

# Analyzing EPL Spending and Performance

## Abstract

This project aims to analyze the relationship between financial decisions and team performance in the English Premier League (EPL). Specifically, I investigated how spending habits, squad characteristics, and transfer activity influence a team's total number of wins over the past six seasons. While some clubs consistently invest large sums in players and infrastructure, others operate with more modest budgets. The objective is to determine whether greater spending directly correlates with on-field success, and to identify clubs that have overperformed or underperformed relative to their investment levels.

I first compiled and merged two main datasets: one containing seasonal team performance statistics such as wins, final league position, net spend, and transfer income from 2019–2025; and another providing data on squad attributes like average age and total players signed. To ensure consistency, I only included clubs that appeared in at least five of the six seasons to eliminate teams frequently relegated or promoted.

Key transformations included summing total wins and average squad characteristics over the six-season span, and calculating a new metric called *Relative Spending*, which measures each club's net spend in a given season divided by the league's average net spend. These preparations allowed us to create a unified modeling-ready dataset that fairly compared teams regardless of market size. In addition, I also created a new variable called *Spending Efficiency*, calculated by dividing each club's average relative spending by its total squad value. This metric helped evaluate which teams achieved more wins per financial input, enabling a fairer assessment across clubs of varying wealth.

# 1 Clustering Analysis

To better understand how financial strategies translate into on-field success, I applied **K-means clustering** to identify distinct groups of EPL teams based on their performance metrics and financial behavior. Clustering allowed us to explore patterns in how spending and squad composition relate to long-term winning outcomes. I explored three clustering scenarios:

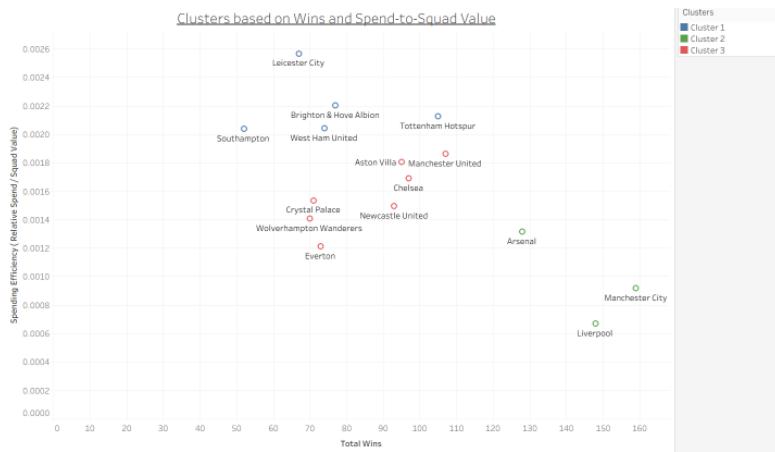


Figure 1: **Cluster 1 – ”Efficient” Clubs.** This cluster includes teams like Leicester City, Brighton, and West Ham who achieved competitive win totals despite modest spending. These clubs demonstrate strong *spending efficiency*, outperforming expectations relative to their financial investment and squad value.

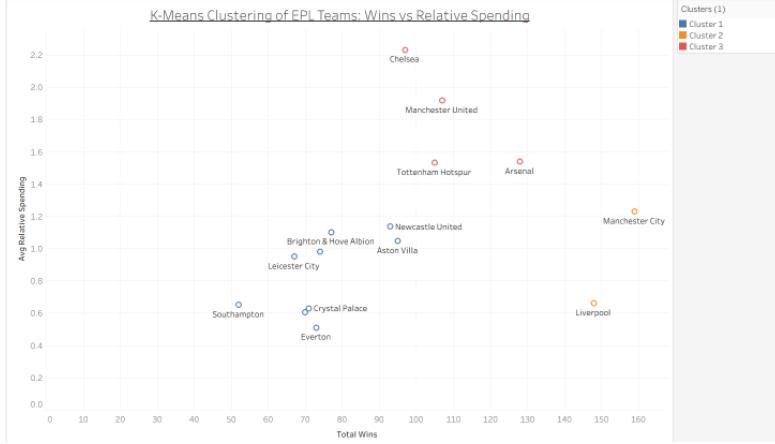


Figure 2: **Cluster 2 – ”Over-Investing” Clubs.** This group features clubs like Chelsea, Manchester United, and Arsenal, who achieved high win totals but also had the highest relative spending levels. While successful on the pitch, these clubs may not be as financially efficient as others, relying on heavy investment to achieve results.

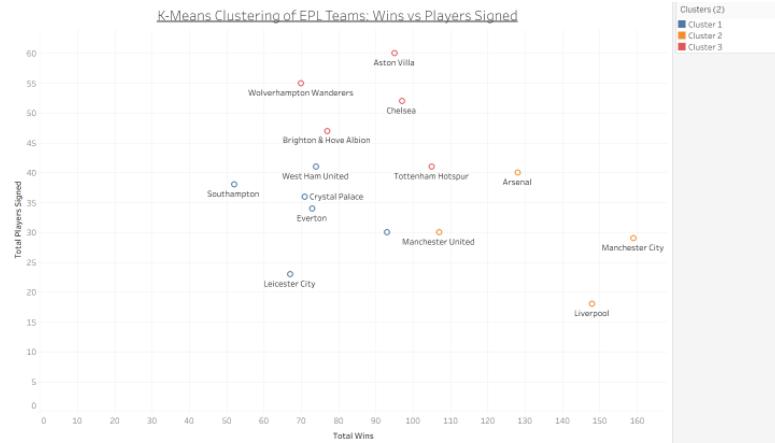


Figure 3: **Cluster 3 – ”High Turnover, Low Efficiency” Teams.** This cluster highlights clubs such as Chelsea, Aston Villa, and Wolves who have signed a large number of players but haven’t achieved significantly higher win totals. This suggests a potential lack of stability or inefficiency in transfer strategies.

## 2 Modeling

To evaluate how financial and squad-related variables influence team performance in the English Premier League, we built three Poisson regression models. Poisson regression was selected since the target variable, *Total Wins*, represents count data (non-negative integers), making it well-suited for modeling via a Poisson distribution.

The first model used *Average Relative Spending* as the sole predictor. This variable was statistically significant at the 0.005 level ( $p < 0.001$ ), indicating that teams spending more relative to the league average tend to win more games. The model yielded a residual deviance of 3957.1 and an AIC of 7254.8.

The second model included both *Average Relative Spending* and *Average Squad Age*. Both predictors were statistically significant ( $p < 0.005$ ). Spending continued to show a positive association with wins, while average age exhibited a negative effect, suggesting older squads win slightly fewer games. The residual deviance dropped to 3894.3 and AIC improved to 7194.

The final model added a third predictor: *Total Players Signed*. All three variables were statistically significant ( $p < 0.001$ ). Notably, the number of players signed had a negative coefficient (Estimate = -0.0128), implying that frequent roster turnover might harm team consistency and performance. This model had the best fit, with a residual deviance of 2927.2 and AIC of 6228.9, suggesting it best explains variation in total wins.

## 3 Conclusion

My analysis demonstrated that *Average Relative Spending* was a significant and consistent predictor of team success in the English Premier League. Across all Poisson regression models, higher spending relative to the league average was strongly associated with increased total wins, underscoring the importance of strategic financial investment.

Clustering and modeling also revealed that several lower-spending teams (e.g., Brighton, Brentford) outperformed their financial peers. These findings highlight that smart squad management and efficient spending can help clubs compete with more resource-rich teams.