Simulating the Boston Celtics' Playoff Success Using Python and Arena

ISYE6644 Simulation - Project Topic 32

Group 56: Jonas Matos & Maekala Turner

Abstract

This project investigates NBA game dynamics through a performance equation integrating Net Rating (NetRtg), Fatigue, and Home Court Advantage (HCA). Using Monte Carlo simulations, we explore the Boston Celtics' 2023–24 season, offering insights into key performance drivers under varying conditions. NetRtg measures team efficiency, Fatigue reflects rest-related performance changes, and HCA accounts for the benefits of playing at home. Initial findings highlighted variability in outcomes, but challenges with implementing HCA and over-weighted regression coefficients for Fatigue revealed limitations in the current model.

Background

Professional basketball outcomes are influenced by statistical metrics, player performance, and contextual factors. Net Rating (NetRtg) is a widely used indicator of team efficiency, but it fails to capture the full complexity of variables like Fatigue and HCA. This project uses Monte Carlo simulations to explore these dynamics for the 2023–24 Boston Celtics dominating season. By integrating statistical analysis and performance modeling, the goal is to uncover insights into how these factors interact to shape game outcomes.

Methodology

Performance Equation

At the core of the simulation is the performance equation, which captures the inherent variability in basketball outcomes and models a team's expected performance in each game:

$$\text{Performance} = \text{NetRtg} - \text{Fatigue} + \text{HomeCourtAdvantage (HCA)}$$

Each component of this equation reflects a critical factor influencing game outcomes:

1. **Net Rating (NetRtg):** The baseline measure of a team's efficiency, calculated as the difference between offensive and defensive ratings per 100 possessions. NetRtg serves as the primary indicator of team strength, directly derived from historical game data.
2. **Fatigue:** Fatigue accounts for reduced player performance due to cumulative minutes played and limited rest between games. It is modeled using a regression-based approach:$\text{Fatigue} = \alpha \cdot \text{MIN} - \beta \cdot \text{DaysRest} + \epsilon$, but sensitivity to these values caused challenges during simulations.
3. **Court Advantage (HCA):** Reflects the psychological and logistical advantages of playing at home, such as fan support and reduced travel stress. The model calculates HCA as the difference in NetRtg between home and away games, using historical data for each team. HCA is represented as a random variable sampled from a normal distribution.

Variable Distribution Fitting

Each of the three key variables were fitted to their appropriate distributions. The net rating distributions for all teams were analyzed to assess their normality. Visualizations and fitted normal curves for each team's net rating showed a strong alignment with the characteristics of a normal distribution, supporting its use in the performance equation.

