

```
In [1]: import time
import pandas as pd
import numpy as np
```

```
In [2]: from selenium import webdriver

chrome_path = "C:/Program Files/ChromeDriver/chromedriver.exe"
```

```
In [8]: link = "https://www.knvb.nl/competities/eredivisie/uitslagen"

data = {
    "HomeTeam": [],
    "AwayTeam": [],
    "Score": []
}

driver = webdriver.Chrome(chrome_path)
driver.get(link)
for i in driver.find_elements_by_class_name("row"):
    count = 0
    for team in i.find_elements_by_class_name("team"):
        count += 1
        if count == 1:
            data["HomeTeam"].append(team.text)
        else:
            data["AwayTeam"].append(team.text)

    data["Score"].append(i.find_element_by_class_name("center").text)
```

```
In [9]: eredivisie_matches = pd.DataFrame.from_dict(data)
```

```
In [10]: eredivisie_matches.head(1000)
```

```
Out[10]:
```

	HomeTeam	AwayTeam	Score
0	Feyenoord	RKC Waalwijk	3-0
1	AZ	Heracles Almelo	5-0

	HomeTeam	AwayTeam	Score
2	PEC Zwolle	FC Groningen	1-0
3	sc Heerenveen	Sparta Rotterdam	1-2
4	VVV-Venlo	FC Emmen	0-4
...
301	Heracles Almelo	ADO Den Haag	2-0
302	FC Emmen	VVV-Venlo	3-5
303	FC Twente	Fortuna Sittard	2-0
304	PEC Zwolle	Feyenoord	0-2
305	sc Heerenveen	Willem II	2-0

306 rows × 3 columns

```
In [34]: len(eredivisie_matches)
```

```
Out[34]: 306
```

```
In [27]: European_Leagues = pd.read_csv("C:/Users/Gebruiker/Desktop/Engineering/Voetbal-Predictor/Code/Goaly/Data/European_Lea
European_Leagues.head()
```

```
Out[27]:
```

	Country	Organisation	Affiliation	Competitions
0	Austria	Österreichische Fußball Bundesliga	2005	Bundesliga \r\nErste Liga
1	Azerbaijan	Azerbaijan Professional Football League	2013	Premier League \r\nI Division \r\nCup \r\nSupe...
2	Belgium	Pro League	2005	Pro League \r\nSuper Cup
3	Czech Republic	Ligová Fotbalová Asociace	2018	Czech First League \r\nCzech National Football...
4	Denmark	Divisionsforeningen	2005	Superliga \r\n1. division \r\n2. division (Øst...

```
In [28]: European_Leagues['PrimaryLeague'] = European_Leagues["Competitions"].str.split(" \r\n", n=1, expand=True)[0]
```

In [29]: European_Leagues

Out[29]:

