

WEB SCRAPING AND SOCIAL MEDIA SCRAPING

Final Project Description and Analysis

Temmuz Yavuzer (444130)

Merve Gulsum Kiratli (431674)

Scraping NFL Player Stats

Description of the Topic and the Web Page

The National Football League (NFL) is a professional American football league that consists of 32 teams. The NFL is one of the major North American professional sports leagues and the highest professional level of American football in the world. Each NFL season begins with a three-week preseason in August, followed by the eighteen-week regular season, which runs from early September to early January. Each team plays seventeen games and has one bye week. Following the conclusion of the regular season, seven teams from each conference (four division winners and three wild card teams) advance to the playoffs, a single-elimination tournament that culminates in the Super Bowl.

Ways To Score In Football	How Many Points
Touchdown	6 Points
Field Goal	3 Points
Safety	2 Points
2 Point Conversion (After a Touchdown)	2 Points
Point After Touchdown (PAT)	1 Point

The goal of the project is to scrape results for NFL player stats, especially for the players combined seasons, age 26 or younger, who were drafted between 1936 and 2021, sorted by descending Passing TD. Because of the points scored by touchdowns could lead to getting 6 to 8 points which are most crucial for a team to win the game.

The data to be scraped includes basic information about the player's career stats, such as Games Played (GS), Approximate Value (AV), Team record in-game started by this QB (QBrec), Percentage of Successful Passes (Cmp%), Yards gained by passing(Yds), Yards gained by pass attempts(Y/A), Passing Touchdowns (TD), Interceptions Thrown (Int), Fantasy Points (FantPt)

```
{0: {'player_name': 'Dan Marino', 'G': '242', 'AV': '216', 'QBrec': '147-93-0', 'Cmp%': '59.4', 'Yds': '61361', 'Y/A': '7.3', 'TD': '420', 'Int': '252', 'FantPt': '3594.5'}, 1: {'player_name': 'Patrick Mahomes', 'G': '63', 'AV': '75', 'QBrec': '50-13-0', 'Cmp%': '66.1', 'Yds': '18991', 'Y/A': '8.1', 'TD': '151', 'Int': '37', 'FantPt': '1450.6'}, 2: {'player_name': 'Peyton Manning', 'G': '266', 'AV': '271', 'QBrec': '186-79-0', 'Cmp%': '65.3', 'Yds': '71940', 'Y/A': '7.7', 'TD': '539', 'Int': '251', 'FantPt': '4686.3'}, 3: {'player_name': 'Matt Leinart', 'G': '182', 'AV': '148', 'QBrec': '86-95-1', 'Cmp%': '63.0', 'Yds': '49995', 'Y/A': '7.3', 'TD': '323', 'Int': '161', 'FantPt': '3137.7'}, 4: {'player_name': 'Drew Bledsoe', 'G': '194', 'AV': '139', 'QBrec': '98-95-0', 'Cmp%': '57.2', 'Yds': '44611', 'Y/A': '6.6', 'TD': '251', 'Int': '206', 'FantPt': '2412.0'}, 5: {'player_name': 'Jameis Winston', 'G': '83', 'AV': '64', 'QBrec': '33-44-0', 'Cmp%': '61.2', 'Yds': '20982', 'Y/A': '7.7', 'TD': '135', 'Int': '91', 'FantPt': '1353.8'}, 6: {'player_name': 'Cam Newton', 'G': '148', 'AV': '148', 'QBrec': '148-93-0', 'Cmp%': '61.2', 'Yds': '20982', 'Y/A': '7.7', 'TD': '135', 'Int': '91', 'FantPt': '1353.8'}}
```

The starting page to be used in this project is;

<https://stathead.com/tiny/tTuxM>

then the codes will go to each player profile to scrape their data such as; <https://www.pro-football-reference.com/players/M/MariDa00.htm>

Short description of your scraper mechanics

Selenium:

Code starts from the main page

Then navigate to each player in the first list

After that, it goes to each player's profile

Scrape their career stats

Print it on the console as a dictionary

Scrapy:

Code starts from the main page

Takes all of the players urls and store it into a text file called info.txt

After that, it goes to each player's profile using those urls

Stores the html files of each player inside of folder called all_data

Scrape their career stats

Print it on the console as a dictionary

BeautifulSoup:

Code starts from the main page ('https://stathead.com/tiny/tTuxM')

Takes all the players urls in main page and store it into an array which called as "links"

Scrape players' career information from the html text of the players urls

Print it on the console as a dictionary

Store this information to the 'player.csv' file

Extremely elementary data analysis

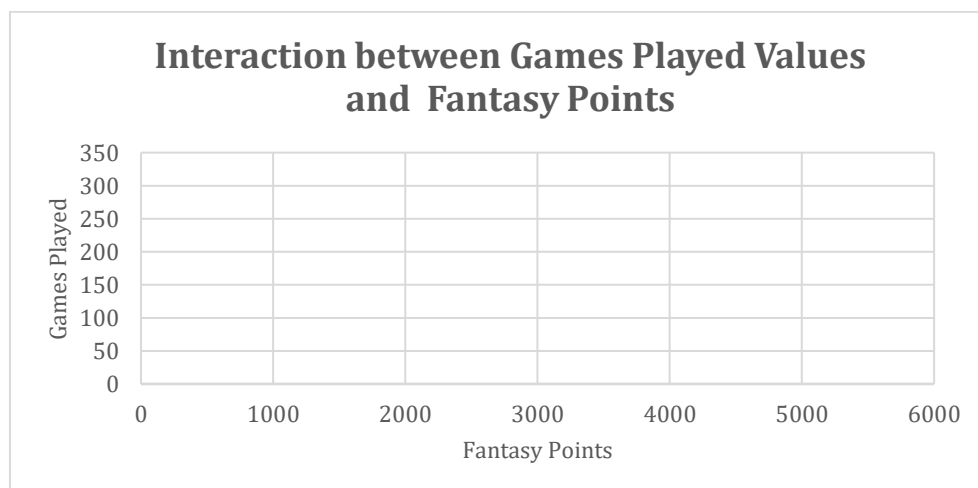
	Name	G	AV	QBrec	Cmp%	Yds	Y/A	TD	Int	FantPt
0	Dan Marino	242	216	147-93-0	59.4	61361	7.3	420	252	3594.5
1	Patrick Mahomes	63	75	50-13-0	66.1	18991	8.1	151	37	1450.6
2	Peyton Manning	266	271	186-79-0	65.3	71940	7.7	539	251	4686.3
3	Matthew Stafford	182	148	86-95-1	63.0	49995	7.3	323	161	3137.7
4	Drew Bledsoe	194	139	98-95-0	57.2	44611	6.6	251	206	2412.0
..
95	Kyle Boller	67	17	20-27-0	56.7	8931	5.9	48	54	478.5
96	Daniel Jones	38	25	12-25-0	62.8	8398	6.6	45	29	561.5
97	Jim McMahon	119	71	67-30-0	58.0	18148	7.1	100	90	1194.3
98	Philip Rivers	244	218	134-106-0	64.9	63440	7.8	421	209	3812.8
99	Pat Haden	65	36	35-19-1	53.6	9296	6.8	52	60	535.7

[100 rows x 10 columns]

When we look at the data that is scraped, we see that the player position that contributes the most to the game is the quarterbacks. The ratio of this is 98 of the total 100 players who help their team to score more points via touchdowns are quarterbacks. The scraped data can help determine which of the players who are still in their careers will be elected to the hall of fame in the future. The best example of this is Patrick Mahomes, as the 2nd player in history to have the most touchdowns for players in the first-four-years of their career. This type of data is widely used by NFL teams. In fact, some teams make their transfers completely by looking at those type of data.

Moreover, when we analyze the relationship between columns, there is a linear trend between the games played and fantasy point. In the data, Fantasy Points is calculated as shown below:

- 1 point per 25 yards passing
- 4 points per passing touchdown
- -2 points per interception thrown
- 1 point per 10 yards rushing/receiving
- 6 points per TD
- 2 points per two-point conversion
- -2 points per fumble lost



This graph is totally logical because it is most likely happen that which is when people play more games, they have more chances to increase their fantasy points.

Division of the Work

Analysis and short description - Merve Gulsum Kiratli

BeautifulSoup - Merve Gulsum Kiratli

Scrapy - Temmuz Yavuzer

Selenium - Temmuz Yavuzer