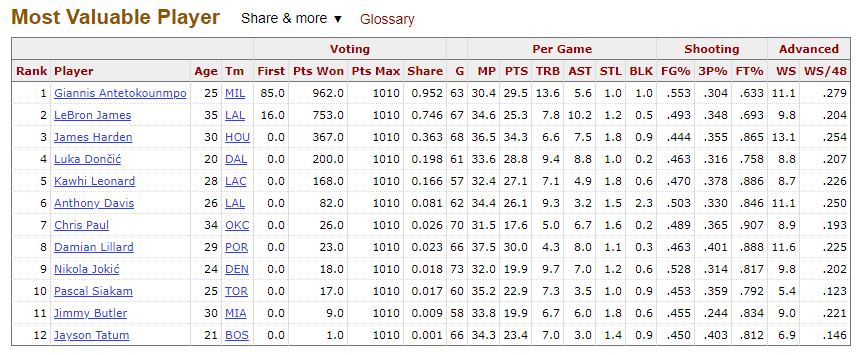
1. What is the main question that you are trying to answer? Be as specific and detailed as possible.

Every season in the NBA multiple players will receive votes for most valuable player (MVP). The question I am trying to answer is if I can use players’ and teams’ metrics to predict who wins MVP each year. Specifically, I want to be able to see if out of the players that received MVP votes, I can predict which one receives the highest share of votes.

1. What does your data look like?

I will be using two types of tables. The first is the table of the players who received MVP votes each year and the second is the table of the team metrics for that year. Here are examples from this past season (2019-2020):





* Where are you getting your data?
  + I am getting my data from basketball-reference.com
* What is the target?
  + The target is MVP voting share. I’ll specifically be using past data to predict this past year’s results. So the training data will be from years 2000-2019 and the testing data will be the 2019-2020 season.
* What are the features?
  + The features are most of the individual player and team stats from those tables. I will drop some of them that won’t be useful, such as the Games Behind (GB) metric for the teams.
* Are you planning on creating / engineering new features?
  + I don’t plan on creating many new features since I want to use numbers that are already being gathered by sports analysts so I’ll be primarily combining features from the two types of tables and dropping some that aren’t necessary or that represent the same thing (e.g. win and loss columns represent the same information as the W/L% column).
* How many observations do you / will you have?
  + I plan on using information from the 19 years prior to this past season. With around 12 players receiving votes each year, I should have around 230 observations.
* How does the data that you are collecting help you answer your question of interest?
  + The data I am using will hopefully provide the features that are most important when it comes to deciding the value of a player. Using these features I should be able to predict which player is deemed the “most valuable”.