Determinants of Soccer Players' Market Values in Major European Leagues*

A Comparative Linear Regression Analysis Across Five Leagues

John Zhang

November 27, 2024

Table of contents

*Code and data are available at: https://github.com/Clearsky21z/Player_Market_Value_Analysis

1 Introduction

The global soccer transfer market is a dynamic and high-stakes environment where players' market values are central to team-building strategies, investments, and negotiations. Understanding what drives these values is a critical question for analysts, clubs, and stakeholders. While common factors such as goals scored or club reputation are widely acknowledged to affect valuations, their relative importance likely varies across leagues with differing levels of competition, financial resources, and playing styles.

This paper focuses on five major European soccer leagues—Premier League(England), La Liga (Spain), Serie A (Italy), Bundesliga (Germany), and Ligue 1 (France). These leagues are not only among the most competitive but also vary in terms of audience size, financial backing, and player scouting pipelines, making them ideal for a comparative study of market value determinants.

Using player data from the 2023/24 season, this study develops league-specific linear regression models with market value as the dependent variable. Key predictors include age, goals, assists, club ranking, national team ranking, minutes played, and position. The analysis reveals both shared and league-specific patterns in the factors influencing market value. For instance, goals and assists consistently emerge as significant predictors, but their impact differs by league. Similarly, younger players tend to command higher valuations universally, with age effects most pronounced in Serie A.

These findings contribute to our understanding of how soccer markets function across different leagues and provide actionable insights for clubs and agents. The structure of this paper is as follows: Section 2 describes the dataset and variables, Section 3 outlines the modeling approach, Section 4 presents the results, and Section 5 discusses the implications and limitations of the findings.

2 Data

2.1 Dataset Overview

The data includes information on players from the top five European leagues during the 2023/24 season. Variables of interest are:

- **Market Value**: Player's estimated value in Euros.
- **Age**: Player age at the start of the season.
- **Goals and Assists**: Indicators of offensive contributions.
- **Club Ranking**: Rank of the player's club based on league and international performance.
- **National Team Ranking**: Ranking of the player's national team (1 to 210).
- **Minutes Played**: Total minutes played during the season.
- **Position**: Categorized as Defender (DF), Midfielder (MF), Forward (FW), and hybrids like FWDF.

2.2 England

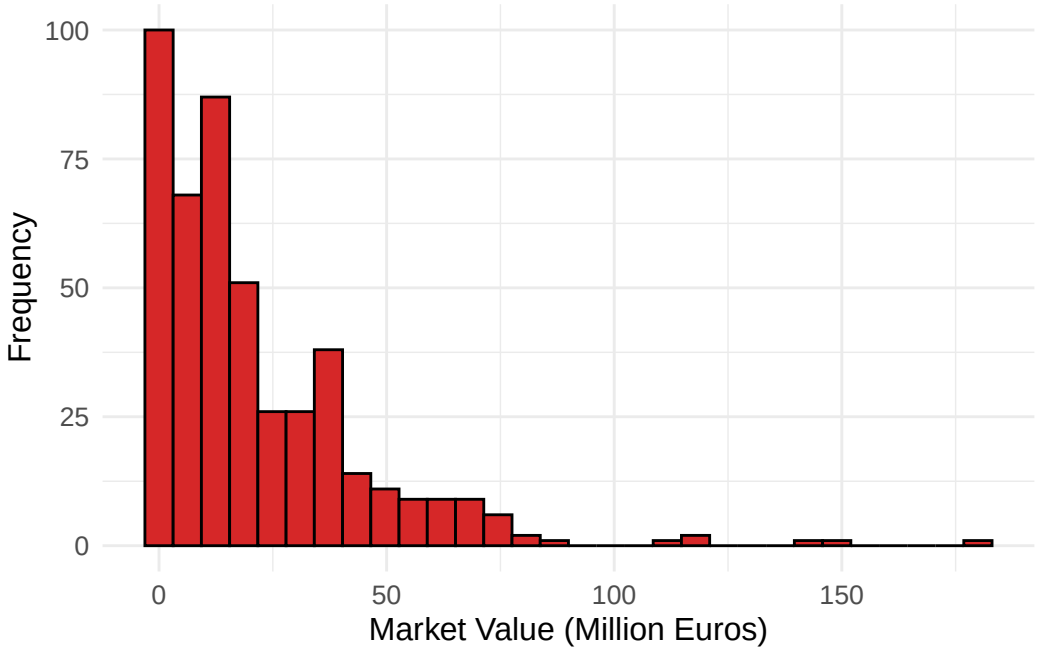

Figure 1: Distribution of Market Value (in Million Euros)

