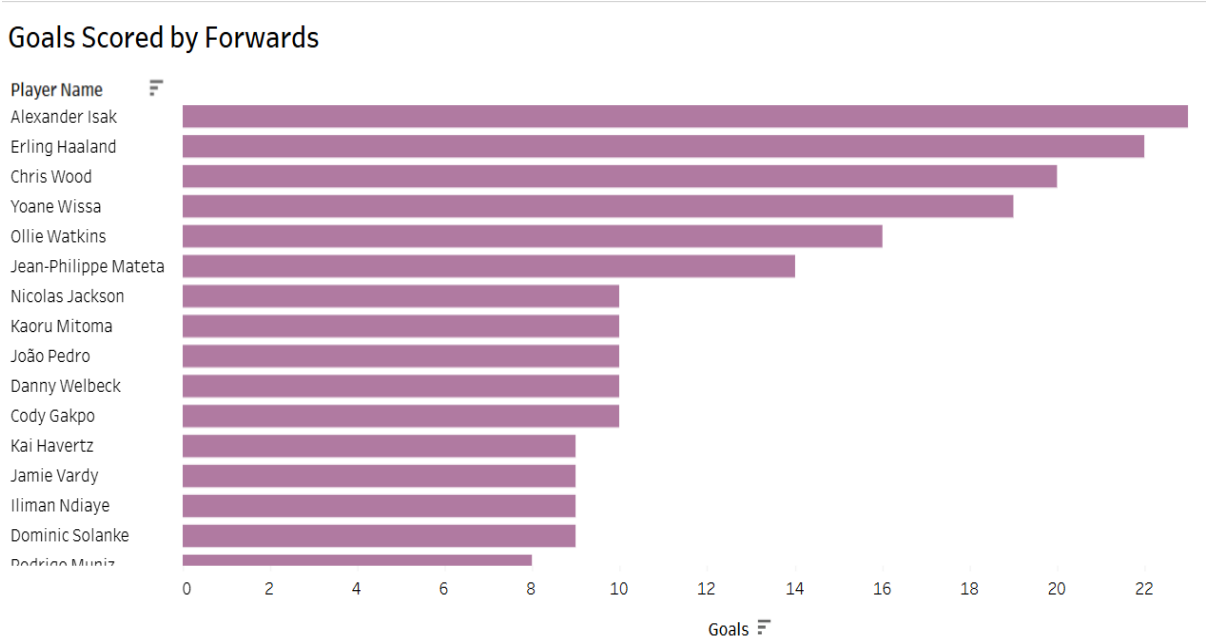

Figure 5. Goals Scored by Forwards.

Figure 5 shows the number of goals scored by each forward, an important metric for identifying the best-performing forwards this season.

5. Team Performance

The final part of this analysis focuses on overall team performance to understand which clubs were the most dominant throughout the season in attack, defence, and consistency. Analysing team-level metrics allows us to identify the strongest teams, understand how their success was achieved, and compare their playing styles.

This analysis will help highlight not only the teams competing for the title but also those that struggled or were at risk of relegation. Understanding the statistical differences between these teams can explain league outcomes and reveal key factors behind their performance.

To analyse attacking performance, we will use total team goals and assists to show how effective each team was at creating and converting chances.

For defensive performance, we will use goals conceded and clean sheets to assess how strong each team is at preventing goals. Teams with lower goals conceded and higher clean sheets are likely to be those competing at the top of the table.

Possession and build-up play will also be examined using successful pass %, touches, and progressive carries, helping us understand how teams control games and transition from defence to attack.

By comparing these core metrics across clubs, we are able to identify the most balanced teams, the strongest attacking sides, and those who rely heavily on defensive structure. This will be represented through a combination of bar charts in the dashboard, giving a clear visual overview of team strengths and weaknesses throughout the season.