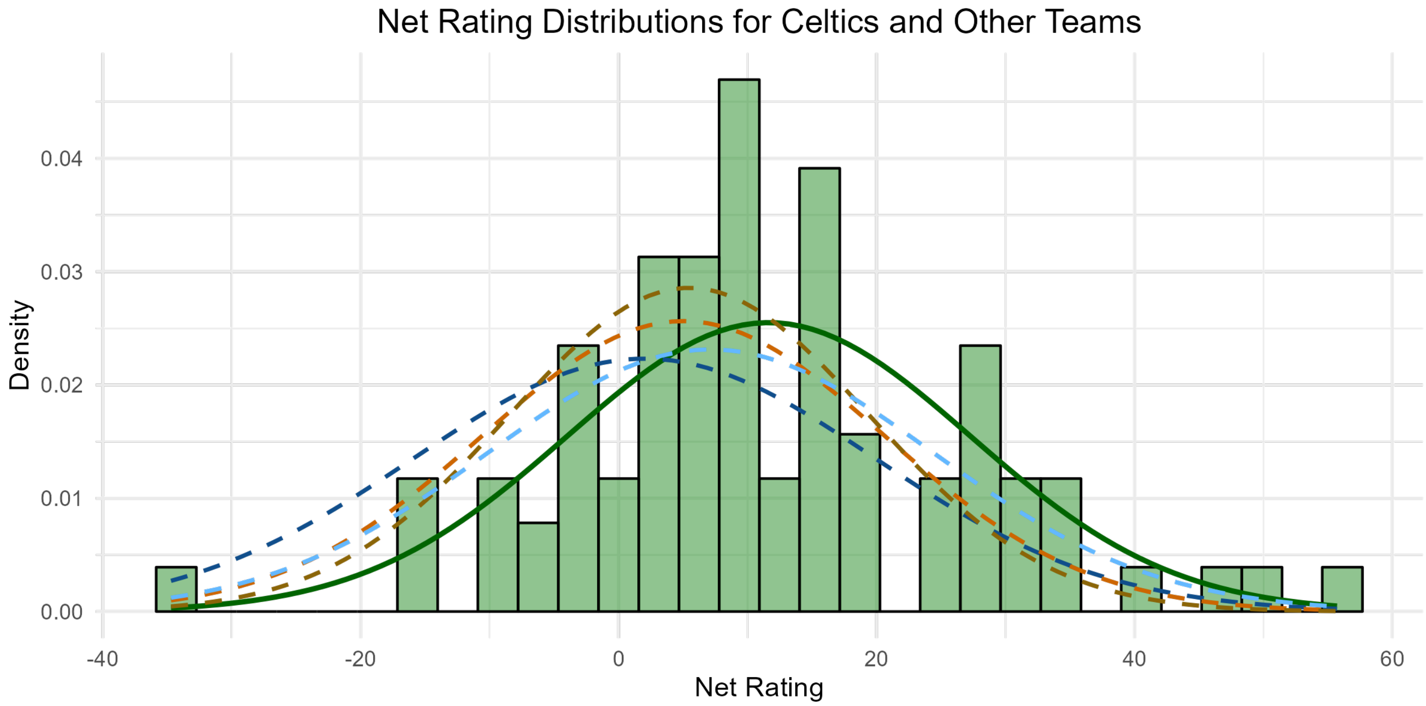

Figure 1: Net Rating Distributions

To validate the fatigue variable, a Shapiro-Wilk test was performed on the residuals of the regression models for each team. The test assessed whether the residuals followed a normal distribution, which is a critical assumption for the reliability of the fatigue calculation. The following table contains the test results:

Team	**Shapiro-Wilk**	**p-value**
Celtics	0.97726	0.1535
Knicks	0.98966	0.7602
Mavericks	0.98743	0.6102
Thunder	0.97909	0.2013
Nuggets	0.97134	0.06316

The Shapiro-Wilk test results indicate that the residuals for all teams do not significantly deviate from normality ($p > 0.05$ for all). Visual diagnostics further support these findings particularly for the Nuggets, confirming that the regression-based fatigue calculation is valid for integration into the performance model.

To validate the robustness of the Home Court Advantage (HCA) calculations, a two-sample t-test was performed to compare the mean Net Rating (NetRtg) for games played at home versus away for each team. This test evaluates whether the differences in performance across locations are statistically significant, supporting the inclusion of HCA as a variable in the simulation model. The table below summarizes the t-test results for each team:

Team	Home Mean	Away Mean	P-Value	Significant ($p < 0.05$)
Celtics	15.4049	7.8854	0.0287	Yes
Knicks	6.2829	3.7659	0.4674	No
Mavericks	3.2976	0.7024	0.5144	No
Thunder	12.6610	1.6073	0.0031	Yes
Nuggets	10.00000	0.9390	0.0028	Yes

The p-values indicate that the difference in NetRtg between home and away games is statistically significant for the Celtics, Thunder, and Nuggets, confirming that Home Court Advantage (HCA) plays a meaningful role for these teams. For the Knicks and Mavericks, the differences are not statistically significant, suggesting a less pronounced effect of HCA in their cases. However, Home Court Advantage is a well-documented factor in basketball performance, particularly in high-stakes scenarios like playoffs. Thus, while the effect varies across teams, HCA will be implemented for all teams in the simulation model to ensure consistency and account for its potential impact under different conditions. This approach ensures that the model remains comprehensive and adaptable, even if HCA is more influential for some teams than others.

Weighting of Variables

To simulate game outcomes realistically, weights are assigned to each component in the performance equation, reflecting the relative impact of each variable. These weights are normalized to sum to 1, ensuring consistency across simulations, and multiple weight configurations are tested to assess the model's sensitivity and robustness. Weights reflect the relative impact of each variable, with NetRtg as the dominant factor.

