NBA Analysis

Adonis Shareef
Fall Semester
https://github.com/adonisShareef/DataScience

Which Domain?

What domain is this data going to come from? Please list 10 references (with a brief annotation) to use to make sense of what you're doing with these data.

My domain is the sport of basketball and more specifically I will be targeting the NBA as it is the most competitive league in the world. The goal of this project will be to analyze current and past players to find the true GOAT or who has been the most efficient player thus far. I will be looking to use multiple data sets to do my analysis on because I want to make sure that I have a big enough pool of players not just from a certain year range.

1. https://onlinedsa.merrimack.edu/nba-analytics-changing-basketball/
2. https://www.delltechnologies.com/en-us/perspectives/tech-takes-the-court-nba-teams-turn-to-data-and-analytics-for-a-competitive-edge/
3. https://www.nba.com/stats/
4. https://projects.fivethirtyeight.com/nba-player-ratings/
5. https://medium.com/fastbreak-data/classifying-the-modern-nba-player-with-machine-learning-539da03bb824
6. https://www.thespax.com/nba/using-machine-learning-to-classify-nba-players-part-ii/
7. https://medium.com/@jbdave112/using-data-to-rank-the-top-15-players-in-nba-history-1862f15f3a0f
8. https://nbamath.com/progressive-goat-rankings/
9. https://www.sportingnews.com/us/nba/news/michael-jordan-vs-lebron-james-goat-debate/sl8xdozy5u1m1s4t5m3npeqo1
10. https://www.nba.com/stats/alltime-leaders/

Which Data?

What is the dataset you'll be examining? Please provide a codebook if there is one or a link to the dataset as well as a detailed description.

1. https://www.kaggle.com/jacobbaruch/basketball-players-stats-per-season-49-leagues
    - This data is not only NBA data but in other leagues so it has a lot less bias than just the NBA so I chose this one because I think I can get more accurate stats because an international data might have different stats than just straight from the NBA.
2. https://www.kaggle.com/justinas/nba-players-data
    - Another data set for NBA stats but this time just data with a good number of players not just the top players with few missing values so this will still be a good source to take from.

Research Questions? Benefits? Why analyze these data?

How are you proposing to analyze this dataset? This is about your approach. Here, you'll be proposing your research questions as well as justifications for why you'd offer these data in this way.

Who is the Greatest of All Time (GOAT) in the NBA?

I plan to use the stats from the NBA and other sources such as Olympic play to also determine the best of the best in the most competitive basketball league in the world. Using all stats like points, assists, rebounds, plus/minus stats and time played to determine who was the best. If a player has one great year but not career wise, then that will affect their ranking because the time you spend dominating also counts just like in other sports like boxing.

What Method?

What methods will you be using? What will those methods provide in terms of analysis? How is this useful?

Just from preliminary research I know I will be using clustering and from that cluster I could then use regression to determine who will come out on top almost like a tournament but instead of plaything the games their stats will be what takes them through. This will be useful because from the clusters I will be able to drop the players not in the highest cluster because those cannot win but then I just need to isolate the best players and from there start my tournament type of regression which will eventually lead me to the GOAT.

Potential Issues?

What challenges do you anticipate having? What could cause this project to go off schedule?

Some issues can be the data itself because with basketball sometimes the stats are a bit off from reality someone is always there to interpret the game and referees can say things you do in the game do not count or foul out. This is just a part of the game but as time went on there has been more and more player safety so things like free throws should not count as much as there were not as many in the past eras. Although their efficiency rating could counter act that we will just have to see. Not finding enough data that is unique will also be a problem because the players are going to be the same there is a limited amount of data to pull from so tracking down some more data might be a bit difficult.

Concluding Remarks

Tie it all together. Think of this section as your final report's abstract.

With the data I have collected so far and approach with model building and then evaluation I will be able to answer the long-awaited question of who the GOAT in basketball is. There will be some challenges but nothing that would keep me from completing the project within the given time frame. As the start of the NBA season is approaching us this is a good time to decide this because the stats are already marked in stone from previous seasons and some players will be returning, others have made it to retirement so they can be judged and go into this conversation, the goal being to no longer have this be a debate but fact.