Predicting NBA Player Longevity

CS 365 Project Proposal by Judson Marble

The National Basketball Association (NBA) is one of the most watched sports leagues in the world, with a wide cultural reach that extends beyond just sports. To viewers, athletes are not just athletes, but icons and celebrities. However, to NBA sports franchises, athletes are assets. Assets that can be bought and traded as teams search for wins. The abundance of advanced stats and analytics for NBA players allows teams to use predictive statistics to help them find players with the best value for a team. This project aims to investigate the specific desirable value of longevity. Key research questions include: How do physical characteristics like height and weight affect a player's career span? Is there a difference in career longevity between defensive specialists and high-scoring players? How does injury history influence a player's tenure in the league? By analyzing these factors, we intend to develop predictive models that estimate an NBA player's career length, offering valuable insights for team management and player development strategies.

Data Set:

https://www.basketball-reference.com contains comprehensive statistics on every player in NBA league history. For our purposes, we will use a limited data set of all retired players in the league since the year 2000. We want to use data from players of the same era, since advancements in medicine and changes in the physicality of the game could skew our analysis. Data points such as career length, physical characteristics, and performance metrics will be used to develop our analysis.

Peer Reviewed Sources:

1. Sports Analytics - Evaluation of basketball players and team performance
    - https://www-sciencedirect-com.ezproxy.bu.edu/science/article/pii/S0306437920300557?via%3Dihub
2. A Data Science approach analyzing the Impact of Injuries of Basketball Player and Team Performance
    - https://www-sciencedirect-com.ezproxy.bu.edu/science/article/pii/S030643792100020X?via%3Dihub
3. A deep learning approach to injury forecasting in NBA basketball
    - https://bu.primo.exlibrisgroup.com/discovery/fulldisplay?docid=cdi_openaire_primary_doi_3244af9df8c2527a69681dbacd3756ca&context=PC&vid=01BOSU_INST:BU&lang=en&search_scope=MyInst_and_CI_NoLondon&adaptor=Primo%20Central&tab=Everything&query=any,contains,basketball%20analytics&offset=10
4. A supervised learning model to identify the star potential of a basketball player
    - https://bu.primo.exlibrisgroup.com/discovery/fulldisplay?docid=cdi_unpaywall_primary_10_1111_exsy_12772&context=PC&vid=01BOSU_INST:BU&lang=en&search_scope=MyInst_and_CI_NoLondon&adaptor=Primo%20Central&tab=Everything&query=any,contains,basketball%20analytics&offset=30

Method:
The project will employ a combination of statistical analyses and machine learning techniques to identify patterns and predict player longevity. We will use Python as our main programming language, utilizing libraries such as pandas for data manipulation and NumPy for numerical calculations. Our analysis will begin with exploratory data analysis to identify potential correlations and trends, followed by feature selection to determine the most influential factors affecting career length. Predictive models, including regression analysis and decision trees, will be developed to forecast player longevity. The effectiveness of these models will be evaluated using metrics such as R-squared for regression models and accuracy for classification models.

Expected Outcomes:
Significant correlations between a player's physical attributes, performance metrics, and their career longevity are expected. The project aims to identify key predictors of career length, enabling us to create a predictive model with the capability to forecast an NBA player's career trajectory.