NBA Attendance Analysis Report

Daniel Rojas

Table of Contents

Executive Summary

This report presents an in-depth analysis of factors influencing NBA game attendance, focusing on results from a logistic regression model. The analysis highlights the key variables that impact game attendance, including the day of the week, the month of the season, and team performance metrics, particularly for the away team. The findings show that attendance peaks during weekends and in the final months of the season, especially in April. These insights provide actionable recommendations for optimizing NBA operations, including game scheduling, marketing efforts, and resource allocation.

Introduction

Background

The NBA's revenue streams are significantly influenced by game attendance, which drives ticket sales, concessions, merchandise, and broadcast deals. Understanding the factors that contribute to higher game attendance is critical for the NBA's strategic planning, particularly in scheduling, marketing, and resource management. This report aims to provide data-driven insights into what influences game attendance, helping the NBA make informed decisions that can maximize revenue and enhance fan engagement.

Project Objectives

The primary objective of this project was to identify the factors that lead to above-average. By analyzing historical data, the project sought to uncover patterns and predictors that could inform strategic decisions on game scheduling, marketing, sponsorships, and event planning.

Data Overview

Data Sources

The dataset for this analysis was sourced from a comprehensive SQLite database provided on Kaggle. The database includes various tables with information on NBA games, teams, players, and more, covering NBA data from as far back as 1946. Key tables used in the analysis include:

- **Game Table**: Contains detailed information about each game, including dates, locations, and teams involved.
- **Game Summary Table**: Provides summarized statistics for each game, such as points scored by home and away teams.
- **Player Table**: Includes information on players who participated in the games.
- **Team Table**: Provides details on NBA teams, including geographic and historical data.

Data Preparation

Before proceeding with the analysis, the data underwent several preprocessing and transformation steps to ensure it was clean, accurate, and suitable for modeling:

1. **Data Cleaning**: The data was initially cleaned to handle missing values, remove irrelevant or redundant features, and correct any inconsistencies. For instance, games with incomplete records or irregular seasons were excluded to maintain data integrity.
2. **Feature Engineering**: New features were created to capture potential influences on attendance. These included:
    - **Attendance Ratio**: A metric calculated as the ratio of actual attendance to the arena's capacity for home games.
    - **Distance Traveled by the Away Team**: Calculated based on the latitude and longitude of the teams' home cities.
    - **Previous Season Playoff Participation**: A binary indicator of whether the home or away team participated in the previous season's playoffs.

3. **Data Transformation**: The dataset was transformed to focus on the most relevant variables likely to impact game attendance. These included:
    - **Day of the Week and Month**: Categorical variables representing the timing of the game.
    - **Team Performance Metrics**: Such as season win percentage for both home and away teams, points scored, and plus-minus statistics.
    - **Geographical and Historical Team Data**: Including the location of the teams and their historical performance.
4. **Merging Data**: Various tables were combined to create a comprehensive dataset for analysis. This involved merging data on game details, team statistics, and additional contextual factors like travel distance and playoff participation.
5. **Exporting Data**: The final combined dataset was exported for further analysis and modeling. This dataset formed the basis for all subsequent analysis and model development, providing a rich and detailed view of the factors influencing NBA game attendance.

Methodology

Analytical Tools and Techniques

The analysis was conducted using Python, with libraries such as pandas, scikit-learn, and matplotlib for data manipulation, modeling, and visualization. SQL was used for data extraction and initial exploration.

Model Development

A logistic regression model was chosen for its interpretability and ability to handle binary outcomes, such as predicting whether a game will have above-average attendance. The model development process involved the following steps:

- **Data Filtering:** During stakeholder interviews an emphasis was placed on analyzing the modern iteration of the league and so only game data after the year 2000 was used and irregular seasons identified during EDA were excluded to not influence the model.
- **Pipeline Construction**: A pipeline was built to streamline data preprocessing, including scaling and dimensionality reduction, and to automate the modeling process.
- **Hyperparameter Tuning**: A grid search and 5-fold cross-validation was employed to optimize the model's hyperparameters. Parameters such as the type of scaler, regularization strength (C), and penalty type were tuned to enhance model performance.

