# Predicting Soccer Player Market Value

Dav King, Luke Thomas, Thomas Barker, Harry Liu

# Introduction and Motivation

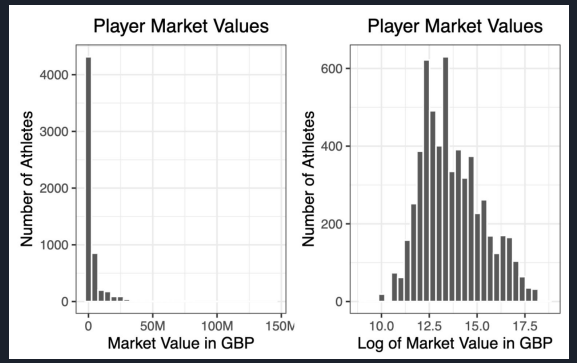
- Tens of thousands of soccer players are transferred on the international transfer market each year
- Players often cost clubs millions of dollars and they play an integral role in team success
- Identifying undervalued or overvalued players is desirable from a management perspective

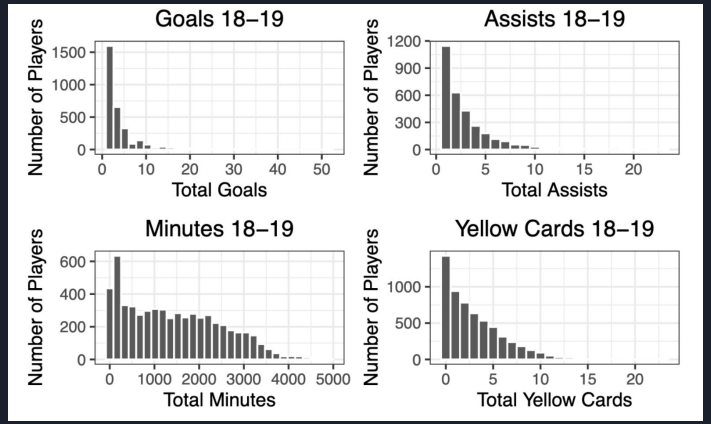
# Data Introduction

- Data downloaded from Kaggle
- Original data collected from TransferMarkt
- Combined three datasets
- Predictors from on-field performance include each player's total goals, assists, minutes played, and yellow cards during the 2018-19 season
- Other relevant predictors include each player's position, nationality, age, and height
- Response variable is market value (GBP) at the end of the 2018-19 season

# Exploratory Data Analysis

- Moderate correlations between minutes played, goals, and assists
  - Consider interaction effects
- On-field performance predictors very right-skewed, with most observations by 0
  - Consider log-transform
- Player market value also very right skewed
  - Consider log-transform

# Final Model

$$\hat{final\_value} = e^{13.7} \times e^{0.00049 \times age\_in\_days} \times e^{0.0602 \times total\_goals} \times e^{0.0671 \times total\_assists} \times e^{0.00062 \times total\_minutes} \times e^{-0.00787 \times total\_yellow} \times e^{0.00988 \times height\_in\_cm} \times e^{0.35308 \times citizenship\_France} \times e^{0.45405 \times citizenship\_Spain} \times e^{-0.21465 \times citizenship\_Other} \times e^{-0.02893 \times position\_Defender} \times e^{0.00323 \times position\_Goalkeeper} \times e^{0.11777 \times position\_Midfield}$$

| term | estimate | std.error | statistic | p.value | conf.low | conf.high |
|---|---|---|---|---|---|---|
| (Intercept) | 13.72449 | 0.01826 | 751.58388 | 0.00000 | 13.68869 | 13.76029 |
| age_in_days | -0.00049 | 0.00001 | -36.45455 | 0.00000 | -0.00051 | -0.00046 |
| total_goals | 0.06020 | 0.00689 | 8.73898 | 0.00000 | 0.04670 | 0.07371 |
| total_assists | 0.06710 | 0.01012 | 6.62931 | 0.00000 | 0.04726 | 0.08695 |
| total_minutes | 0.00062 | 0.00003 | 21.94142 | 0.00000 | 0.00056 | 0.00067 |
| total_yellow | -0.00787 | 0.00833 | -0.94541 | 0.34450 | -0.02419 | 0.00845 |
| height_in_cm | 0.00988 | 0.00134 | 7.34644 | 0.00000 | 0.00724 | 0.01251 |
| country_of_citizenship_France | 0.35308 | 0.10759 | 3.28187 | 0.00104 | 0.14216 | 0.56400 |
| country_of_citizenship_Spain | 0.45405 | 0.10250 | 4.42998 | 0.00001 | 0.25311 | 0.65500 |
| country_of_citizenship_other | -0.21465 | 0.07754 | -2.76832 | 0.00566 | -0.36666 | -0.06264 |
| position_Defender | -0.02893 | 0.05151 | -0.56164 | 0.57439 | -0.12991 | 0.07205 |
| position_Goalkeeper | 0.00330 | 0.08308 | 0.03971 | 0.96832 | -0.15958 | 0.16618 |
| position_Midfield | 0.11777 | 0.05404 | 2.17930 | 0.02936 | 0.01182 | 0.22372 |

# Interesting Findings

- All effects except for total yellow cards, being a defender, or being a goalkeeper were significant
- Being younger, taller, and responsible for more goals, assists, and minutes all predicted an increase in a player's market value
- Model performed well on testing data - $R^2$ = 0.391 (0.411 on training data) and RMSE = 1.264 (1.213 on training data)

# Conclusions and Future Work

- The ability to predict a player's market value is extremely useful for soccer clubs
- Significant effects:
  - Minutes played
  - Total goals
  - Total assists
  - Age
  - Height
  - Midfielder
  - Nation of citizenship (Brazil, France, Spain)
- Limitations:
  - Skewed explanatory variables
  - Statistical assumptions
- Future work:
  - Multiple seasons
  - More predictors

# Citation

Data Source:

Cariboo, David. *Football Data from Transfermarkt.* (data file). Kaggle, 2022. Web. 03 Nov 2022. https://www.kaggle.com/datasets/davidcariboo/player-scores