Total Team Spending and Profitability Results in a Higher League Ranking in the English Premier League*

An Analysis of Economic Predictors on League Rank in the 2023-2024 Season

Isabella MacLean

April 19, 2024

This project investigates how to formulate a linear regression model based on three explanatory variables: match attendance, total payroll, and market value, in order to try and predict a teams finishing place in the standings. Utilizing R's lm function, this paper analyzes the relationship between these variables and teams total points accumulated during the English Premier League 2023-2024 season. Findings suggest that each explanatory variable significantly influences a team's league position (total points), highlighting the association between team success and team wealth.

Table of contents

*Code and data are available at: https://github.com/Bellamaclean7/English_Premier_League_Economic_Predictors_on_League_Rank_2023-2024_Season.git.

1 Introduction

You can and should cross-reference sections and sub-sections. We use R Core Team (2023) and Wickham et al. (2019).

The remainder of this paper is structured as follows. Section 2….

You can and should cross-reference sections and sub-sections. We use R Core Team (2023) and Wickham et al. (2019).

The remainder of this paper is structured as follows. Section 2….

You can and should cross-reference sections and sub-sections. We use R Core Team (2023) and Wickham et al. (2019).

The remainder of this paper is structured as follows. Section 2….

You can and should cross-reference sections and sub-sections. We use R Core Team (2023) and Wickham et al. (2019).

The remainder of this paper is structured as follows. Section 2….

2 Data

The analysis dataset used for this paper is comprised of five datasets that were compiled during the data cleaning process. The first dataset, Table 1, contains every teams average home matchday attendance for the 2023-2024 season, which was collected from Football Web Pages (2023). The second dataset, Table 2, contains every teams current market value, which

was collected from Transfermarkt (2023a). The third dataset, Table 3, contains the current total payroll for each team in the English Premier League, and was collected from Spotrac (2023a). The forth dataset, Table 4, contains the total amount spent in transfer fees prior to the start of the 2023-2024 season for each team in the English Premier League, and was collected from Spotrac (2023b). The final dataset, Table 5, contains the total points earned by each team in the English Premier League during the 2023-2024 as of Matchday 30, and was retrieved from Transfermarkt (2023b).

Table 1: The dataset containing average home matchday attendance for each team in the English Premier League during the 2023-2024 season.

Team	Home Matchday Attendance
Manchester City F.C.	53,194
Manchester United F.C.	73,523
Arsenal F.C.	60,213
Chelsea F.C.	39,626
Liverpool F.C.	54,672
Aston Villa F.C.	41,783
Tottenham Hotspur F.C.	61,524
West Ham United F.C.	62,463
Newcastle United F.C.	52,158
Everton F.C.	39,063
Crystal Palace	24,797
Nottingham Forest F.C.	29,356
Fulham F.C.	24,290
Brighton & Hove Albion	31,517
AFC Bournemouth	11,085
Brentford F.C.	17,081
Burnley F.C.	21,168
Wolverhampton Wanderers F.C.	31,291
Sheffield United F.C.	30,269
Luton Town F.C.	11,113

Table 2: The dataset containing the current market value of each team in the English Premier League.

Team	Market Value
Manchester City F.C.	£1.27B
Manchester United F.C.	£734M
Arsenal F.C.	£1.12B
Chelsea F.C.	£928M
Liverpool F.C.	£921M
Aston Villa F.C.	£646M
Tottenham Hotspur F.C.	£777M
West Ham United F.C.	£447M
Newcastle United F.C.	£638M
Everton F.C.	£345M
Crystal Palace	£405M
Nottingham Forest F.C.	£370M

Fulham F.C.	£338M
Brighton & Hove Albion	£505M
AFC Bournemouth	£353M
Brentford F.C.	£426M
Burnley F.C.	£265M
Wolverhampton Wanderers F.C.	£340M
Sheffield United F.C.	£144M
Luton Town F.C.	£125M

Table 3: The dataset containing the current total payroll for each team in the English Premier League.

