Exploratory Data Analysis
English Premier League 2024/25 Player and Team Performance Analysis

Problem Statement

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 will focus 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.

Business Impact

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.

Dataset Details

File Name: epl_player_stats_24_25
Description: The data contains 562 players with a rich set of statistical features, offering deep insights into offensive, defensive, passing, goalkeeping, and possession-related metrics.
Dataset Details: 563 Rows and 57 Columns
Source: Kaggle - English Premier League - Player Stats - 24/25

Key Columns Used in the Analysis

The problem statement outlines the importance of analysing the statistics of the EPL players to understand the player and team performance to help fans and professionals determine the best players throughout the 2024-25 season.

The key columns that can help us determine the best players are:

- Club: It shows the club each player is playing for.
- Position: It shows the position of each player.
- Minutes: It shows how many minutes played by each player throughout the season.
- Goals: It shows how many goals were scored by each player.
- Assists: It shows how many assists were given by each player.
- Saves, Punches, High Claims, Goals prevented, Clean Sheets: Those columns can help us determine the best goalkeepers throughout the season.
- Defensive Contributions, Ground Duels Won%, Aerial Duels Won%: Those columns can help us determine the best defenders throughout the season.
- Carries, Progressive Carries, Through Balls, Successful Final Third Passes %, Successful Passes %, and Touches: Those columns can help us determine the best midfielders throughout the season.
- Shots and Big Chances Missed: Those columns can help us determine the best forwards throughout the season.

Analysis List

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

The analysis will be 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'll use 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, as shown in the figure below.

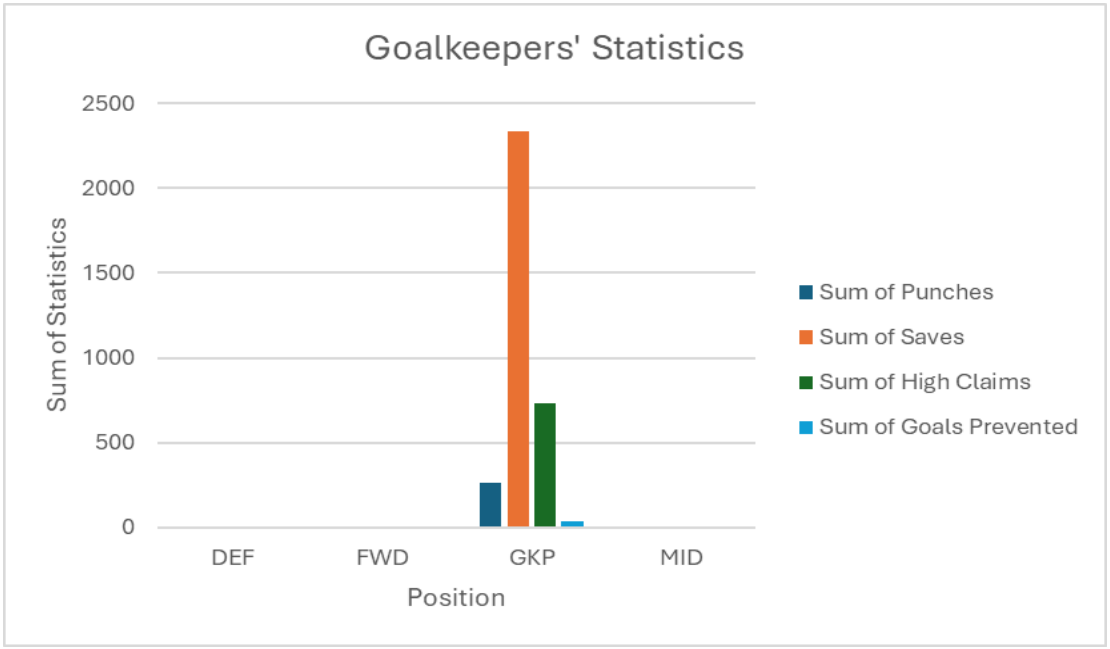

Figure 1. Goalkeepers' Statistics.

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

2. Defenders' performance:

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

We will use 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 will also use the Clean Sheets column to show which defenders helped their team keep a clean sheet during matches. In addition, we will include 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. These comparisons are shown in the figures below.

