

Shao Ru Zhang
CIS 3120- Programming for Analytics
Homework #1
6th March 2021

The purpose of this homework is to create a data driven solution in order for a marketing company to know if varsity swimmers are on average taller than their volleyball counterparts. Our goal is to utilize the Baruch men's and women's swimming and volleyball teams' roster pages to web scrape for their height and find the average heights in centimeters.

My approach to this is to first create a function called web scrape since we need to pass multiple website URLs through it. I started using one URL to scrape from, since it is easier to test. I used the men's swimming and diving roster website, right clicked and highlighted the first player's height, and clicked inspect. I noticed that the tag is span, and the class is called "sidearm-roster-player-height". I used the beautiful soup python package to parse the HTML webpage, so it is more readable. Within beautiful soup, I used the .find_all function to extract only the tag labeled span and a class named "sidearm-roster-players height". Next, I created a for loop, so it will extract all the height text from within that class and append it to the empty list. Then, I used list comprehension to split the list further into two separate lists. One list will only contain the feet and the other list contains the inches. It utilizes the ' within the first list to know where to split, for every ft' inch' within the list. For the inches list, we were still left with the " character, so I replaced the character with an empty space. Both lists are now parsed down to a simple list. I then used another for loop (nested), to run through both lists and convert them to integers, allowing me to multiply them by 30.48 and 2.54 to convert them into their respective cm. Finally, I found the sum of both lists, then divided them by the length of each list to find the average. Afterwards, I added the sum of both lists since they were originally part of the same height in one list, and found the average height of all players. Utilizing this function, I created a list of all the websites we must scrape from. Subsequently, I used another list comprehension to run each URL in the websites list through the function to find average height for each team.

My results were an average of 180.23 cm for the men's swimming team, 186.76 cm for men's volleyball team, 163.90 cm for women's swimming team, and 167.10 cm for women's volleyball team. In the men's team, the volleyball team is on average 6.53 cm taller than the swimming team. In the women's team, the volleyball team is on average 3.2 cm taller than the swimming team. In general, the average volleyball player is taller than the swimming team. This makes sense and is in line with my hypothesis, you need more speed in swimming rather than height since it's a race. In volleyball, taller players are needed to spike, attack, and block the ball successful over the net. The nets are high, so it will be more advantageous to be taller when you are in the front. Of course, we still have to keep in mind that these results may differ when scraping more data from different varsity teams.