



Soccer Team

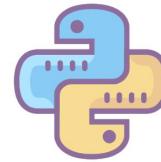


Scraping Team



Scraping Overview

- Goal: Obtain data for the analytics team to use
- Focused on specific data points: goal %, total goals, passes,etc.
- Scrapped data dating back 5 years up to 2018 season
- Used tools such as Python, Pandas, and BeautifulSoup





Louis

- Scraped soccerment.com
- Used a list of URLs to get data from different years into one csv file
- Would replace URL stat value with whatever data analytics team needed
- Created csv files for 15 different stats

```
urls = [requests.get("https://analytics.soccerment.com/en/League/premier_league/players/stat/Successful%20Dribbles?season=2021"),
        requests.get("https://analytics.soccerment.com/en/League/premier_league/players/stat/Successful%20Dribbles?season=2020"),
        requests.get("https://analytics.soccerment.com/en/League/premier_league/players/stat/Successful%20Dribbles?season=2019"),
        requests.get("https://analytics.soccerment.com/en/League/premier_league/players/stat/Successful%20Dribbles?season=2018")]

counter = 0;
with open('SuccessfulDribbles.csv', 'w', newline='') as file:
    writer = csv.writer(file)
    xvalue = ["Name", "Team name", "Successful Dribbles", "Year"]
    writer.writerow(xvalue)
    for url in urls:
        soup = BeautifulSoup(url.text, 'html.parser')

        alldata = soup.find("div", attrs = {'class': 'card mb10'})
        num1= alldata.find("div", attrs = {'class': 'top_element'}).text
        li = num1.split('\n')

        name = li[1]
        teamName=li[3]
        data = li[5]
        if not li[5].isnumeric():
            data = 0
        value = [name, teamName, data, 2021 - counter]

        print(value)
        writer.writerow(value)

        temp = alldata.find_all("tr")
```



Yongha

540	541	Jordan Zemura	zw ZIM	DF	Boumbohem	23-112	19	17	1454	0	1	1	0	0	1	0	0.4	19
541	542	Oleksandr Zinchenko	ua UKR	DF	Arsenal	26-081	19	18	1519	1	2	3	1	0	2	0	1.2	13
542	543	Hakim Ziyech	mas MAR	FWMF	Chelsea	29-352	13	6	572	0	1	1	0	0	1	0	0.6	14
543	544	Kurt Zouma	fr FRA	DF	West Ham	28-130	16	15	1296	1	0	1	1	0	0	1	0.0	11.0
544	545	Matteo Politano	no NOR	MF	Acquafoglia	24-070	26	26	3126	0	6	15	0	0	3	0	7.0	8.0

```
545 rows × 18 columns

In [24]: df.columns
Out[24]: Index(['List', 'Player', 'Nation', 'Pos', 'Squad', 'Age', 'Matches', 'Starts',
       'Min', 'Goals', 'Assists', 'G+A', 'non-PK', 'PK', 'yellow', 'red', 'xG',
       'xAG'],
      dtype='object')

In [25]: df = df.rename(columns = {
       'Rk' : 'list',
       'Player' : 'name',
       'NP' : 'match',
       'G+PK' : 'non-PK',
       'CrdY' : 'yellow',
       'CrdR' : 'red',
       'Gls' : 'goals',
       'Ast' : 'assists',
      })
In [28]: df = df.drop(columns = ['List'])
In [29]: df.to_csv('fbref.csv')
```

- Scrapped fbref.com
 - Imported the csv file as a table into the Mysql database
 - Used panda to clean data and improve the readability of the data



Sherry

Player Goalkeeping 2022-2023 Premier League [Advanced Goalkeeping](#) Share & Export

Rk	Player	Nation	Pos	Squad	Age	Born	Playing Time				Performance			
							MP	Starts	Min	90s	GA	GA90	SoTA	Saves
1	Alisson	BRA	GK	Liverpool	30-190	1992	28	28	2,520	28.0	33	1.18	115	83
2	Alphonse Areola	FRA	GK	West Ham	30-042	1993	5	2	309	3.4	7	2.04	16	9
3	Kepa Arrizabalaga	ESP	GK	Chelsea	28-189	1994	22	22	1,935	21.5	19	0.88	81	64
4	Gavin Bazunu	IRL	GK	Southampton	21-049	2002	30	30	2,700	30.0	51	1.70	113	65
5	Asmir Begović	BIH	GK	Everton	35-294	1987	1	1	90	1.0	0	0.00	4	4
6	Martin Dúbravka	SVK	GK	Newcastle Utd	34-085	1989	1	0	67	0.7	0	0.00	5	5
7	Ederson	BRA	GK	Manchester City	29-236	1993	29	29	2,610	29.0	27	0.93	62	36

```
data_19_22 = pd.DataFrame()

for year in ['2019', '2020', '2021', '2022']:
    data = pd.read_csv(f'f'{year}.csv')
    data['year'] = year
    data_19_22 = pd.concat([data, data_19_22])

0.3s
```