Team	Active Players	Forwards	Midfielders	Defensmen	Goalkeepers	Total
Manchester City F.C.	24	£27,300,000	£91,052,000	£63,440,000	£9,620,000	£191,412,000
Manchester United F.C.	26	£47,780,000	£58,500,000	£63,955,000	£10,400,000	£180,635,000
Arsenal F.C.	25	£43,160,000	£63,460,000	£48,306,000	£11,180,000	£166,106,000
Chelsea F.C.	28	£23,764,000	£59,540,000	£59,800,000	£6,240,000	£149,344,000
Liverpool F.C.	24	£41,912,000	£39,780,000	£42,380,000	£11,440,000	£135,512,000
Aston Villa F.C.	26	£25,220,000	£32,710,000	£40,200,000	£14,040,000	£112,170,000
Tottenham Hotspur F.C.	26	£34,060,000	£36,240,000	£24,270,000	£8,580,000	£103,150,000
West Ham United F.C.	23	£12,948,000	£42,900,000	£29,120,000	£9,828,000	£94,796,000
Newcastle United F.C.	28	£16,180,000	£37,180,000	£28,184,000	£6,240,000	£87,784,000
Everton F.C.	24	£18,486,000	£28,715,000	£23,552,000	£7,280,000	£78,033,000
Crystal Palace	29	£9,100,000	£26,780,000	£28,860,000	£8,080,000	£73,210,000
Nottingham Forest F.C.	28	£18,850,000	£16,120,000	£21,840,000	£7,020,000	£63,830,000
Fulham F.C.	24	£15,210,000	£21,476,000	£19,970,000	£5,980,000	£62,636,000
Brighton & Hove Albion	29	£22,230,000	£16,770,000	£18,000,000	£3,380,000	£60,380,000
AFC Bournemouth	24	£10,400,000	£10,998,000	£12,740,000	£4,004,000	£43,342,000
Brentford F.C.	30	£8,840,000	£13,260,000	£13,260,000	£3,900,000	£39,260,000
Burnley F.C.	31	£9,308,000	£12,480,000	£10,192,000	£3,120,000	£36,140,000
Wolverhampton Wanderers F.C.	20	£11,440,000	£8,840,000	£13,000,000	£2,270,000	£35,550,000
Sheffield United F.C.	28	£7,358,000	£8,190,000	£11,518,000	£4,290,000	£31,876,000
Luton Town F.C.	27	£6,110,000	£9,360,000	£4,290,000	£2,990,000	£23,940,000

Table 4: The dataset containing the total amount spent in transfer fees prior to the start of the 2023-2024 season for each team in the English Premier League.

Team	Total Transfer Fees
Manchester City F.C.	£151,100,000.00
Manchester United F.C.	£121,700,000.00
Arsenal F.C.	£226,600,000.00
Chelsea F.C.	£449,100,000.00
Liverpool F.C.	£172,000,000.00
Aston Villa F.C.	£97,400,000.00
Tottenham Hotspur F.C.	£231,300,000.00
West Ham United F.C.	£135,800,000.00
Newcastle United F.C.	£145,200,000.00
Everton F.C.	£37,500,000.00
Crystal Palace	£67,800,000.00

Nottingham Forest F.C.	£93,870,000.00
Fulham F.C.	£298,900,000.00
Brighton & Hove Albion	£107,350,000.00
AFC Bournemouth	£87,070,000.00
Brentford F.C.	£67,850,000.00
Burnley F.C.	£107,050,000.00
Wolverhampton Wanderers F.C.	£44,000,000.00
Sheffield United F.C.	£41,850,000.00
Luton Town F.C.	£7,600,000.00

Table 5: The dataset containing the total points earned by each team in the English Premier League during the 2023-2024 as of Matchday 30.

