

Sport Analytics Data for Player Performance and Financial Risk Management

Gorngris Kositsakul

Oundle School, Oundle, Peterborough, UK

Email: gorn.kositsakul@gmail.com

How to cite this paper: Kositsakul, G. (2025). Sport Analytics Data for Player Performance and Financial Risk Management. *Journal of Financial Risk Management*, 14, 270-284.
<https://doi.org/10.4236/jfrm.2025.143015>

Received: July 30, 2025

Accepted: August 29, 2025

Published: September 1, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This study investigates the relationship between various player characteristics and the performance metrics in professional football using sport data analytics. Utilising a data set of over 10,000 observations, the analysis employs regression models to identify the significant determinants of player performance and value. Results highlight that most factors are positively correlated and that subjective opinions on metrics which normally influence player performance and values are supported by objective evidence. These insights demonstrate the positive impacts of utilizing sport data analytics in enhancing decision-making in football. The findings also provide insights into how objective performance data can support better financial decisions in football.

Keywords

Player Performance, Player Valuation, Data Analytics, Regression

1. Introduction

Data analytics is becoming increasingly important in the sports industry. Brentford and Brighton are two outstanding examples of football teams which have indicated that the data-driven research plays a key role in team success.

This paper will investigate the connections between variables within the game of football through the use of data analysis and modelling. Some questions that could be answered with the data are: how much does age affect a player's goal-scoring ability in the real world? Does the number of yellow cards received correlate with the number of clean sheets achieved by a team?

It is important to understand these relationships because football clubs today have to face a lot of competition and high costs. Data analytics can help teams make better decisions on the pitch and make smarter investments off the pitch by

finding every small advantage. Also, using sport data analytics gives us an objective way to look at things that people typically argue over in the game.

By applying rigorous quantitative methods, this study demonstrates how traditional assumptions about player performance and value can be tested and supported with empirical evidence. This analytical approach enables football clubs to make more informed decisions regarding player recruitment, development, and tactical planning. As data-driven practices continue to evolve, the capacity to derive meaningful insights from large and complex datasets will remain a crucial element in enhancing decision-making processes and sustaining a competitive advantage in sports such as professional football.

The purpose of this paper is to analyse how different player characteristics affect their performance metrics in professional football. By using data analytics and regression analysis, this study provides insights into which factors strongly influence player performance and valuation. The research confirms that there are positive links between the key sport analytics indicators (Goals, Clean sheets, Age, Position, Yellow cards, Second yellow cards, and Days injured) and performance and market value (Current and Highest player value, and Minutes played).

The paper makes several contributions to the literature. Utilising one of the largest datasets, the study examines multiple performance indicators based on over 10,000 player observations. This paper integrates a wide range of characteristics and combines multiple methodologies to provide more robust evidence.

The practical implications of this study are relevant to a wide range of stakeholders within the football industry. Club managers and scouts can apply the findings to improve recruitment processes by identifying players whose characteristics are shown to have a measurable impact on performance and market value. These insights can help coaches and team analysts tailor training plans that focus on the factors that are most likely to affect performance. Players and agents can also benefit from knowing which performance indicators have the biggest impact on a player's value, allowing them to focus on areas that will improve their career prospects and marketability. These data-driven insights may assist investors and club executives in making smart financial choices, making sure that transfer fees and salaries are in line with expected performance returns. For example, while negotiating player contracts, decision-makers might use the statistical evidence to see if the demands are reasonable and fair. Overall, this study supports a more evidence-based approach to running football teams by using mathematical models to back up assumptions. This helps clubs and their stakeholders make better choices in order to stay ahead of their competition in a sport that is becoming more data-driven.

The findings from this paper offer valuable insights into the use of objective performance metrics in managing financial decisions in the sport industry, specifically football.

This paper proceeds as follows: Section 2 provides background information on the existing literature, Section 3 highlights the empirical findings, and Section 4 concludes the paper.

2. Literature Review

In literature, many papers explore how different factors can affect player performance. Recent studies suggest that sports analytics has gotten more advanced by using both traditional statistical models and new AI methods to acquire a better picture of how players and teams are doing. Regression models remain widely used to assess how specific actions like passing, shooting, or fitness affect success (Frick, 2011; Cornforth et al., 2015; Jana & Hemalatha, 2021). For example, Frick (2011) discovered that recent player performance can affect salaries more than talent alone, while Cornforth et al. (2015) showed that regression can predict match results through fitness analysis.

Machine learning and AI have expanded these capabilities. Studies have demonstrated how algorithms can classify player positions (García-Aliaga et al., 2021), track health outcomes (Yang et al., 2024), and optimize tactics and talent identification (Wisdom & Javed, 2023; Aliyarov et al., 2023). Chang et al. (2024) found that frequent match participation benefits players with strong offensive skills. Baboota and Kaur (2019) and Aswin et al. (2024) highlighted machine learning's potential to reduce bias and improve predictions, though Baboota and Kaur note that more data is still needed.