data_19_22
0.3s

Rk	Player	Nation	Pos	Squad	Age	Born	MP	Starts	Min	...	CS	CS%	PKatt	PKA	
0	1	Alisson	br BRA	GK	Liverpool	30-190	1992	28	28	2520	...	11.0	39.3	3.0	2.0
1	2	Alphonse Areola	fr FRA	GK	West Ham	30-042	1993	5	2	309	...	1.0	50.0	2.0	2.0
2	3	Kepa Arrizabalaga	es ESP	GK	Chelsea	28-189	1994	22	22	1935	...	9.0	40.9	0.0	0.0
3	4	Gavin Bazunu	ie IRL	GK	Southampton	21-049	2002	30	30	2700	...	4.0	13.3	2.0	1.0
4	5	Asmir Begović	ba BIH	GK	Everton	35-294	1987	1	1	90	...	1.0	100.0	0.0	0.0
...	

- Acquired goalkeeping data in CSV format from fbref.com
- Merged data from different seasons into a single dataset



Data Storage

- Data was stored in a MySQL database
- Separate tables for each website containing all of the scraped data

(MySQL 5.7.39-google-log) WSA/Soc...

Soccer-2023

Select Database Structure Content Relations Triggers Table Info Query Table History Users Console

Player	season	Pos	Squad	GA	GA90	SoTA	Saves	Save%	CS	CS%	PKS
Allison	2021/2022	GK	Liverpool	24	0.67	99	76	75.8	20	55.6	0
Kepa Arrizabalaga	2021/2022	GK	Chelsea	2	0.5	13	11	84.6	2	50	0
Daniel Bachmann	2021/2022	GK	Watford	28	2.33	72	46	65.3	0	0	0
Jack Butland	2021/2022	GK	Crystal Palace	13	1.53	28	16	53.6	1	12.5	100
Karl Darlow	2021/2022	GK	Newcastle Utd	15	1.87	37	23	64.9	0	0	0
Martin Dubravka	2021/2022	GK	Newcastle Utd	35	1.35	106	72	67.9	8	30.8	0
Ederson	2021/2022	GK	Manchester City	26	0.7	83	57	69.9	20	54.1	0
Lukasz Fabianski	2021/2022	GK	West Ham	50	1.35	162	111	70.4	8	21.6	60
Alvaro Fernandez	2021/2022	GK	Brentford	24	2	69	47	66.7	1	8.3	0
Fraser Forster	2021/2022	GK	Southampton	35	1.84	101	68	67.3	3	15.8	0
Ben Foster	2021/2022	GK	Watford	49	1.88	116	67	60.3	4	15.4	25
David de Gea	2021/2022	GK	Manchester Utd	57	1.5	180	124	70	8	21.1	40
Vicente Guaita	2021/2022	GK	Crystal Palace	33	1.12	116	83	74.1	11	36.7	0
Angus Gunn	2021/2022	GK	Norwich City	22	2.44	57	34	64.9	1	11.1	33
Tim Krul	2021/2022	GK	Norwich City	62	2.14	165	105	66.7	5	17.2	0
Bernd Leno	2021/2022	GK	Arsenal	9	2.25	19	10	52.6	1	2.5	0
Hugo Lloris	2021/2022	GK	Tottenham	40	1.05	137	98	72.3	16	42.1	0
Emiliano Martinez	2021/2022	GK	Aston Villa	48	1.33	140	93	68.6	11	30.6	20
Alex McCarthy	2021/2022	GK	Southampton	27	1.59	71	44	62	5	29.4	0
Edouard Mendy	2021/2022	GK	Chelsea	31	0.91	102	72	75.5	14	41.2	0
Ilan Meslier	2021/2022	GK	Leeds United	79	2.1	220	142	66.4	5	13.2	0
Jordan Pickford	2021/2022	GK	Everton	58	1.66	170	116	66.5	7	20	0
Nick Pope	2021/2022	GK	Burnley	47	1.31	165	119	74.5	9	25	0
Aaron Ramsdale	2021/2022	GK	Arsenal	39	1.15	128	90	73.4	12	35.3	0
David Raya	2021/2022	GK	Brentford	27	1.13	103	76	77.7	8	33.3	0
Jose Sa	2021/2022	GK	Wolves	40	1.1	160	121	80.6	11	29.7	0
Robert Sanchez	2021/2022	GK	Brighton	42	1.14	141	99	72.3	11	29.7	0
Kasper Schmeichel	2021/2022	GK	Leicester City	58	1.57	187	128	72.7	7	18.9	22.2
Fredie Woodman	2021/2022	GK	Newcastle Utd	12	3	22	9	54.5	0	0	33