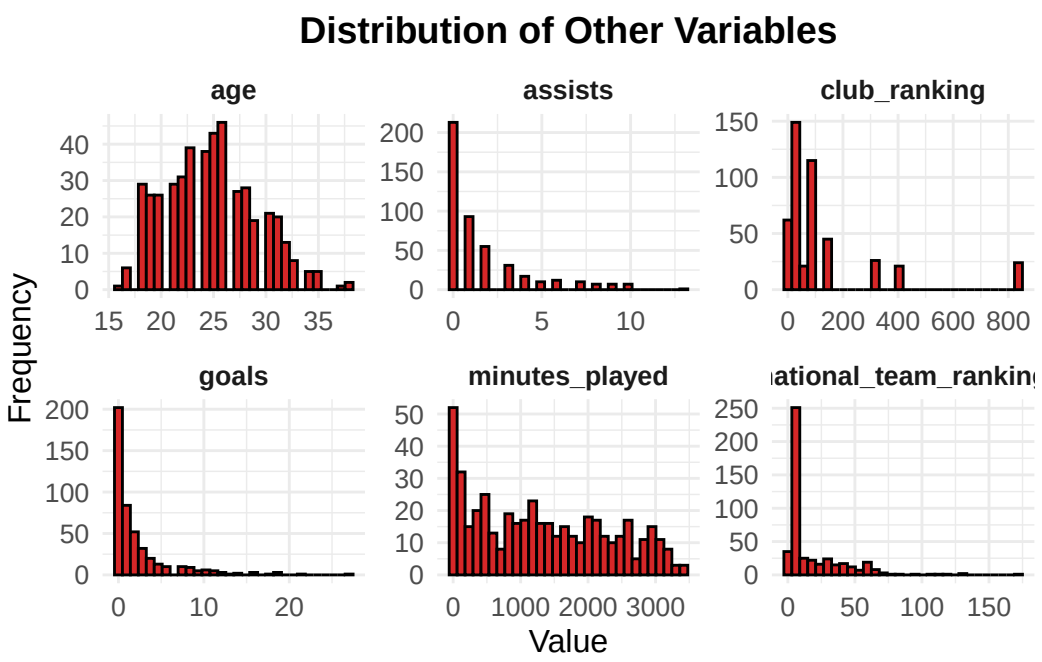

Figure 2: Distribution of other Variables

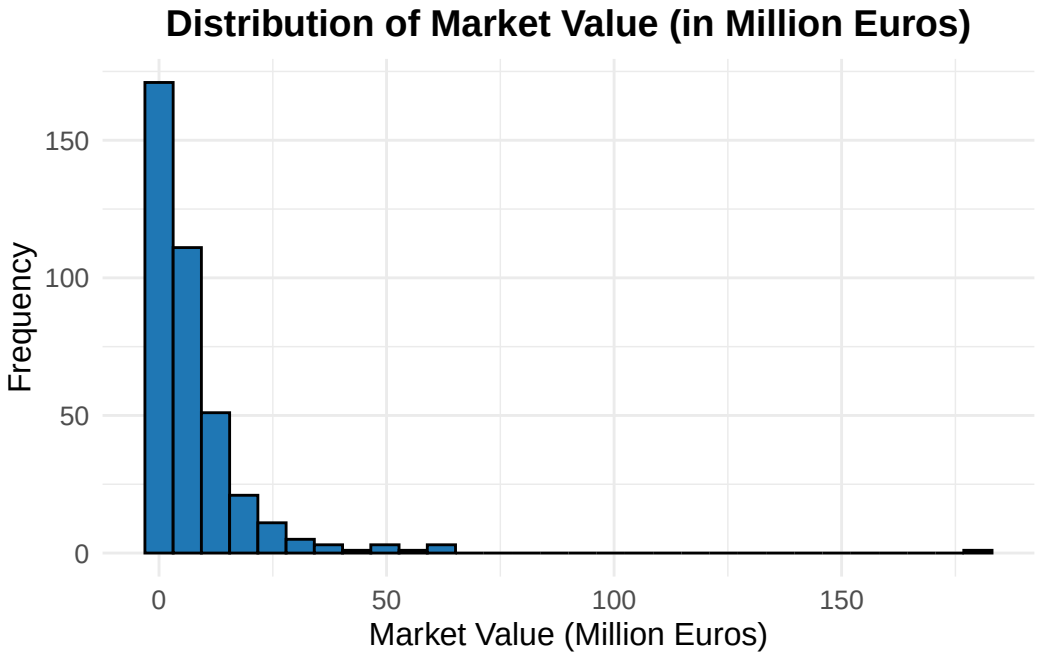

Figure 3: Distribution of Market Value (in Million Euros)

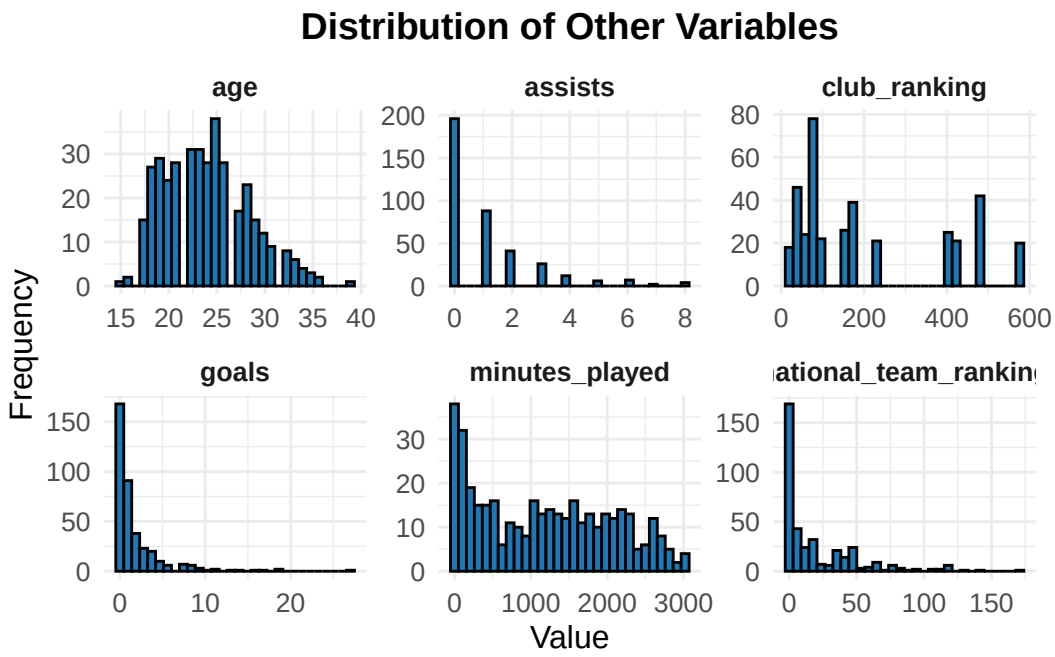

Figure 4: Distribution of other Variables

2.3 France

2.4 Germany

2.5 Italy

2.6 Spain

3 Model

In this study, we constructed a linear regression model for each of the five major European football leagues to analyze the determinants of player market value. The dependent variable, market value (Market Value$_i$), was scaled to millions of Euros for easier interpretability. The model incorporates a comprehensive set of predictors, including player performance metrics, contextual factors, and positional variables, to capture the multi-dimensional aspects influencing market valuation.

The general form of the model is as follows:

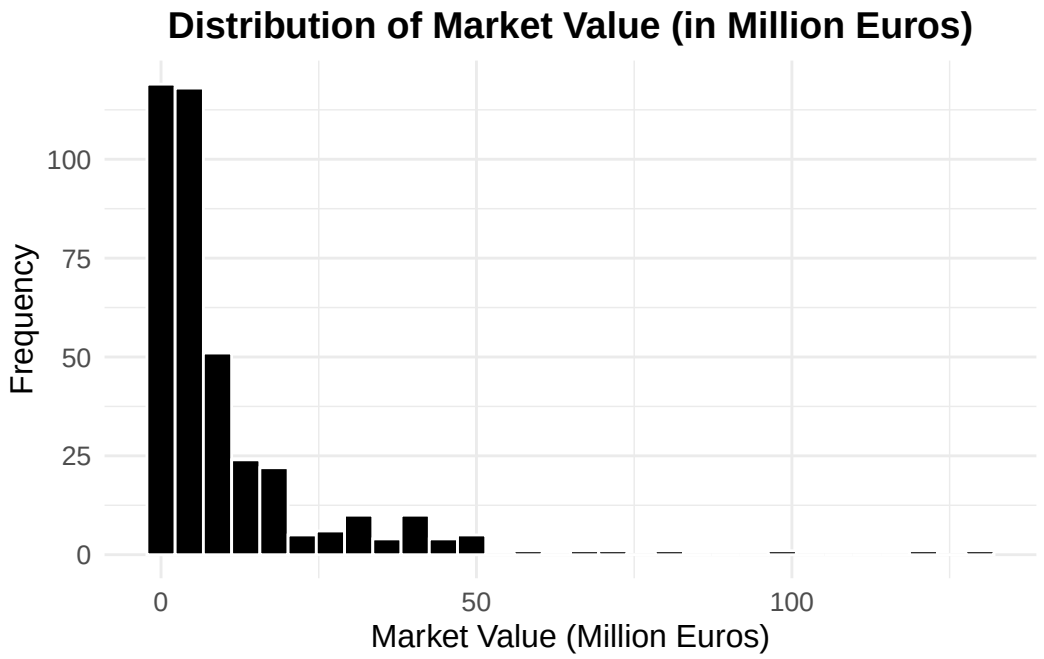

Figure 5: Distribution of Market Value (in Million Euros)

