

# Advanced Analytics for NBA Game Outcomes and Score Prediction

## Introduction:

The National Basketball Association (NBA) is not only a spectacle of athletic prowess but also a dynamic narrative of performance data that unfolds with each game. In the data-rich environment of professional basketball, the ability to predict game outcomes and team scores stands as a pinnacle of analytical endeavor. This project embarks on the challenge of developing a sophisticated classification model and predictive modeling framework using historical NBA data. The objective is to forecast the winners of NBA games and to predict team scores with a high degree of accuracy. Such models have the potential to revolutionize strategic decision-making within the NBA, influence betting markets, and elevate fan engagement to new heights.

## Dataset Description:

The dataset, sourced from Kaggle and originally compiled from the official NBA stats website (<https://www.nba.com/stats>), provides an exhaustive statistical overview of NBA games, players, and team performances. The data spans from 2003 to 2020, offering a comprehensive temporal coverage that captures the strategic evolution of the league across different seasons. Focused on the geospatial context of the USA, where the NBA championship has been a centerpiece of the sports culture, the dataset includes:

- games.csv: A historical log of games detailing dates, participating teams, and scores.
- games\_details.csv: In-depth player statistics for each game, essential for performance analysis.
- players.csv: Details of NBA players, providing a reference for individual tracking.
- ranking.csv: Daily league rankings, offering insights into team standings over time.
- teams.csv: Information about NBA teams, including location and organizational details.

## Project Goal:

This project aims to use machine learning to forecast NBA game winners and team scores, drawing on data like team history, player performance, and rankings. The resulting models will classify winners and predict scores, aiding teams, analysts, and fans, while enhancing our grasp of basketball through data-driven insights. Our analysis will strive to translate rich historical NBA data into actionable insights, uncovering the drivers of team outcomes.

## Smart Questions:

- i. Trend Analysis Over the Season:
  - Which player performance metrics most strongly correlate with changes in team rankings throughout the season, and do these correlations suggest causation or simply association?
- ii. Impact of Back-to-Back Games:

- How does player efficiency rating (PER) and injury incidence rate differ in back-to-back games compared to non-consecutive games, and what does this imply about player fatigue and risk?

iii. Game Winner Classification:

- Can integrating dynamic variables such as player morale (e.g., streaks, media sentiment) and real-time conditions (e.g., travel delays, weather) into the historical data enhance the accuracy of game-winner predictions?

iv. Team Score Prediction:

- How do unexpected events like key player substitutions and mid-season trades affect the predictive accuracy of our model for team scores, and can the model be made adaptive to such changes?

**Team02:**

ANERI GIRISHBHAI PATEL, G40408020, [anerigirishbhai.patel@gwu.edu](mailto:anerigirishbhai.patel@gwu.edu)

SHRIHAN THOKALAH, G36452583, [shrihan.thokala@gwu.edu](mailto:shrihan.thokala@gwu.edu)

SAMIKSHA RAMNATH BURKUL, G46017190, [samiksharamnath.burkul@gwu.edu](mailto:samiksharamnath.burkul@gwu.edu)

SWATHI K R, G48428717, [swathi.kannettiramanareddy@gwmail.gwu.edu](mailto:swathi.kannettiramanareddy@gwmail.gwu.edu)

**Dataset Link:**

<https://www.kaggle.com/code/hqfang/nba-game-wins-prediction>

**Github repo:**

<https://github.com/Shrihan999/Data-Mining-Project.git>