Team	Points
Manchester City F.C.	73
Arsenal F.C.	71
Liverpool F.C.	71
Aston Villa F.C.	63
Tottenham Hotspur F.C.	60
Newcastle United F.C.	50
Manchester United F.C.	50
West Ham United F.C.	48
Chelsea F.C.	47
Brighton & Hove Albion	44
Wolverhampton Wanderers F.C.	43
Fulham F.C.	42
AFC Bournemouth	42
Crystal Palace	33
Brentford F.C.	32
Everton F.C.	27
Nottingham Forest F.C.	26
Luton Town F.C.	25
Burnley F.C.	20
Sheffield United F.C.	16

The datasets were then combined during the cleaning process to create one master dataset for analysis. The merging process was conducted on the common column `team`, ensuring that data from each dataset corresponding to a specific team was consolidated into a single row within the analysis dataset. This methodology allowed for a unified view of each team's `matchday attendance`, `market value`, `total team points`, `total payroll`, and `total transfer spending`, facilitating a multifaceted analysis of team performance and financial metrics.

As part of the data cleaning process, specific columns containing numeric values but stored as character strings—due to the inclusion of commas as thousands separators and currency symbols—were converted to their appropriate numeric data types. This conversion involved stripping the character columns 'average_home_matchday_attendance', 'total_wage_bill',

and 'transfer_fees' of non-numeric characters (i.e., commas and the pound sterling symbol) and then casting them to integer or numeric types as contextually appropriate, as well as applying the same scaling methods to each variable. This step was essential for enabling quantitative analysis of these variables.

Furthermore, to streamline the dataset and focus the analysis on key variables of interest, certain columns deemed extraneous to the core analytical objectives were removed. These included detailed payroll information per position (`forwards`, `midfielders`, `defensmen`, `goalkeepers`) and the column `active.players`. The removal of these columns served to reduce the dataset's complexity, facilitating a more focused and manageable analysis of the relationships between team performance metrics and financial expenditures.

The resulting dataset for analysis is shown in Table 6.

Table 6: The dataset containing the total points earned by each team in the English Premier League during the 2023-2024 as of Matchday 30.

Team	Home Matchday Attendance	Market Value	Points	Total Payroll	Total Transfer Fees
Manchester City F.C.	53194	1.27000	73	0.191412	0.15110
Manchester United F.C.	73523	0.73425	50	0.180635	0.12170
Arsenal F.C.	60213	1.12000	71	0.166106	0.22660
Chelsea F.C.	39626	0.92830	47	0.149344	0.44910
Liverpool F.C.	54672	0.92140	71	0.135512	0.17200
Aston Villa F.C.	41783	0.64620	63	0.112170	0.09740
Tottenham Hotspur F.C.	61524	0.77730	60	0.103150	0.23130
West Ham United F.C.	62463	0.44660	48	0.094796	0.13580
Newcastle United F.C.	52158	0.63770	50	0.087784	0.14520
Everton F.C.	39063	0.34540	27	0.078033	0.03750
Crystal Palace	24797	0.40470	33	0.073210	0.06780
Nottingham Forest F.C.	29356	0.36965	26	0.063830	0.09387
Fulham F.C.	24290	0.33800	42	0.062636	0.29890
Brighton & Hove Albion	31517	0.50510	44	0.060380	0.10735
AFC Bournemouth	11085	0.35280	42	0.043342	0.08707
Brentford F.C.	17081	0.42608	32	0.039260	0.06785
Burnley F.C.	21168	0.26510	20	0.036140	0.10705
Wolverhampton Wanderers F.C.	31291	0.33970	43	0.035550	0.04400
Sheffield United F.C.	30269	0.14375	16	0.031876	0.04185
Luton Town F.C.	11113	0.12510	25	0.023940	0.00760

The analysis dataset has 6 columns:

1. The `Team` column, which denotes the name of each football club in the English Premier League.
2. The `Home Matchday Attendance` column, which denotes the average number of spectators attending home matches of the football club. This helps understand the club's popularity, fan base size, and infrastructure.
3. The `Market Value` column, which represents the total market value of the team in billions of pounds. It indicates the financial strength and quality of the club.