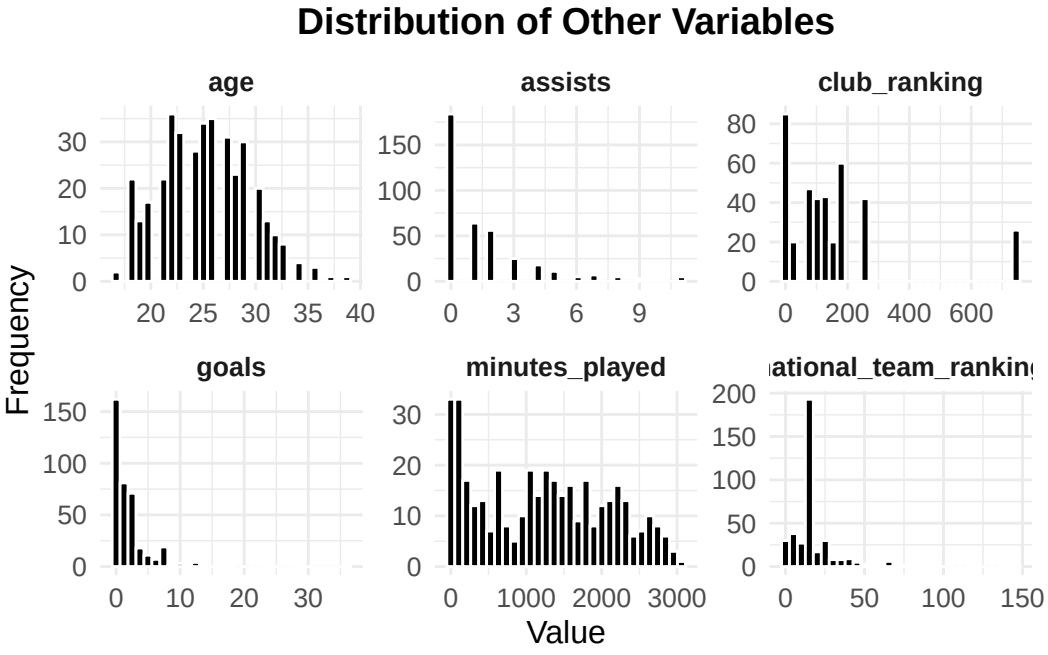

Figure 6: Distribution of other Variables

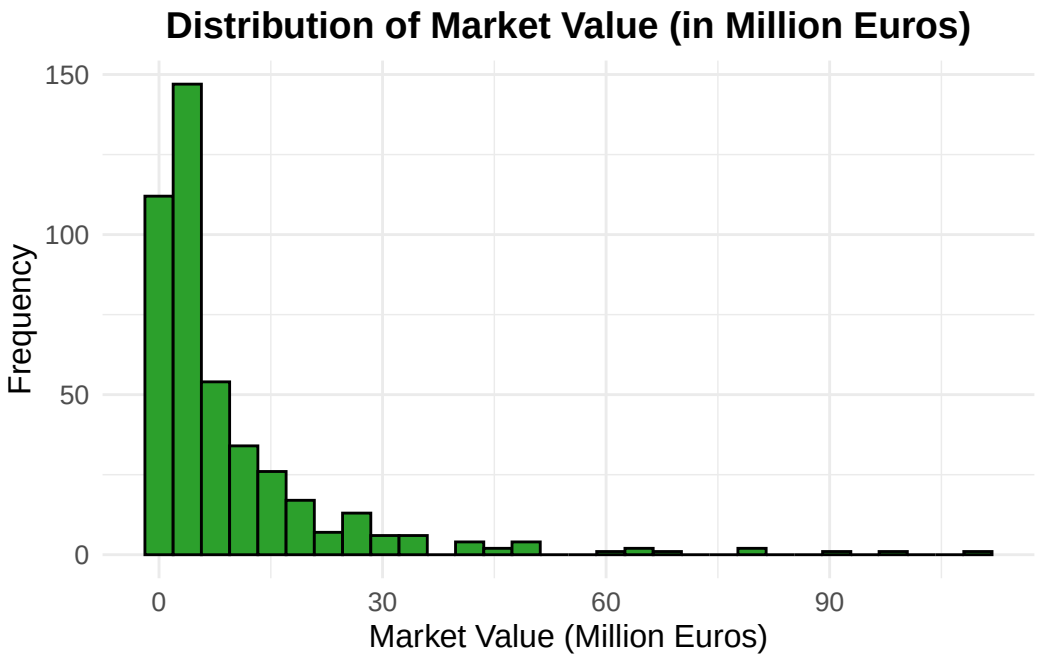

Figure 7: Distribution of Market Value (in Million Euros)

