Keylia Carter

IST 652

Mini Project 2

**Web Scraping: An Analysis on the New England Patriots Roster v. Philadelphia Eagles**

**Section I: Web Scraping**

For this assignment, I used web scraping techniques in Python to parse the HTML data for the following teams: [New England Patriots](https://www.espn.com/nfl/team/roster/_/name/ne/new-england-patriots) and [Philadelphia Eagles](https://www.espn.com/nfl/team/roster/_/name/phi/philadelphia-eagles). After parsing the HTML data, I created a list that stores each of the children links found on the parent domain page. I then used the inspection tool for each URL to determine the class of each player which was also stored into a list then converted to a data frame. Before I could convert the list to a data frame, I created a for loop that iterated through my player list, withdraw every 8th element, and stored each player information into an empty list. The reason we decided to parse every 8th element is because there are 8 elements for each player found in the respective lists with the first one being a blank piece of information. After successfully storing the data, I began the data cleaning process.

**Section II: Data Cleaning**

In the data cleaning process, I started by transforming three columns. First, I removed all the characters found in the Wt (Weight) column which would ensure I could run summary statistics in the analysis portion. I also converted the Wt column to as type integer. The next transformation took place on the Ht (Height) column by replacing the punctuations with empty spaces. The last transformation took place on the Age column. I replaced any row with missing data or “--” with the mean imputation of the age column which happened to be 27.

**Section III: Information on the Data**

Each dataset contains 7 columns such as: Name, Pos (position), Age, Ht (height), Wt (weight), EXP (expected points), and College. There are 15 different positions such as: C (corner), CB (cornerback), DE (defensive end), DT (defensive tackle), G (guard), LB (linebacker), LS (long snapper), OT (offensive tackle), P (punter), PK (place kicker), QB (quarterback), RB (running back), S (safety), TE (tight end), WR (wide receiver)

**Section IIII: Data Exploration**

In my analysis, I ran a few descriptive statistics on both teams and found the average age of a New England Patriots player is 26 years old, the mean height is 6 ft.3 in, and the average weight is 242.5 pounds. In comparison, the average age of a Philadelphia Eagles player is 25 years old, the average height is 6ft 2 in, and the average weight is 244.67 pounds.

Next in my analysis, I began grouping by different attributes to determine the following for the New England Patriots and the Philadelphia Eagles:

1. **Average age for each position on the Patriots**

**Figure 1: Table 1:**

Chart, box and whisker chart

Description automatically generatedA picture containing text, meter, device

Description automatically generated

1. **Count of Players who Attended Each Universities for Patriots**

**Table 2: Table 3:**

A picture containing text

Description automatically generated **Text

Description automatically generated**

1. **Mean Weight for Each Position for Patriots**

**Figure 2: Table 4:**

**Chart, bar chart

Description automatically generated**Table

Description automatically generated

1. **Average Age for Each Position on the Eagles**

**Figure 3: Table 5:**

Chart, box and whisker chart

Description automatically generatedTable

Description automatically generated

1. **Count of Players who Attended Each Universities for the Eagles**

**Table 6: Table 7:**

**Text

Description automatically generated with low confidence Text

Description automatically generated with medium confidence**

1. **Mean Weight for Each Position for the Eagles**

**Figure 4: Table 8:**

**Chart, bar chart

Description automatically generated**Table

Description automatically generated

**Section V: Data Analysis**

Based on my findings, the Place Kicker position has the highest average age out of all the 15 positions for the New England Patriots while the position on the Philadelphia Eagles with the highest average age are the Punters and Long Snapper. I can infer these positions have higher age averages is because these positions do not encounter as much physical contact like tackling compared to other positions like defensive ends. Therefore, because there is less tackling for the three positions, they are more likely to play football for a longer period of time and refrain from retiring at early ages because their body does not experience as much injury and trauma.

Additionally, there is an outlier found in the Patriots Wide Receiver position at 37 years of age which is surprising because according to FantasyPros, “[wide receivers begin to decline in performance] around 33 or 34 years old. There’s also an outlier found in the Eagles’ Defensive End position at 36 years old and according to FantasyPros, “the average defensive end career lasts 3.3 years then are forced out the game—seeing retirement at 27 years old on average.”

Moving on, I then took a look at the count of players that attended each college for both teams. I found that the Patriots players mostly attended University of Alabama and LSU. In comparison, players on the Eagles mostly attended University of Alabama and University of Oklahoma. These findings do not surprise me because most drafted NFL players come from top perspective schools like Alabama, LSU, and Oklahoma. In fact, according to Jayna Baradhl, “Alabama tied with Georgia with 10 picks from the 2023 NFL draft, leading all other schools” ([The Athletic, 2023](https://theathletic.com/4471828/2023/04/29/nfl-draft-school-players-drafted/)).

Lastly, the position with the highest average weight from the New England Patriots is Offensive Tackle which is the same for the Philadelphia Eagles. These findings are not surprising considering the position and responsibility of these players. The job of the offensive tackle is to block players on the defensive line and prevent opposing players from reaching the Quarterback this way the Quarterback safely makes a pass or handoff.

**Section VI: Challenges and Limitations**

For this project, I faced a few challenges and limitations. My first encounter with a challenge began with struggling to parse every nth attribute for the players from each roster. I realized I was trying to parse every 7th element when it was supposed to be every 8th because there was an empty space that divided each player in the list. Another challenge I faced was casting some of the columns as floats and integers. I struggled with this because the data was imported from a list and Python would send an error message since the original form was in a string. One limitation I faced was not utilizing the “EXP” column because some of the data was a mix between integers and characters so I wouldn’t be able to parse each players expected point information. I also did not have time to look up this information for all ~185 players from both teams therefore I decided it was best to exclude this column from my analysis.