Not all studies agree that AI alone is the answer. Berrar et al. (2019) argued that domain knowledge and creative feature engineering often matter more than the choice of algorithm. Min et al. (2008) supported this by showing that hybrid systems can predict performance even with smaller datasets. Maszczyk et al. (2014) compared AI and traditional models, concluding that AI can outperform regression for sports predictions.

Existing research also highlights a wide range of variables used to measure player performance, team success, and athlete health. Examples include years of exercise (Yang et al., 2024), shot selection such as long shots (Jana & Hemalatha, 2021), spatial metrics like the average number of players near the ball (Min et al., 2008), passing accuracy for long passes (Thunberg Kalt, 2024), and goals scored (Chang et al., 2024).

Different regions apply analytics in unique ways. Chu & Wang (2019) showed that Major League Baseball teams using analytics more heavily had greater playoff success, though unpredictable factors limited forecasts. Toma & Campobasso (2023) highlighted how analytics can also reveal economic trends, such as income inequality in European football. Mehta et al. (2024) found that coaches and analysts value different types of data, revealing that subjective preferences still shape how analytics are used in practice.

Other studies have introduced new methods and variables. Auer & Hiller (2015) used game-theoretic measures to complement traditional stats, while Thunberg Kalt (2024) used cluster analysis and detailed passing data to study team tactics. Ramirez et al. (2017) combined regression and correlation to show the impact of defenders and midfielders on match outcomes. Sarlis & Tjortjis (2020) showed how basketball teams can use analytics to optimize rotations and predict top per-

formers. tables

3. Empirical Evidence

3.1. Data and Methodology

Table 1 in the appendix presents various sports analytics data sourced from kaggle.com. It contains detailed information on over 10,000 professional football players from various European leagues and seasons. The data include both biographical variables (e.g., height, age) and performance metrics (e.g., goals per 90 minutes, assists, yellow cards, minutes played, injuries, player value).

Before analysis, the dataset was cleaned to ensure accuracy and consistency. Observations with missing values in key variables were removed. Outliers, such as extreme values in player value or minutes played, were examined and excluded if they resulted from data entry errors. All variables were carefully labeled, and some were encoded to enable regression analysis.

Table 1. Descriptive statistics.

Variables	Obs	Mean	Std. Dev.	Min	Max
height	10,754	181.24	6.97	156	206
age	10,754	26.042	4.778	15	43
appearance	10,754	36.407	26.527	0	107
goals	10,754	0.126	0.236	0	11.25
assists	10,754	0.087	0.143	0	4
yellowcards	10,754	0.19	0.432	0	30
secondyellowcards	10,754	0.005	0.025	0	1
redcards	10,754	0.007	0.081	0	6.923
goalsconceded	10,754	0.132	0.442	0	9
cleansheets	10,754	0.045	0.924	0	90
minutesplayed	10,754	2470.789	2021.703	0	9510
days injured	10,754	117.962	175.207	0	2349
games injured	10,754	15.826	23.384	0	339
award	10,754	1.961	3.744	0	92
current value	10,754	3,620,000	9,100,000	0	1.800e+08
highest value	10,754	6,150,000	13,400,000	0	2.000e+08
position encoded	10,754	2.713	0.986	1	4
winger	10,754	0.308	0.461	0	1

The sport analytics variables include the height of a player, the age of a player, number of games played by a player, goals scored by a player per 90 minutes, assists by a player per 90 minutes, number of times a player's been given a yellow card per 90 minutes, number of times a player's been given 2 yellow cards per 90 minutes, goals conceded only by a goalkeeper per 90 minutes, number of times a

goalkeeper doesn't concede per 90 minutes, number of minutes played by a player, number of days injured in the player's career, number of games in which a player didn't play due to injury, number of awards won by a player, current value of a player according to transfermarkt in Euros, highest valued price of a player in Euros, position of the player (Goalkeeper, 1, Defender, 2, midfielder, 3, Attack, 4), dummy variable 1 for winger and 0 for not a winger.

3.2. Descriptive Statistics

Table 1 provides descriptive statistics for various football analytic variables used in the analysis. The height has 10,754 observations with a mean of 181.24 cm and a standard deviation of 6.97, ranging from a minimum of 156 cm to a maximum of 206.

The age has 10,754 observations with a mean of 26.042 years and a standard deviation of 4.778, ranging from a minimum of 15 years old to a maximum of 43 years old.

The appearance has 10,754 observations with a mean of 36.407 games and a standard deviation of 26.527, ranging from a minimum of 0 to a maximum of 107.

The goals have 10,754 observations with a mean of 0.16 goals per 90 minutes and a standard deviation of 0.236, ranging from a minimum of 0 goals to a maximum of 11.25.

The assists have 10,754 observations with a mean of 0.087 per 90 minutes and a standard deviation of 0.143, ranging from a minimum of 0 to a maximum of 4.

The yellowcards have 10,754 observations with a mean of 0.19 per match and a standard deviation of 0.432, ranging from a minimum of 0 to a maximum of 30.

3.3. Findings

The scatter plot **Figure 1** shows an interesting connection between the number of goals scored and the number of minutes played by each footballer. It is clear that the number of goals scored positively correlates with the number of minutes played, suggesting that the higher number of minutes contributes to more goals.