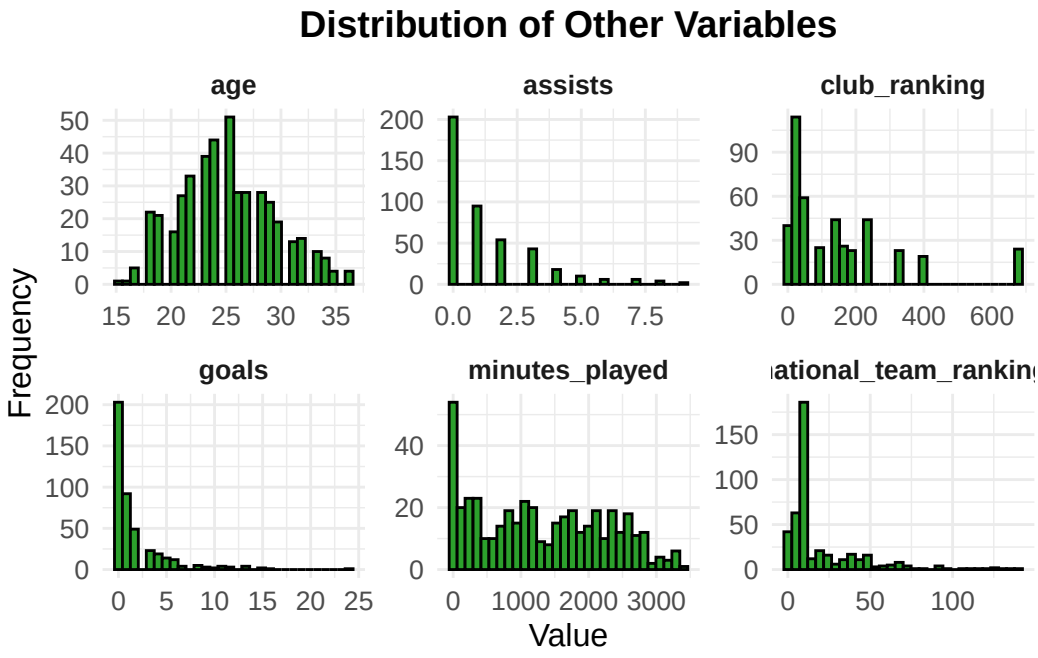

Figure 8: Distribution of other Variables

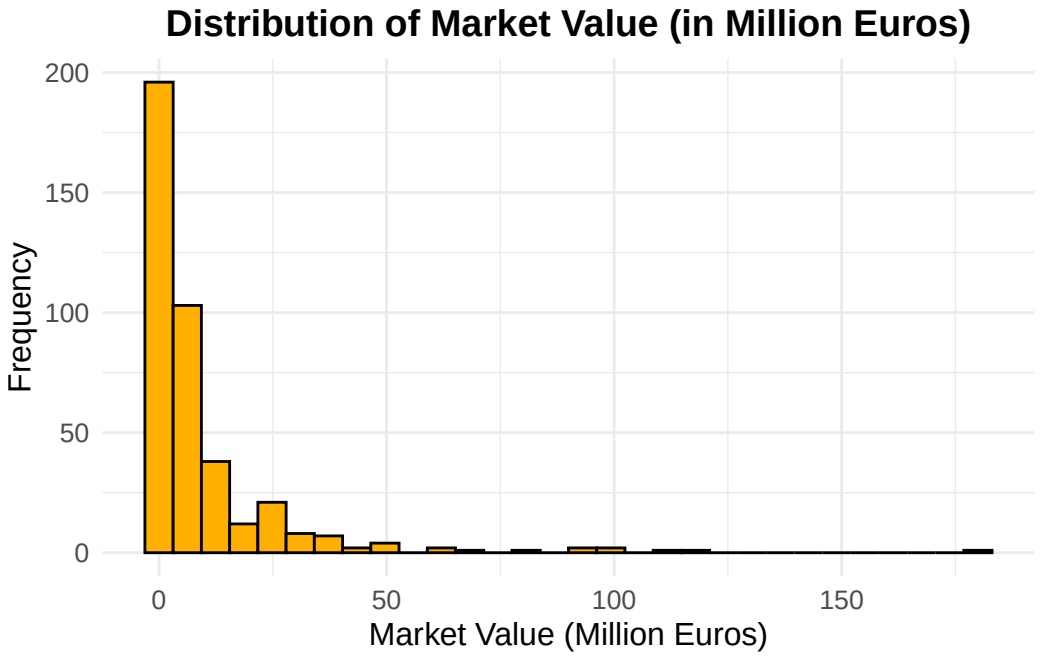

Figure 9: Distribution of Market Value (in Million Euros)

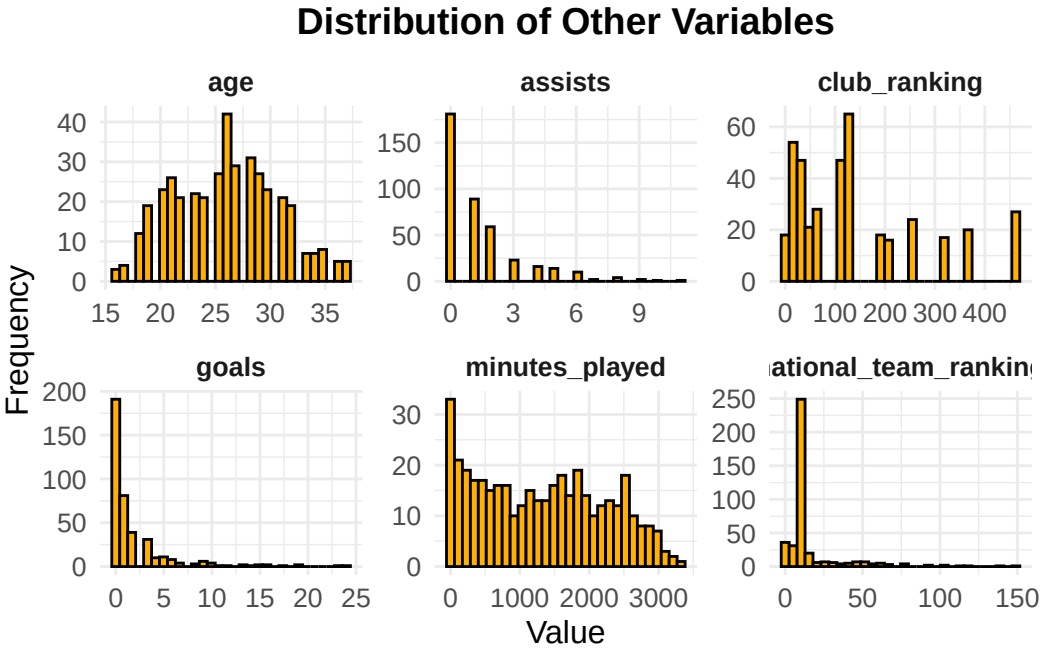

Figure 10: Distribution of other Variables

$$\begin{aligned}\text{Market Value}_i = \beta_0 &+ \beta_1(\text{Age}_i) + \beta_2(\text{Goals}_i) \\ &+ \beta_3(\text{Assists}_i) + \beta_4(\text{Club Ranking}_i) \\ &+ \beta_5(\text{National Team Ranking}_i) + \beta_6(\text{Minutes Played}_i) \\ &+ \beta_7(\text{Position}_i) + \epsilon_i\end{aligned}$$

where: - β_0: The intercept, representing the baseline market value when all predictors are zero. - β_1: The coefficient for age, showing how market value changes with each additional year of age. - β_2: The coefficient for goals, representing the increase in market value per additional goal scored. - β_3: The coefficient for assists, reflecting the market value impact of providing assists. - β_4: The coefficient for club ranking, with better (lower) rankings associated with higher market values. - β_5: The coefficient for national team ranking, highlighting the effect of international representation. - β_6: The coefficient for minutes played, indicating the impact of game time on market value. - β_7: The set of coefficients for positional categories, capturing positional differences in market valuation relative to a baseline. - ϵ_i: The residual error, accounting for unobserved factors affecting market value.

