

Sahil Rajapkar
CIS3120 - Avinash Jairam
12/21/21

Project 2

For this project, it was amazing to see how integrated technologies could be. A few things that went into the implementation and success of this project include, but are not limited to, the following: web scraping, pandas data frames, APIs, data visualization, and much more. Some things going into the project that I had to consider were what my motivations were for this study, what data sources could be utilized and what were some justifications for them, what my findings suggested, how/why they were significant, and some visualizations to represent the overall idea and purpose. But, before I dive into everything, I would like to begin by speaking about my motivation.

Ever since this class started, I had seen data related to the weather, Spotify, and other topics of such but none tailored to sports, though an area of great interest. I chose to do work on the NBA and its players from past and present because I simply enjoy watching and playing the sport. But, I thought, what if I can do some analytics, what if I could create some valuable insight, what if I could do something that could help society. I knew that this would help me because I would be learning to program as well as conducting meaningful and applicable research and analytics. Hence, this just a bit that created a motivation for me. Other than that, I used some data sources such as tables listed on Wikipedia that I'd like to dive into next.

So, in the beginning, it was very difficult to scrape the data from Wikipedia, but I ended up doing so for the following categories: the 50 greatest NBA players in history, the 75th-anniversary team of NBA players, and a list of the NBA's greatest career scoring leaders. I chose these because I knew they had relevant data to my research and ones that also had some similarities. Whether it was through common players or statistics, I knew that I could make a viable product and project out of these categories. But, even within them, I knew I wanted to use the following columns to use later on with my NBA API: the player's full name, their position, how many points they have generated in their basketball career, and the number of All-NBA titles they've received.

The first thing that started my work with the NBA API was having a merged data frame to work with to make requests. This merged data frame was called data frame one. From this, I made sure to import all corresponding libraries and then create a data frame for the players within the NBA API. Then, I created data frame two which had new data that were joined by the commonalities in the merged data, and the NBA API data which would then be used to merge with just the merged data to create data frame three (see Figure 3.0). On this new data frame, I conducted descriptive statistics using the `.describe()` method and also converted it into a CSV file which showed the new data vividly in a new table that was created by yours truly and not shown on any webpage particularly. Nevertheless, I'd like to discuss my findings and why they're significant.

Regarding my findings, I found them to be especially interesting. On page 3 (Figure 3.1), you can see the result I received after using the `.describe()` method. Unfortunately, instead of 100 players that were supposed to be received, only 50 players were considered hence our count of 50. These were the 50 top players in NBA history. But, the significance of this is interesting because some things that stood out from these statistics were that the average amount of points for these players was roughly 25,000 total points, the maximum total points were close to 38,000 points, and the standard deviation was 4618 points. This goes to show that the length of a player's career does impact the number of points they earn, especially for those All-Time NBA titles to accompany them.

One of the first graphs I wanted to create would question if the number of all-time NBA titles would determine a player's points. The answer is yes! Yes, it does! But, do more all-NBA titles mean more NBA points? For this, the answer is no, not necessarily. As seen in Figure 3.2, statistics show that there is no correlation between titles and points because a player with 12 titles has fewer points than that who has only 6, half the amount! This goes to show it truly is subject to randomness and that numbers are determined primarily by the person and their pursuits as well as perhaps their position.

Next, I calculated the total amount of points by a player's position, rather than a random sample of the player's position. What do I mean by this? Well, I wanted to see if there were certain positions in the NBA that, historically, have produced more points for a team versus others. And, the answer was that there was! From my random sample of players and the use of the `.plot()` method (in Figure 3.3), I came to notice that the average guard (whether point guard or shooting guard) made significantly more points than all other positions. I wasn't very surprised to see this because players such as Stephen Curry and Michael Jordan have almost always been the leading scorer on their teams. While this may be true, does that put the title of a "guard" in the most frequent position of players in NBA history, or is there another position that takes up that spot.

The answer is the forward! Whether it be the small forward or power forward, from my random sample of players, it can be seen that "forward" is the position that takes up the majority of all player's positions of the top 50 in NBA history, with the center coming second, and the guard coming in third (as seen in Figure 3.4 as well).

In conclusion, I went about plotting my last sample through a pie chart, and the previous two with a bar chart and scatter plot. I found this to be helpful as I learned to program using different visualization techniques, but, more importantly, was able to conduct data analytics on NBA data. This was from the beginning what motivated me in pursuing this project and through that, I was to use different data sources with justifications and create findings that could be meaningful for NBA teams. Nevertheless, if there are a few key takeaways, they should dive into story-telling and always be ready for any opportunity! Because, to all NBA coaches and players, do you believe in me?

	Name	Position_x	Points_x
0	Kareem Abdul-Jabbar	C	38387
1	Nate Archibald	G	16481
2	Paul Arizin	F	16266
3	Charles Barkley	F	23757
4	Rick Barry	F	18395
5	Elgin Baylor	F	23149
6	Dave Bing	G	18327
7	Larry Bird	F	21791
8	Wilt Chamberlain	C	31419
9	Bob Cousy	G	16960

Figure 3.0

The description statistics of DF3:

	Name	Position_x	Points_x
count	10	10	10.0
unique	10	3	10.0
top	Charles Barkley	F	21791.0
freq	1	5	1.0

Figure 3.1

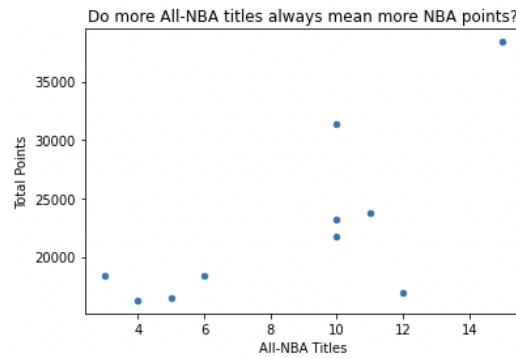


Figure 3.2

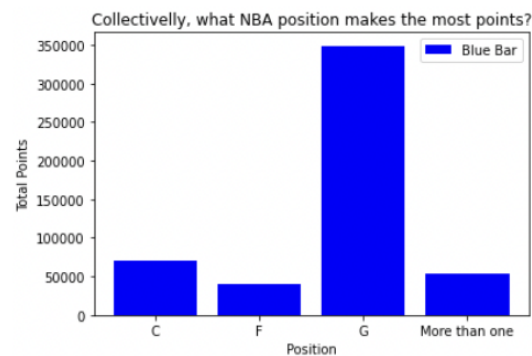


Figure 3.3

Of all the positions, which has highest amount of NBA players within the top 50 all-time?

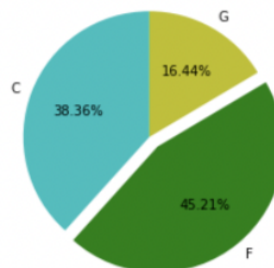


Figure 3.4