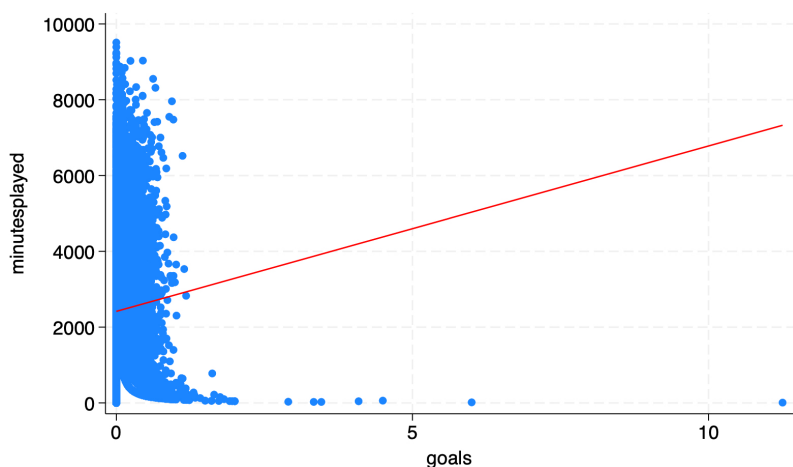


Figure 1. Scatter plot graphics of minutes played vs goals scored per 90 minutes.

The scatter plot **Figure 2** shows an insightful connection between the number of goals scored and the value of each footballer. It is clear that the number of goals scored positively correlates with the highest market value of players, suggesting that players who score a lot of goals are valued very highly.

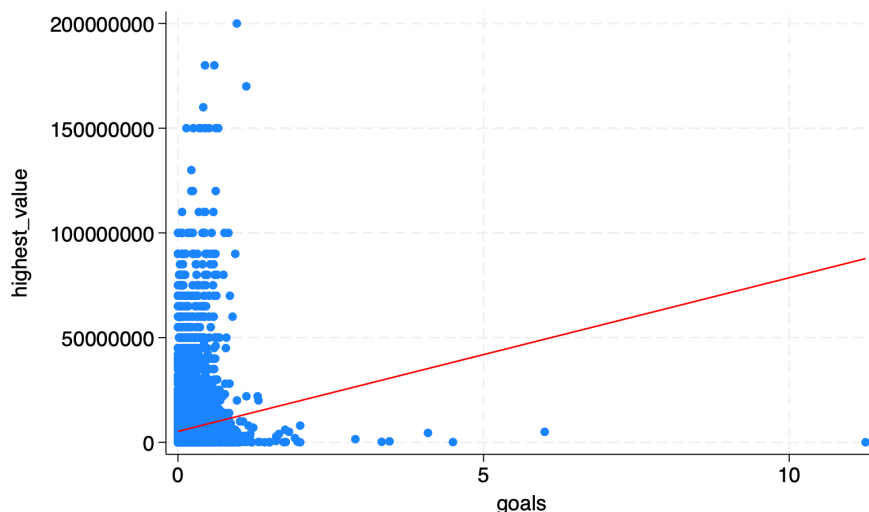


Figure 2. Scatter plot graphics of the highest value of a player vs goals scored per 90 minutes.

The scatter plot **Figure 3** shows the link between the number of goals scored and the current value of each footballer. It is clear that the number of goals scored positively correlates with a player's current value, suggesting that the higher number of goals scored is associated with a higher current value.

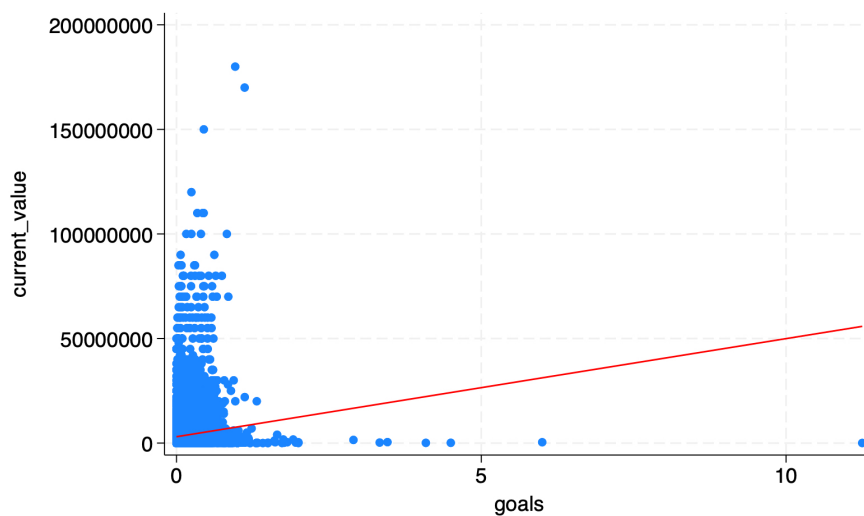


Figure 3. Scatter plot graphics of the current value of a player vs goals scored per 90 minutes.

The scatter plot **Figure 4** shows a connection between the number of games a player is injured for and the number of minutes they have played. It is clear that

the number of games where the player is injured correlates positively and strongly with the number of minutes played, suggesting that the higher number of minutes played contributes to more injuries.

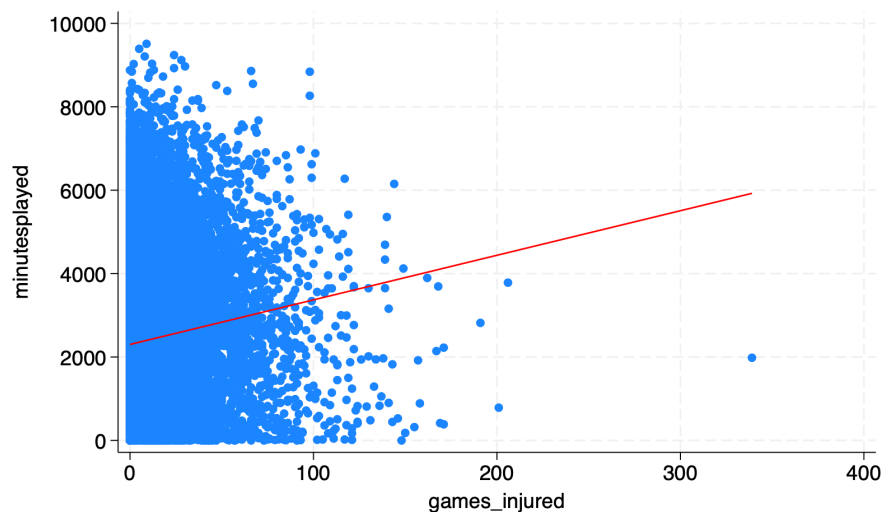


Figure 4. Scatter plot graphics of the number of minutes played vs games injured.

The scatter plot **Figure 5** shows a link between the number of goals scored and the position played by each footballer. There is a strong positive correlation between the number of goals and where the player plays, illustrating that the higher the player is up the pitch *i.e.* striker, the more goals they will score.

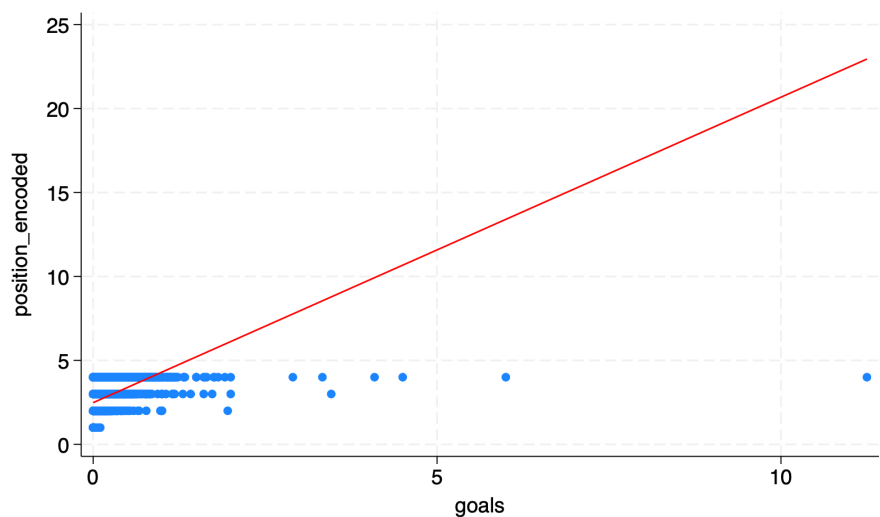


Figure 5. Scatter plot graphics of the position of the player (1 being the keeper and 4 the attacker) vs goals scored.

The scatter plot **Figure 6** displays an association between the number of days a player is injured for and the number of minutes played by each footballer. It is clear that the number of days injured positively correlates with the number of minutes played, suggesting that more minutes lead to longer injury time.

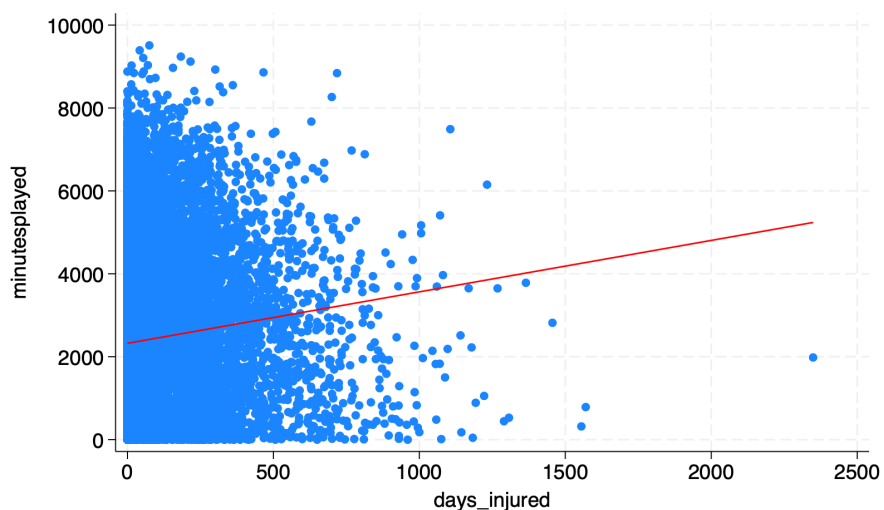


Figure 6. Scatter plot graphics of the position of the minutes played vs days injured.

The scatter plot **Figure 7** shows an interesting relationship between the number of yellow cards received and the number of clean sheets achieved by each goal-keeper. There is a slightly negative correlation between number of yellow cards and clean sheets, suggesting that the more yellow cards the less clean sheets, or in other words more games where a goal is conceded.

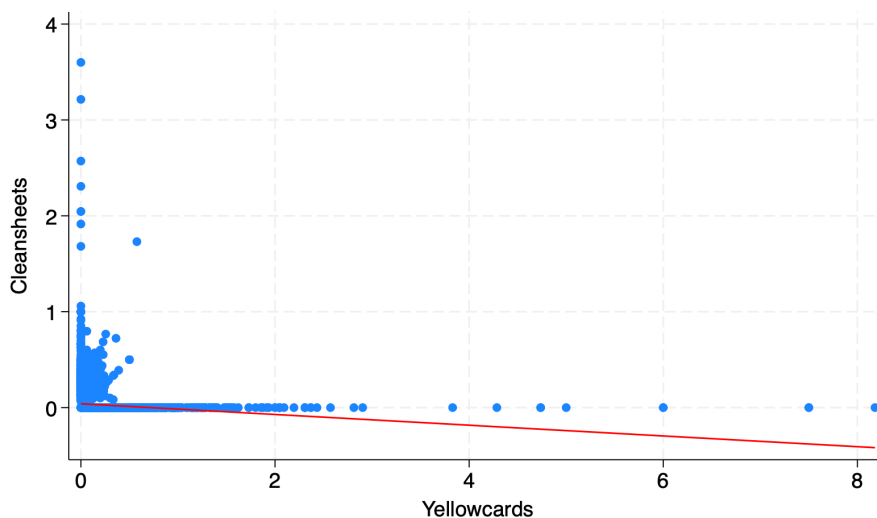


Figure 7. Scatter plot graphics of the number of clean sheets vs number of yellow cards.

The scatter plot **Figure 8** shows an interesting connection between the number of clean sheets and the age of a football player. The diagram shows that the number of clean sheets achieved by a goalkeeper positively correlates with their ages. This suggests that the older the player, the more clean sheets they have.

The scatter plot **Figure 9** presents a relationship between the number of yellow cards and second yellow cards received by each football player. There is a weakly positive correlation between receiving the yellow card once and again for the sec-

ond time in a match. This illustrates that one yellow card has a high chance of leading to a second.

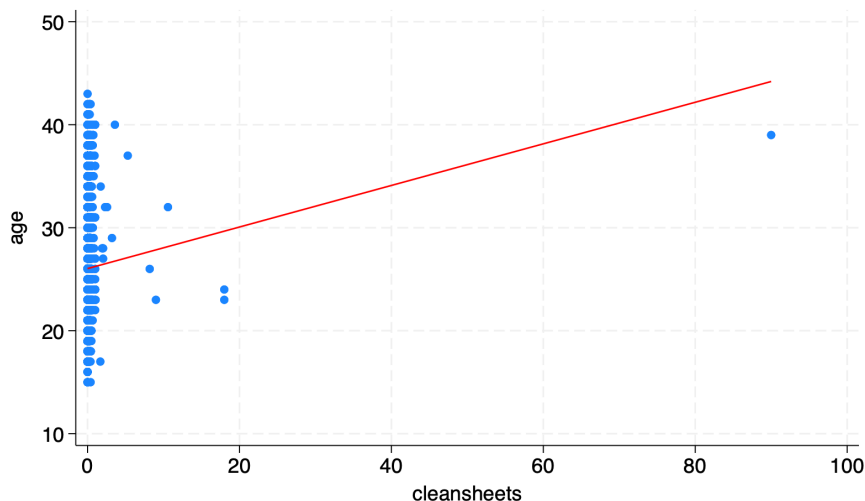


Figure 8. Scatter plot graphics of the number of the player’s age vs number of clean sheets.

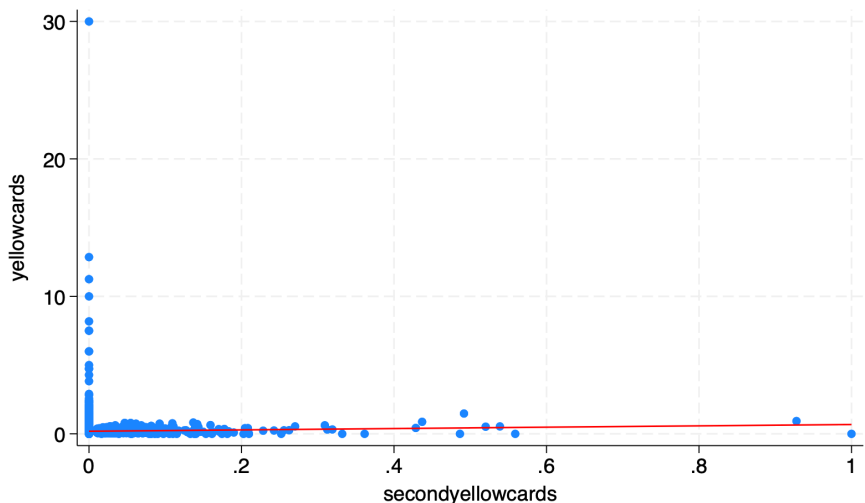


Figure 9. Scatter plot graphics of the number of yellow cards vs the number of second yellow cards.