4. The `Points` column, which represents the total points a team has accumulated over the season in the league. Points are awarded based on wins and draws.
5. The `Total Payroll` column, which details the total annual wages paid by the club to its players, given in pounds. It reflects the financial commitment of the club towards its playing squad.
6. The `Total Transfer Fees` column, which denotes the total amount spent by the club on transfer fees in pounds during the transfer window for the given year. It indicates the club's investment in new players for the season.

From here I began examining the relationships between English Premier League teams' league performance (expressed in points accumulated) and various financial indicators during the 2023-2024 season. The following four graphs (Figure 1, Figure 2, Figure 3, and Figure 4) presents a scatter plot paired with a linear regression trend line to evaluate the correlation between league points and one of the four key economic metrics presented in Table 6.

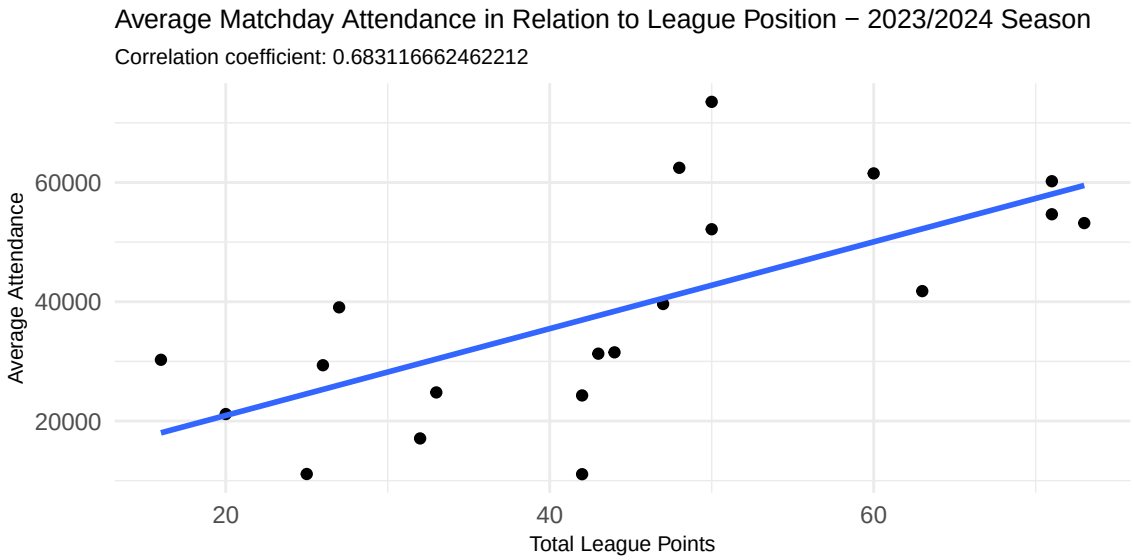

Figure 1: Scatter plot illustrating the relationship between total league points and average matchday attendance during the 2023-2024 season.

In Figure 1, Figure 2, Figure 3, and Figure 4, each point point on the graph represents a team, plotted with their total points on the x-axis and their average home attendance on the y-axis. The trend line in Figure 1, added using linear regression analysis, suggests a correlation between team performance and matchday turnout. The correlation coefficient of 0.683116662462212 indicates a positive association with moderate correlation, where teams with higher league points tend to have greater average attendance. Figure 2 linear regression trend line suggests a correlation between a team's financial expenditure on payroll and their success in the league. A correlation coefficient of 0.788682340622421 reflects a positive relationship, implying that teams with higher payrolls may be more likely to accumulate greater

