main

December 18, 2025

1 Main Summary and Discussion of Results

Group 16: Carly Tran, Sami Leong, Joshua Gonzalez, Ziyu Ge

1.1 Assumptions

When analyzing this dataset, we had to make a few core assumptions that helped justify our analysis.

We analyze publicly available NBA game-level data from the 2024–2025 season, which was not collected under experimental conditions. Because of this, our work is exploratory and focuses on simple statistical relationships rather than causal inference. We use summary statistics, correlations, and linear regression to examine how common performance metrics relate to player consistency, shooting efficiency, and team outcomes.

1.2 Results

This project investigated the NBA 2024-2025 season with player and team performance in basketball. It focused on consistency, efficiency, and contributions to wins. In each notebook, we focused on the player and aimed to identify which athletes are most consistent in scoring, rebounding, and assisting by calculating summary statistics and gathering correlations across different game categories. These were visualized through distributions with boxplots and histograms. We also explored how a player's minutes relate to their scoring, determining the optimal number of minutes that maximizes on-court performance through scatterplots and efficiency metrics.

At the team level, the analysis examines which statistical categories, beyond points, are most important in contributing to wins. We aggregate individual player statistics to compute team per-game averages for rebounds, assists, turnovers, and shooting percentages, and investigate how these factors correlate with team success. Additionally, we looked at predictors of shooting efficiency (FG%, 3P%, FT%) using regression models, correlation analyses, and clustering to identify the variables most strongly associated with shooting outcomes.

Overall, the analysis combines descriptive statistics, visualizations, and regression techniques to provide a clear picture of both individual and team performance factors in the NBA. Our data was cleaned and taken from Kaggle, an online datahub. Therefore, we have minimal control over the original acquisition of it. Despite this, we made assumptions that the games were independent, and metrics were comparable across players and teams. We aimed to fully address our curiosities in sports metrics through this exploratory analysis and gather insightful data when looking at the NBA.

1.3 Author Contributions

Carly Tran: Assisted with the creation of different notebooks for each question. Completed Q0 and Q1, where I imported the data and analyzed player's consistency based on points, rebounds, and assists. Also completed Q5 to model FG%, 3P%, and FT%.

Sami Leong: Organized and created repo to begin project. Completed Q2 to analyze player points to minutes played. Created myst.yml and binder link, launched myst site to GitHub pages.

Joshua Gonzales: Completed Question 3 analysis to identify which non-point statistics most influence team wins. Utilized matplotlib functions to visualize correlations.

Ziyu Ge: Completed Question 4 by aggregating player-level data into team-level game statistics and modeling win/loss outcomes using logistic regression with cross-validation. Interpreted correlations and model coefficients.

1.4 Bibliography / References

“‘{bibliography}