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:

* À décrire

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
2. 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

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.