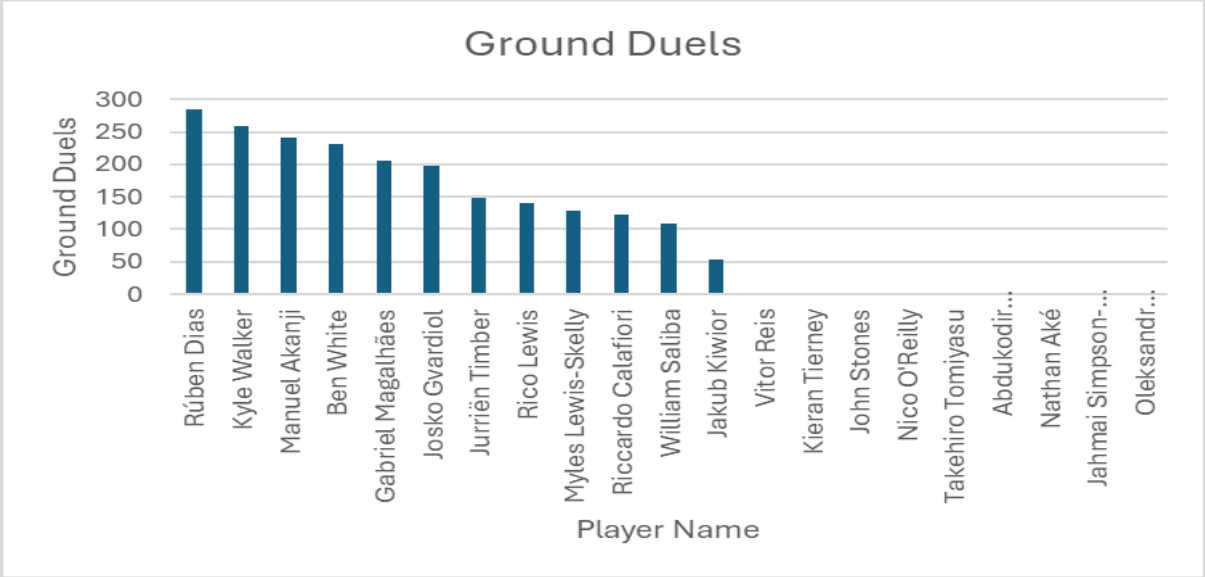

Figure 2. Ground Duels for Each Player in Arsenal and Manchester City.

