**Final Project Report**

IST 652

Chang Liu

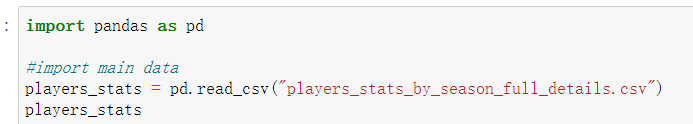
2020/12/7

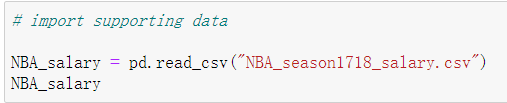
**Data and source**

I wanted to work with some data which relate to my interest, so I found a dataset from Kaggle: <https://www.kaggle.com/jacobbaruch/basketball-players-stats-per-season-49-leagues>. This data set contains data from 1999 to 2020. It includes players’ name, team, height, weight, shoots made, total shoots fired, and etc. Also, I found a supporting data set <https://www.kaggle.com/koki25ando/salary> from Kaggle too. This supporting dataset contains each player’s salary from 2017-2018 season.

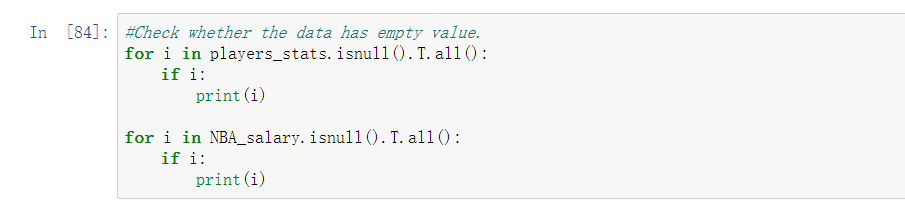
**Data cleaning, summarizing and combining**

First, import the main data and the supporting data.

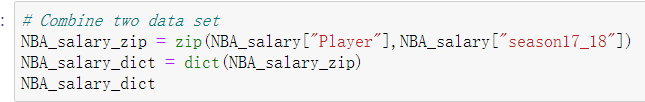


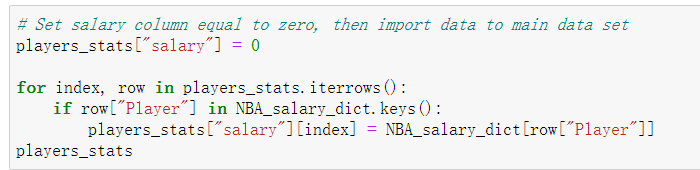


Then, let’s check whether the data contains empty values:

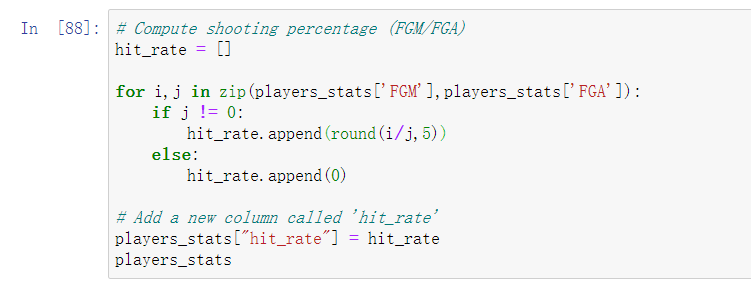


Next, combine two data set.



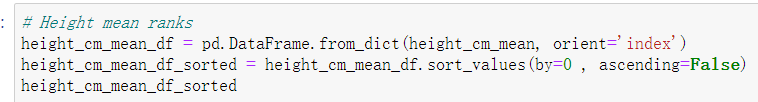


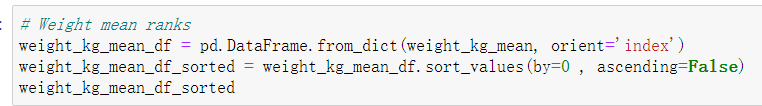
Add a column called “hit\_rate” which is the shooting percentage (FGM/FGA)

