Kabita Gurung

3120 - Project #1 - CIS 3120 Project #1

Date: 10-21-2023

*Introduction*

This project is designed to make scraper function take a list of URLs as input, scrapes information from these URLs, specifically athlete names and their heights, and returns a Pandas DataFrame with this information. It also calculates the mean height of the athletes and replaces any missing (NaN) values with this mean value. I also used other functions to find the top 5 tallest and shortest players on all four categories. I then sent all the four dataframes to csv file. After that, I made a dataframe to compare all four dataframes with their average. Lastly, I created bar plot to show average heights for all four dataframes.

*Techniques Used*

After analyzing the urls, I moved to collect and clean the data by creating scraper function and used for loop to make request to that url’s server. Then, I used Beautiful soup and imported raw html. I moved to use soup. find\_all to find sidearm-table-player-name tags and started for loop to extract text and appending it to a names with name.get\_text() to to extract text from the HTML and .strip('\n') to remove the newline characters. After extracting heights, I converted the heights to inches and append to the heights list. I then moved to create build dataframe and find the average to replacing the nans.

-*Methodology*

I used data collection method by using Beautiful soup by scraping the data from the HTML. I used data analysis and cleaning by extracting and cleaning the data to append it to 4 dataframes and further cleaned it by finding the average to fill the empty data. I then sampled and reviewed the results and then using the dataframe to filter the data as needed such as smallest and largest players. After reviewing the results, I created a new dataframe to compare the Average results. I then use bar plot for data visualization and to show the compared results and build report based on the result.

Here are some of the findings.

|  |  |  |
| --- | --- | --- |
| Team Category | All Rosters Avg ht. (inches) | Baruch’s Avg height (inches) |
| Womens Swim Team | 66.0 inches | 63.85 inches |
| Womens Volleyball Team | 68.0 inches | 67.25 inches |
| Mens Volleyball | 71.0 inches | 72.47 inches |
| Mens Swimming | 72.0 inches | 70.33 inches |

*Discussion* – As you can see the findings, Women’s Volleyball Team have higher average height compared to Swim Team and it is same with Baruch’s team as Ill which was generated on the first homework assignment. But on the contrary, Men’s Swimming team players have higher average height by an inch but not in Baruch where Men’s Volleyball team have higher average. So, in conclusion Women’s average height in Volleyball team are higher across different colleges.

Below is the bar plot to compare the average heights on all four categories.

A graph of a bar graph

Description automatically generated