

Influences on Soccer Player Value

An Analysis of Transfermarkt and Europe's Top 5 Leagues

Devin Williams

April 29th, 2025

ECON 5253

University of Oklahoma

Table of Contents

Introduction

- Significance of The World's Game
- Operational Definitions
- Margin of Error

Data

- TransferMarkt and API

Methods

- Finding Connections
- Club Categories
- Regression Equations

Findings

- Regression Findings
- Peak Age

Conclusion

Introduction

Significance of the World's Game

- 1.5 Billion people watched the World Cup Final in 2022 (Summerscales, 2023).
- Highest paid player in Europe - Robert Lewandowski (Chalke, 2025).
- Other players worldwide make more, but Europe is the highest level of competition.

Operational Definitions

- Transfer/"Sell"
- Transfer Value/Market Value
- Release Clause
- "The Big 5"
- Club
- Relegation/Promotion
- Transfermarkt

€10,960,000,000

Amount spent in JUST 2024 on Player Transfers (Poli et al., 2024).

Margin of Error - Neymar Jr.


$$\frac{\text{Market value (100m)}}{\text{True price (222m)}}$$


- A lot of different factors can impact true sale price.
- Transfermarkt can't quantify everything.

Margin of Error - Frankie De Jong


$$\frac{\text{Market value (85m)}}{\text{True price (86m)}}$$


- Many causes, evaluations are close to being accurate selling prices.
- This study looked to understand factors to the website Transfermarkt's player evaluations, while evaluating country of origin biases.

Data

Transfermarkt and API

- Transfermarkt is a site where people "go to find, and discuss, information about soccer players" (Smith, 2021).
- The website is a leader in player and club evaluations. (Smith, 2021).
- Absence of a algorithm (Smith, 2021).
- This project utilizes a third-party Github repository that allows access to an unofficial Transfermarkt API (Allegretti, 2023).
- This API was used to download data related to player demographics, club and league information, performance metrics, and current market evaluations.

Methods

Finding Connections

- Data was collected to understand what impacts market evaluation.
- Systematic comparisons were done of market values across continents, performance metrics, positions, and age groups.
- Player's age was squared to evaluate the non-linearity of age in the sample.
- Log-linear regression models were done for each position group.

Club Categories

- Clubs are put into 5 different categories.
- Elite Clubs, Strong Clubs, Mid-Table Clubs, Lower-Tier Clubs, and Other Clubs.
- Teams are split based on Rescaled squad value (40%), maximum player value(30%), squad size(20%), and youthfulness(10%) of the squad.
- Z-scores used to standardize the average squad value within each league.
This is meant to allow fairer comparison across leagues.

Tier	Examples
Tier 1: Elite Clubs	 Real Madrid,  Barcelona,  Bayern Munich
Tier 2: Strong Clubs	 AS Monaco,  Newcastle United,  Atletico Madrid
Tier 3: Mid-Table Clubs	 Aston Villa,  Crystal Palace,  Brentford FC
Tier 4: Lower-Tier Clubs	 Girona FC,  Leicester City,  Stade Reims
Tier 5: Other Clubs	 Real Betis,  FC Nantes,  FC Union Berlin

