Maxime Duverger, Maxime Thuillier, Bastien Her, Adrien Mareel

Python Project Report

NBA Scrapping

Group 8 December 2022

Summary:

1. Functionalities and how to use
2. Web scrapping methods
3. Data analysis & conclusions
4. Drawbacks & upgrades
5. Difficulties encountered

1) Functionalities of our app and how to use

Our app is composed of a single main.py file which is used to deal with scrapping, rendering data graphs. While running the main.py file, the program asks you which task you want to do. Those tasks are each one of the five plotting options asked in the project requirements. When one task is done the program comes back to the main function and asks again what to do. Therefore, while the program is running (not stopped), the program keeps looping on the select task function.

Our app can be used to see 5 different data & plots:

-The first one is just a plot linking the time played with the score for each player

-The second one is an heatmap linking again for each player the 3 points marked and the assist

-The third one is sorting all the player thanks to their team with a same color and is linking their points made for each one

-The fourth one is joining the two information 3points attempts and 3 points made, under the form of a kde graph

-The last one is linking the number of dunks tried and the number of the rebound done for each player in the team of UTAH

1. Web scrapping methods

The web scrapping part of our app is made with selenium.

We first go to the official NBA website to see the stats of the teams playing in the current season (here 2022/2023), we get the data stored in a table on the official website and put it in a 2D array, then we organize the array by adding a string key to all the types of stats to make it clear to finish we store the data in a json file, then we repeat the same process for the players with a different array containing different data, we organize them like we did for the team data and finally we store the player data in another separate json file.

1. Data analysis & conclusions

Our first function analyzes the average points scored per time played for each player. We see that the increase in playing time almost linearly increases the number of baskets scored, a logical and predictable result.

Our second function analyzes the 3 points scored per team based on assists. We see that for the player the correlation between the 2 stats is 0.34 so it’s not related.

Our third function analyzes the points scored by team based on free throws scored. We can say that there are about 7 times more points than baskets scored in free throws. There is not team that is above the other it’s just some players that stand out.

Our fourth function analyzes the 3 points attempted per team based on 3 points scored. Obviously, the more attempts you make in 3 points the more points you get in average, and we can see a line in this kind of graphs.

Our fifth function analyzes the team rebounds based on dunks attempted. We can see a correlation between these two stats that is logical because when you try to make a dunk even at a pro level you take risk to do a rebound.

in conclusion we can say that some statistics are not particularly interesting to consider (for example the 3 points scored based on the assists), however thanks to the analyzes that we have made we could use the correlations of some of them to optimize the forecasts of some meetings or even player training.

1. Drawbacks & upgrades

Getting the player data takes a lot of time (3 to 5 minutes)

We only get the data concerning the current season so our app currently can’t use data from previous seasons

Getting more data about players, like opening each file of player to get the age, their height, weight... to get more general tendance and more interesting graphs

Using more "team" info, because we mostly use the player's data and not the team's data that much

Finding idea of link, most of them were kind of obvious correlations

5) Difficulties encountered

We had a hard time to organize well the data before storing it in the json files and then use them.

We also tried to use the nba\_api but quickly discarded this idea has we often got timeout while getting the data, making the web scrapping part without using it was easier for us.