

CIS 3120 Homework #1 Web Scraping

A marketing company would like to know if varsity (college) swimmers are (on average) taller than their volleyball counterparts. You have been asked to create a data driven solution in order to answer this question.

The following web pages contain the roster of the Bearcats' men's and women's swimming and volleyball teams.

Men's Swimming Team

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

Men's Volleyball Team

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

Women's Swimming Team

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

Women's Volleyball Team

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

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

1. Scrape the heights of all the players on the men's swimming team and find the average.
2. Scrape the heights of all the players on the men's volleyball team and find the average.
3. Scrape the heights of all the players on the women's swimming team and find the average.
4. Scrape the heights of all the players on the women's volleyball team and find the average.
5. Compare the averages between the two men's teams. Write a few lines about your findings.
6. Compare the averages between the two women's teams. Write a few lines about your findings.
7. Are you able to determine whether, in general, if the average swimmer is taller than the

average volleyball player? Write a few lines about your findings.

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.

Note:

The tasks listed here **span** many different topics in python. (There's a huge clue in the previous sentence!)

Submission

Submit your code and one page report via Blackboard.

Due: 10/16/2021 11:59PM EST.