Quantifying the Impact of Performance Statistics and Personal Attributes on Football Attackers' Market Value

Career goals and assists, league strength, playing position significantly influence market value

MOTIVATION/RESEARCH QUESTION

Football player valuations guide decision-making and financial planning for clubs and agents, informing player transfers and contract negotiations. This research explores the effect of various factors on an attacker's value, aiming to answer the question:

How do a football attacker's performance statistics and personal attributes influence their market value?

DATA COLLECTION

The analysis dataset, consisting of multiple CSV files with information on competitions, games, clubs, and players, is published on Kaggle (Cariboo, 2024). The data was collected by scraping the website Transfermarkt, which contains historical football statistics, player valuation, and attributes. We fit a model with variables hypothesized to affect player value to answer our research question.

#	Player	Age	Nat.	Club	Market value
1	Erling Haaland Centre-Forward	24			€200.00m
2	Vinicius Junior Left Winger	24			€200.00m
3	Kylian Mbappé Centre-Forward	25			€180.00m
4	Lamine Yamal Right Winger	17			€150.00m

REFERENCES

- Cariboo, D. (2024). *transfermarkt-scraper* [Code repository]. GitHub. https://github.com/dcaribou/transfermarkt-scraper
- Cariboo, D. (2024). *Football Data from Transfermarkt*. Kaggle. https://www.kaggle.com/datasets/davidcariboo/player-scores/data
- Transfermarkt. (n.d.). *Transfermarkt*. https://www.transfermarkt.com

METHODS

- Fit a linear regression model with a dataset with 1,582 of the most valuable football attackers that play in Europe.
- Initially, the response variable is the market value, with 9 dataset variables as predictors.
- Checked conditional mean response and predictors. Evaluated residual plots for linearity, constant variance, and uncorrelated errors. Assessed multicollinearity with VIF and normality with a Q-Q plot.
- Applied transformations to address skewness and non-constant variance in residual plots: Used log transformation on the response variable and season goals to stabilize variance; Box-Cox transformations on season minutes and league coefficient to stabilize variance and linearize relationships.
- Conducted ANOVA and t-tests to identify significant predictors. Excluded insignificant predictor season assists and validated the reduced model with a partial F-test.

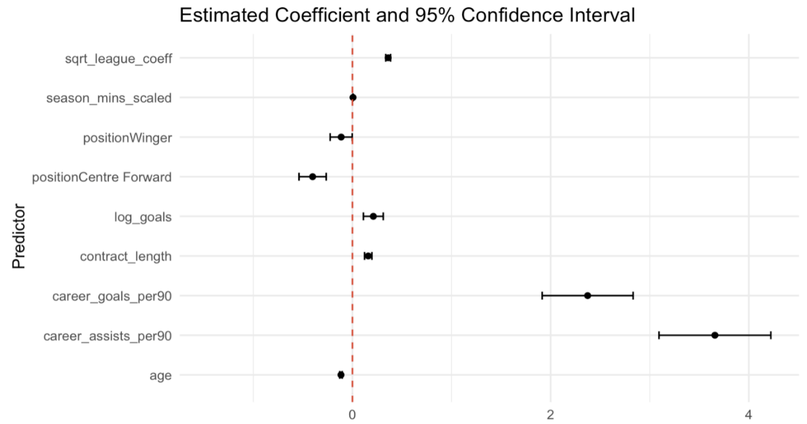
$$\log(\text{MarketValue}) = \beta_0 + \beta_1 \cdot \text{Age} + \beta_2 \cdot \text{Position (Winger)} + \beta_3 \cdot \text{Position (Centre Forward)}$$
$$+\beta_4 \cdot \text{Contract Length} + \beta_5 \cdot \sqrt{\text{League Strength}} + \beta_6 \cdot \log(\text{Goals Scored})$$
$$+\beta_7 \cdot \text{Minutes Played}^{2/3} + \beta_8 \cdot \text{Career Goals}/90 + \beta_9 \cdot \text{Career Assists}/90 + \epsilon$$

- Constructed confidence intervals for coefficients and mean response to quantify precision. Built prediction intervals to assess the accuracy of individual predictions.

CONCLUSIONS

- **Career statistics** (goals and assists per 90 minutes) have the most significant impact on market value, emphasizing the importance of consistent long-term performance.
- **Current performance** has less effect: log of season goals impacts value, but minutes played have minimal influence, and assists were insignificant; likely due to sample size and team dependency.
- **Player attributes** show that attacking midfielders are more valuable than wingers and centre forwards with equivalent attributes; as well, value increases in stronger leagues and decreases with age and shorter contracts.

RESULTS



Confidence intervals reflect the precision of our model coefficient estimates.

Coefficient	Value
β_0: Intercept (represents when all predictors are 0 or baseline)	-1.030
β_1: Age (in years)	-0.115
β_2: Position (Winger) (compared to baseline position, Att. Midfielder)	-0.114
β_3: Position (Centre Forward) (compared to baseline position)	-0.402
β_4: Contract length (years until expiry)	0.159
β_5: Square root of league strength coefficient	0.361
β_6: Log of goals scored in the current season	0.211
β_7: Minutes played in the current season to the power of 2/3	0.006
β_8: Career goals scored per 90 minutes	2.373
β_9: Career assists per 90 minutes	3.658

Coefficient value: the effect of each predictor on the response variable, which is the log of market value in millions. Example interpretation: a 0.1-unit increase in career goals per 90 corresponds to a 0.237 increase in log market value (97% increase in market value): a player with 1 more goal every 10 90-minute career matches becomes twice as valuable.

LIMITATIONS

- 27% of variance in market value is unexplained; factors like wages, injury history, and economic conditions could be explored further.
- Market value data from Transfermarkt may not fully reflect player worth, as player valuations vary by club due to tactical fit, familiarity, and salary demands.
- Broadly grouping positions oversimplifies tactical differences; future research with granular classifications could provide greater insight.