Analytics Team



Initial Approach

- Originally we brainstormed how to best quantify transfer impact in the Premier League. We decided to develop our own player score for player performance in a season.
- Scores were derived by adding up a combination of raw and advanced statistics and applying weights based on the importance of the statistics to their position.
- Players are then compared against the average at their position to determine their final player score for a season.



Positions & Weights

- With 11 men in the squad, the goalkeeper is obviously going to have a much different role than a striker.
- Thus, one of our challenges was to create our own weighting system based on the data and the position.
- For example, the role of a striker is to score, so each goal scored by a striker carried the weight of 18, whilst a goal scored by a center back, a defender, had a weight of 5.
- All of the outfield positions was based on the same statistics (assists, shots on target, aerials won), while goalkeepers had their own set of statistics (clean sheets, saves, crosses stopped %).



Important Statistics

- **Wingers:** goals, assists, chances created, expected goals
- **Striker:** goals, assists, successful dribbles, expected goals
- **Attacking Mid:** goals, assists, chances created, successful crosses
- **Defending Mid:** interceptions, tackles won, goals, assists
- **Left/Right Back:** assists, successful crosses, chances created, expected assists
- **Center Back:** interceptions, aerials won, tackles, won, clearances
- **Goalkeepers:** goals conceded per 90, post shot expected goals, save %, saves, clean sheets



Overall Analysis

- We have three different rankings of teams for each season
 1. Overall team score
 2. Overall team score of transfer players
 3. Percentage of team score that came from transfer players
- For each season we have also identified the top 10 individual transfer players and the top 5 individual loan players

2018-19

	WSA	EPL	WSA	EPL
1.			11.	
2.			12.	
3.			13.	
4.			14.	
5.			15.	
6.			16.	
7.			17.	
8.			18.	
9.			19.	
10.			20.	

Top 10 Transfers

1. Leander Dendoncker
2. Raul Jimenez
3. Ben Foster
4. Andre Gomes
5. Fabian Schar
6. Michy Batshuayi
7. Lukasz Fabianski
8. Federico Fernandez
9. Ruben Vinagre
10. Fabian Balbuena

Top 5 Loan Transfers

1. Danny Ings
2. Youri Tielemans
3. Mateo Kovacic
4. Nathaniel Clyne
5. Sergio Rico



2019-20

WSA

EPL

1.		
2.		
3.		
4.		
5.		
6.		
7.		
8.		
9.		
10.		

WSA

EPL

11.		
12.		
13.		
14.		
15.		
16.		
17.		
18.		
19.		
20.		

Top 10 Transfers

1. Pepe Reina
2. Sam Byram
3. Jordan Ayew
4. Tomas Soucek
5. Jetro Willems
6. Erik Pieters
7. Ben Osborn
8. Djibril Sidibe
9. David Luiz
10. Harry Wilson

Top 5 Loan Transfers

1. Ryan Bennett
2. Dean Henderson
3. Dani Ceballos
4. Ondrej Duda



2020-21

WSA

EPL

1.		
2.		
3.		
4.		
5.		
6.		
7.		
8.		
9.		
10.		

WSA

EPL

11.		
12.		
13.		
14.		
15.		
16.		
17.		
18.		
19.		
20.		

Top 10 Transfers

1. Joel Veltman
2. Joachim Andersen
3. Ademola Lookman
4. Abdul-Nasir
5. Abdoulaye Doucoure
6. Ozan Kabak
7. Antonee Robinson
8. Jesse Lingard
9. Dale Stephens
10. Vladimir Coufal

Top 5 Loan Transfers

1. Rayan Ait Nouri
2. Gareth Bale
3. Alphonse Areola
4. Okay Yokuslu
5. Ainsley Maitland-Niles



2021-22

WSA

EPL

1.		
2.		
3.		
4.		
5.		
6.		
7.		
8.		
9.		
10.		