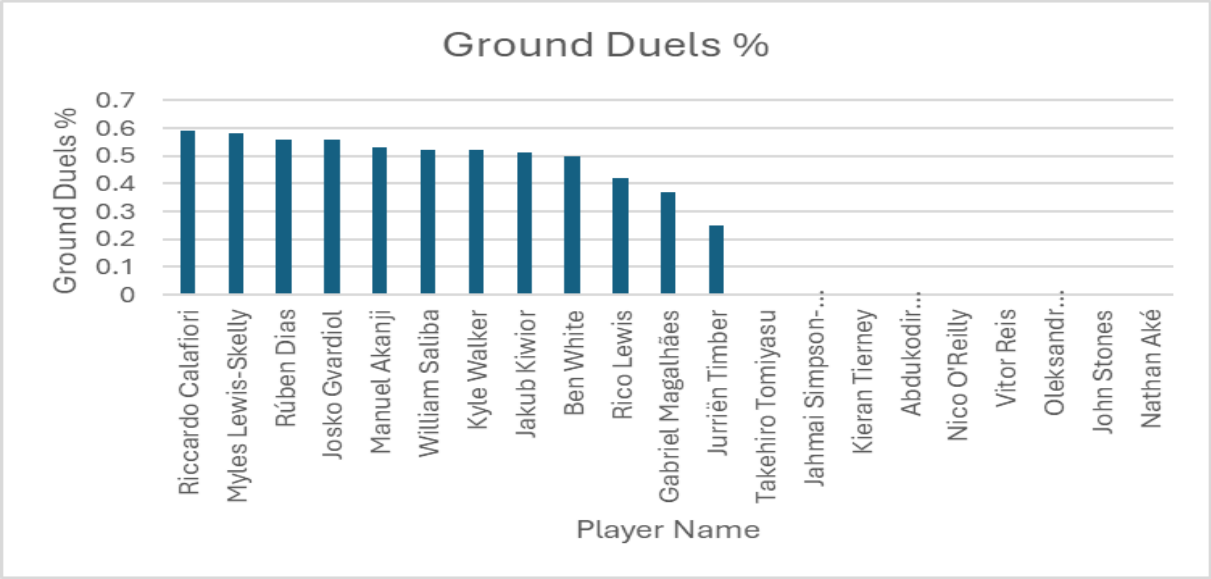

Figure 3. Successful Ground Duels % for Each Player in Arsenal and Manchester City.

We will also consider 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 bar and line charts in the dashboard, helping the audience see which defenders had performed the best that 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, as shown in the figures below.

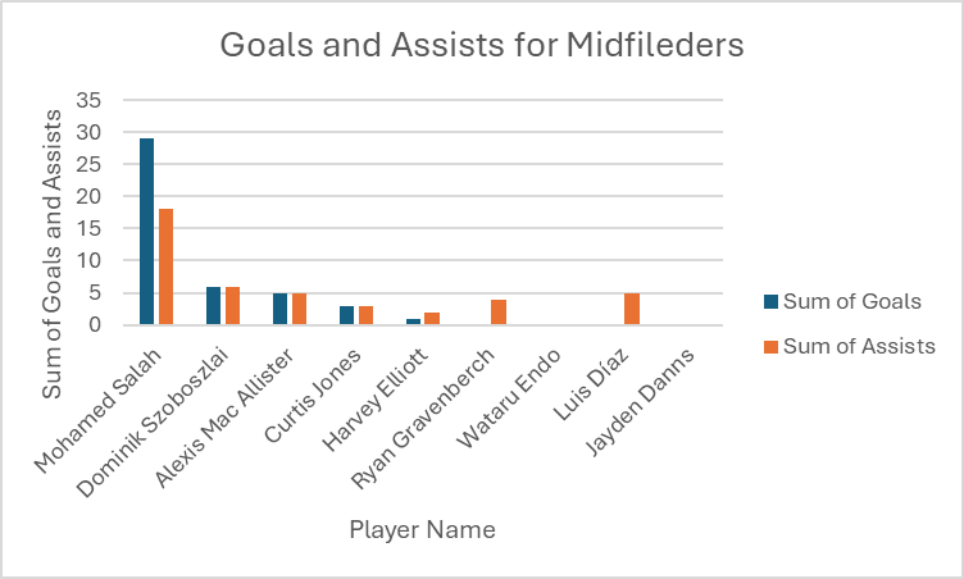

Figure 4. Goals and Assists for Each Midfielder Playing for Liverpool.

