Project 2 Proposal

NBA Player of the Week Historical Analysis

W200 Section 2, Group 1

Team Members: Anusha Praturu, Tom Goter, Noah Pflaum
Github Repo: https://github.com/UCB-INFO-PYTHON/Project2_Section2_Group1

Datasets

The primary dataset our project intends to analyse is from Kaggle and contains all the NBA player of the week award from the 1984/85 season through the 2017/18 season. It contains some additional information regarding the players' height, weight, age, team, position and a couple other attributes.

Additional datasets may be considered as the analysis evolves. For example, rather than looking at absolute physical characteristics of players, we may normalize it per season in which case we'd need a dataset that provides data on all players in the league that year, not just the players of the week. We can get this data from NBA.com's API, scrape it from basketball-reference.com or tap into other basketball resources. Sentiment analysis or search data related to a player from Twitter and Google respectively may also be considered to see how player of the week awards affect their public perception.

Primary dataset URL: https://www.kaggle.com/jacobbaruch/nba-player-of-the-week
Secondary dataset URL: https://github.com/swar/nba_api

Core Questions:

- How have the physical characteristics of "star" NBA players changed as the game has evolved over the past three decades?
    - Consider external factors such as changes to the rules which may have led to shifts in star player attributes.
    - Horace Grant's blog post suggest the game has transitioned from a big man's game to "small ball" and we can see how the data compares to his claims.
- How does NBA player of the week awards correlate with team performance?

Data Strategy

From our primary dataset, we plan to use most of the player information columns, including their height and weight, their age/length of career, and position to see if we can find discernible trends or patterns over the years. We will also try to use some calculated columns, such as length of time with current team and time in season when the award was granted, among

others. We will use the player name and team name columns to join the data with supplementary datasets from the NBA API. These additional datasets will add depth to the Player of the Week data set and allow us to look at player career stats, team stats, and playoff info in addition to the limited player info detailed in the Player of the Week dataset.

Plans for Final Report

The final report on this exploratory data analysis will be organized into an estimated seven sections.

- **Abstract/Executive Summary** - This section will provide a brief overview of the primary question that was asked, the data that was used and the most interesting insights.
- **Introduction/Questions/Purpose** - This section will discuss both the primary questions that were asked at the beginning of the analysis and how this question (and the supplemental questions) evolved as the exploratory data analysis progressed.
- **Datasets** - This section will describe in detail the primary dataset from Kaggle that was leveraged for the analysis, as well as any secondary sources from the NBA API that were included and how the datasets were used in conjunction. It will also detail data cleansing efforts and the strategy used to identify which columns and datasets were relevant to the research question(s).
- **Analysis Assumptions -** This section will provide information on the assumptions made during the data analysis and provide context about the scope of the analysis.
- **Discussion of Analysis -** This will be the main body of the data analysis report where the authors will walk the audience through the data analysis, providing visual representations of the data (tables and charts) as necessary to convey interesting findings. This section will focus on primary questions of the analysis; however, any interesting supplemental insights will also be discussed.
- **Conclusions -** This section will summarize the key findings from the exploratory data analysis. The conclusions from this section will be similar to those provided in the Executive Summary, but slightly more detail may be provided here.
- **Area for Continued Investigation -** This section will describe potential areas for further investigation which were outside the scope of this analysis, as well as additional questions that were raised and anomalies that were noted during the data exploration phase. It is intended to be used as the starting point for for further analysis, either for the authors or other interested parties.

Example Plots