WSA

EPL

11.		
12.		
13.		
14.		
15.		
16.		
17.		
18.		
19.		
20.		

Top 10 Transfers

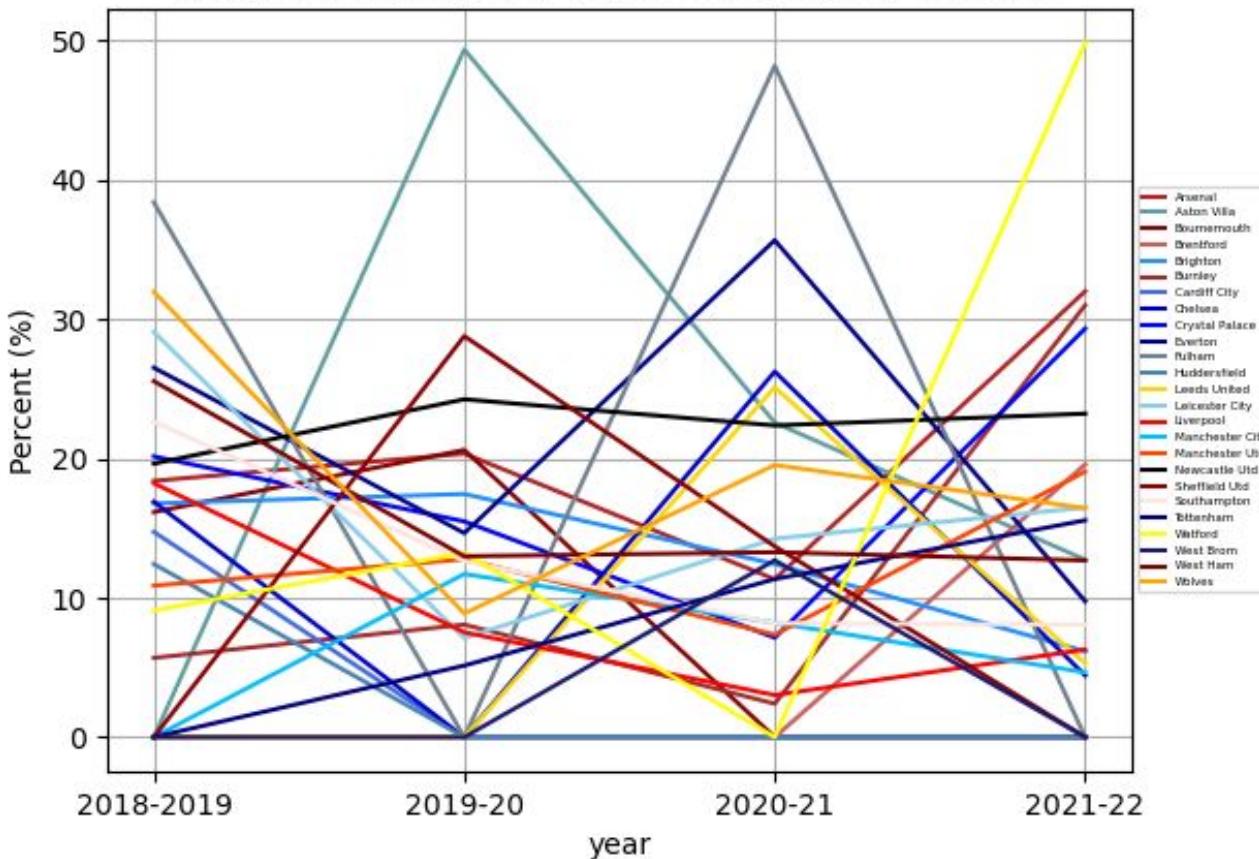
1. Alvaro Fernandez
2. Craig Dawson
3. Demarai Gray
4. Moussa Sissoko
5. Emmanuel Dennis
6. Connor Roberts
7. Mathias Normann
8. Pierre Lees-Melou
9. Dan Burn
10. Hassane Kamara