Regression Equations

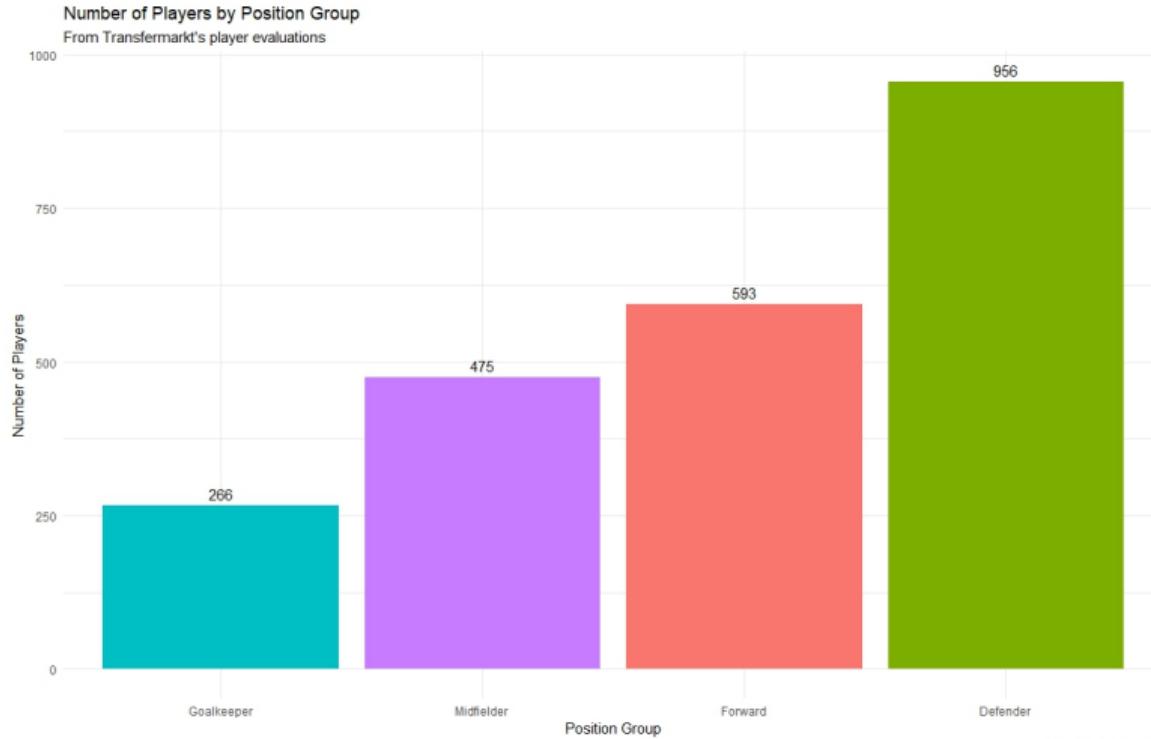
$$\begin{aligned}\ln(\text{MarketValue}_i) = & \beta_0 + \beta_1 \text{Continent}_i + \beta_2 \text{Age}_i \\ & + \beta_3 \text{AgeSquared}_i \\ & + \beta_4 \text{Appearances}_i \\ & + \beta_5 \text{GoalContributions}_i^* \\ & + \beta_6 \text{League}_i + \beta_7 \text{ClubTier}_i + \varepsilon_i\end{aligned}\tag{1}$$

*Not included in the goalkeeper model

Where:

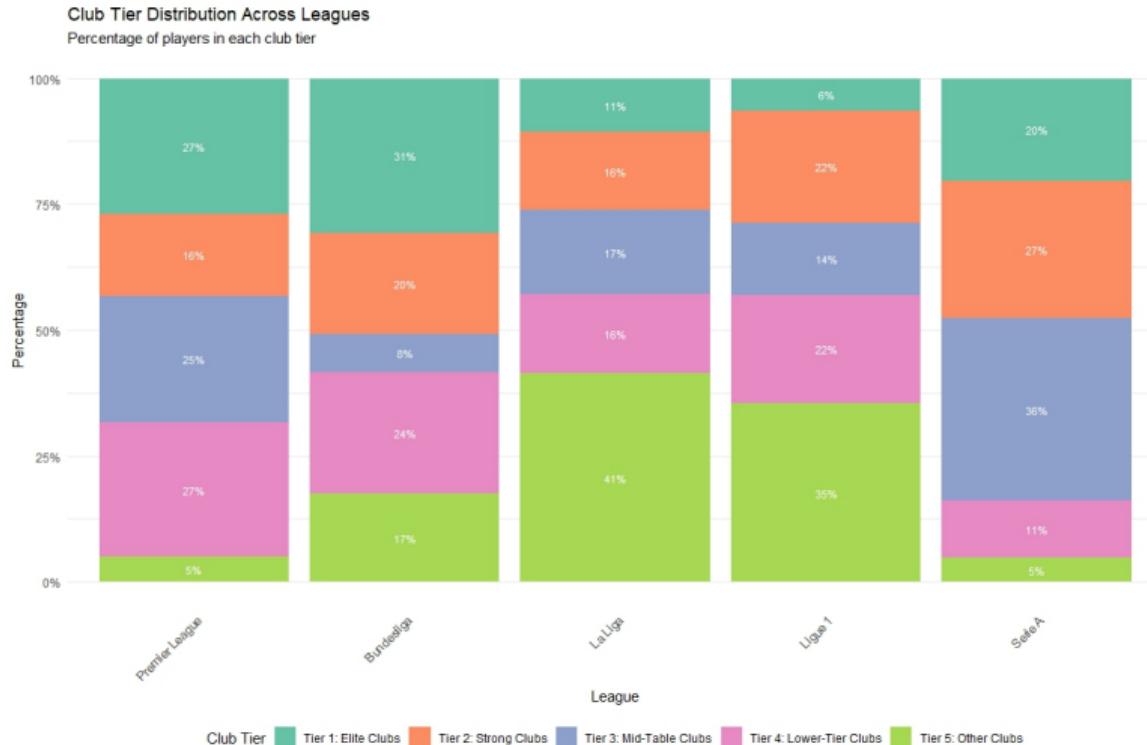
- MarketValue_i = Value in € (By position)
- Continent_i = Origin (ref: Europe)
- Age_i = Age in years
- AgeSquared_i = Age^2
- Appearances_i = Games played in last 2 years
- $\text{GoalContributions}_i$ = Goals+assists in last 2 years
- League_i = League (ref: EPL)
- ClubTier_i = Club tier (ref: Elite)
- ε_i = Error term

Player Distribution by Position Group



Source: Transfermarkt

Club Tier Distribution Across Leagues



Findings

Regression Findings

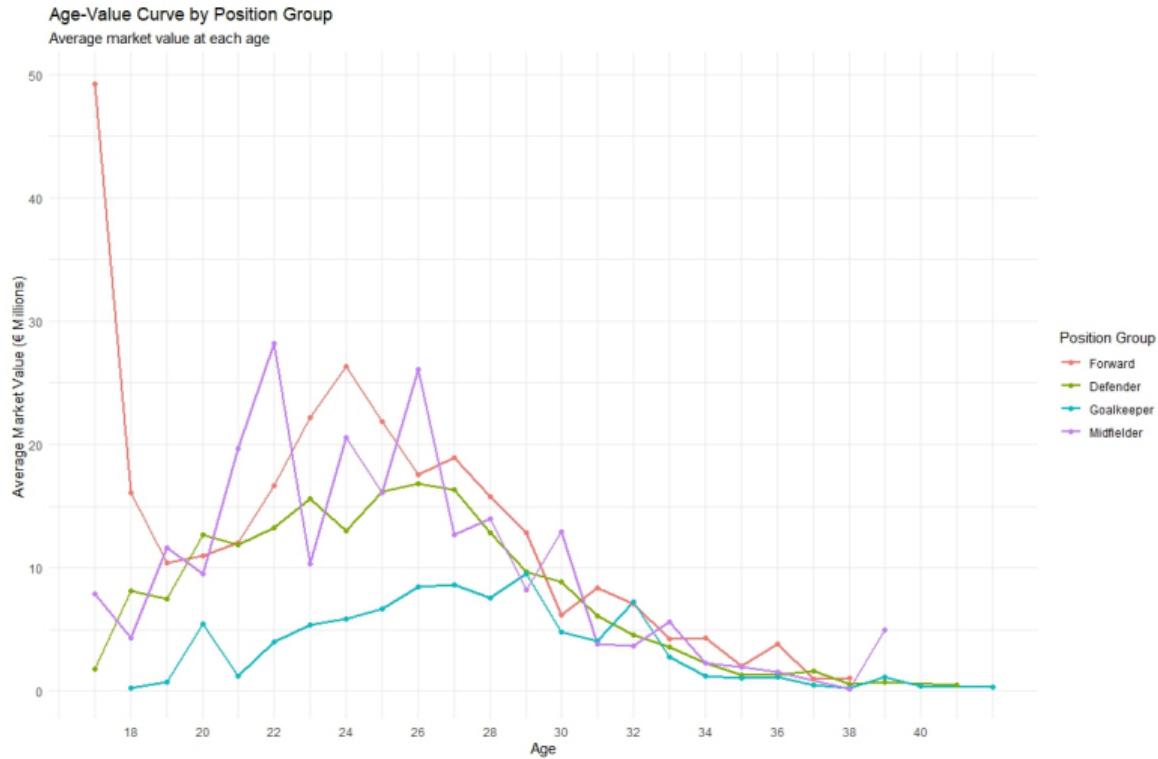
Variable	Forward (N = 593)	Midfielder (N = 475)	Defender (N = 956)	Goalkeeper (N = 266)
Continent (Africa)	0.041	-0.104	0.176*	-0.079
Continent (S. America)	0.425***	0.553***	0.378***	0.132
Age	0.784***	1.103***	0.763***	0.842***
Age ²	-0.016***	-0.022***	-0.015***	-0.015***
Appearances	0.022***	0.025***	0.024***	0.043***
Goal Contributions	0.023***	0.014**	0.009	-
League (Bundesliga)	-1.004***	-0.800***	-1.024***	-0.903***
League (La Liga)	-0.785***	-0.609***	-0.834***	-0.210
League (Ligue 1)	-0.697***	-0.349*	-0.890***	-0.491*
Club Tier (2)	-0.345**	-0.451***	-0.652***	-0.519*
Club Tier (3)	-0.750***	-0.989***	-1.133***	-0.962***
Club Tier (4)	-0.922***	-1.110***	-1.227***	-0.866***
Adjusted R ²	0.677	0.621	0.624	0.636

Significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Reference: Europe (Continent), Premier League (League), Tier 1 Elite Clubs (Club Tier)

Interpretation: Percentage change in market value = $(e^\beta - 1) \times 100\%$
where e is the base of the natural logarithm (approx. 2.718) and β is the coefficient

Age-Value Relationship



Peak Market Value Age by Position

Peak Age (Years)

Position	Peak Age
Forwards	24.7
Midfielders	25.5
Defenders	24.7
Goalkeepers	28.3

Calculated as: $\text{PeakAge} = -\frac{\beta_{Age}}{2 \times \beta_{Age^2}}$

Conclusion

Conclusion

- Market value follows a clear age curve across all positions:
 - The positive coefficient for age and negative coefficient for age-squared indicates an inverted U-shaped relationship.
- Premier League commands a substantial premium:
 - Players in all other leagues are valued much lower.
 - Bundesliga shows the biggest discount: $(e^{-1.004} - 1) \times 100\% = -63.4\%$ for forwards
- South American players carry a significant premium:
 - Being from South America increases market value by 45-74% for outfield players
 - For midfielders: $(e^{0.553} - 1) \times 100\% = 73.8\%$
 - For forwards: $(e^{0.425} - 1) \times 100\% = 53.0\%$
 - For defenders: $(e^{0.378} - 1) \times 100\% = 45.9\%$
- Model explanatory power is strong:
 - The R^2 values around 0.62-0.68 indicate these factors collectively explain roughly two-thirds of the variation in player market values.
 - Implications: Despite not using an algorithm, over 60% of Transfermarkt's evaluations can be explained through collectible statistics.

References

References

- Allegretti, F. (2023). transfermarkt-api: API service to get data from Transfermarkt.
- Chalke, R. (2025). Robert lewandowski salary: How much does barcelona star earn per week and annually in laliga?
- Poli, R., Ravenel, L., and Besson, R. (2024). Global economic analysis of the transfer market (2015-2024) - cies football observatory.
- Smith, R. (2021). The wisdom of the crowd. The New York Times. August 12th, 2021.
- Summerscales, R. (2023). Fifa world cup final beat super bowl lvi by more than one billion viewers in tv ratings.