Final Report : Impact of Euroleague Statistics on Season Standings

Harun Can Yurdagül - 32092 DSA210 - Fall 2025 Term Project

---

1. Motivation

The main motivation of this project is to examine the relationship between EuroLeague team statistics and final season standings, and to identify which statistical factors significantly contribute to team success. By applying data science techniques and machine learning methods, this study aims to uncover patterns that have a meaningful impact on EuroLeague team standings.

Understanding these relationships can provide deeper insights into match outcomes and overall team performance, particularly from the perspective of basketball fans and who closely follow the EuroLeague.

2. Data Source and Collection

The dataset used in this project consists of team-level statistics from the EuroLeague basketball. The data includes various performance metrics related to offensive performance, defensive performance, and overall team efficiency collected over a full season.

All data was obtained from the publicly available website HackaStat.eu, which provides detailed statistical information for EuroLeague teams.

Data collection was carried out by selecting relevant team statistics from the source, and organizing them into structured datasets. These datasets were loaded into python for extended data analysis.

3. Data Analysis

The exploratory data analysis phase focused on understanding the structure of the dataset and identifying relationships between team-level performance metrics and total wins. The analysis began with basic descriptive statistics and visualization of the target variable to examine its distribution across all team-season observations.

Correlation analysis was then applied to quantify the relationship between total wins and numerical performance metrics. Pearson correlation coefficients were computed for all relevant variables, and features were ranked based on the magnitude of their correlations with wins. The most strongly correlated metrics were selected for further analysis and hypothesis testing.

Hypothesis Testing

For the selected performance metrics, statistical hypothesis testing was conducted to evaluate their relationship with total wins. Pearson correlation tests were applied, and statistical significance was assessed after controlling for multiple comparisons using False Discovery Rate correction.

Total Wins Prediction

Regression-based models were applied to predict total wins using selected team-level performance metrics. Models were trained using features identified during the EDA stage, and model performance was evaluated using train-test split and standard regression metrics.

4. Key Findings

- Offensive performance metrics show the strongest association with team success in the EuroLeague.
- Total points scored has the highest positive correlation with total wins, indicating that overall scoring output is a key determinant of season performance.
- Three-point makes and field goals made are also strongly correlated with total wins, highlighting the importance of shooting efficiency.
- Defensive rebounds contribute positively to team success by increasing possession control and limiting opponents second-chance opportunities.
- Interestingly Free throws made have a moderate but statistically significant relationship with total wins, suggesting that efficiency in free-throw scoring plays a meaningful role over a full season.

5. Limitations and Future Work

One limitation of this study is that the analysis is based on aggregated team-level statistics, which may not fully capture in-game dynamics or situational factors such as injuries, player rotations, or coaching strategies. Additionally, the dataset is limited to a fixed number of seasons, which may restrict the generalizability of the findings across different eras of the EuroLeague.

For future work, the analysis could be extended by incorporating player-level data to better understand individual contributions to team success. Including advanced metrics, such as efficiency ratings or possession-based statistics, may further improve the analysis. Moreover, applying more advanced machine learning models and expanding the dataset with additional seasons could enhance the predictive power of the study.

6. Conclusion

This project analyzed the relationship between EuroLeague team-level performance statistics and season-long success using data science and machine learning techniques. Through exploratory data analysis, hypothesis testing, and predictive modeling, key performance metrics associated with total wins were identified.

The results indicate that offensive efficiency metrics play a dominant role in determining team success, while defensive and efficiency-based statistics also contribute meaningfully. The combination of statistical analysis and machine learning methods provides a structured and interpretable approach to understanding team performance in the EuroLeague.

AI Disclosure

AI tools (ChatGPT) were used to assist with text refinement, report structuring, and Markdown formatting in accordance with the course guidelines. AI assistance was also used to improve the clarity and academic tone of written sections.

All data collection, data analysis, hypothesis testing, machine learning implementation, and interpretation of results were performed independently by me. The use of AI tools was limited to writing support and did not contribute to the development of the analytical methods or code ideas.