Luke Gibson - Visualization Creator, Analyst
Tarun Malik - Analyst, Writer
Joshua May- Project Management, Data Administration, Webmaster, Analyst, Writer

INST 462 Final Project

Introduction

Our group focused on NBA statistics starting first by building a dataset, then exploring it for meaning. There have been debates about which era of basketball was better, who is the greatest to play the game, and what physical features favor players the most. These are the types of questions this site attempts to address.

We obtained our data by using Python to scrape Basketball-Reference.com and gather as many of the basic statistics(height, weight, points, etc.) we could. Starting with a complete list NBA players since the league's inaugural season. Our Python script resulted in 2608 rows and over 20 columns of data for the group of athletes. There is a lot to look at, and there is a tale to be told of high heights, low lows, and how both have their advantages in the history of the NBA;

Data

Data from Basketball-Reference is collected by a group of people who started the site and is constantly being updated by their team. They collect data from official NBA scorecards and scheduling but it is not exactly readily available for use. Our first challenge was creating a list of all the player names we wanted to scrape statistics for. This was an issue because we wanted a good spread of each era, but fortunately the list of all-time NBA players was easily acquired and we had a solid foundation to start on. As we dug deeper and wanted to see more, we encountered questions that required more data wrangling to be discussed below.

As we began looking at the data, more questions arose. We wondered if there was a way to look at the schools players had attended and plot out some numbers. This required finding a separate dataset for universities that included location details. Prioritizing and filtering what schools would be looked at was also important, as our set of players included over 300 schools, the colleges dataset had over 5000, and the names of schools in either were not uniform. Our solution was to only extract information for schools with 10 or more players in our dataset. This whittled the list down to approximately 80 academic institutions, that we audited for matching naming conventions. This made it much easier to merge the datasets using Python.

Delivery

We focused efforts on developing a web site to display our visuals which in turn creates a path for the reader to follow. It dually creates a medium that can convey all sorts of media, like this report itself, data visualizations, video, and more. We wanted to find a story within the NBA statistics that has written itself over time by some of the greatest athletes in the past 100 years. We relied on physical attributes such as height to try and make sense of what it takes to be an NBA player, and how height affects the success of a player statistically. We created different pages for our site, each with its own variable of analysis or a story point, using a global menu to make it easier for visitors to navigate the site for an engaging, user-friendly experience.

Findings

Of the things that stood out from our analysis, one thing speeds to the front of the line, if not captures attention before it bounces down the court. The skills of a player that has made it to the NBA may be predicted by their height. We found that taller players were more likely to have strong rebounding numbers, while shorter players were more likely to have strong assist numbers. This is shown through height distribution charts of subsets of data pertaining to a given statistic. With that, we also found that wins contributed by a player were only slightly correlated with taller heights.

https://inst462proj.herokuapp.com/