	Country	Organisation	Affiliation	Competitions	PrimaryLeague
0	Austria	Österreichische Fußball Bundesliga	2005	Bundesliga \r\nErste Liga	Bundesliga
1	Azerbaijan	Azerbaijan Professional Football League	2013	Premier League \r\nI Division \r\nCup \r\nSupe...	Premier League
2	Belgium	Pro League	2005	Pro League \r\nSuper Cup	Pro League
3	Czech Republic	Ligová Fotbalová Asociace	2018	Czech First League \r\nCzech National Football...	Czech First League
4	Denmark	Divisionsforeningen	2005	Superliga \r\n1. division \r\n2. division (Øst...	Superliga
5	England	The Premier League	2005	Premier League	Premier League
6	Finland	The Finnish Football League Association	2005	Veikkausliiga \r\nLiigacup	Veikkausliiga
7	France	Ligue de Football Professionnel	2005	Ligue 1 \r\nLigue 2 \r\nTrophée des Champions	Ligue 1
8	Germany	DFL Deutsche Fußball Liga GmbH	2005	Bundesliga \r\n2. Bundesliga\r\nDFL-Supercup	Bundesliga
9	Greece	Super League Greece	2005	Super League	Super League
10	Israel	Israeli Professional Football Leagues	2014	Premier League \r\nLiga Leumit	Premier League
11	Italy	Lega Nazionale Professionisti Serie A	2005	Serie A \r\nCoppa \r\nSupercoppa	Serie A
12	Kazakhstan	Professional Football League of Kazakhstan	2015	Premier League \r\nFirst League \r\nCup \r\nSu...	Premier League
13	Norway	Norsk Toppfotball	2007	Eliteserien \r\n1. divisjon	Eliteserien
14	Netherlands	Eredivisie CV	2005	Eredivisie	Eredivisie
15	Poland	Polish Professional Football League	2007	Ekstraklasa \r\nSuper Cup	Ekstraklasa
16	Portugal	Liga Portuguesa de Futebol Profissional	2005	Primeira Liga \r\nSegunda Liga \r\nTaça da Liga	Primeira Liga
17	Romania	Liga Profesionist? de Fotbal (2009  2011) \r\n2014		Liga 1 \r\nCupa Ligii	Liga 1
18	Russia	Russian Football Premier League	2007	Championship \r\nSuper Cup	Championship
19	Scotland	Scottish Professional Football League	2005	Premiership \r\nChampionship \r\nLeague 1 \r\n...	Premiership
20	Serbia	Serbian Superliga	2010	Super Liga	Super Liga
21	Spain	Liga Nacional de Fútbol Profesional	2005	La Liga \r\nSegunda División	La Liga
22	Sweden	Swedish Elite Football	2005	Allsvenskan \r\nSuperettan	Allsvenskan
23	Switzerland	Swiss Football League	2005	Super League \r\nChallenge League	Super League

	Country	Organisation	Affiliation	Competitions	PrimaryLeague
24	Ukraine	Ukrainian Premier League	2009	Championship \nSuper Cup	Championship

```
In [30]: baselink = "https://www.skysports.com/"
secondary_part = "-results/2020-21"

data = {
    "HomeTeam": [],
    "HomeScore": [],
    "AwayTeam": [],
    "AwayScore": []
}

for i in European_Leagues.PrimaryLeague:
    driver.get(baselink + i.replace(" ", "-").lower() + secondary_part)
    time.sleep(1)

    for el in driver.find_elements_by_class_name("fixres__item"):
        count = 0

        for scores in el.find_elements_by_class_name("matches__teamscores-side"):
            count += 1
            if count == 1:
                data["HomeScore"].append(scores.text)
            else:
                data["AwayScore"].append(scores.text)

        count = 0
        for teams in el.find_elements_by_class_name("swap-text__target"):
            count += 1
            if count == 1:
                data["HomeTeam"].append(teams.text)
            else:
                data["AwayTeam"].append(teams.text)
```

```
In [31]: matches = pd.DataFrame.from_dict(data)
matches.head()
```

```
Out[31]:
```

	HomeTeam	HomeScore	AwayTeam	AwayScore
--	----------	-----------	----------	-----------

	HomeTeam	HomeScore	AwayTeam	AwayScore
0	1. FC Union Berlin	2	RB Leipzig	1
1	Bayern Munich	5	FC Augsburg	2
2	Borussia Dortmund	3	Bayer Leverkusen	1
3	Cologne	1	Schalke	0
4	Eintracht Frankfurt	3	SC Freiburg	1

```
In [33]: len(matches)
```

```
Out[33]: 2400
```

```
In [47]: matches.to_csv('matches.csv', index=False)
```

```
In [32]: matches.HomeTeam.value_counts()
```

```
Out[32]:
Manchester City      2236
Manchester City         4
Everton              4
Aston Villa          4
West Bromwich Albion 4
...
Vitesse Arnhem       1
Toulouse             1
Strasbourg           1
Brest                1
PEC Zwolle           1
Name: HomeTeam, Length: 84, dtype: int64
```

```
In [35]: matches.AwayTeam.value_counts()
```

```
Out[35]:
Brighton and Hove Albion 2236
Brighton and Hove Albion 4
Newcastle United        4
Tottenham Hotspur       4
Wolverhampton Wanderers 4
...
Verona                  1
Udinese                 1
```

```
Paris Saint-Germain      1
Granada                  1
Nimes                    1
Name: AwayTeam, Length: 87, dtype: int64
```

In []: