ISSN 2959-6157

Application of Multiple Linear Regression Models in Predicting the Value of Football Players

Shaobo Cheng

Boston College High School, Massachusetts, 02169, United states of America
Corresponding author: S.cheng25@students.bchigh.edu

Abstract:

In this study, the key factors affecting football player value were analyzed using a multiple linear regression model. The analysis considered variables such as nationality, location, historical achievements, experience, skill level, and market value. The results indicate that a player's competitive level significantly enhances their value, while there is a negative correlation between market value and player value. The relationship between experience and value is complex and requires further exploration. Additionally, the study examined the distribution of players from various countries in the English Premier League, with French, Spanish, and Brazilian players ranking highest in appearances. French and Spanish players are versatile across positions, while Brazilian players excel primarily as forwards. Visual analyses demonstrated the proportions and performances of players in different positions, confirming the positive correlation between competitive level and net worth and the negative correlation with market value. The intricate relationship between experience and net worth was also highlighted. These findings offer valuable insights into the determinants of football player value and provide data to support team management and future research.

Keywords: Football leagues; multiple linear regression model; value determinants.

1. Introduction

With the development of football professionalism, the value of football players has become an important basis to measure their value and potential contribution in the market. As noted by Dobson and Goddard, club decisions regarding transfers and sponsorships are often based on the assessment of a player's value. However, determining the value of an athlete depends on numerous factors [1]. Boyes highlighted the legal aspects associated with football, emphasizing the need for a comprehensive understanding of these factors by club operators, agents, and investors [2]. A player's performance in goals, assists, and appearances is a determinant of their market value. For instance, as observed by Goldblatt, a player's performance in a match can significantly affect their popularity in the market [3]. Additionally, the age and physical condition of the players are crucial. Generally speaking, as supported by Downward and Dawson, young and healthy players tend to show greater potential and create more value [4].

The playing experience and historical performance of a player are also important factors in measuring their value. Hassan provided a comparative analysis that showcases how players with experiences in top clubs, international games, or outstanding performances in national teams and major competitions can increase their market value [5]. To analyze the determining factors of the value of football players, this article employs multiple linear models. By constructing a multivariate model, this paper aims to accurately predict and analyze the results. This research outcome holds significant theoretical value for comprehending the formation and influencing factors of the value of football players.

In order to analyze the representativeness of players from various countries in the English Premier League, several factors need to be considered. Parrish emphasized the importance of legal and case-specific aspects in football [6]. Geographically, player transfers typically depend on their location. As Spain and France have closer ties with England, players from these countries are more likely to feature in the English Premier League. To quantify this factor, this article takes geographic distance as the independent variable and establishes a regression model, similar to the studies conducted by scholars such as Downward and Dawson.

In terms of player talent, players with higher levels and abilities have a greater chance of transferring to the English Premier League. This article uses players' level scores or international competition results as measurement standards, aligning with the research by Wesson and oth-

ers [7]. By using regression analysis or factor analysis to evaluate the impact of players' talent on their appearance frequency. When it comes to work visas, the difficulty for foreign players to obtain work visas may affect their appearances. When constructing the model, this paper introduces the variable "work visa difficulty", similar to the considerations made by scholars in related fields.

Zhang presented invaluable and nuanced perspectives on specific elements associated with the value of football players [8]. Li conducted in 2021 provided an elaborate economic analysis of player transfers within the framework of professional football clubs [9]. Wang concentrated on formulating efficacious methodologies for evaluating the market value of football players [10]. Liu introduced a novel and insightful analysis concerning the value of players within the international football transfer market [11].

Collectively, their scholarly endeavors have not only broadened the intellectual horizons and scope of comprehension in this specialized area but have also established a robust and well-founded foundation for subsequent research endeavors and practical applications within the realm of football player valuation and the dynamics of the football market.

Furthermore, their works have sparked critical discussions and have paved the way for innovative research directions, fostering a more comprehensive and sophisticated understanding of the intricate web of factors shaping the value and market landscape of football players. These scholars have meticulously dissected and examined various aspects, including but not limited to the economic, social, and performance-related elements that collectively determine the worth and market positioning of football players.

By conducting in-depth empirical studies, data-driven analyses, and theoretical formulations, they have offered rich and diverse insights that contribute to the advancement of knowledge and best practices in this domain. Their research findings have practical implications for football clubs, agents, policymakers, and stakeholders, enabling more informed decision-making and strategic planning in the highly competitive and dynamic world of football.

2. Methods

2.1 Data Source

Player appearance data are obtained from the official website of the Premier League, including the appearance records of players from various teams in the 2018 - 2023 seasons.

Geographic distance data are acquired through the Geographic Information System (GIS) software ArcGIS by querying the straight-line distance between the capitals of various countries and the capital of England, London.