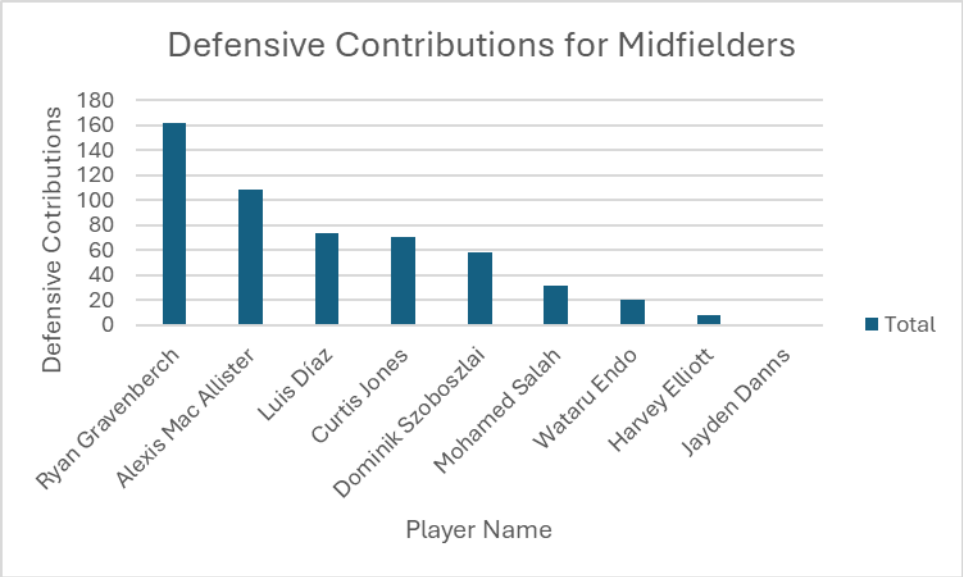

Figure 5. Defensive Contributions for each Midfielder Playing for Liverpool.

These two figures show that the top-performing midfielders in attack are different from those who perform best defensively. However, some midfielders manage to balance both roles effectively. This highlights the importance of analysing both attacking and defensive metrics when evaluating midfielders.

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.

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.

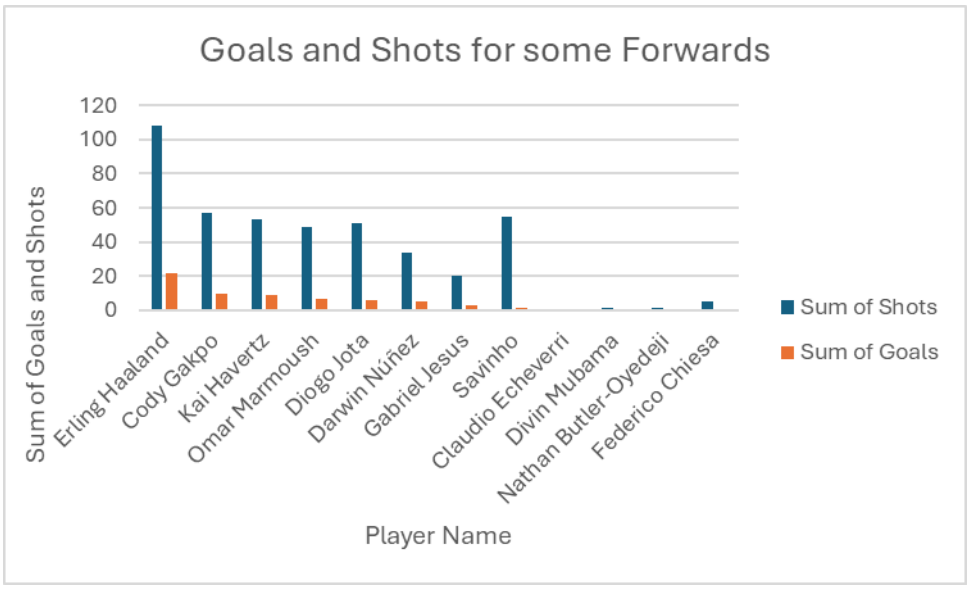

Figure 6. Goals and Shots for Some Forwards from Different Teams.

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. Efficiency will be highlighted by analysing goals per 90 minutes, which helps identify which forwards make the most impact relative to their playing time.

Analysing and comparing these metrics for each forward will allow us to create clear bar and scatter charts in the dashboard, helping the audience determine which forwards were the most dangerous, efficient, and valuable to their teams throughout the 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. Figure 7 will display a comparison between goals for each club, highlighting which teams are clinical finishers.