3.1 Variable Justification

The predictors included in the model were carefully selected based on their relevance to market valuation processes in professional football:

1. **Age**: Younger players typically have higher market values due to their potential for future development.
2. **Goals and Assists**: These are key indicators of a player's offensive contribution, which are highly valued across all leagues.
3. **Club Ranking**: Players at higher-performing clubs (lower rankings) tend to have greater exposure and higher valuations.
4. **National Team Ranking**: Representing a higher-ranked national team enhances player marketability and perceived skill.
5. **Minutes Played**: A proxy for fitness and importance to the team, minutes played captures game-time contributions.
6. **Position**: Positional variables were included to account for role-specific valuation differences, such as the higher premium on forwards compared to defenders.

3.2 Modeling Process

The models were implemented using the R programming language, specifically leveraging the `lm()` function for linear regression. For each league, data were split into training and testing sets to validate model performance. A stratified train-test split (80% training, 20% testing) was performed using the `caret` package to ensure balanced representation of market value distributions across the sets.

3.3 Model Validation

To evaluate model performance, predictions were generated on the test set, and Mean Squared Error (MSE) was computed for each league. MSE provides an average measure of prediction error, quantifying the extent to which observed market values deviate from predicted values. The MSE values across the five leagues are as follows: - England: 265.27 - France: 100.49 - Germany: 215.51 - Italy: 109.95 - Spain: 122.17

The variation in MSE reflects differences in data variability and the complexity of valuation processes in each league. For instance, the relatively high MSE in England suggests greater variability in market values, likely influenced by the financial power and diversity of clubs in the Premier League.

3.4 Alternative Models and Future Directions

While linear regression provides a transparent and interpretable framework, alternative approaches could enhance predictive accuracy: - **Ridge Regression**: To address multicollinearity, ridge regression could be employed to shrink coefficient estimates, reducing variance. - **Random Forests or Gradient Boosting**: Nonlinear methods could capture complex interactions between predictors, potentially improving fit for leagues with high variability. - **Bayesian Regression**: A Bayesian approach would allow for the incorporation of prior knowledge and explicit uncertainty quantification for coefficients.

4 Results

This section summarizes the key findings from the league-specific regression analyses, emphasizing only significant predictors of market value. Insignificant estimates, such as national team rankings and most position-specific coefficients, are excluded from this section but are included in the appendix.

The intercepts, which represent the baseline market value for a player with average characteristics, vary significantly across leagues. The Premier League (England) exhibits the highest intercept (41.057 million Euros), as shown in Figure 11, highlighting the league's financial dominance. Conversely, Ligue 1 (France) has the lowest intercept (22.946 million Euros), reflecting its relatively limited financial capacity.

Age negatively affects market value across all leagues, reflecting the premium placed on youth. Figure 12 reveals that La Liga (Spain) places the strongest emphasis on younger players, with a coefficient of -1.179 million Euros, followed by the Premier League (-1.137 million Euros). Ligue 1 (-0.593 million Euros) places the least emphasis on age.

Goals and assists are consistent predictors of market value across leagues. As shown in Figure 13, the impact of goals is highest in the Premier League (2.993 million Euros per goal) and