Table 2 examines the relationship between the dependent variable, minutes played, and several independent variables, including goals, yellow cards, assists, games injured, and age. Each column represents a different regression model, incorporating various combinations of these independent variables to understand the impact of independent variables on the dependent variables. The p values indicate the statistical significance of coefficients where $p < 0.01$, $p < 0.05$, and $p < 0.1$ (indicating ***, **, and *, respectively). The positive coefficient of the “goals” variable suggests that scoring more goals is associated with more minutes played. The coefficient of goals is significant in all models. For “yellow cards”, the negative and significant coefficient indicates that receiving more yellow cards re-

duces the minutes played due to the disciplinary actions or performance issues. For the “assists” variable, positive coefficients suggest that players with more assists tend to play significantly more minutes because they are the key to more goals being scored. For the “games injured” variable, positive but small coefficients imply a minor increase in minutes played with more games injured. For the “age” variable, a positive coefficient in model 5 suggests that all the players may play more minutes, perhaps due to more experience or being cheaper to acquire to play for a team.

Table 2. Regression analysis for minutes played.

	(1)	(2)	(3)	(4)	(5)
VARIABLES	minutes played	minutes played	minutes played	minutes played	minutes played
goals	436.5** (170.8)	432.2** (170.6)	294.6** (150.1)	261.1* (142.5)	315.1** (151.1)
yellow cards		-182.4*** (50.7)	-181.8*** (51.3)	-189.4*** (52.0)	-190.7*** (52.0)
assists			1036.2*** (252.9)	964.8*** (242.9)	1032.2*** (248.1)
games_injured				10.3*** (0.9)	5.9*** (0.9)
age					63.1*** (4.3)
Constant	2416.0*** (28.5)	2451.1*** (30.4)	2378.2*** (31.3)	2226.4*** (32.4)	642.1*** (110.3)
Observations	10,754	10,754	10,754	10,754	10,754
R-squared	0.0	0.0	0.0	0.0	0.0

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3 examines the relationship between the dependent variable, highest_value, and several independent variables, including goals, awards, position_encoded, assists, and age. Each column represents a different regression model, incorporating various combinations of these independent variables to understand the impact of independent variables on the dependent variables. The p values indicate the statistical significance of coefficients where $p < 0.01$, $p < 0.05$, and $p < 0.1$ (indicating ***, **, and *, respectively). For the “goals” variable, the coefficients are consistently positively significant, indicating that scoring more goals is associated with a higher value for the player. For the “awards” variable, we can see a positive and highly significant coefficient across all models, therefore suggesting that receiving more awards significantly increases a player’s value. For the “position_encoded” variable, we can see a positive and significant relationship, indicating that higher encoded positions (e.g. 4 being a striker, 3 a midfielder, 2 a

defender, and 1 being a goalkeeper) mean a higher value for the player. In other words, a striker is more likely to be worth more than a goalkeeper. For the “assists” variable, we can see positive and significant coefficients in models 4 to 6, showing that achieving more assists enhances the player’s value. The coefficient of the “age” variable, is negative and significant in model 6, implying that all of the player’s values tend to be lower as they get older.

Table 3. Regression analysis for highest_value.

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	hv	hv	hv	hv	hv	hv
goals	7.334*** (1.880)	4.384*** (1.184)	2.784*** (1.019)	2.414*** (0.916)	2.414*** (0.916)	2.327*** (0.898)
award		1.863*** (0.084)	1.866*** (0.084)	1.847*** (0.082)	1.847*** (0.082)	1.934*** (0.092)
position_encoded			0.877*** (0.133)	0.587*** (0.126)	0.587*** (0.126)	0.496*** (0.126)
assists				6.988*** (1.213)	6.988*** (1.213)	6.864*** (1.201)
age						-0.185*** (0.027)
Constant	5.232*** (0.238)	1.949*** (0.194)	-0.235 (0.338)	0.027 (0.326)	0.027 (0.326)	4.938*** (0.693)
Observations	10,754	10,754	10,754	10,754	10,754	10,754
R-squared	0.017	0.285	0.289	0.294	0.294	0.297

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

We applied log transformation to the monetary variables (highest_value). The main results remain similar and robust.