Top 5 Loan Transfers

1. Ademola Lookman
2. Christian Romero
3. Juraj Kucka
4. Philippe Coutinho
5. Hwang Hee-chan

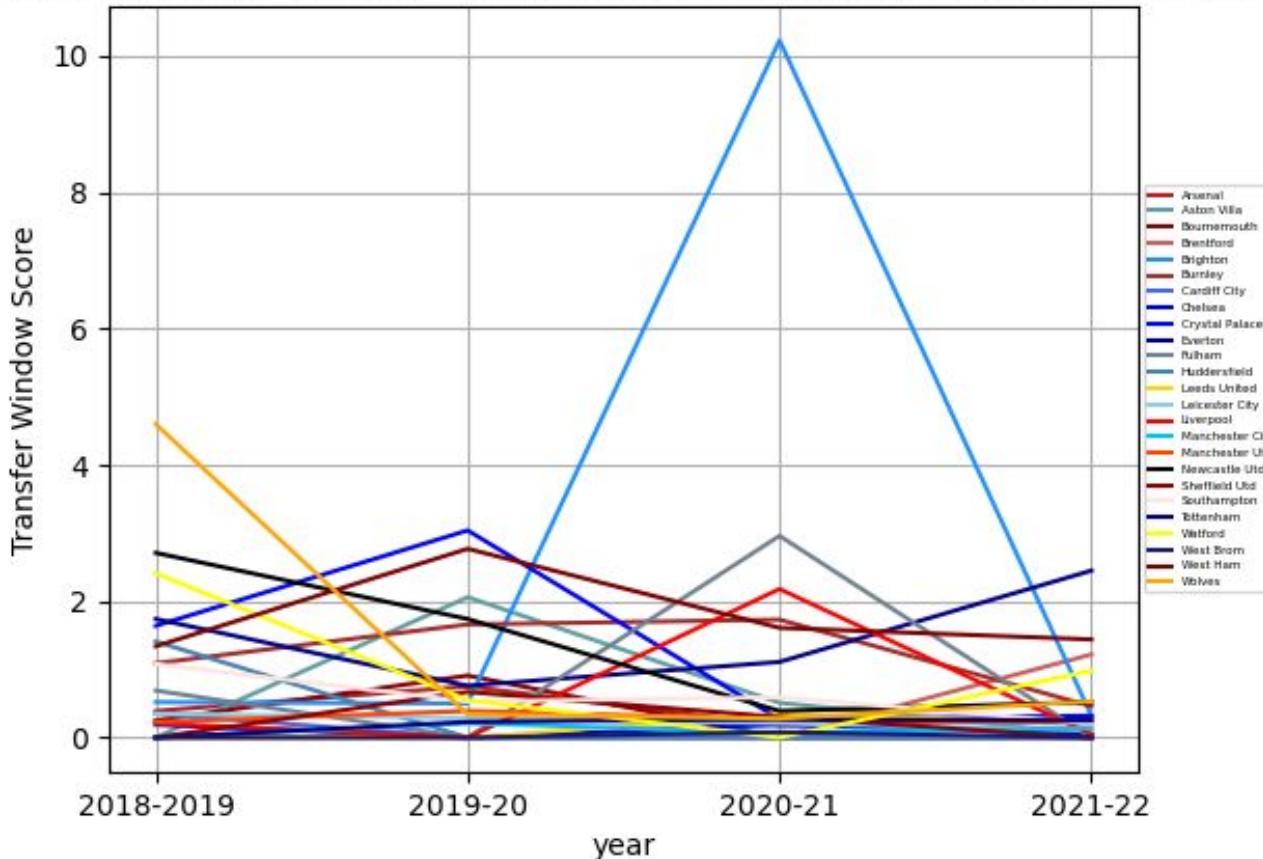


Premier League Pct of Team Score by New Transfers





Premier League Transfer Window Scores considering Total Money Spent





Initial Takeaways

- Our club rankings are decently accurate
 - 25% on champions, 69% on UCL spots, 58% on relegation spots
 - Definitely some outliers (Arsenal 17th, Brighton in 3rd, Newcastle relegated, etc.)
- Our best transfers by value include very few big names
 - Mainly lesser-known players who were purchased for less than 5 mil
 - Still shows which clubs are making smart purchases
- The graphs look a bit messy but the information is interesting to follow
 - Brighton is an outlier in second graph



What's next?

- Possible summer improvements:
- Adding more statistics for each player (many were members only on certain sites)
- Using Tensorflow to determine our weights
- Making sure every new transfer is accounted for
- Looking at free transfers and separating paid loans from transfers
- Cleaning up graphs & visuals to make more appealing to the eye