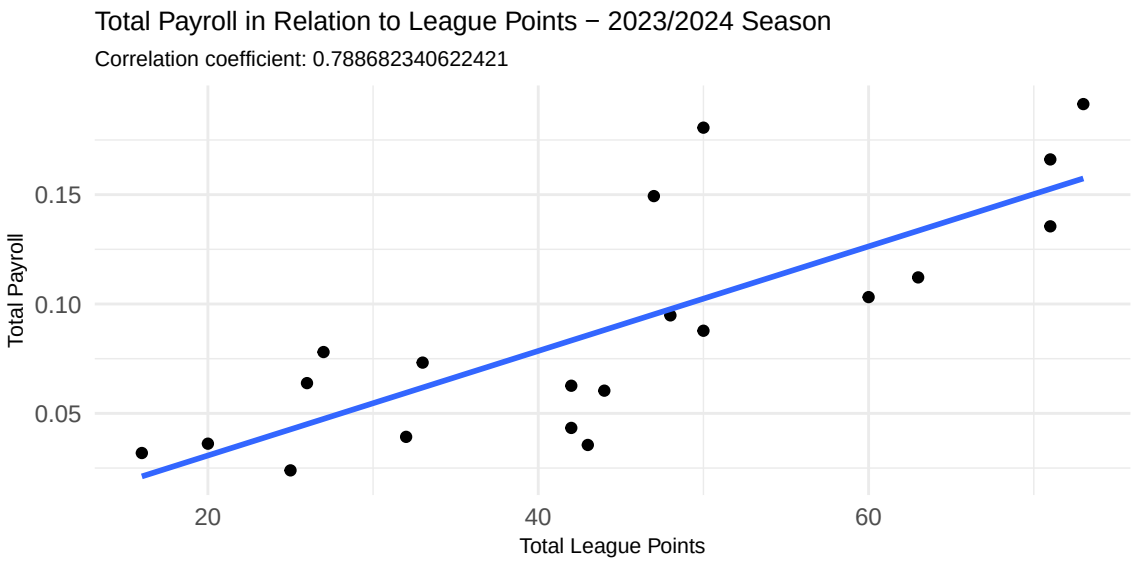

Figure 2: Scatter plot depicting the association between total league points and the total payroll for teams in the 2023-2024 season.

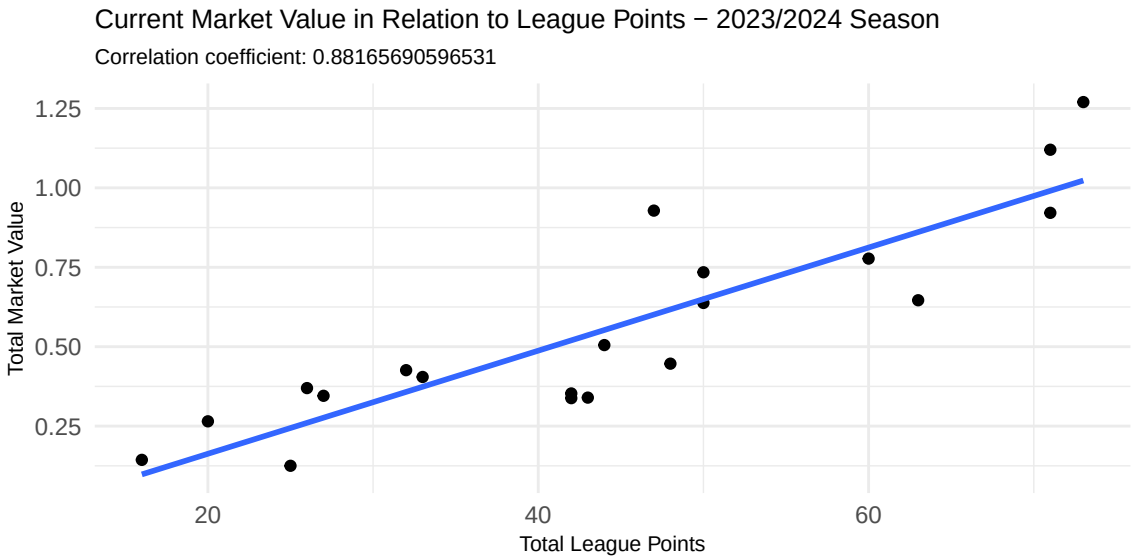

Figure 3: Scatter plot demonstrating the correlation between the current market value of teams and their accumulated league points for the 2023-2024 season.