Variable	Weight
NetRtg	0.55
Fatigue	0.30
HCA	0.15

Monte Carlo Simulation Framework

The Monte Carlo simulation framework modeled thousands of game outcomes by sampling values for Fatigue and Home Court Advantage (HCA) from their distributions. Game performance for both teams was calculated using the performance equation, and a win was recorded if the Celtics' score exceeded their opponent's. This process was repeated thousands of times to capture variability and ensure robust results.

Arena Logic Development

To refine the simulation model, we used Arena, a discrete-event simulation tool, to structure game dynamics and define key variables before transitioning to Python. Each basketball game was modeled as an entity flowing through modules. The **Create Module** initialized playoff matchups for the best-of-seven series, ensuring progression until one team achieved four wins. The **Assign Module** updated attributes like Net Rating (NetRtg), Fatigue, and Home Court Advantage (HCA) as the series advanced. The **Process Module** calculated performance scores using the performance equation, while the **Decision Module** determined outcomes and updated series records. The **Record Module** tracked results, and the **Dispose Module** finalized and stored outcomes for analysis.

Main Results & Findings

The Monte Carlo simulation consistently resulted in series sweeps (4-0 wins), revealing imbalances in the model's structure. The fatigue component, modeled as:$\text{Fatigue} = \alpha \cdot \text{MIN} - \beta \cdot \text{DaysRest} + \epsilon$, dominated game outcomes due to its high sensitivity to regression coefficients (α and β), often overshadowing other variables. Additionally, Home Court Advantage (HCA), implemented as a random variable sampled from historical home/away NetRtg differences, had minimal influence on outcomes. Combined, these imbalances heavily favored the higher-rated team, leaving little room for variability and consistently predicting dominant performances by the Celtics.

Discussion & Limitations

This project highlights both the strengths and challenges of using simulations to model NBA game dynamics. While the framework successfully integrated key variables like NetRtg, Fatigue, and HCA, limitations in the model's assumptions and implementation affected its realism:

1. **Fatigue Sensitivity**: The regression-based fatigue model disproportionately influenced game outcomes, often overwhelming the contributions of NetRtg and HCA. This suggests a need for alternative approaches, such as dynamic fatigue adjustments that consider game-specific contexts.

2. **HCA Representation**: The simplistic implementation of HCA as a random variable failed to capture its nuanced impact, such as travel schedules, fan intensity, or playoff-specific pressures. Future work should refine HCA modeling to better reflect these contextual factors.

3. **Simulation Imbalance**: The model's weight configuration and variable sensitivities led to unrealistic sweep scenarios, highlighting the need for improved balance among variables. Exploring alternative distributions or re-calibrating weights could help introduce more competitive variability.

Despite these challenges, the project demonstrates the utility of simulation-based approaches for uncovering insights into playoff dynamics. Addressing these limitations by incorporating player-level data, refining variable representations, and leveraging machine learning techniques could enhance the model's accuracy and applicability. These improvements would make the framework more robust for strategic decision-making in sports analytics.

Conclusions

This project demonstrated the potential of simulation to analyze complex systems like NBA game outcomes, focusing on the Boston Celtics' playoff performance. By integrating key factors such as Net Rating (NetRtg), Fatigue, and Home Court Advantage (HCA) into a performance equation, we modeled game dynamics with statistical rigor. However, challenges in implementing HCA and the outsized impact of regression-based Fatigue coefficients revealed areas where the model requires refinement. The Monte Carlo simulation framework, supported by Arena logic, provided valuable insights but consistently predicted series sweeps, highlighting imbalances in variable weighting and sensitivity.

One key takeaway was the need for adjustments to better account for competitive variability and the interplay of critical variables. The model's sensitivity to Fatigue coefficients often overshadowed the influence of HCA, while the lack of player-level data limited its granularity. Future work could address these issues by incorporating player-specific metrics, such as individual workload or injury status, and refining the representation of HCA to reflect contextual factors like travel or fan engagement. Advanced techniques, such as machine learning, could dynamically adjust weights based on game context, enhancing prediction accuracy. Beyond basketball, this methodology offers adaptability to other sports or competitive scenarios. Overall, while the project highlighted the value of data-driven approaches in sports analytics, it also emphasized the need for continued refinement to achieve more balanced and realistic predictions.