

Boston Celtics Data Analysis

COMP3125 Individual Project

Michael Yentin
School of Computing and Data Science

I. INTRODUCTION

The Boston Celtics are one of the most storied franchises in professional basketball, with a rich history of championships, legendary players, and cultural significance in the NBA. Understanding the factors that contribute to their performance is both a compelling sports analytics problem and a valuable case study in data-driven decision-making. From 2000 to 2025, the Celtics have experienced periods of dominance, rebuilding phases, and significant organizational changes, including coaching transitions, star player acquisitions, and impactful trades. Analyzing this 25-year span provides a unique opportunity to identify the key performance indicators and strategic decisions that have shaped the team's success.

Existing research in sports analytics has shown that both team-level metrics such as win percentage, offensive/defensive ratings, and net rating and player statistics can be strong predictors of season outcomes. Studies have also highlighted the role of roster stability, coaching strategies, and mid-season adjustments in influencing playoff success. By combining historical performance data with contextual information on roster moves and leadership changes, analysts can better understand not just what drives wins, but how those drivers evolve over time.

This project seeks to examine Boston Celtics performance trends from 2000 to 2025, determine which statistics most strongly correlate with winning seasons, and evaluate the effects of major organizational shifts. Ultimately, the analysis will aim to predict future win percentages and postseason qualification using statistical modeling, providing insights for both basketball enthusiasts and data science practitioners.

DATASET

A. Boston Celtics Basketball-Reference Dataset

Basketball-Reference.com, a respectable and well-known sports statistics supplier run by Sports Reference LLC, provided the dataset used in this analysis, with SportRadar contributing data. A reliable source of information on professional basketball, Basketball-Reference is renowned for its precision, thoroughness, and historical breadth. Updated frequently, most recently in August 2025, the Boston Celtics team dataset covers every season from the team's first 1946–47 season to the next 2025–26 campaign. Official NBA game logs, team summaries, and sophisticated statistical computations were used to create the dataset, guaranteeing a trustworthy and authoritative account of team performance over time.

B. Characteristics of the datasets

Including key performance indicators like games played, wins, losses, win-loss percentage, standings, the Simple Rating System (SRS), pace, offensive and defensive ratings, net ratings, playoff results, coaching staff, team executives, attendance data, and arena information, the dataset is arranged as a statistical record for each season. Both numerical and categorical variables are displayed in a tabular format, usually in HTML but also exportable to CSV. For instance, ratings are presented as scores per 100 possessions, win-loss percentages are stated as decimal fractions, and tempo is evaluated in possessions per game. All these units are clearly defined. The dataset, which includes 80 seasons of data, can be cleaned by managing missing values for statistics that were not recorded in previous eras, translating season labels into datetime variables, and converting percentages to standard % format, but for this specific analysis, we did not have to add any missing values to statistics that were previously missing. Joining keys like team name and season year are used when integrating with other sources, such as player rosters or game-level statistics. To improve the depth of insights derived from the data, other categories can be developed for analysis, such as identifying "elite seasons" according to a net rating criterion. For analytical work, transparency and dependability are guaranteed by this meticulous structure and degree of detail.

<https://www.basketball-reference.com/teams/BOS/>

II. METHODOLOGY

Descriptive + Predictive Analytics:

Used to identify Boston Celtics' player statistics, organizational changes, and seasonal performance trends from 2000 to 2025 to identify patterns that result in winning seasons and postseason success, and to project future performance outcomes using past data. Python's pandas, numpy, seaborn, and matplotlib libraries were used to generate visualizations.

III. RESULTS

Boston Celtics Performance Analysis

a) Win Percentage Over Time (2000–2025): Time series shows fluctuations in win percentage across seasons. Peaks occurred in seasons like 2007–08, 2008–09, 2010–11, 2016–17, 2017–18, and recent years (2021–2024), while notable dips occurred in 2006–07, 2013–14, and 2018–19. Overall, there's an upward trend in recent years, especially under the most recent coaching regimes.

Identify applicable funding agency here. If none, delete this text box.

b) Correlation of Team Statistics with Win Percentage: Correlation heatmap reveals that *Point Differential* has the strongest positive correlation with win percentage (0.97), followed by *Offensive Rating* (0.52) and *Defensive Rating* (0.47). Relative pace has the strongest negative correlation (-0.46), suggesting that faster relative pace does not necessarily equate to higher win percentages for the Celtics.

c) Impact of Coaching Changes on Season Outcomes: Analysis of coaching transitions shows mixed short-term impacts on both win percentage and point differential. Some transitions, like from Jim O'Brien to Doc Rivers, coincided with performance drops, while others, like Brad Stevens' early years, aligned with upward trends. Vertical dashed lines mark coach change years, with green/red bars showing win% relative to average and a blue line tracking point differential.

d) Actual vs Predicted Point Differential (Regression Analysis):

A scatter plot comparing actual vs. predicted point differentials shows that predictions generally follow the trend of actual outcomes, but with some variance. The red dashed line (ideal fit) indicates that while the model captures the general relationship, residual errors exist—particularly in extreme cases.

IV. DISCUSSION

The inherent limitations of measures based on box scores are a significant shortcoming of Basketball-Reference. The statistics primarily documents offensive performance, but it rarely or never documents defensive contributions including positioning, help defense, communication, and subtle "hustle" plays. Since contributions are sometimes shared or attributed in an inaccurate manner, even crucial defensive acts like defensive rebounds and blocks might not accurately represent a player's true influence. Additionally, intangible elements like rotations, shot deterrent, and defensive disruption cannot be taken into consideration by box-score data. Future developments might use more detailed, tracking-based data that records the temporal and spatial dimensions of player movement and defense in order to close this gap. Beyond what can be seen in box scores, advanced player-tracking data (such as defensive proximity, contests, and

rotation timing) can provide quantitative insight into defensive efficacy. By capturing the entire range of contributions, both offensive and defensive, such data, when accessible, would enhance this study.

V. CONCLUSION

This study used statistical correlations, historical patterns, coaching change effects, and predictive modeling to analyze the Boston Celtics' performance from 2000 to 2025. Win percentage has shown steady development and competitive stability in recent seasons, after cyclical periods of success and decline, according to the data. Point difference was found to be the best predictor of win % (0.97) by correlation analysis, confirming its importance as a crucial performance indicator. Offensive and defensive ratings also had a positive impact, and relative tempo had a negative correlation. While coaching is important, roster quality and organizational stability are still crucial, as evidenced by the mixed short-term consequences of coaching changes, some of which resulted in quick declines and others that correlated with increases. Point differential prediction modeling showed a respectable level of accuracy, while there was a noticeable variation in the extreme results. These results demonstrate the usefulness of predictive analytics in predicting team success and have practical ramifications for front offices, analysts, and supporters. They offer data-driven recommendations for roster construction, strategic planning, and establishing reasonable expectations during coaching changes.

ACKNOWLEDGMENT

Thank you Dr. Pang for such a wonderful summer semester and allowing us to have some freedom in choosing our own topic for individual study.

REFERENCES

- [1] "Boston Celtics Franchise Index," *Basketball-Reference.com*. <https://www.basketball-reference.com/teams/BOS/>