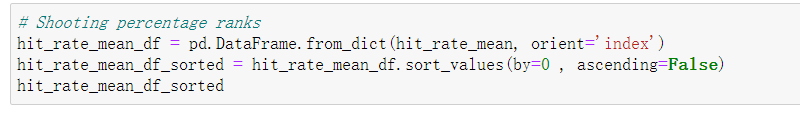


Summarize player’s data according to different basketball league



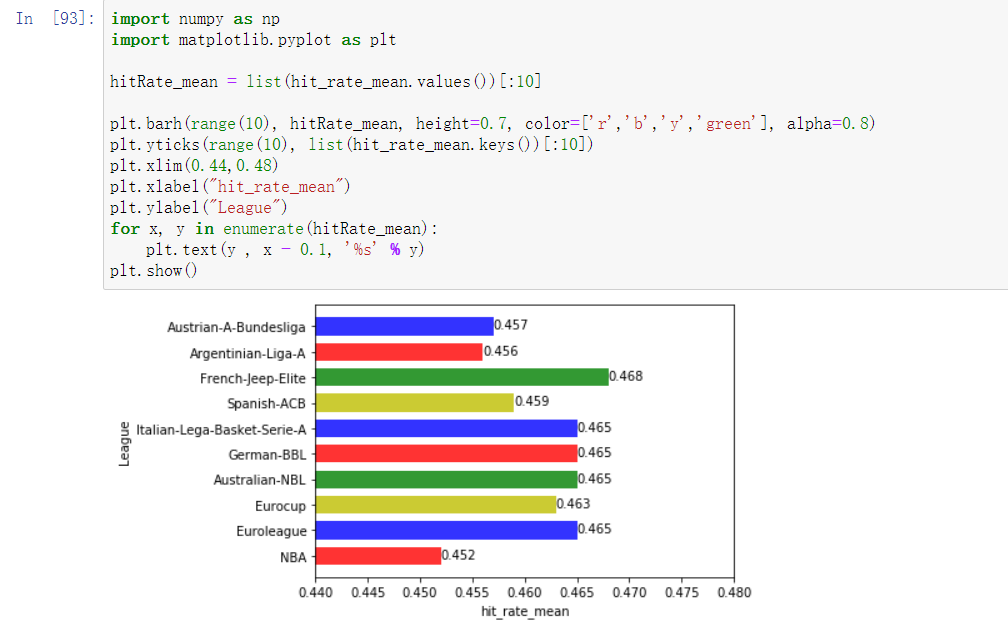


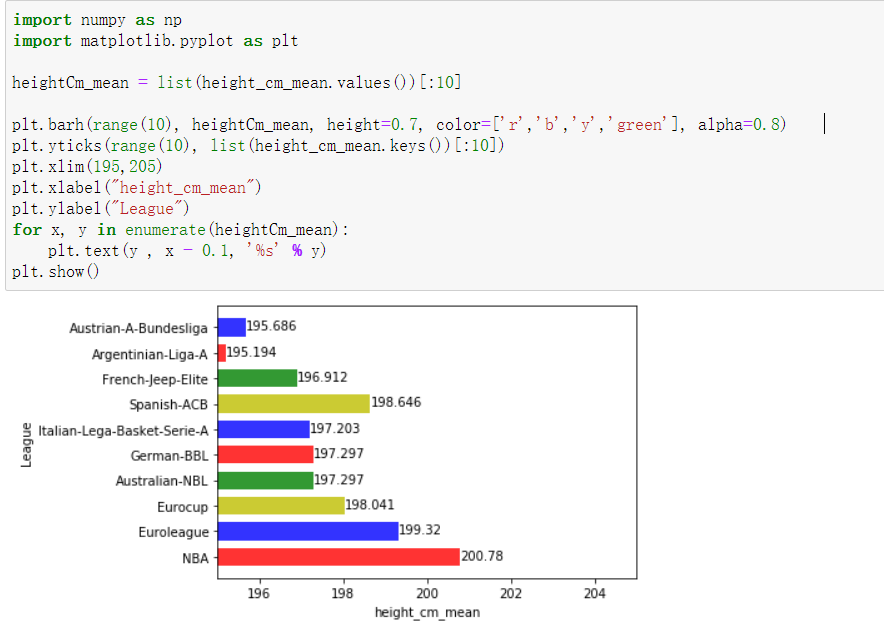


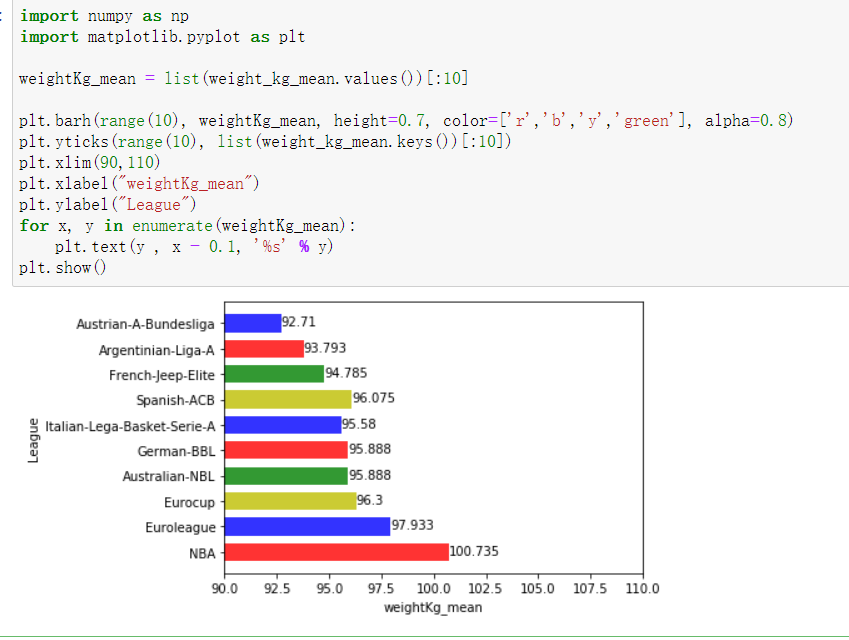


**Question one**

Why do U.S. basketball player perform better than other leagues? Is it because that players made more shots? Is it because players’ somatic data are better?