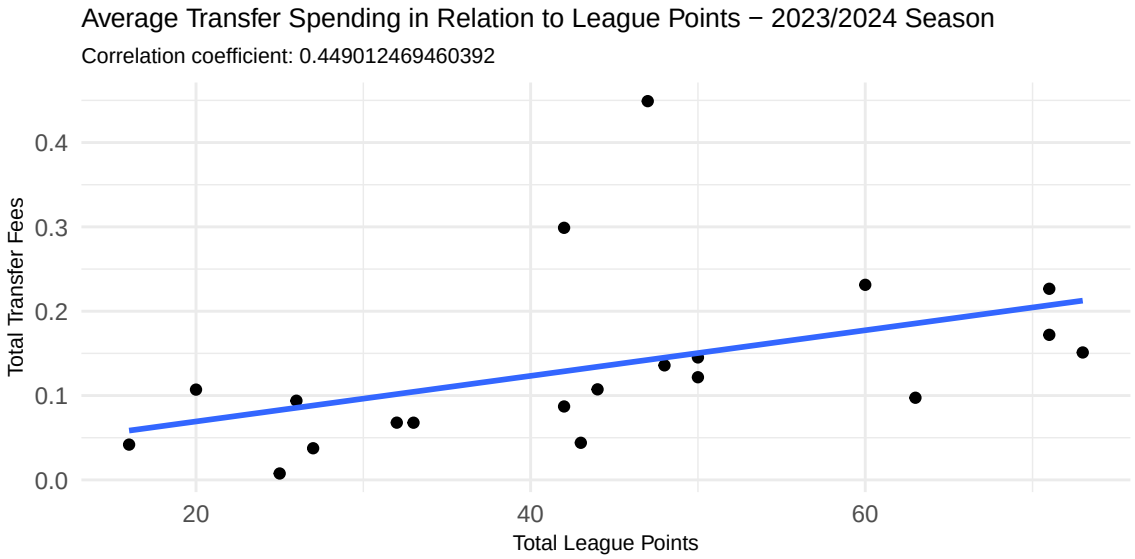

Figure 4: Scatter plot illustrating the relationship between teams total transfer spending and their corresponding league points for the 2023-2024 season.

league points. Figure 3 demonstrates a pronounced upward trend represented by the linear regression line, with a correlation coefficient of 0.88165690596531, indicative of a strong positive relationship. This suggests that teams with higher market values tend to amass more league points, underlining a potential link between financial strength and on-field success. Figure 4 trend line indicates a moderate positive correlation, with a correlation coefficient of approximately 0.449012469460392. This suggests that higher league points tend to coincide with increased transfer spending, although the relationship is not strongly linear.

When running simple linear regression with `Points` ~ `Each Dependent Variable`, you're looking at the relationship between each variable in isolation with `Points`. The positive trend and strong correlation coefficients found in Figure 1, Figure 2, Figure 3 when analyzed in isolation with `Points`, suggests that as `Market Value`, `Total Payroll`, and `Average Matchday Attendance` of a team increases (independently from each other), so do the points accumulated.

3 Model

3.1 Linear Regression

A linear regression model is a statistical model for linear relationships between variables given by

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + ... + \beta_p X_{i,p} + \varepsilon_i \quad \text{for } i = 1, ..., n. \tag{1}$$

As shown in equation (1) a linear regression model has the following components:

- Y_i: The dependent variable for the i^{th} observation. This is the response variable that is trying to be predicted or explained, dependent on the corresponding explanatory variables $X_{i,1}, X_{i,2}, ..., X_{i,p}$.
- β_0: The intercept of the regression line. It represents the expected value of Y_i when all the independent variables (X) are equal to 0.
- $\beta_1, \beta_2, ..., \beta_p$: The coefficients of the model. Each β_j (for $j = 1, ..., p$) represents the expected change in Y_i for a one-unit change in the j^{th} independent variable, $X_{i,j}$, holding all other variables constant.
- $X_{i,1}, X_{i,2}, ..., X_{i,p}$: The independent variables (also called predictors or explanatory variables) for the i^{th} observation. These variables are used to predict the value of the dependent variable.
- ε_i: The error term for the i^{th} observation. It represents the difference between the observed value of the dependent variable and the value predicted by the model. It's assumed to be randomly distributed with a mean of 0.
- $i = 1, ..., n$: This indicates that the equation applies to each observation in the dataset, from the first ($i = 1$) to the n^{th} (the last observation), where n is the total number of observations.