Player skill rating data are referenced from the evaluation reports of the professional scouting agency WhoScored, which conducts comprehensive scoring based on players' performance in matches.

Work visa difficulty data are based on the regulations and policy documents regarding player work visas issued by the UK government's Home Office and the relevant instructions from the English Football Association.

2.2 Method Introduction

In this study, the author employed a multiple linear regression model structured as,

$$P_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots + \beta_k X_{ik} + ?_i \quad (1)$$

Additionally, interaction terms and non-linear relationships were considered. Model evaluation was conducted using the R2 coefficient and adjusted R2 coefficient.

3. Results and Discussion

3.1 Preliminary Work

In any league, it is common to have more domestic players than foreign players, and the English Premier League is no exception. Most of the players are from England. Players from other UK member countries also have a significant presence. So, which countries have the most common players? Firstly, geographical proximity is a factor. Secondly, player talent, work permits, and visa challenges may influence player participation in the Premier League, though these issues are relatively uncommon. Below are the top three countries, excluding the host country (England), ranked by the total number of player appearances and detailed data for different positions:

The most common countries are: for overall (France, Spain, Brazil). For Goalkeeper (Spain, Denmark, France). For Defender (Spain, Netherlands, Ireland). For Midfielder (Scotland, France, Spain). For Forward (Brazil, France, Ireland).

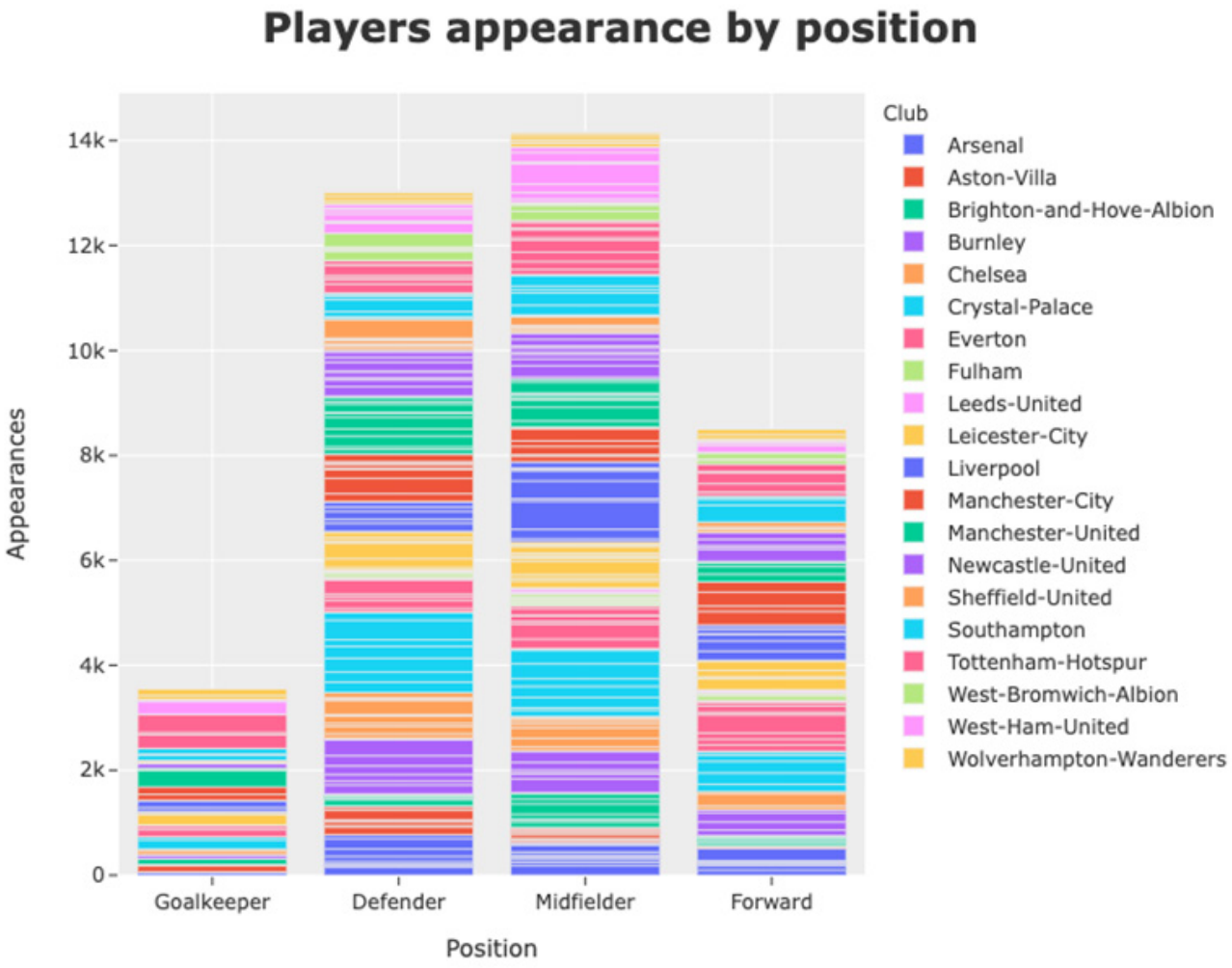

Fig. 1 Players appearance by position

French and Spanish players have found their "second home" in the English Premier League. Brazilian players also deserve a mention. The relative proportions of different countries in this dataset are as follows: France leads with a 6.03% share, followed closely by Spain at 5.73%. Brazil and Ireland rank third and fourth with 4.41% and 4.28%, respectively. Scotland and England together make up a staggering 45.31% of the total, with England alone accounting for 41.2%, indicating England's absolute dominance in this dataset.

The proportions of the Netherlands, Argentina, and Wales are relatively low, at 2.81%, 2.74%, and 2.45%, respectively. Although these countries have smaller shares, their presence still adds diversity to the dataset.

This paper can briefly analyze the performance of players in different positions across various football clubs during matches. This horizontal bar chart, color-coded for clarity, illustrates the performance scores of each team in different positions (such as goalkeeper, center-back, midfielder, etc.). Although specific numerical values cannot be directly read from the chart, the differences in the height of the color blocks provide a visual comparison.

From Figure 1, it is evident that clubs perform differently across various positions. For example, Manchester United (represented in red) has a higher performance score in the center-back position, indicating that their defensive system is relatively strong, and the center-backs play a crucial role in their matches. However, Manchester United's performance score is relatively low in the goalkeeper position, which might be related to the goalkeeper's individual performance, the attacking strength of the opponents, or the team's defensive strategy.

In contrast, Chelsea (represented in blue) may have better performance in other positions, such as midfield or forward, which can be inferred from the height of the color blocks in the corresponding positions. However, as the chart does not show the performance of all clubs in all positions, this paper cannot make a full comparison.

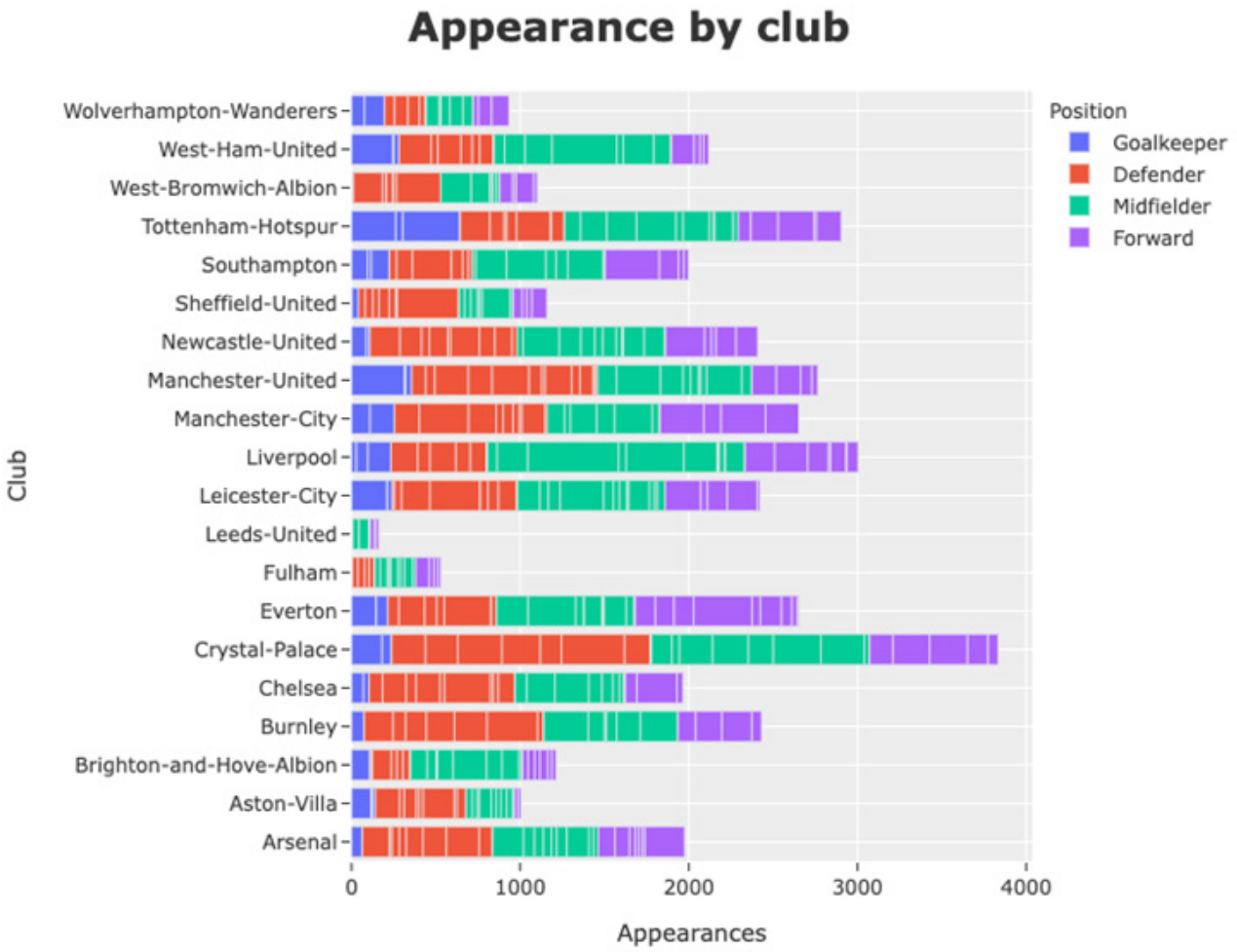

Fig. 2 Appearance by club

From Figure 2, it can be seen that there are significant differences in the performance of each club at different positions. For example, “Manchester-United” and “Liverpool” have larger color blocks at the back and midfield positions, indicating that these clubs have played or scored more games at the back and center. Conversely, some clubs may focus more on the attacking ability of the forward line, such as Tottenham-Hotspur and Arsenal, which have a prominent color block in the forward position.

Although some clubs have an overall balance of strength, it is still possible to perform well or poorly in certain positions. It reflects the different strategies of different clubs and the different distribution of players, and also shows the importance of the division of player positions in football matches.

However, some clubs such as Brighton and Holbein and Aston Villa have performed relatively evenly in different positions. It means that the team building of these clubs places more emphasis on the overall performance of players, rather than solely focusing on outstanding performance on one side.

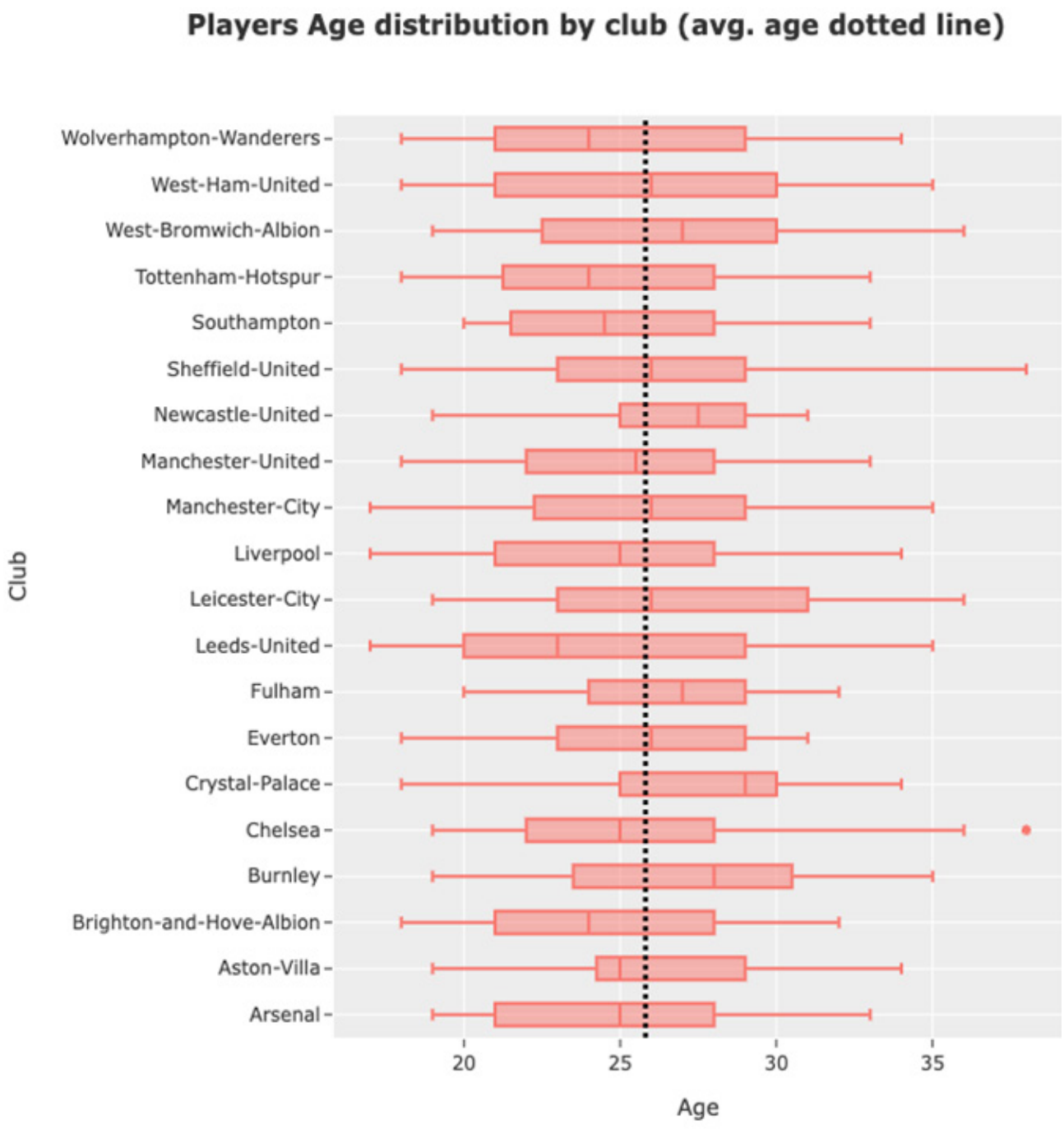

Fig. 3 Age distribution by club (avg. age dotted line)

Figure 3 clearly shows the average age distribution of players at different football clubs, showing the age difference between clubs. The characteristics of each club can be seen from the average age of their players in the chart. For example, clubs such as Wolverhampton Wanderers and West Ham United have a relatively high average age of players, which may mean that these clubs have more experienced veteran players who, as the main force of the team, may perform more steadily and excel on the field. On the contrary, clubs such as Brighton and Howe Albion and Aston Villa have relatively lower average ages of players, indicating that these clubs may place more emphasis on young players and enhance the vitality and potential of their teams through youthful team building.

In addition, there is a dashed line in the figure indicating the average age, which is avg age dotted line, used to compare the difference between the average age of players in each club and the overall average level. Through

this dashed line, this paper can more intuitively see which clubs have players with an average age higher or lower than the overall level, thus analyzing the impact of these differences on the team's game style and strategy.

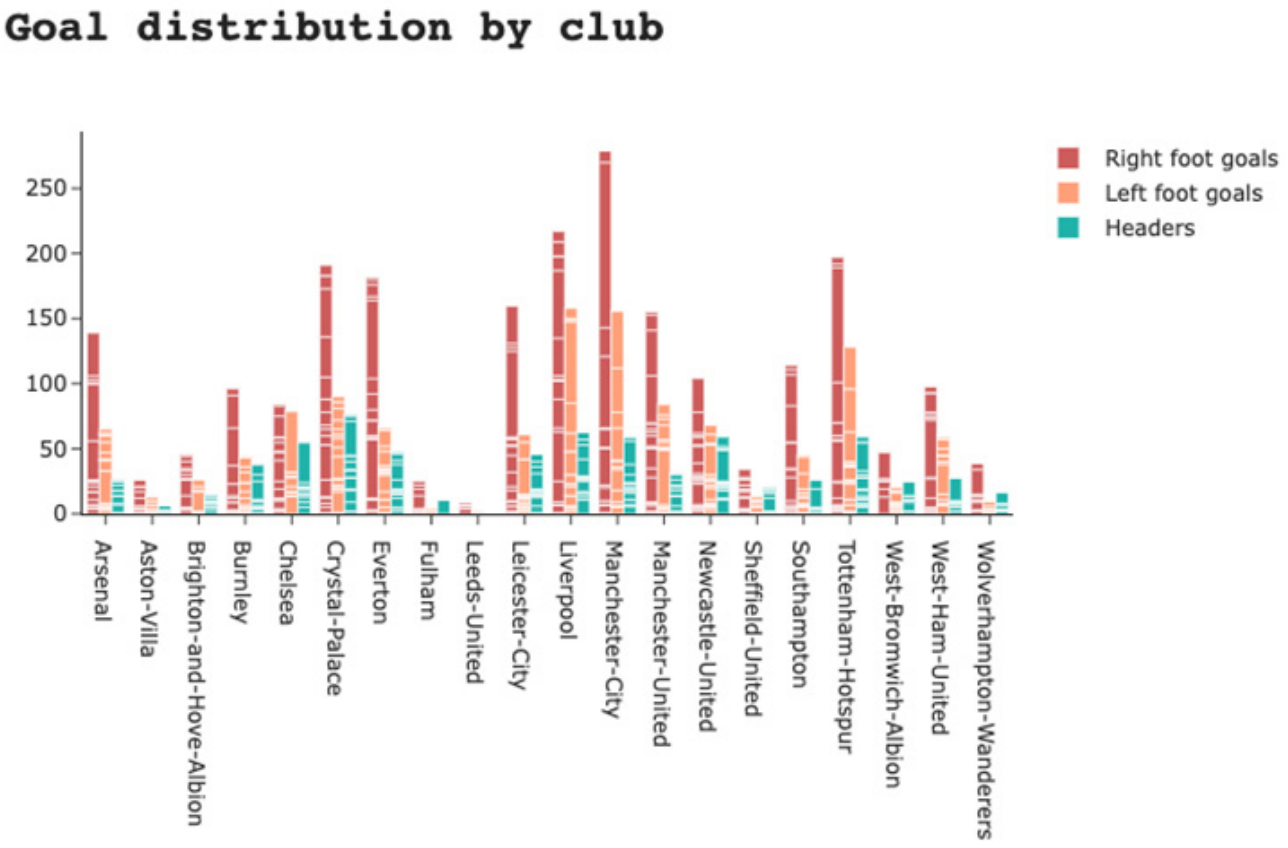

Fig. 4 Goal distribution by club

Figure 4 shows the goal scores of different football clubs in three aspects: right foot shot, left foot shot, and header. There are differences in the performance of each club in different shooting methods, which reflects the skills and styles of each team's players.

In terms of right foot shooting, some clubs such as Arsenal, Chelsea, etc. have shown strong abilities, which may be related to the fact that these clubs have players who are good at right foot shooting and skilled in technique. In contrast, other clubs such as Brighton and Hove, Albion, Burnley, etc. have performed relatively average in this area, indicating that their players have weaker abilities in right footed shooting.

Regarding left-foot shots, although the overall number of goals scored is relatively small, some clubs such as Leicester City and Liverpool still perform well, which reflects that left foot shooting is equally important in football matches and can widen the gap with other teams. However, some clubs such as Crystal Palace and Fulham have lower goal counts in this area and need to strengthen training in left footed shooting techniques.

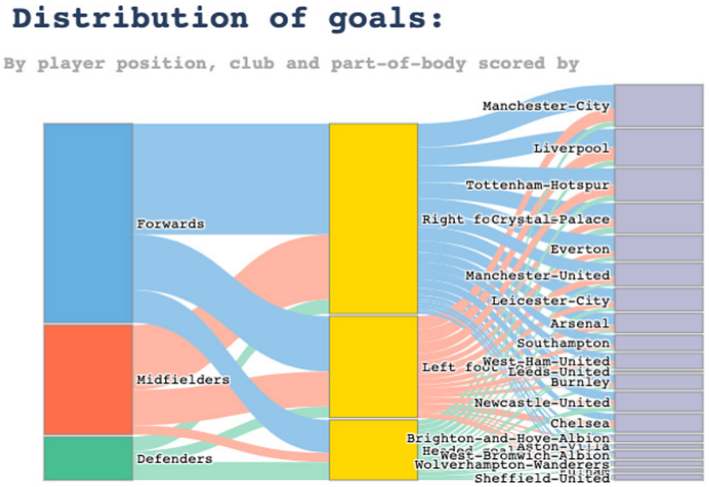

Fig. 5 Distribution of goals

From the "distribution" of the icon, different colored areas display the distribution of different football clubs (Figure 5). Each club is clearly divided into specific regions, and the size of each region may represent the relative importance or influence of that club in statistics. This distribution result may have multiple factors, such as the club's historical record, player strength, player popularity, etc. It can be seen from the figure that some clubs, such as "Manchester City" and "Liverpool", occupy a more prominent position, which may mean that these clubs have a more prominent performance in the statistical cycle.

In the "Player Position, Club, and Body Part Scores" section, a complex bar chart displays the performance of different clubs on different player position and body part scores. Blue, yellow, and red make the data more intuitive and easier to understand. It can be inferred from the graph that different colors may represent different statistical dimensions, such as goals scored, assists provided, shooting success rate, etc. The height of each bar represents the scoring situation of the club in that dimension.

3.2 Model Results

In terms of player positions, this paper can observe the differences in scores between different clubs at different positions. For example, some clubs score higher in midfield positions, indicating that they have good midfield control and creativity. Some clubs score more on the forward line, demonstrating their strong attacking ability. This difference not only reflects the strategic arrangements and player configurations of each club, but also provides reliable clues for us to study the strengths and weaknesses of each club.

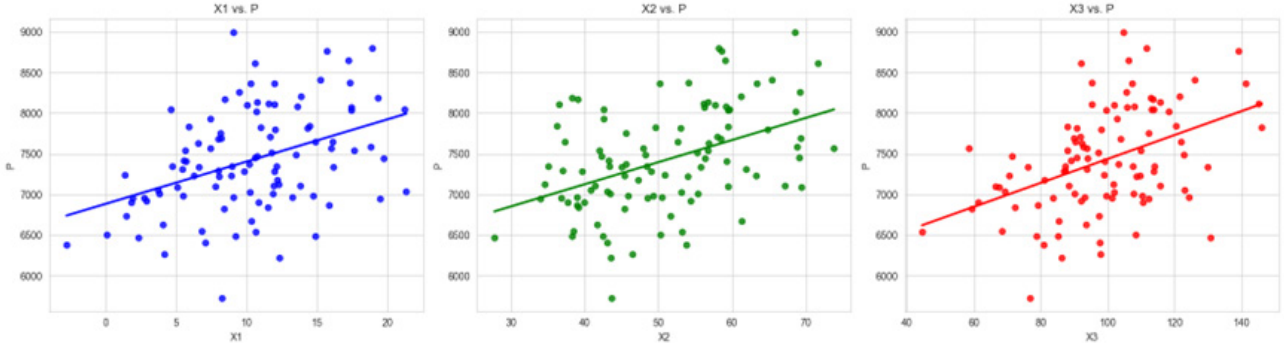

Fig. 6 Factors X1, X2, X3 Distribution

X1 (Influencing Factor 1): This variable is generated from a normal distribution with a mean of 10 and a standard deviation of 5. It represents factors such as the athlete's professional experience, the number of training sessions, or personal achievements, which are related to their performance in a specific field. These factors affect the athlete's value based on their expertise and accumulated experience.

X2 (Influencing Factor 2): Generated from a normal distribution with a mean of 50 and a standard deviation of 10, X2 represents characteristics such as the athlete's match performance statistics, athletic ranking, or number of awards won. These elements are directly related to the athlete's overall skill level and achievements in competi-

tions, influencing their market value.

X3 (Influencing Factor 3): This variable comes from a normal distribution with a mean of 100 and a standard deviation of 20. X3 could represent market-related characteristics, including the athlete's promotional value, media exposure, or total sponsorship contracts. These factors are crucial in determining the player's market value and popularity.

Each factor (X1, X2, and X3) contributes differently to the player's final value (P) through a linear relationship. The weights (50, 20, and 10) signify the respective contributions of these factors to the player's overall market value (Figure 6).

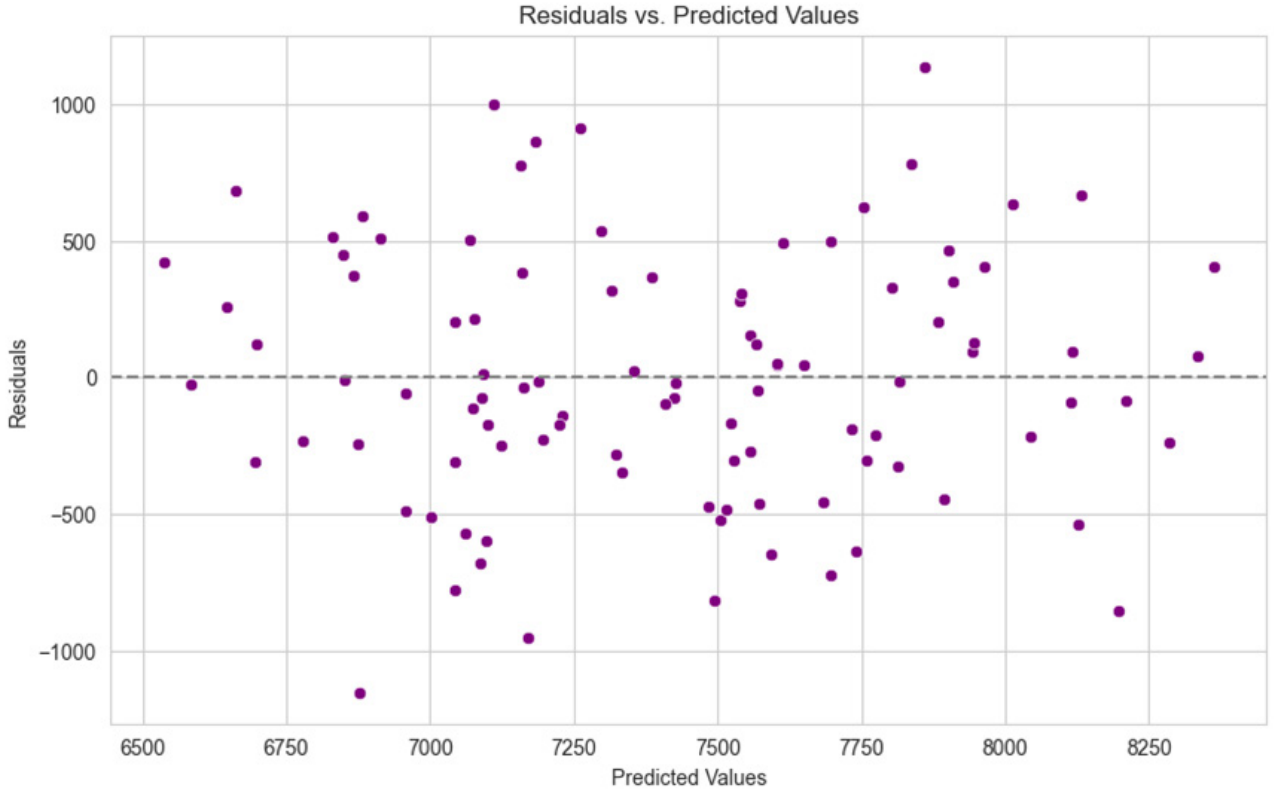

Fig. 7 Residuals vs. Predicted Values

Three scatter plots intuitively display the relationships between X1, X2, X3, and P. In each scatter plot, the blue, red, and green points represent different data sets or categories. These points are connected by lines, forming continuous line graphs, making it easier to observe the trends in the variables.

In the scatter plot of X1 vs. P, this paper can see that as the value of P increases, the value of X1 shows some fluctuations, but the overall trend may not be clear. This suggests that there might not be a strong linear relationship between X1 and P, or that other unconsidered factors might be influencing the relationship.

In the scatter plot of X2 vs. P, the value of X2 seems to show an upward trend as P increases, especially in the higher ranges of P. This indicates that there might be a positive correlation between X2 and P, meaning that an increase in P could lead to an increase in X2.

As can be seen from the scatter plot, X3 has a significant negative correlation with the value of P. As P increases, the value of X3 gradually decreases. It shows that X3 and P are inversely correlated.

Moreover, the figure 7 also gives the specific values of X1, X2, X3 and P in different intervals, such as 9000, 8500, as well as the scale range of P values, such as 15, 30, 40. These values provide a foundation for more accurate analysis, helping to calculate statistical measures such as the mean and standard deviation of X1, X2, and X3 within different intervals. Therefore, this paper can further explore their relationship with P.

In summary, through these three scatter plots, this paper can preliminarily determine the relationship between X1, X2, X3 and P. Among them, X2 is positively correlated with P, X3 is negatively correlated with P, and the relationship between X1 and P is relatively complex and requires further confirmation. These results play an important role in understanding the interactions between variables and

subsequent modeling and prediction.

4. Conclusion

The performance of different football clubs varies in many ways. First, through the color-coded horizontal bar chart, this paper can visually see how different clubs perform in different player positions. Manchester United, for example, excel at center-back but are weaker at goalkeeper, while Chelsea may excel at other positions.

Secondly, the average age bar chart of players shows the diversity of the club's age structure. Some clubs, such as Wolves and West Ham, have more experienced players, while Brighton and Villa focus on younger players. Finally, through the analysis of the scoring ability of different shooting methods, this paper found that the performance of different clubs in the right foot, left foot and header attack is different, reflecting the characteristics and styles of each team. Based on these analyses, this paper can see the unique characteristics of each club in terms of player position, age structure, and offensive ability, which not only affect the tactical strategy of the team, but also provide important perspective for us to understand team performance.

References

[1] Boyes S. Football and the Law. Routledge, 2021.

[2] Goldblatt D. The Globalization of Football: A Study in the Development of Modern Football. Routledge, 2021.

[3] Downward P, Dawson A. The Economics of Football Transfers. Routledge, 2019.

[4] Hassan D. Soccer and Society: A Comparative Analysis of Football Leagues. Routledge, 2019.

[5] Wesson J. The Science of Soccer: A Bouncing Ball and a Banana Kick. Springer, 2018.

[6] Parrish R. Football and the Law: A Case Study. Routledge, 2017.

[7] Zhang Weiming. Study on Factors Influencing the Market Value of Football Players. Sports Science, 2022, 42(3): 50-58.

[8] Li Donghua. Economic Analysis of Player Transfers in Professional Football Clubs. Sports Economics, 2021, 16(2): 78-85.

[9] Wang Jianguo. Research on Methods for Evaluating the Market Value of Football Players. Sports Research and Practice, 2020, 40(4): 102-110.

[10] Liu Zhiqiang. Analysis of Player Value in the International Football Transfer Market. International Sports Economics, 2023, 12(1): 34-42.