Basketball is a sport competition, there has been an intense confrontation. From these three graphs, we can see that NBA player’s body data is the best of all the leagues. Surprisingly, that shooting percentage of NBA is not dominant, but maybe this resulted to opposability which makes players harder to score.

**Question two**

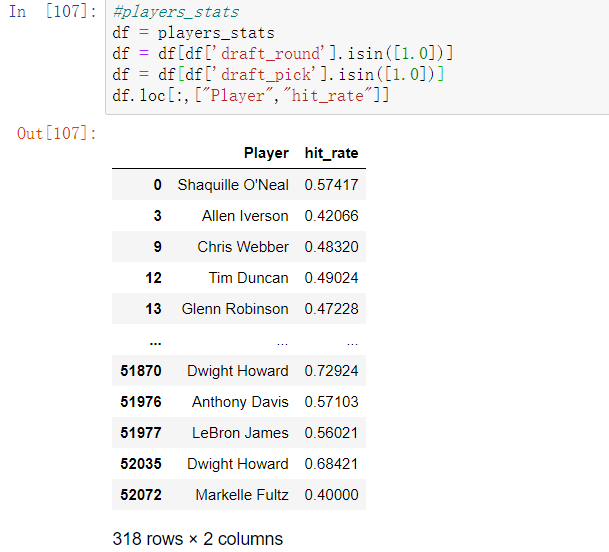
Which 10 players have the highest salary (2017-2018)? How were those players being drawn in the beginning?