3.2 Model set-up

$$\begin{aligned} \text{pts}_i = \beta_0 &+ \beta_1 \times \text{average\_home\_matchday\_attendance}_i \\ &+ \beta_2 \times \text{total\_wage\_bill}_i + \beta_3 \times \text{market\_value}_i + \varepsilon_i \end{aligned} \tag{2}$$

Equation (2) can be explained as follows:

- pts_i: The dependent variable for the i^{th} observation. This is the outcome variable that the model is trying to predict or explain, which in this context could represent points (or any other metric of success) associated with each observation.
- $\beta_1, \beta_2, \beta_3$: The coefficients of the model. Each of these coefficients represents the expected change in pts_i for a one-unit increase in their respective independent variable, assuming all other variables are held constant.
  - β_1 is associated with $\text{average\_home\_matchday\_attendance}_i$, indicating how changes in home matchday attendance are expected to affect pts_i.

Table 7: Summary of Linear Model Coefficients

[H]

	Estimate	StdError	tValue	Pr
(Intercept)	14.0159614	4.6112152	3.039537	0.0078076
average_home_matchday_attendance	0.0002593	0.0001778	1.458001	0.1641895
total_wage_bill	-139.8905906	113.2953089	-1.234743	0.2347529
market_value	58.6176203	15.3445374	3.820097	0.0015074

- β_2 corresponds to total_wage_bill_i, reflecting the impact of the total wage bill on pts_i.
- β_3 is linked with market_value_i, showing how the market value is predicted to influence pts_i.

3.3 Model justification

The linear regression model given by Equation (2), derived from the foundational principles of linear regression outlined in Equation (1), provides a transparent mechanism to quantify the impact of various factors on the total points accumulated by EPL teams (a direct correlation to team rank in league standings). The model will aid in gaining insights into how attendance, total payroll, and market value could affect performance metrics. The model is operationalized using the `lm` function in `R`. A strong positive relationship is expected between the dependent variable (`Points`) and three independent variables (`Home Matchday Attendance`, `Total Payroll`, and `Market Value`). The higher the three independent variables are, the more points there should be accumulated by the team.

3.4 Model Prediction

A positive relationship is hypothesized between the dependent variable (`Points`) and the three independent variables (`Home Matchday Attendance`, `Total Payroll`, and `Market Value`). The underlying assumption is that higher values for these independent variables should correlate with an increased number of points earned by a team. Such a correlation would suggest that greater fan engagement (as reflected in matchday attendance), higher investment in player talent (as indicated by the payroll), and a more substantial market presence (denoted by market value) are all strategic levers that could lead to better performance in the league.

4 Results

Table 7 shows the table of the coefficients after running the linear model summary.

Note: Residuals: Min 1Q Median 3Q Max -10.8121 -6.0644 -0.9837 5.1376 15.9644

Residual standard error: 8.261 on 16 degrees of freedom Multiple R-squared: 0.8057, Adjusted R-squared: 0.7693 F-statistic: 22.12 on 3 and 16 DF, p-value: 6.159e-06

5 Discussion

5.1 First discussion point

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

B.2 Diagnostics

References

Football Web Pages. 2023. "Premier League Attendances 2022-2023." https://www.footballwebpages.co.uk/premier-league/attendances/2022-2023.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Spotrac. 2023a. "English Premier League Payroll." https://www.spotrac.com/epl/payroll/.

———. 2023b. "English Premier League Transfer Fees." https://www.spotrac.com/epl/payroll/.

Transfermarkt. 2023a. "Premier League - Market Values." https://www.transfermarkt.us/premier-league/daten/wettbewerb/GB1.

———. 2023b. "Premier League Rankings." https://www.transfermarkt.us/premier-league/spieltag/wettbewerb/GB1/plus/?saison_id=2023&spieltag=30.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.