MiG Shyaka

ECON 8320

In this project, we were tasked with collecting and organizing information about student-athlete NIL deals found on various websites. We focused on two websites, https://nilcollegeathletes.com/ and https://www.on3.com/os/, to gather data on each athlete's school, sport, sponsorships, and social media accounts.

To accomplish this, we used the Python libraries BeautifulSoup and Pandas to scrape and organize the data. We looped through 420 pages to collect athlete information and created a dataframe to store the data. We then created a new column for the athlete link and social media links and scraped sponsors and social media links for each athlete.

The final output of the code was a dataframe with columns for the athlete's name, sponsors, university, sport, Instagram, and Twitter. We exploded the sponsors column to ensure that each sponsor had its own row, and we excluded the Details column to create a more streamlined output.

One of the difficulties we encountered during the process was the issue of duplicate athlete names. Since the athlete link was created based on the first and last name, we had to ensure that we were linking to the correct athlete profile page. This required extra steps to verify the correct athlete link and ensure that the data scraped was accurate. Another difficulty of this process was knowing which HTML tags to use to extract the desired information. It required careful inspection of the website's HTML code to identify the relevant tags and their attributes. Additionally, some information was not readily available in the table on the main page, requiring further web scraping of individual athlete pages.

Despite these challenges, the resulting dataframe provided some interesting insights into the world of student-athlete NIL deals. For example, the top 5 universities with the most athletes with deals were all from the BigTen conference: Ohio State University, Michigan State University, University of Michigan, Penn State University, and University of Wisconsin. This suggests that universities with larger athletic programs and fan bases may have a competitive advantage in securing sponsorship deals for their athletes.

Furthermore, we observed that some athletes had multiple sponsorships, with some having up to 7 different sponsors. This highlights the potential financial opportunities available to student-athletes as a result of the recent NIL policy changes. In addition to these insights, the data also showed that certain sports had higher counts of student-athletes with NIL deals than others. Football, baseball, soccer, lacrosse, and basketball were the top sports with high counts of student-athletes with NIL deals, which is not surprising given the popularity and visibility of these sports

In the future, to improve this code, we could use the multiprocessing package to speed up the web scraping process. Additionally, we could consider expanding our data sources to include other websites that report on NIL deals for student-athletes.

In conclusion, this project provided valuable insights into the world of NIL deals for student-athletes and demonstrated the use of web scraping and data analysis techniques to collect and organize information. We learned how to extract valuable data from multiple websites and organize it into a cohesive output, while also overcoming challenges such as duplicate athlete names.