Table 4(a) and **Table 4(b)** examine the relationship between the dependent variable, position_encoded, and several independent variables, including goals, yellowcards, assists, appearance, height, award, highest_value, and current_value. Each column represents a different regression model, incorporating various combinations of these independent variables to understand the impact of independent variables on the dependent variables. The p values indicate the statistical significance of coefficients where $p < 0.01$, $p < 0.05$, and $p < 0.1$ (indicating ***, **, and *; respectively). For the “goals” variable, the coefficients are consistently positively significant, indicating that scoring more goals is associated with a higher value of position encoded (e.g. goals are scored more by strikers than goalkeepers). For the “yellow cards” variable, we can see a slightly positive and significant coefficient across all models, therefore suggesting that receiving more yellow cards correlates with an encoded position of lower values, potentially reflecting more aggressiveness

Table 4. (a) Regression analysis for position_encoded; (b) Regression analysis for position_encoded.

(a)				
	(1)	(2)	(3)	(4)
VARIABLES	position_encoded	position_encoded	position_encoded	position_encoded
goals	1.820*** (0.375)	1.824*** (0.375)	1.603*** (0.341)	1.584*** (0.339)
yellow cards		0.189*** (0.064)	0.190*** (0.063)	0.192*** (0.063)
assists			1.662*** (0.223)	1.620*** (0.215)
appearance				0.002*** (0.001)
Constant	2.485*** (0.046)	2.448*** (0.047)	2.331*** (0.036)	2.277*** (0.026)
Observations	10,754	10,754	10,754	10,754
R-squared	0.189	0.196	0.251	0.253
(b)				
	(1)	(2)	(3)	(4)
VARIABLES	position_encoded	position_encoded	position_encoded	position_encoded
goals	1.554*** (0.333)	1.563*** (0.336)	1.557*** (0.335)	1.560*** (0.336)
yellow cards	0.164*** (0.055)	0.164*** (0.055)	0.164*** (0.055)	0.164*** (0.055)
assists	1.143*** (0.179)	1.156*** (0.180)	1.140*** (0.179)	1.145*** (0.179)
appearance	0.003*** (0.001)	0.003*** (0.000)	0.003*** (0.000)	0.003*** (0.000)
height	-0.046*** (0.001)	-0.046*** (0.001)	-0.046*** (0.001)	-0.046*** (0.001)
award		-0.010*** (0.002)	-0.015*** (0.003)	-0.018*** (0.003)
highest_value			0.001*** (0.000)	0.001*** (0.000)
current_value				-0.001*** (0.000)
Constant	10.641*** (0.224)	10.634*** (0.225)	10.670*** (0.226)	10.656*** (0.226)
Observations	10,754	10,754	10,754	10,754
R-squared	0.354	0.355	0.356	0.357

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

and fouls closer to the goals by defenders and midfielders. For the “assists” variable, we can see a positive and significant relationship, showing that assists tend to be made by higher encoded positions (e.g. 4 being a striker, 3 a midfielder, 2 a defender, and 1 being a goalkeeper). The coefficient of the “appearance” variable is slightly positive and significant, indicating that more appearances are correlated with lower encoded positions (e.g. a goalkeeper will play more than strikers). The coefficient of the “height” variable is negative and significant, implying that taller players are more likely to occupy lower encoded positions like goalkeepers and defenders. The coefficient of “awards” is negative and significant, suggesting an inverse relationship between awards and encoded positions, which might be a reflection of the distribution of awards amongst different position types. The coefficient of “highest_value” is positive and significant, indicating that higher encoded positions tend to be valued more. For “current_value”, the coefficient is negative and significant, suggesting that players with high current values may occupy positions with lower encoded values.

Although some outcomes (minutes played) are count variables and others (position_encoded) are ordinal, OLS was chosen for simplicity and interpretability. To ensure robustness, we also tested alternative specifications using Poisson and ordered logit models. The results remained consistent in direction and significance.

4. Conclusion

This paper demonstrates that sport data analytics can be effectively used to evaluate player performance and their corresponding values in football. We have been able to identify statistical factors that correlate with one another to determine if actions made by a player influence their future performances and valuations. Most of the variables positively correlate with each other except for player’s values that go against age and position, the position of players that goes against height and number of awards, as well as the number of yellow cards received in comparison to the number of minutes played. The study shows the importance of data analytics in today’s world for linking performance metrics together in order to develop strategies to enhance player performance.

4.1. Recommendation for Future Studies

Future research could expand on this study in several ways. Firstly, additional and more complicated variables, such as Expected Goals and the value of each possession, can be added to see how each performance indicator changes relative to new data. Secondly, Artificial Intelligence based models can be used to help dive deeper into the causes of each event, provide more comprehensive data and refine the data set in order to identify outliers that might not directly impact each performance indicator. Future studies could also explore the psychological factors or other aspects of players that might link to their performance and overall value.

4.2. Limitations of This Study

League-mix biases, potential omitted variables (team strength, salary), and correlations do not establish causality.

