

In a previous assignment we compared volleyball players' heights with those of swimmers. However, our analysis was restricted to teams belonging to a single campus. Thus, our sample space was quite small and our findings may not have been accurate!

Consequently, why not expand our sample space? The objective of this project is the similar to that of Homework #1. However, we are analyzing more data.

The CUNY Athletic Conference (CUNYAC) has 9 participating colleges. We will scrape the heights of men and women athletes from the volleyball and swimming teams from 5 colleges: Brooklyn College, Baruch College, York College, Queens College, and John Jay College.

Below are links to the various rosters.

Volleyball

Brooklyn College Men's Volleyball Team

<https://www.brooklyncollegeathletics.com/sports/mens-volleyball/roster/2019>

Brooklyn College Women's Volleyball Team

<https://www.brooklyncollegeathletics.com/sports/womens-volleyball/roster/2019>

Baruch College Men's Volleyball Team

<https://athletics.baruch.cuny.edu/sports/mens-volleyball/roster>

Baruch College Women's Volleyball Team

<https://athletics.baruch.cuny.edu/sports/womens-volleyball/roster>

York College Men's Volleyball Team

<https://yorkathletics.com/sports/mens-volleyball/roster>

John Jay College Women's Volleyball Team

<https://johnjayathletics.com/sports/womens-volleyball/roster>

Swimming

Brooklyn College Men's Swimming Team

<https://www.brooklyncollegeathletics.com/sports/mens-swimming-and-diving/roster>

Brooklyn College Women's Swimming Team

<https://www.brooklyncollegeathletics.com/sports/womens-swimming-and-diving/roster>

Baruch College Men's Swimming Team

<https://athletics.baruch.cuny.edu/sports/mens-swimming-and-diving/roster>

Baruch College Women's Swimming Team

<https://athletics.baruch.cuny.edu/sports/womens-swimming-and-diving/roster>

York College Men's Swimming Team

<https://yorkathletics.com/sports/mens-swimming-and-diving/roster>

Queens College Women's Swimming Team

<https://queensknight.com/sports/womens-swimming-and-diving/roster>

The height of each player is listed on all web pages.

1. Scrape data and compile a dataframe of all the names and heights of the players on the men's swimming team
2. Scrape data and compile a dataframe of all the names and heights of the players on the women's swimming team
3. Scrape data and compile a dataframe of all the names and heights of the players on the men's volleyball team
4. Scrape data and compile a dataframe of all the names and heights of the players on the women's volleyball team
5. Find the average height in each of the 4 dataframes (so you should have 4 averages in total)
6. List the names and the heights of the 5 tallest and the 5 shortest swimmers and volleyball players for both the men's and women's teams. That is you must have 8 lists in total: tallest men swimmers, tallest men volleyball players, tallest women swimmers, tallest women volleyball players, shortest men swimmers, shortest women volleyball players, shortest women swimmers, shortest women volleyball players,
7. Are you able to determine whether, in general, if the average swimmer is taller than the average volleyball player? Compare your findings in this project to those in homework #1. Write a 1 page report describing this.

Hints:

Inspect the html on each page listed above. Determine which tag and class point to the players' heights. Configure your web scraper accordingly. Follow the steps used in:

https://github.com/avinashjairam/avinashjairam.github.io/blob/master/project_example.ipynb

After you have scraped the heights and have stored them in different lists, you may have to convert the data (the heights) from strings to a numeric type (and then perhaps to centimeters or meters?) to find the average.

You may have to use a separate dataframe for each roster and then merge them into a single dataframe. For example, there are 3 rosters provided for athletes from the men's swimming team. Create 3 dataframes (one for each roster) and then merge the 3 into a single dataframe. This will allow you to easily find the average, etc. Repeat for the other categories.

Note:

The tasks listed here span many different topics in python. (There's a huge clue in the previous sentence!) This clue may not apply to all the rosters!

Submission

Submit your code and one page report via Blackboard.

Due: 04/10/2021 11:59PM EST.

There will be no extensions of the deadline regarding this project.

LATE SUBMISSIONS WILL RECEIVE A 30% PENALTY!

All submissions are final.

You have approximately one month to complete this project. START YOUR WORK EARLY!