Milestone 1 Outline

Chase Stockwell, Thomas Huang

<https://github.com/chas3stockwell/chas3stockwell.github.io>

<https://www.sports-reference.com/cbb/seasons/2001.html>

***Background/ Current Literature***

There has been a significant amount of research into the prediction of basketball players based on their collegiate years into their overall significance in the NBA arena. Due to the massive market that the NBA now possesses as a center point of entertainment in the US and around the world, the predictive ability of the best agents and teams are constantly being improved to increase their use of draft positions.

However, these tools are not widely available to the public for personal use. Although online discussion can help contribute and shape perspective of player rankings and success, there has been less noted focus on the longevity of player performance in the NBA. Common methods for predicting ability of players were using the NBA formula for productivity. However, many of these statistics are used to make predictions for early success in their career and their ability to establish a draft position. Our study, rather, will look at the longitudinal data retroactively to determine the longevity of their time in the league and how this is correlated to aspects taken from their time in collegiate basketball.

***Objective***

We intend to create using data science an algorithm that is able to extract the stats that are most significant in indicating and/or predicting longitudinal success in the professional scene among college players. We define **success** as the following:

1. Longevity of NBA career relative to position
2. Amount of money earned on the player’s contract
3. Overall ‘basketball statistics’ that would be considered successful (points, assists, steals, blocks).
   1. Wins
   2. College Statistics
   3. Physical Qualities

***Databases***

|  |  |  |
| --- | --- | --- |
| Hoops Hype | Salaries of individual players 2000-2020 | <https://hoopshype.com/salaries/2000-2001/> |
| Bart Torvik | On-court stats of individual players | <http://barttorvik.com/trankpre.php> |

1. Hoops Hype Database
   1. Will require data scraping
   2. Salaries of NBA Players
   3. Details the salaries of individual players in the NBA from years 2000 to 2020. This is raw data, so the salary data that will be used as one factor of many to determine success of players. Salaries will also be taken in relevance to salary caps and other monetary statistics for the NBA that year as inflation as well as the compensation of players has changed significantly over the years.
2. Bart Torvik Database
   1. Will not require data scraping - contacted Bart Torvik, will send raw data in csv files.
   2. Statistics of individual players in college years.
   3. Details include all on-court stats, school, physical characteristics, etc.

***Methods***

1. *Data Scraping* - we want to data scrape the databases above from their websites.
   1. We intend to use methods learnt in class to achieve this
   2. For more complicated extractions, we will use Stack Overflow and other modules online to learn how to compile the data.
   3. We may also interact with software API’s to gain more information on the data, using methods as we have learned in lecture.
2. Python + Pandas to interpret and analyze csv raw data
   1. We intend to use the same techniques as performed in class to create DataFrames of the tables.
   2. Some SQL may be used to join and merge together tables for a clearer picture whatever inferential statistic we are pursuing.
3. Statistics Modeling to analyze data
   1. General Stats Description: Mean, Medians, etc… by quartile groups of “success” groups
   2. Assignment of success scores of current NBA players based on various factors.

***Impact***

The **inspiration** behind this project is that the arguably most influential college basketball player within the last 20 years drafted on the New Orleans team this year, and we want to make sure that the indicators of his success in the NBA look promising.

But this project has much more breadth. This could examine:

* **Coaches & Scouts** that use general success statistics to access overall talent.
* **Video Game creators** to decide what number to quantify the skill of the player in sports video games such as the *NBA 2k Series* and the *NBA Live Series*
* **Advertisers** to figure out which potential superstars to target for a cheap price before they become superstars.
* **Small Market Teams** who can elect to select value players that have an effective skill-to-dollar-paid ratio.

***Conflicts of Interest***

There are no conflicts of interests. This project was not sourced nor paid for by any participating basketball organization nor any subjects of the study This will be unbiased, objective research only using quantifiable statistics to make empirical assessments.

Sources:

Coates, Dennis, University of Maryland, ResearchGate, *The Length and Success of NBA Careers: Does College Production Predict Professional Outcomes?* Oct 11th, 2019