Acknowledgements

I would like to thank Dr. Abdullah Yalaman for guiding me through the process of conducting this research.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- Aliyarov, C., Rikhsivoev, M., Arabboev, M., Begmatov, S., Saydiakbarov, S., Nosirov, K., Vakhkhobov, S. et al. (2023). Artificial Intelligence in Performance Analysis of Football Matches and Players. *Bulletin of TUIT Management and Communication Technologies*, 3, 1-7.
- Aswin, P. S., Sajeev, A., Madhusudhanan, A., Moideen, N., & Suhail, C. R. (2024). Revolutionizing Football Management: A Data-Driven Approach with Random Forest Regressor. *International Journal of Engineering Research and Applications*, 4, 1-6.
- Auer, B. R., & Hiller, T. (2015). On the Evaluation of Soccer Players: A Comparison of a New Game-Theoretical Approach to Classic Performance Measures. *Applied Economics Letters*, 22, 1100-1107. <https://doi.org/10.1080/13504851.2015.1005809>
- Baboota, R., & Kaur, H. (2019). Predictive Analysis and Modelling Football Results Using Machine Learning Approach for English Premier League. *International Journal of Forecasting*, 35, 741-755. <https://doi.org/10.1016/j.ijforecast.2018.01.003>
- Berrar, D., Lopes, P., & Dubitzky, W. (2019). Incorporating Domain Knowledge in Machine Learning for Soccer Outcome Prediction. *Machine Learning*, 108, 97-126. <https://doi.org/10.1007/s10994-018-5747-8>
- Chang, V., Sajeev, S., Xu, Q. A., Tan, M., & Wang, H. (2024). Football Analytics: Assessing the Correlation between Workload, Injury and Performance of Football Players in the English Premier League. *Applied Sciences*, 14, Article 7217. <https://doi.org/10.3390/app14167217>
- Chu, D. P., & Wang, C. W. (2019). Empirical Study on Relationship between Sports Analytics and Success in Regular Season and Postseason in Major League Baseball. *Journal of Sports Analytics*, 5, 205-222. <https://doi.org/10.3233/jsa-190269>
- Cornforth, D., Campbell, P., Nesbitt, K., Robinson, D., & Jelinek, H. F. (2015). Prediction of Game Performance in Australian Football Using Heart Rate Variability Measures. *International Journal of Signal and Imaging Systems Engineering*, 8, 80-88. <https://doi.org/10.1504/ijssise.2015.067072>
- Frick, B. (2011). Performance, Salaries, and Contract Length: Empirical Evidence from German Soccer. *International Journal of Sport Finance*, 6, 87-118.
- García-Aliaga, A., Marquina, M., Coterón, J., Rodríguez-González, A., & Luengo-Sánchez, S. (2021). In-Game Behaviour Analysis of Football Players Using Machine Learning Techniques Based on Player Statistics. *International Journal of Sports Science & Coaching*, 16, 148-157. <https://doi.org/10.1177/1747954120959762>
- Jana, A., & Hemalatha, S. (2021). Football Player Performance Analysis Using Particle

- Swarm Optimization and Player Value Calculation Using Regression. *Journal of Physics: Conference Series*, 1911, Article ID: 012011. <https://doi.org/10.1088/1742-6596/1911/1/012011>
- Maszczyk, A., Golaś, A., Pietraszewski, P., Roczniok, R., Zając, A., & Stanula, A. (2014). Application of Neural and Regression Models in Sports Results Prediction. *Procedia—Social and Behavioral Sciences*, 117, 482-487. <https://doi.org/10.1016/j.sbspro.2014.02.249>
- Mehta, S., Furley, P., Raabe, D., & Memmert, D. (2024). Examining How Data Becomes Information for an Upcoming Opponent in Football. *International Journal of Sports Science & Coaching*, 19, 978-987. <https://doi.org/10.1177/17479541231187871>
- Min, B., Kim, J., Choe, C., Eom, H., & (Bob) McKay, R. I. (2008). A Compound Framework for Sports Results Prediction: A Football Case Study. *Knowledge-Based Systems*, 21, 551-562. <https://doi.org/10.1016/j.knosys.2008.03.016>
- Ramirez, M. C., Vilorio, A., Muñoz, A. P., & Posso, H. (2017). Application of Multiple Linear Regression Models in the Identification of Factors Affecting the Results of the Chelsea Football Team. *International Journal of Control Theory and Applications*, 10, 7-13.
- Sarlis, V., & Tjortjis, C. (2020). Sports Analytics—Evaluation of Basketball Players and Team Performance. *Information Systems*, 93, Article ID: 101562. <https://doi.org/10.1016/j.is.2020.101562>
- Thunberg Kalt, C. (2024). *Cluster Analysis on Football Teams Performance Data*. <https://www.diva-portal.org/smash/get/diva2:1833447/FULLTEXT01.pdf>
- Toma, P., & Campobasso, F. (2023). Using Data Analytics to Capture the Strategic and Financial Decision-Making of Europe's Top Football Club. *Technological Forecasting and Social Change*, 186, Article ID: 122116. <https://doi.org/10.1016/j.techfore.2022.122116>
- Wisdom, C., & Javed, A. (2023). *Machine Learning for Data Analytics in Football: Quantifying Performance and Enhancing Strategic Decision-Making*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4558733
- Yang, H., Zhang, S., Zhang, J., & Wang, C. (2024). Evaluating the Performance of Athletes in Various Sports Using Data Mining and Big Data Analytics. *Soft Computing*, 28, 2875-2890. <https://doi.org/10.1007/s00500-023-09620-9>