NBA Games Data: Predicting the Characteristics of Basketball Winners

• • •

Minh Bui, Bryant Hicks, Cassandra Cohen

Project Motivation

The motivation behind this project is to make novel discoveries within NBA statistics. Questions such as the following are intended to be answered:

- What characteristics are predictors for winning games in the NBA?
- Which teams have had the most success in basketball?
- Which players recorded the best statistical performances?
- What stats correlate the highest with win rate?
- Does defense win championships?
- Can "Jump Shooting Teams" win championships?

Prior Work

A great deal of prior work has been done on this subject by the NBA itself.

The Golden State Warriors were one of the first teams to recognize that three point shooting holds a strong statistical correlation with winning.

With this knowledge, the Warriors built a dynasty and proved that it is possible to win championships by shooting a high volume of three pointers.

Sports media outlets, such as ESPN, also hire statisticians to discover patterns and trends in the NBA.

Sports betting utilizes statistical analytics to predict the winners of sporting events.

Datasets

- The data is provided by Kaggle and the NBA stats website and is available at the following link:
  - https://www.kaggle.com/datasets/nathanlauga/nba-games?resource=download
- There are multiple datasets that include statistics from NBA games since 2004
  - One dataset includes hundreds of thousands of tuples detailing player statistics organized by game.
  - There's also a dataset that records which team won each game and the overall stats for that game.
  - A common attribute (game id) ties the above two datasets together
- Dataset is stored locally on Minh's own personal computer

Proposed Work

- Data cleaning/reduction
  - Remove columns containing attributes that are deemed irrelevant for answering the proposed questions
    - Example: The date on which games were played
  - Removal of rows containing data offering little insight, such as stats for players who hardly played any games.
- Exploratory data analysis
  - Which team statistics are the best predictors of winning percentage?
- Market basket analysis
  - Show which players participated in the most games.

List of Tools

- Python (including libraries such as pandas and numpy)

- Jupyter Notebook

- Github

Evaluation

- The datamining results will be evaluated by comparing training and test sets to measure the predictive accuracy of the developed machine learning models.
- Compare whether discovered results either affirm or contradict common basketball knowledge.
    - Is LeBron James truly the greatest player since Michael Jordan, or could it be someone else, like Steph Curry or Tim Duncan?
    - Which team was the greatest dynasty of the era: the Golden State Warriors or the San Antonio Spurs?
    - Can you win championships as a "Jump Shooting Team" that shoots large quantities of threes?