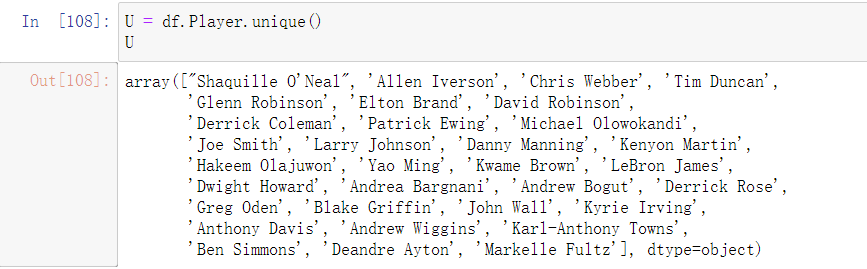


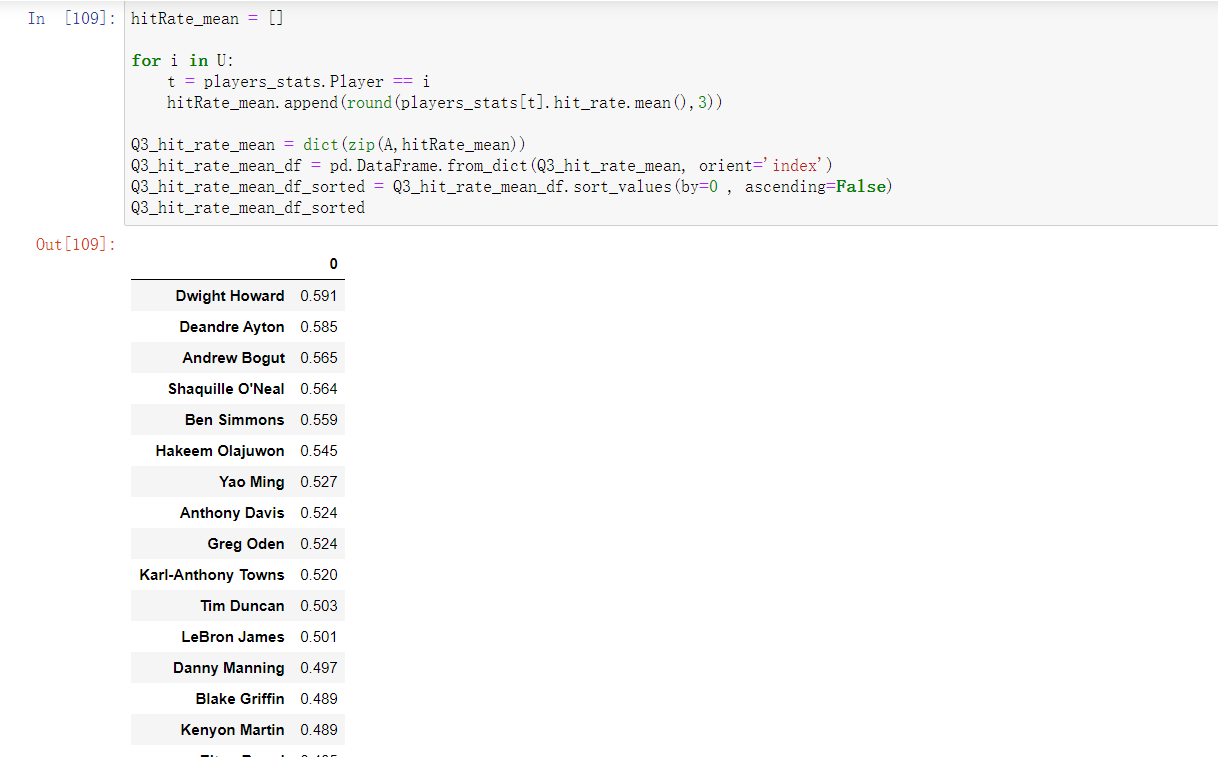
From the result, we see that Stephen Curry made the most, followed by LeBron James and Paul Millsap. When came to NBA, most of these players were being drawn during the 1st Round of pick, but not always the1st pick. I guess talent determine the baseline of a player, and the practice and workout are more important. Even 2nd round picked player (in this case, Millsap) can also make more money than 1st round picked players.

**Question three**

How does those Number One picked player perform?







This outcome is very interesting. I ranked the hit rate from high to low and we see that the top several Very First Picked player are Center, for example, Dwight Howard, Andrew Bogut, and Shaquille O’Neal. Maybe Centers are always close to the net, so that their shots are easier to go in. Also, this result tells us that, teams should pick Center players, because their shooting percent are higher than other positions.

Thank you

-The End-