DSC540Brian_Reppeto_540_week_7_8_Milestone_3

February 3, 2024

0.0.1 DSC 540 Week 2 Data Wrangling with Python:

0.0.2 Project: Milestone 3

0.0.3 Author: Brian Reppeto 2/1/2024

[23]:
```
# import libraries

import pandas as pd
from urllib.request import urlopen
from bs4 import BeautifulSoup
```

[24]:
```
# create a list to store all player stats

all_stats = []
```

[25]:
```
# scrape the data using a for loop for the range of years wanted

for year in range(2017, 2024):
    url = "https://www.pro-football-reference.com/years/{}/fantasy.htm".
 ↪format(year)
    html = urlopen(url)
    soup = BeautifulSoup(html, 'html.parser')

    # extract headers from the second row of the table

    headers = [th.getText() for th in soup.findAll('tr')[1].findAll('th')]
    headers = headers[1:]  # Exclude the first (0 index) column header

    # extract player statistics from the table

    rows = soup.findAll('tr', class_=lambda table_rows: table_rows != "thead")
    player_stats = [[td.getText() for td in rows[i].findAll('td')]
                    for i in range(len(rows))]
    player_stats = player_stats[2:]

    # create a df for the current year

    stats = pd.DataFrame(player_stats, columns=headers)
```

```
    # add a new column 'Year' and assign the year pulled to each row
    stats['Year'] = year
    # print the first 5 player statistics for the year
    print(f"Player statistics for the year {year}:\n{stats.head()}")
    # append the current year stats to the overall list
    all_stats.append(stats)
    # close the HTML connection
    html.close()
```

```
Player statistics for the year 2017:
           Player   Tm FantPos Age   G  GS Cmp Att Yds TD  … 2PM 2PP FantPt  \
0   Todd Gurley*+  LAR      RB  23  15  15   0   0   0  0  …              319
1  Le'Veon Bell*+  PIT      RB  25  15  15   0   0   0  0  …              257
2    Kareem Hunt*  KAN      RB  22  16  16   0   0   0  0  …              242
3   Alvin Kamara*  NOR      RB  22  16   3   0   0   0  0  …   1          239
4   Melvin Gordon  LAC      RB  24  16  16   0   0   0  0  …              230

     PPR   DKPt   FDPt  VBD PosRank OvRank  Year
0  383.3  391.3  351.3  192       1      1  2017
1  341.6  349.6  299.1  130       2      2  2017
2  295.2  302.2  268.7  115       3      3  2017
3  320.4  327.4  279.9  112       4      4  2017
4  288.1  294.1  259.1  103       5      5  2017

[5 rows x 33 columns]
Player statistics for the year 2018:
                Player   Tm FantPos Age   G  GS  Cmp  Att   Yds  TD  … 2PM  \
0        Todd Gurley*+  LAR      RB  24  14  14    0    0     0   0  …   3
1      Saquon Barkley*  NYG      RB  21  16  16    0    0     0   0  …   1
2  Christian McCaffrey  CAR      RB  22  16  16    1    1    50   1  …
3        Alvin Kamara*  NOR      RB  23  15  13    0    0     0   0  …   3
4    Patrick Mahomes*+  KAN      QB  23  16  16  383  580  5097  50  …   1

  2PP FantPt    PPR   DKPt   FDPt  VBD PosRank OvRank  Year
0        313  372.1  379.1  342.6  181       1      1  2018
1        295  385.8  391.8  340.3  163       2      2  2018
2        279  385.5  392.5  332.0  146       3      3  2018
3        273  354.2  360.2  313.7  141       4      4  2018
4        417  417.1  437.1  429.1  134       1      5  2018
```

```
[5 rows x 33 columns]
Player statistics for the year 2019:
                  Player   Tm FantPos Age   G  GS  Cmp  Att   Yds  TD  …  \
0  Christian McCaffrey*+  CAR      RB  23  16  16    0    2     0   0  …
1        Lamar Jackson*+  BAL      QB  22  15  15  265  401  3127  36  …
2         Derrick Henry*  TEN      RB  25  15  15    0    0     0   0  …
3            Aaron Jones  GNB      RB  25  16  16    0    0     0   0  …
4       Ezekiel Elliott*  DAL      RB  24  16  16    0    0     0   0  …

  2PM 2PP FantPt    PPR   DKPt   FDPt  VBD PosRank OvRank  Year
0   1        355  471.2  477.2  413.2  215       1      1  2019
1            416  415.7  429.7  421.7  152       1      2  2019
2            277  294.6  303.6  285.6  136       2      3  2019
3            266  314.8  322.8  290.3  125       3      4  2019
4            258  311.7  319.7  284.7  117       4      5  2019

[5 rows x 33 columns]
Player statistics for the year 2020:
            Player   Tm FantPos Age   G  GS Cmp Att Yds TD  … 2PM 2PP  \
0  Derrick Henry*+  TEN      RB  26  16  16   0   0   0  0  …   1
1    Alvin Kamara*  NOR      RB  25  15  10   0   0   0  0  …
2     Dalvin Cook*  MIN      RB  25  14  14   0   0   0  0  …   3
3  Davante Adams*+  GNB      WR  28  14  14   0   0   0  0  …
4   Travis Kelce*+  KAN      TE  31  15  15   1   2   4  0  …   1

  FantPt    PPR   DKPt   FDPt  VBD PosRank OvRank  Year
0    314  333.1  341.1  323.6  184       1      1  2020
1    295  377.8  383.8  336.3  165       2      2  2020
2    294  337.8  346.8  315.8  164       3      3  2020
3    243  358.4  362.4  300.9  117       1      4  2020
4    208  312.8  316.8  260.3  117       1      5  2020

[5 rows x 33 columns]
Player statistics for the year 2021:
              Player   Tm FantPos Age   G  GS  Cmp  Att   Yds  TD  … 2PM  \
0  Jonathan Taylor*+  IND      RB  22  17  17    0    0     0   0  …
1      Cooper Kupp*+  LAR      WR  28  17  17    0    1     0   0  …   1
2     Deebo Samuel*+  SFO      WR  25  16  15    1    2    24   1  …
3         Josh Allen  BUF      QB  25  17  17  409  646  4407  36  …   2
4      Austin Ekeler  LAC      RB  26  16  16    0    0     0   0  …   2

  2PP FantPt    PPR   DKPt   FDPt  VBD PosRank OvRank  Year
0         333  373.1  381.1  353.1  176       1      1  2021
1         295  439.5  442.5  367.0  163       1      2  2021
2         262  339.0  347.0  300.5  132       2      3  2021
3   1     403  402.6  426.6  417.6  126       1      4  2021
4         274  343.8  352.8  308.8  120       2      5  2021
```

```
[5 rows x 33 columns]
Player statistics for the year 2022:
                Player   Tm FantPos Age   G  GS  Cmp  Att   Yds  TD  … 2PM  \
0     Patrick Mahomes*+  KAN      QB  27  17  17  435  648  5250  41  …   1
1         Josh Jacobs*+  LVR      RB  24  17  17    0    0     0   0  …
2  Christian McCaffrey*  2TM      RB  26  17  16    1    1    34   1  …
3        Derrick Henry*  TEN      RB  28  16  16    2    2     4   1  …
4    Justin Jefferson*+  MIN      WR  23  17  17    2    2    34   0  …   1

  2PP FantPt    PPR   DKPt   FDPt  VBD PosRank OvRank  Year
0   2    416  417.4  435.4  428.9  136       1      1  2022
1        275  328.3  335.3  301.8  127       1      2  2022
2        271  356.4  362.4  313.9  123       2      3  2022
3        270  302.8  311.8  286.3  122       3      4  2022
4        241  368.7  371.7  304.7  119       1      5  2022

[5 rows x 33 columns]
Player statistics for the year 2023:
                 Player   Tm FantPos Age   G  GS  Cmp  Att   Yds  TD  …  \
0  Christian McCaffrey*+  SFO      RB  27  16  16    0    0     0   0  …
1          CeeDee Lamb*+  DAL      WR  24  17  17    0    0     0   0  …
2             Josh Allen  BUF      QB  27  17  17  385  579  4306  29  …
3          Tyreek Hill*+  MIA      WR  29  16  16    0    0     0   0  …
4           Jalen Hurts*  PHI      QB  25  17  17  352  538  3858  23  …

  2PM 2PP FantPt    PPR   DKPt   FDPt  VBD PosRank OvRank  Year
0             324  391.3  399.3  357.8  157       1      1  2023
1   1         268  403.2  411.2  335.7  131       1      2  2023
2       3     393  392.6  420.6  410.6  122       1      3  2023
3             257  376.4  380.4  316.9  120       2      4  2023
4             357  356.8  382.8  371.8   89       2      5  2023

[5 rows x 33 columns]
```

Step 1 Concat all df's into one df

[26]:
```
# concat all df's into one df for all years

all_years_stats = pd.concat(all_stats, ignore_index=True)
```

[27]:
```
# print the new df

print(all_years_stats)
```

```
              Player   Tm FantPos Age   G  GS Cmp Att Yds TD  … 2PM 2PP  \
0      Todd Gurley*+  LAR      RB  23  15  15   0   0   0  0  …
1     Le'Veon Bell*+  PIT      RB  25  15  15   0   0   0  0  …
2       Kareem Hunt*  KAN      RB  22  16  16   0   0   0  0  …
```

```
3      Alvin Kamara*  NOR      RB  22  16   3   0   0   0  0  …   1
4      Melvin Gordon  LAC      RB  24  16  16   0   0   0  0  …
…                …  …     …  ..  ..  ..  ..  ..  ..  ..  …  ..  ..
4466     Kyle Allen  BUF      QB  27   7   0   0   0   0  0  …
4467    Deon Jackson  3TM      RB  24   4   1   0   0   0  0  …
4468     David Wells  TAM      TE  28   5   0   0   0   0  0  …
4469    James Proche  CLE      WR  27  10   1   0   0   0  0  …
4470    Trent Taylor  CHI      WR  29  17   0   0   0   0  0  …

     FantPt    PPR   DKPt   FDPt  VBD PosRank OvRank  Year
0       319  383.3  391.3  351.3  192       1      1  2017
1       257  341.6  349.6  299.1  130       2      2  2017
2       242  295.2  302.2  268.7  115       3      3  2017
3       239  320.4  327.4  279.9  112       4      4  2017
4       230  288.1  294.1  259.1  103       5      5  2017
…     …    …    …    …  …     …    …   …
4466     -1   -1.3   -1.3   -1.3           83         2023
4467     -1    4.0    6.0    1.5          168         2023
4468     -1    1.0    1.0                 140         2023
4469     -2   -2.0   -1.0   -2.0          242         2023
4470     -2   -2.2   -1.2   -2.2          243         2023

[4471 rows x 33 columns]
```

Step 2 make all the column headings in upper letters

[28]:
```
# make all the column headings in upper letters

all_years_stats.columns= all_years_stats.columns.str.upper()
```

[29]:
```
# head the data frame to see the changed upper column headings

all_years_stats.head()
```

[29]:
```
           PLAYER   TM FANTPOS AGE   G  GS CMP ATT YDS TD  … 2PM 2PP FANTPT  \
0   Todd Gurley*+  LAR      RB  23  15  15   0   0   0  0  …              319
1  Le'Veon Bell*+  PIT      RB  25  15  15   0   0   0  0  …              257
2    Kareem Hunt*  KAN      RB  22  16  16   0   0   0  0  …              242
3   Alvin Kamara*  NOR      RB  22  16   3   0   0   0  0  …   1          239
4   Melvin Gordon  LAC      RB  24  16  16   0   0   0  0  …              230

     PPR   DKPT   FDPT  VBD POSRANK OVRANK  YEAR
0  383.3  391.3  351.3  192       1      1  2017
1  341.6  349.6  299.1  130       2      2  2017
2  295.2  302.2  268.7  115       3      3  2017
3  320.4  327.4  279.9  112       4      4  2017
4  288.1  294.1  259.1  103       5      5  2017
```

```
[5 rows x 33 columns]
```

Step 3

[30]:
```
# removing special characters after the players' names

all_years_stats['PLAYER'] = all_years_stats['PLAYER'].str.replace('[*+]', '',␣
 ↪regex=True)
```

[31]:
```
# head the data frame to see what the new file will look like

all_years_stats.head(5)
```

[31]:

	PLAYER	TM	FANTPOS	AGE	G	GS	CMP	ATT	YDS	TD	…	2PM	2PP	FANTPT	\
0	Todd Gurley	LAR	RB	23	15	15	0	0	0	0	…			319	
1	Le'Veon Bell	PIT	RB	25	15	15	0	0	0	0	…			257	
2	Kareem Hunt	KAN	RB	22	16	16	0	0	0	0	…			242	
3	Alvin Kamara	NOR	RB	22	16	3	0	0	0	0	…	1		239	
4	Melvin Gordon	LAC	RB	24	16	16	0	0	0	0	…			230	

	PPR	DKPT	FDPT	VBD	POSRANK	OVRANK	YEAR
0	383.3	391.3	351.3	192	1	1	2017
1	341.6	349.6	299.1	130	2	2	2017
2	295.2	302.2	268.7	115	3	3	2017
3	320.4	327.4	279.9	112	4	4	2017
4	288.1	294.1	259.1	103	5	5	2017

```
[5 rows x 33 columns]
```

Step 4

[32]:
```
# remove unneeded columns

columns_to_remove = ["CMP", "ATT", "YDS", "TD", "INT","ATT","Y/
 ↪A","TGT","REC","Y/R","FMB","FL","2PM","2PP","PPR","DKPT","FDPT","VBD"]
all_years_stats = all_years_stats.drop(columns=columns_to_remove,␣
 ↪errors='ignore')
```

[33]:
```
# head the data frame to see what the new file will look like

all_years_stats.head(5)
```

[33]:

	PLAYER	TM	FANTPOS	AGE	G	GS	FANTPT	POSRANK	OVRANK	YEAR
0	Todd Gurley	LAR	RB	23	15	15	319	1	1	2017
1	Le'Veon Bell	PIT	RB	25	15	15	257	2	2	2017
2	Kareem Hunt	KAN	RB	22	16	16	242	3	3	2017
3	Alvin Kamara	NOR	RB	22	16	3	239	4	4	2017
4	Melvin Gordon	LAC	RB	24	16	16	230	5	5	2017

Step 5

[34]:
```
# save the cleaned data to a new CSV file

output_file_path = '/Users/brianreppeto/DSC540/scraped_data_cleaned.csv'
all_years_stats.to_csv(output_file_path, index=False)
```

[35]:
```
# head the data frame to see what the new file will look like

all_years_stats.head()
```

[35]:

	PLAYER	TM	FANTPOS	AGE	G	GS	FANTPT	POSRANK	OVRANK	YEAR
0	Todd Gurley	LAR	RB	23	15	15	319	1	1	2017
1	Le'Veon Bell	PIT	RB	25	15	15	257	2	2	2017
2	Kareem Hunt	KAN	RB	22	16	16	242	3	3	2017
3	Alvin Kamara	NOR	RB	22	16	3	239	4	4	2017
4	Melvin Gordon	LAC	RB	24	16	16	230	5	5	2017

Ethical Implications

The use of NFL fantasy stats raises several ethical implications, both in terms of the data collection and the impact on individuals involved. Here are a few of those items:

1. Privacy Concerns for the players around injurys.
2. Accuracy of Stats for indivuals intrested in the sport.
3. Possiblly encouraging gambling.
4. Financial Impact, in terms of the cost to play the game.
5. Exploitation of Athletes. I have heard of atheltes attempting to play as they knew social media would be an issue if they sat the game.

[ ]: