The subject I intend to explore is the NBA. More specifically, I will try to analyze the data to answer one of the most common arguments between fans: who is/was the best NBA player in history?

* Describe the data: What’s its schema? What columns and values does it have?

I will try to combine data from a couple of different sources. One is a Wikipedia page listing the 50 greatest players in NBA history. (The answer has to be one of them, right?) This table lists name, position, points, rebounds, assists, championships won, MVPs won, finals MVP won, and All-Star team awards. Another site, Basketball-Reference.com, has an endless number of tables to search. One of the most valuable resources on the site lists the results of each NBA draft. These tables contain many of the columns as the previously discussed site and adds some advanced metrics, most importantly Win Shares and Win Shares/48 minutes, as well as the columns detailing Draft Round and Pick Number as well as the college attended.

* How did you find it? I understand you will use a search engine but, describing publisher’s methods of exposing the data and how you consume it is really what I am looking for.

I found the Basketball-Reference site by accident. I was watching old basketball games and wondered which players were picked in the 1998 draft before one of my favorite players. This site was the top selection on the Google search. Once there, I saw that there were tabs to select the previous and next years’ draft and I realized I could grab all the tables and combine them into one.

* How do you intend to integrate it (technical steps)? How will you get it into POWERBI?

The first time I visited the Basketball-Reference site I found a download page but I haven’t been able to find it since. If I can’t find it again, I will copy/paste each table into Excel. The site notes that if one uses the data the site should be given credit so I take that as permission to use the data as long as I cite the source. The Wikipedia page should be able to import into PowerBI in the same way as in Lab 8—Importing and Analyzing Data from a Web Page.

* How will you use the data? What questions will it help answer?

I will compare the data to see if the raw numbers present a reasonable conclusion as to who is/was the best player ever. Other questions the data could help answer are who was the best pick of all time. This is not necessarily the best player. A top-10 player picked late could be considered a better pick than a top-10 player picked much earlier. Another question is what is the optimal draft position? The first overall pick is not always the best player in the draft. Which draft position has produced the most and best players?