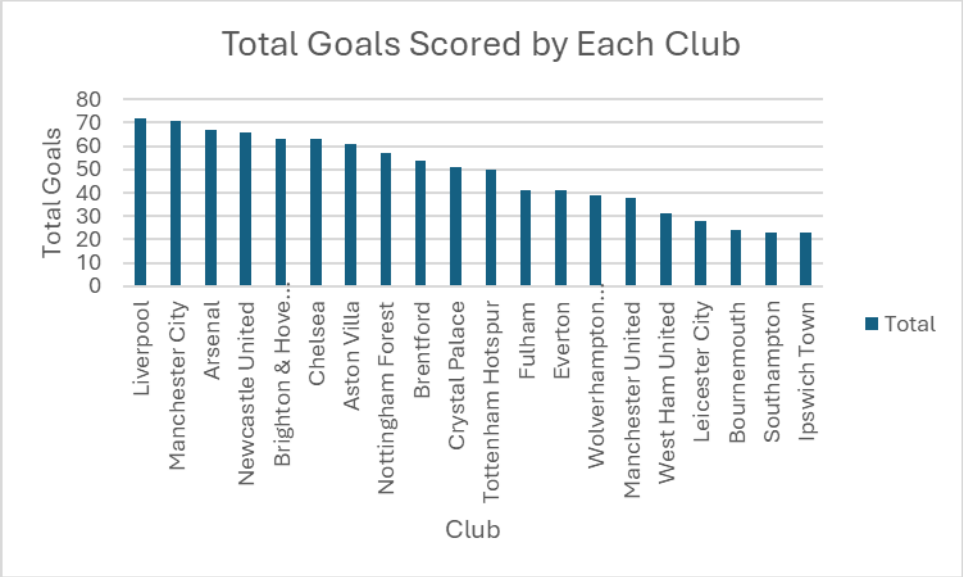

Figure 7. Total Goals Scored by Each Club.

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. Figure 8 will compare goals conceded by each club, highlighting which teams had the strongest and weakest defensive records.

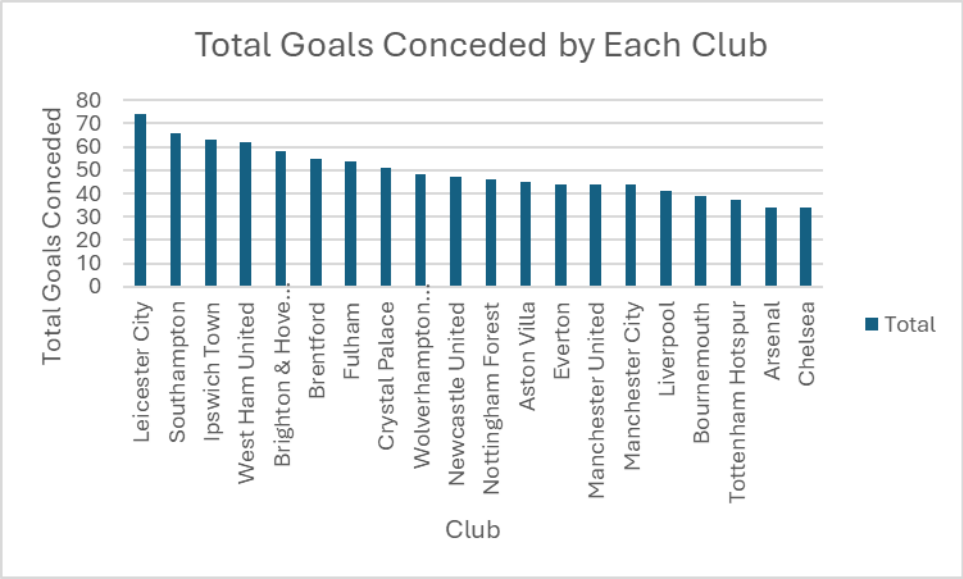

Figure 8. Total Goals Conceded by Each Club.

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.

Conclusion

This exploratory data analysis provided a clear and data-driven understanding of player and team performance in the 2024/25 Premier League season. By analysing goalkeepers, defenders, midfielders, forwards, and overall team statistics, we were able to identify which players had the most impact in their roles and which clubs were the strongest in key areas such as attacking, defending, and possession control.

The results showed that different positions contribute to success in different ways. Goalkeepers were evaluated on their ability to prevent goals, defenders were measured on both defensive stability and attacking support, midfielders were analysed for their creativity and defensive contributions, and forwards were assessed based on their goal-scoring efficiency and involvement in attacking play. At the team level, we could clearly see how attacking output, defensive records, and possession metrics reflected each club's league position and overall consistency.

Overall, this analysis highlights the importance of using performance data rather than opinions to evaluate players and teams. It also provides a strong foundation for selecting the best players in each position and building an ideal formation for the season. These insights can support fans, analysts, and professionals in understanding the key factors that drive success in the Premier League.