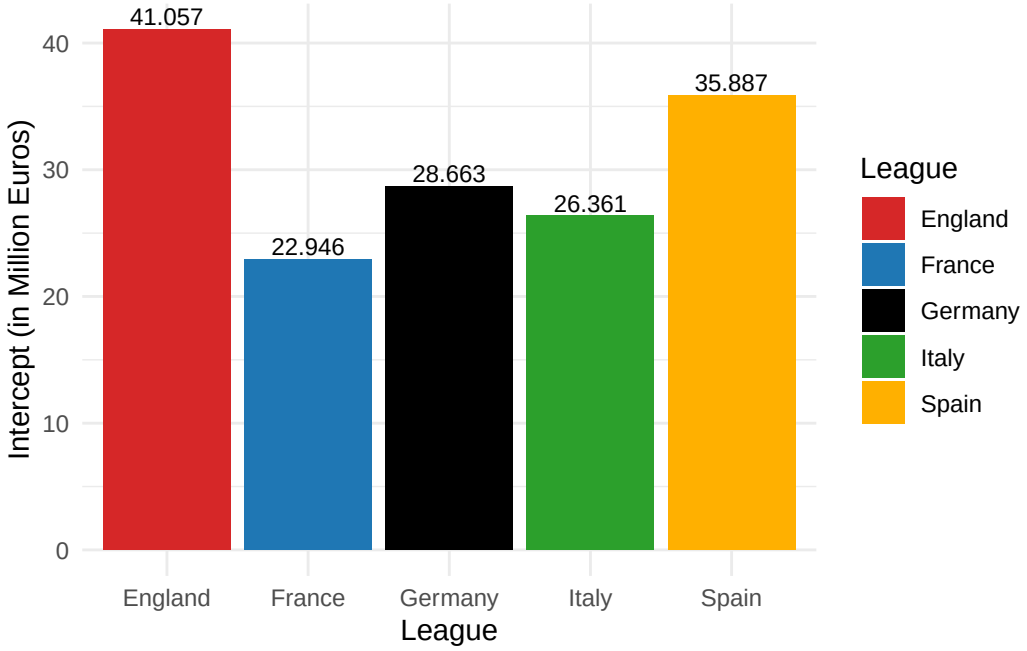

Figure 11: Intercept Estimates Across Leagues

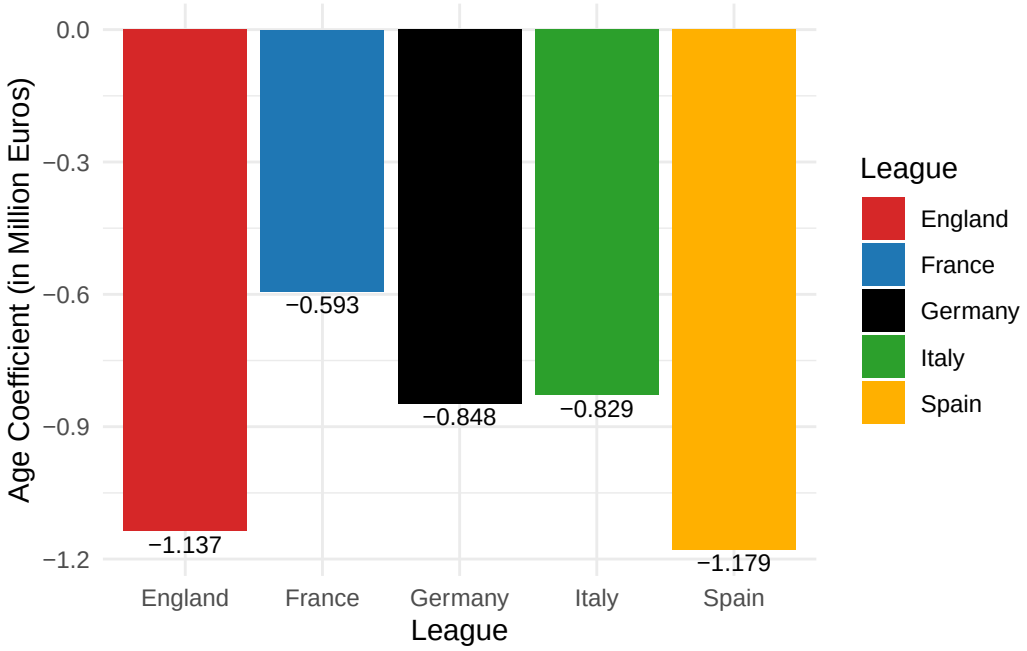

Figure 12: Age Coefficient Estimates Across Leagues

Serie A (Italy) (2.489 million Euros). Figure 14 shows that assists have the strongest effect in the Bundesliga (Germany) (1.828 million Euros), while Serie A demonstrates the smallest effect (0.820 million Euros). These findings emphasize the financial rewards for offensive contributions, particularly in leagues like England and Germany.

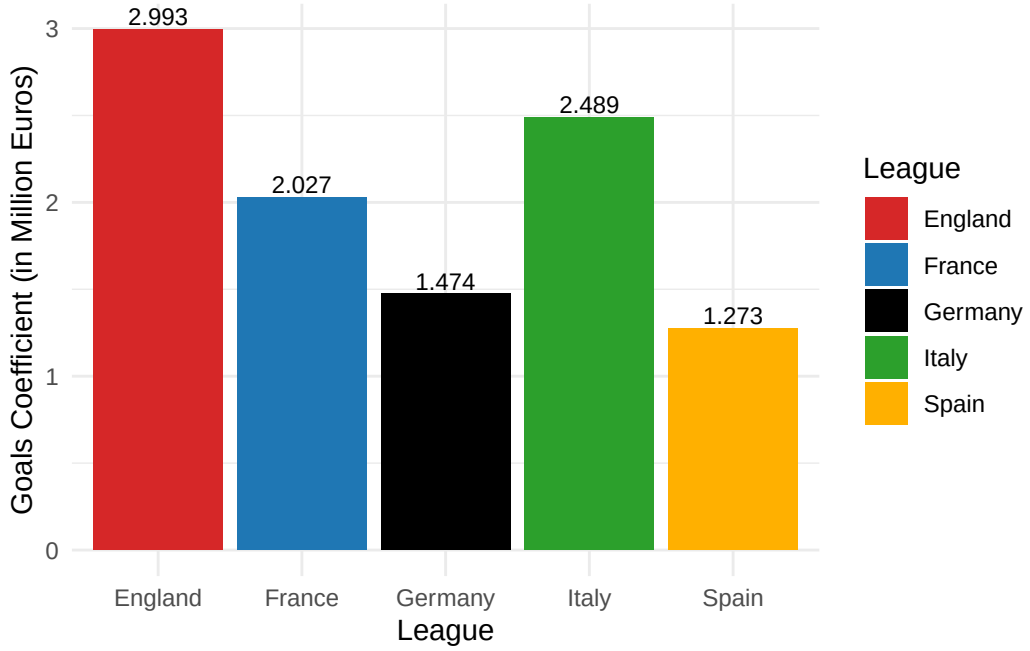

Figure 13: Goal Coefficient Estimates Across Leagues

Club ranking negatively influences market value in all leagues, as better club performance (lower numerical ranking) increases valuations. Figure 15 illustrates that the strongest effect is observed in La Liga (-0.034 million Euros per rank), followed by the Premier League (-0.027 million Euros). This highlights the importance of team performance in player valuation, particularly in Spain.

Minutes played is a significant, though modest, predictor of market value in England, Italy, and Spain. As shown in Figure 16, the coefficients range from 0.004 million Euros in England and Spain to 0.002 million Euros in Italy. These results suggest that playing time moderately impacts market value in these leagues.

Most position-specific coefficients are insignificant across leagues. However, Table 1 highlights some notable findings. In England, forwards (FW) have a significant negative effect on market value (-9.608 million Euros). Similarly, hybrid forward-midfielders (FWMF) have significant negative coefficients in France (-5.011 million Euros), Germany (-5.238 million Euros), and Italy (-4.877 million Euros). In Spain, none of the position-specific estimates are significant, indicating minimal variation in market valuation by position.