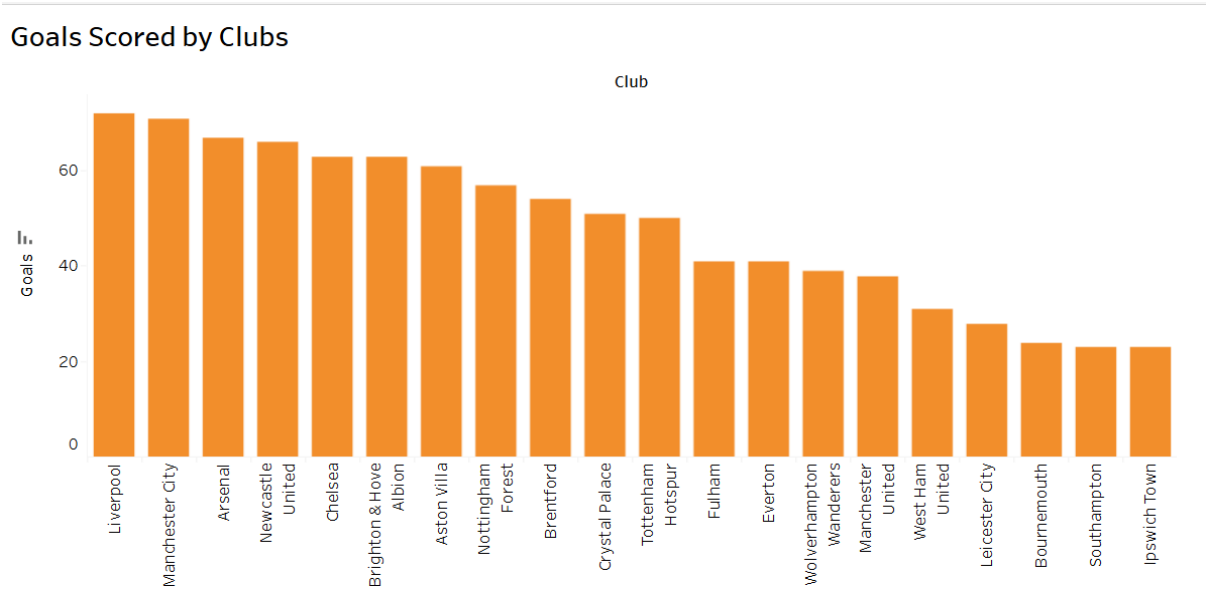

Figure 6. Goals Scored by Clubs.

Figure 6 shows the number of goals scored by each club, an important metric for showing how the clubs were performing this season.

3. Challenges and Solutions:

The main challenge in this project was how to present the data in a way that helps people identify the best players of the season. Since players have different positions and responsibilities on the pitch, comparing them all together in one dashboard wouldn't be fair or meaningful. The best solution was to divide the analysis by position. This allowed fair comparisons between players who share similar roles and made it easier to identify the best performer in each position, making the overall analysis more accurate and insightful.

When analysing the forwards' performance, there was an issue with the total shots on target and the conversion rate. Some players had more shots on target than total shots, which is not possible. This made both metrics unreliable, so I decided to remove these columns from the analysis to maintain accuracy. This is why those two metrics do not appear in the forwards' dashboard.

For the defenders' dashboard, comparing each defensive metric separately (tackles, interceptions, clearances, and blocks) made the dashboard cluttered and difficult to interpret. To solve this, I combined all these metrics into one feature-engineered column called defensive contributions, which represents the total of those four categories. This approach simplified the analysis and made it easier to compare defenders effectively.

These solutions helped create a clean, fair, and meaningful analysis across all positions.

4. Dashboard Description:

The dashboard can be accessed through this link:
English Premier League 2024/25 Player and Team Performance Analysis

The dashboard is divided into six parts: goalkeepers, defenders, defending midfielders, attacking midfielders, forwards, and clubs.

Each part highlights the best players in their positions using multiple metrics, not just one. This allows the audience to compare players and decide who they think performed best based on their own perspective.

The goalkeepers' section includes charts showing clean sheets, total saves, high claims, punches, and goals prevented by each goalkeeper.

The defenders' section includes charts for clean sheets, defensive contributions, ground duels won %, aerial duels won %, and attacking contributions (goals and assists).

The defending midfielders' section shows defensive contributions, ground duels won %, and aerial duels won % for each midfielder.

The attacking midfielders' section displays goals and assists, through balls, progressive carries, and successful pass %.

The forwards' section highlights goals, assists, shots, and big chances missed for each forward.

The clubs' section includes charts showing goals scored and conceded, clean sheets, successful passes, and progressive carries for each club.

The dashboards are interactive and engaging. They include filters that allow users to view players by club, and by selecting a specific player, all of their performance statistics will appear instantly, making it easy to explore and understand their performance throughout the season.

5. Conclusion:

This project helped us understand the performance of Premier League 2024/25 players based on their positions, making it easier to choose the best formation or even the best player in each position using real data. It also allows everyone to make their own judgements, since people have different opinions about what makes a player the best in their role.

The data analysis and dashboards showed that relying only on watching matches or news is not enough to know who the best players are. Data makes the process more insightful and interesting, as it helps explain why certain players deserve to be in the team of the season.

For goalkeepers, the analysis showed that Matz Sels from Nottingham Forest was the best goalkeeper last season. He kept 13 clean sheets, prevented an average of 4.3 goals, and made 120 saves, ranking fourth among goalkeepers in total saves, proving he had an excellent season.

For defenders, the analysis showed that Virgil van Dijk from Liverpool and Nikola Milenkovic from Nottingham Forest were the top two centre-backs. Van Dijk kept 14 clean sheets, while Milenkovic had 13, ranking them as the highest among defenders. Both also recorded very high defensive contributions, placing them in the top six defenders overall.

For midfielders, the analysis revealed that Mohamed Salah from Liverpool was not only the best midfielder but also the best overall player of the season. He scored 29 goals and provided 18 assists, playing a huge role in Liverpool winning the title. He also made 231 progressive carries and achieved a 90% passing accuracy, showing that he contributed to both scoring and playmaking.

For forwards, the analysis showed that Alexander Isak from Newcastle United was the best striker of the season. He scored 23 goals from 99 shots and played a key role in helping Newcastle United qualify for the UEFA Champions League.

6. Future Work:

This project can be useful for coaches, analysts, and scouts to spot top performers and identify areas that need improvement. Football content creators can also use it to support their rankings and discussions with real data. The same approach can be applied to any past or future Premier League season to track player and team performance over time. In the future, we can also analyse the performance of top players across different matches or throughout the entire season to gain deeper and more insightful analysis. Adding predicted metrics like expected goals and expected assists would make the analysis even more exciting and valuable.