Model Evaluation

The model's performance was evaluated using standard metrics, including accuracy, precision, recall, and the confusion matrix. These metrics provided insights into how well the model predicted game attendance and highlighted areas for potential improvement

Analysis and Findings

Model Performance and Evaluation

- **Accuracy**: The logistic regression model achieved an accuracy of approximately 66%, indicating that it is better than random guessing in predicting above-average game attendance.
- **Confusion Matrix**: The model showed more True Negatives (1,699) and True Positives (1,173) than False Positives (646) and False Negatives (854). However, the relatively high number of False Negatives suggests that the model could be improved in predicting games with above-average attendance.
- **Classification Report**: The model displayed higher precision and recall for class 0 (below-average attendance) compared to class 1 (above-average attendance). The macro and weighted averages were around 0.65, consistent with the model's accuracy.

Attendance by Month

The analysis revealed that attendance consistently peaks in April, with an average of 18,039 attendees, suggesting heightened fan engagement as the season concludes. Other months with high attendance include March and February, indicating that fans are particularly interested in games towards the end of the regular season. (Appendix: Figure 1)

Attendance by Day of the Week

Attendance tends to be highest on Sundays (17,847) and Fridays (18,125), with Wednesdays and Saturdays also showing strong numbers. These days are optimal for scheduling high-profile games and promotional events to maximize fan turnout and engagement. (Appendix: Figure 2)

Seasonal Trends in Attendance

Overall attendance has shown a resilient upward trend over the years, with a dip during the 2020 season due to the COVID-19 pandemic. Post-pandemic recovery has been strong, with attendance figures rebounding to pre-pandemic levels, reflecting the enduring popularity of NBA games. (Appendix: Figure 3)

Key Predictors of Attendance

Top Positive Predictors:

- **Season Win Percentage (Away Team)**: The strongest predictor, indicating that games against high-performing away teams attract larger crowds.
- **Day of the Week (Friday, Saturday)**: Games held on these days significantly increase the likelihood of higher attendance.
- **Previous Playoffs (Away Team)**: If the away team participated in the playoffs the previous season, there is a slight increase in the likelihood of above-average attendance, possibly due to their perceived competitiveness or fan interest.

Top Negative Predictors:

- **Months (November, December, January)**: Games held in these months are less likely to have above-average attendance, possibly due to competing holiday activities or the early stage of the season.
- **Distance Traveled (Away Team)**: The greater the distance the away team has to travel, the less likely the game will have above-average attendance, suggesting reduced interest in games with less well-known or local rival teams.
- **Day of the Week (Monday, Tuesday, Wednesday)**: Games held on these midweek days are less likely to have above-average attendance, likely because they are less convenient for fans to attend.

Recommendations

Strategic Scheduling

- **Optimize Game Scheduling**: Prioritize scheduling key games on Fridays, Sundays, and Saturdays, especially in March and April when fan interest is highest. This can maximize attendance and, by extension, revenue from ticket sales and concessions.
- **Target High-Impact Games**: Games involving high-performing away teams should be strategically scheduled for peak days to draw larger crowds.

Marketing Focus

- **Concentrate Marketing Efforts**: Focus marketing campaigns on the months of March and April, capitalizing on the natural increase in fan interest as the season progresses.

Operational Planning

- **Enhance Support for High-Traffic Games**: Allocate additional resources such as security, concessions, and transport logistics for games predicted to have higher attendance. This will ensure a smooth and enjoyable experience for attendees.

Conclusion

This analysis has provided actionable insights into the factors driving NBA game attendance. By leveraging these insights, the NBA can optimize game scheduling, enhance marketing strategies, and improve overall operational efficiency. The logistic regression model developed in this project offers a solid foundation for making data-driven decisions that will help the NBA grow its fan base and maximize game attendance.

Future Work

Further analysis could focus on predicting sellout games and exploring the impact of additional factors such as weather conditions, specific player appearances, and local events. These areas could provide even deeper insights into optimizing NBA operations.