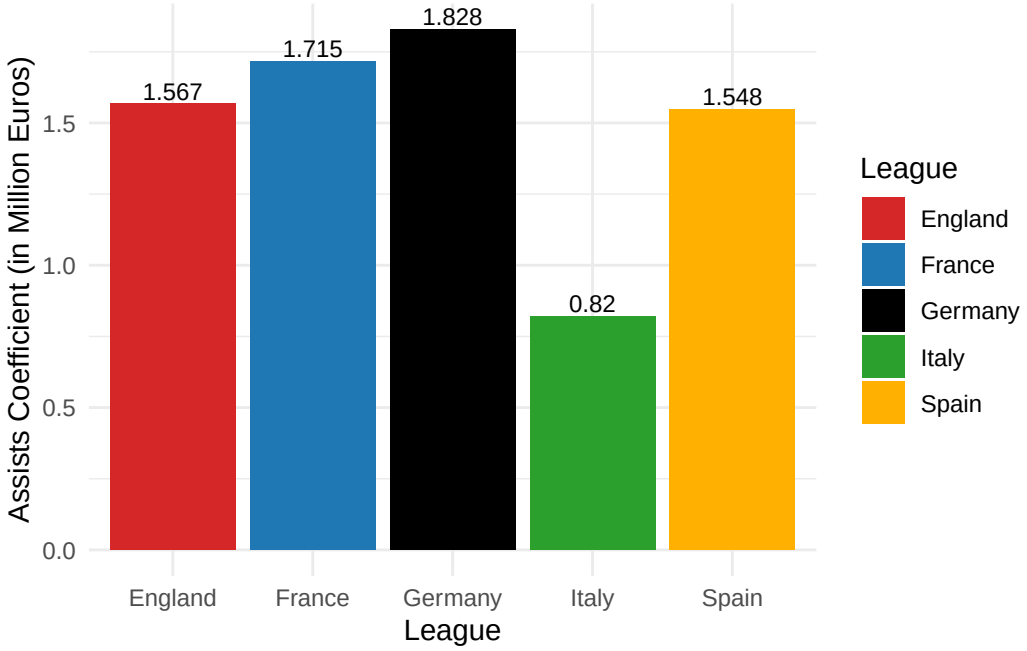

Figure 14: Assists Coefficient Estimates Across Leagues

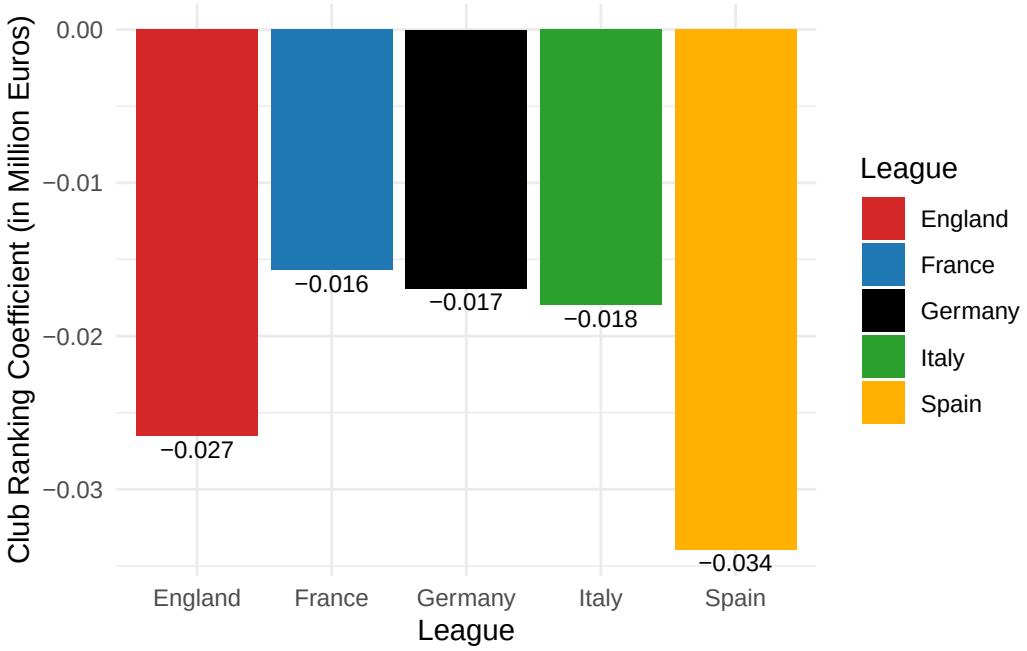

Figure 15: Club Ranking Coefficient Estimates Across Leagues

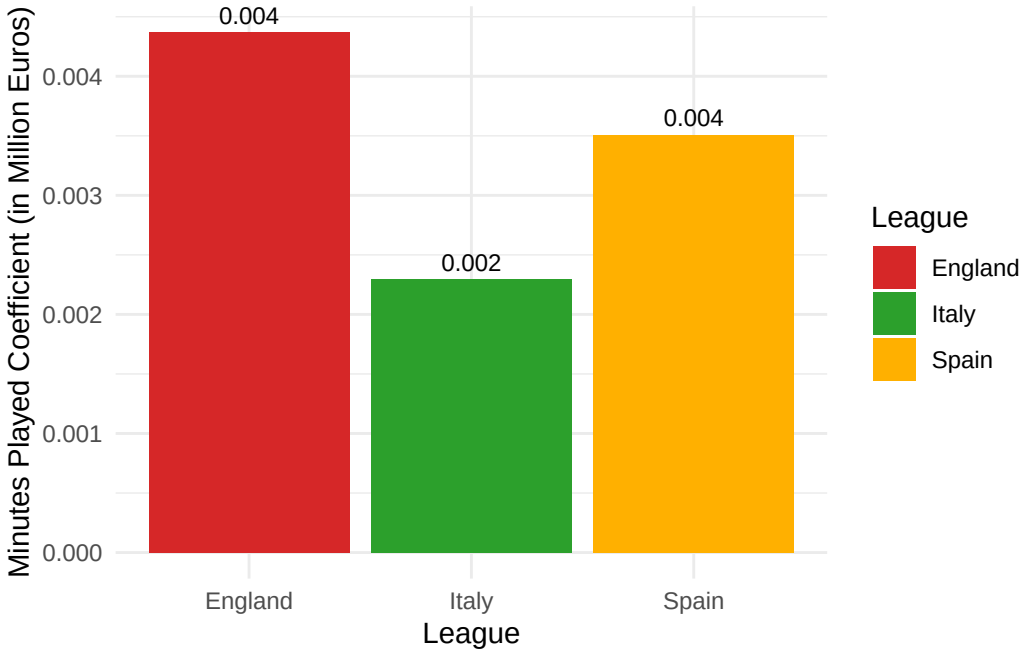

Figure 16: Minutes Played Coefficient Estimates Across Selected Leagues

Table 1: Position Specific Estimates in Selected Leagues

Position	League	Estimate
positionFW	England	-9.608
positionFW	France	-7.041
positionFWMF	France	-5.011
positionFWDF	Germany	-7.835
positionFWMF	Germany	-5.238
positionFWMF	Italy	-4.877

5 Discussion

This paper sought to identify the determinants of soccer players' market values in five major European leagues: the Premier League (England), Ligue 1 (France), Bundesliga (Germany), Serie A (Italy), and La Liga (Spain). By using league-specific linear regression models, we examined how various predictors such as age, goals, assists, club ranking, minutes played, and player position influenced market value. The findings reveal both universal and league-specific patterns in the determinants of market value, providing valuable insights into how soccer markets operate across different contexts.

5.1 Cross-League Differences in Market Value Baselines

One of the most striking findings is the variation in the intercept values across leagues, which reflect the baseline market value for players with average characteristics (or zero predictors). The Premier League (England) has the highest baseline market value (41.057 million Euros), followed by La Liga (Spain) and the Bundesliga (Germany). In contrast, Ligue 1 (France) exhibits the lowest intercept (22.946 million Euros). These results align with existing research, which highlights the financial strength and global appeal of the Premier League compared to other leagues. The disparity in intercepts underscores the role of financial resources, international exposure, and league-specific brand value in shaping player valuations.

5.2 Universal Importance of Goals and Assists

Goals and assists emerged as the most consistent predictors of market value across leagues. The Premier League shows the strongest relationship between goals and market value, followed by Serie A. This finding reflects the premium placed on offensive contributions, particularly in leagues that emphasize high-scoring games or where attacking players are highly valued in the transfer market. Similarly, assists play a crucial role in Germany's Bundesliga, likely reflecting the league's focus on team-oriented and high-tempo playing styles. These results confirm prior research, which identifies offensive output as a key driver of player valuations in soccer.

5.3 Age and the Youth Premium

Age negatively impacts market value across all leagues, with the largest effect observed in La Liga (-1.179 million Euros per year) and the smallest in Ligue 1 (-0.593 million Euros per year). This pattern demonstrates the universal preference for younger players, who are perceived to have greater potential for development and resale value. However, the magnitude of this effect varies by league, possibly due to differences in scouting practices and player development pipelines. For example, Spain's emphasis on nurturing young talent aligns with its stronger age effect, while Ligue 1, known for exporting young players, may rely on other metrics for valuation.

5.4 Club Ranking and Market Value

Club ranking, a proxy for team reputation and performance, shows a significant but modest negative relationship with market value. This finding indicates that players from higher-performing clubs (lower-ranked) tend to have higher valuations. Spain exhibits the strongest club ranking effect (-0.034 million Euros per rank improvement), which may reflect the emphasis placed on club prestige in La Liga's valuation processes. However, the relatively small magnitude across leagues suggests that individual performance metrics, such as goals and assists, are weighted more heavily than team success.

5.5 Position-Based Valuations

The role of player position in determining market value varied significantly across leagues, with most position-specific coefficients being statistically insignificant. However, some notable exceptions were observed. Forwards (FW) in England and France exhibited significant negative effects on market value, with the Premier League showing the strongest effect (-9.608 million Euros). Hybrid positions, such as forward-midfielders (FWMF), were significant in France, Germany, and Italy, indicating that players with versatile roles are valued differently depending on the league. Interestingly, none of the position-specific coefficients in Spain were significant, suggesting that La Liga places less emphasis on positional roles in player valuation and more on overall talent and individual performance. These findings highlight the complexity of valuing players based on position, as tactical demands and league-specific preferences play a significant role.

5.6 Weaknesses of the Study

While the findings provide valuable results, the study has several limitations. First, it does not account for external factors such as player injuries, media influence, and agent negotiations, all of which are known to influence market value. Second, the analysis focuses on a single season (2023/24), limiting its ability to capture temporal trends and variations caused by external events such as the COVID-19 pandemic or changes in league regulations. Third, the use of linear regression assumes a straightforward relationship between predictors and market value, which may oversimplify complex interactions or nonlinear effects (e.g., diminishing returns on goals). Finally, the study's focus on Europe's top five leagues excludes emerging leagues, such as Major League Soccer (MLS) or the Saudi Pro League, where market dynamics may differ significantly.

5.7 Future Directions

The findings of this study have important implications for clubs, agents, and analysts in the soccer industry. Clubs can use these insights to refine their scouting and recruitment strategies by focusing on metrics that are most valued in their specific league, such as goals in the Premier League or assists in the Bundesliga. Agents can leverage the results to highlight the most marketable aspects of their clients during contract negotiations, emphasizing factors like offensive contributions or club prestige. Additionally, analysts and researchers can build on these findings to explore broader trends in player valuations and market dynamics.

Future research could address the limitations of this study by incorporating external variables such as media coverage, player injuries, and sponsorship deals. Expanding the analysis to include data from multiple seasons would provide a more comprehensive understanding of long-term trends and the impact of external shocks, such as economic crises or regulatory changes. Furthermore, the use of advanced machine learning models could capture more

complex relationships between predictors and market value, offering a deeper understanding of the factors that drive player valuations. Including emerging leagues and non-European markets would also provide a more global perspective on the determinants of market value, shedding light on how valuations differ between established and developing soccer ecosystems.

A Appendix

A.1 Dataset and Graph Sketches

A.2 Model Results

Table 2: Regression Results for Premier League (England)

Variable	Estimate	Standard Error	P-Value
(Intercept)	41.057	4.663	0.000
age	-1.137	0.173	0.000
goals	2.993	0.292	0.000
assists	1.567	0.408	0.000
club_ranking	-0.027	0.004	0.000
national_team_ranking	-0.065	0.032	0.044
minutes_played	0.004	0.001	0.000
positionFW	-9.608	2.661	0.000
positionFWDF	-7.253	4.575	0.114
positionFWMF	-3.811	2.300	0.098
positionMF	2.019	2.083	0.333
positionMFDF	-2.042	2.822	0.470

Table 3: Regression Results for Ligue 1 (France)

Variable	Estimate	Standard Error	P-Value
(Intercept)	22.946	3.193	0.000
age	-0.593	0.128	0.000
goals	2.027	0.237	0.000
assists	1.715	0.430	0.000
club_ranking	-0.016	0.003	0.000
national_team_ranking	-0.036	0.018	0.051
minutes_played	0.001	0.001	0.429
positionFW	-7.041	2.085	0.001
positionFWDF	-4.372	2.938	0.138
positionFWMF	-5.011	1.642	0.002
positionMF	-2.938	1.500	0.051
positionMFDF	-1.905	2.148	0.376

Table 4: Regression Results for Bundesliga (Germany)

Variable	Estimate	Standard Error	P-Value
(Intercept)	28.663	3.917	0.000
age	-0.848	0.150	0.000
goals	1.474	0.227	0.000
assists	1.828	0.356	0.000
club_ranking	-0.017	0.004	0.000
national_team_ranking	0.012	0.035	0.726
minutes_played	0.001	0.001	0.152
positionFW	-2.081	2.557	0.416
positionFWDF	-7.835	3.077	0.011
positionFWMF	-5.238	1.845	0.005
positionMF	-1.469	1.817	0.419
positionMFDF	-2.433	2.127	0.253

Table 5: Regression Results for Serie A (Italy)

Variable	Estimate	Standard Error	P-Value
(Intercept)	26.361	3.038	0.000
age	-0.829	0.113	0.000
goals	2.489	0.216	0.000
assists	0.820	0.353	0.021
club_ranking	-0.018	0.003	0.000
national_team_ranking	-0.014	0.020	0.487
minutes_played	0.002	0.001	0.001
positionFW	-2.247	1.694	0.185
positionFWDF	-2.327	3.227	0.471
positionFWMF	-4.877	1.510	0.001
positionMF	-1.911	1.268	0.132
positionMFDF	-1.933	2.183	0.376

Table 6: Regression Results for La Liga (Spain)

Variable	Estimate	Standard Error	P-Value
(Intercept)	35.887	4.505	0.000
age	-1.179	0.160	0.000
goals	1.273	0.296	0.000
assists	1.548	0.502	0.002

Variable	Estimate	Standard Error	P-Value
club_ranking	-0.034	0.006	0.000
national_team_ranking	-0.020	0.037	0.593
minutes_played	0.004	0.001	0.001
positionFW	-4.474	2.788	0.109
positionFWDF	0.062	5.845	0.992
positionFWMF	2.013	2.393	0.401
positionMF	1.170	2.029	0.565
positionMFDF	6.327	3.004	0.036

B References