Appendix

Figure 1. Attendance by Month

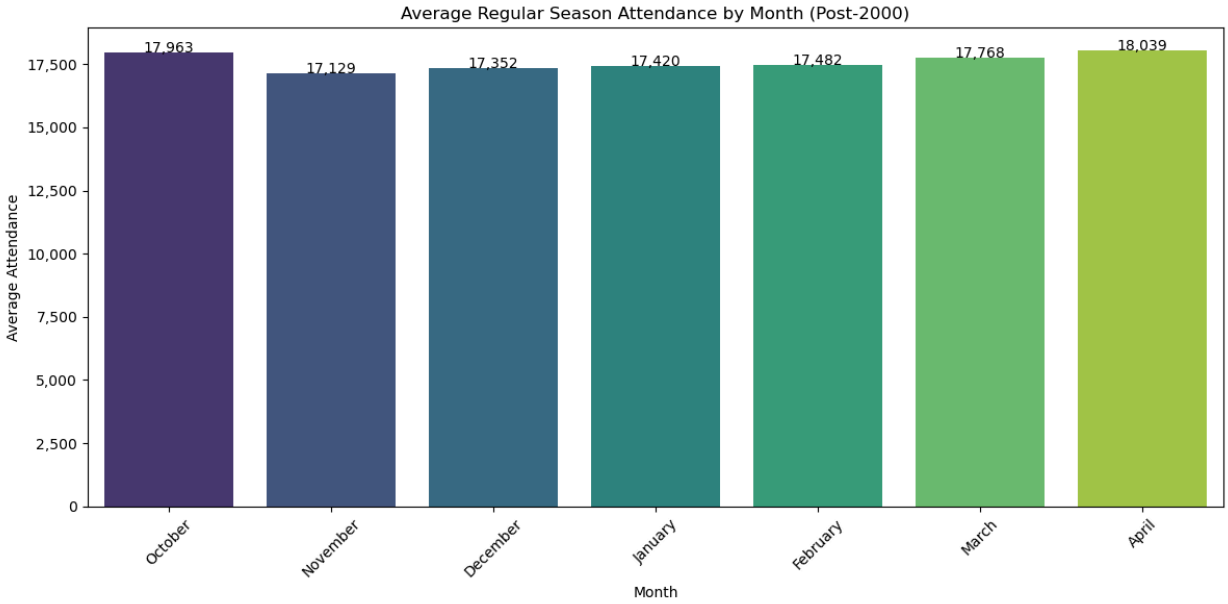

Figure 2. Attendance by Day of the Week

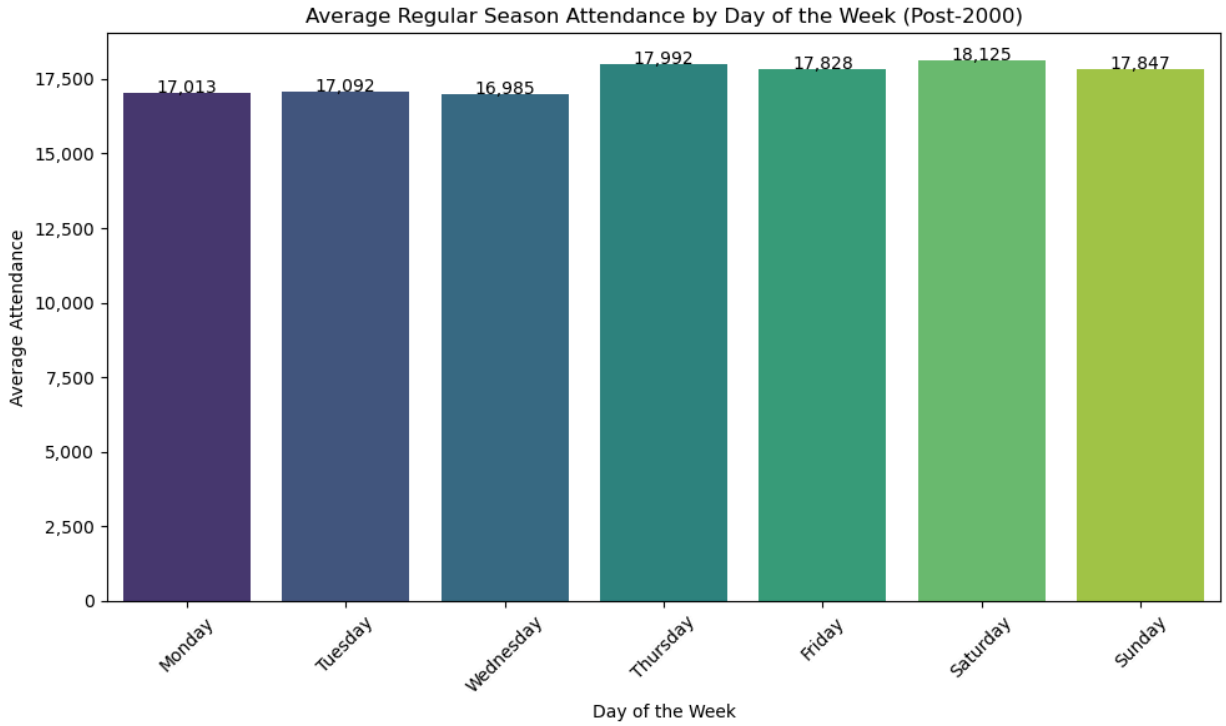

Figure 3. Seasonal Trends in Attendance