

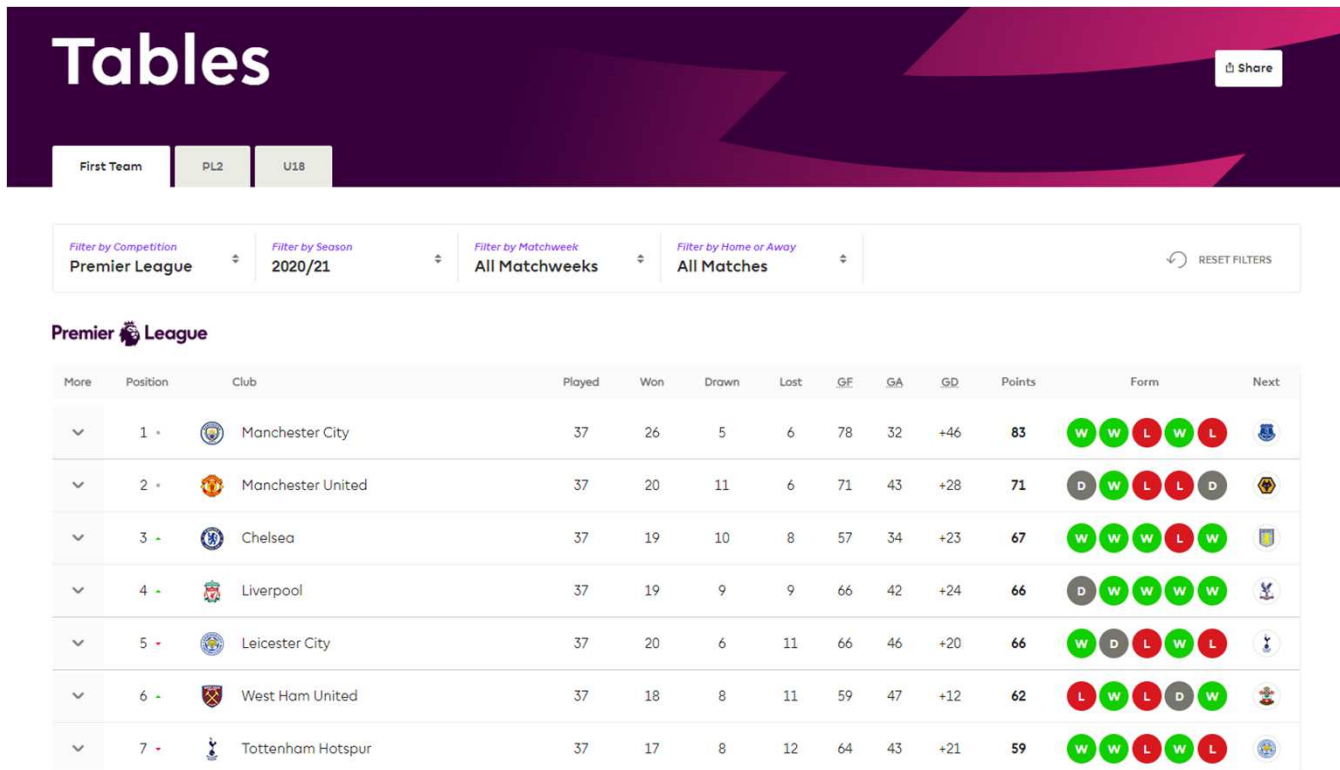
EPL 리그 차트 분석

손진영



Crawling

크롤링 타겟 사이트



The screenshot shows the Premier League Tables page. At the top, there's a 'Tables' header with a 'Share' button. Below it are tabs for 'First Team', 'PL2', and 'U18'. A filter bar allows selection by competition (Premier League), season (2020/21), matchweek (All Matchweeks), and home/away status (All Matches). The main table lists the top 7 teams: Manchester City (1st), Manchester United (2nd), Chelsea (3rd), Liverpool (4th), Leicester City (5th), West Ham United (6th), and Tottenham Hotspur (7th). Each row includes columns for position, club, played, won, drawn, lost, goals for/against/difference, points, form (represented by colored circles), and the next opponent.

More	Position	Club	Played	Won	Drawn	Lost	GF	GA	GD	Points	Form	Next
▼	1 +	Manchester City	37	26	5	6	78	32	+46	83	W W L W L	
▼	2 +	Manchester United	37	20	11	6	71	43	+28	71	D W L L D	
▼	3 +	Chelsea	37	19	10	8	57	34	+23	67	W W W L W	
▼	4 +	Liverpool	37	19	9	9	66	42	+24	66	D W W W W	
▼	5 -	Leicester City	37	20	6	11	66	46	+20	66	W D L W L	
▼	6 +	West Ham United	37	18	8	11	59	47	+12	62	L W L D W	
▼	7 -	Tottenham Hotspur	37	17	8	12	64	43	+21	59	W W L W L	

링크 : <https://www.premierleague.com/tables>

EPL 공식 홈페이지
리그 데이터를 크롤링

이 중 손흥민 선수가 뛰고있는
토트넘을 타겟팅하여
라운드별 순위 분석

Crawling

크롤링

Selenium 을 활용한 크롤링 진행

링크 접속

```
# selenium으로 URL에 접속
from selenium import webdriver
import time
# webdriver 설치 경로 확인
driver =
webdriver.Chrome('C:/Users/jyson/git/learning_python/03_WebConn/4_selenium_class/webdrive
r/chromedriver.exe')

# 최대사이즈
driver.maximize_window()

url = "https://www.premierleague.com/tables"
driver.get(url)
```

해당 페이지의 경우 접속시마다 쿠키 발생

쿠키 제거

```
driver.implicitly_wait(3)
driver.find_element_by_class_name('cookies-notice-accept').click()
driver.find_element_by_id('advertClose').click()
```

Crawling 크롤링

1라운드부터 진행라운드까지 데이터 크롤링

라운드별 데이터 크롤링

```
# 리스트 갯수
ul =
driver.find_elements_by_xpath('/html/body/main/div[2]/div[1]/div[1]/section/div[3]/ul/li'
)
print(len(ul))

round = [] # 라운드
rank = [] # 순위

# 라운드 순회
for i in range(len(ul) - 1):
    time.sleep(1)
    driver.find_elements_by_class_name('current')[2].click()
    time.sleep(1)
    ul[i + 1].click()
    time.sleep(1)
    tr = driver.find_element_by_xpath("//tr[@data-filtered-table-row-name='Tottenham Hotspur']")

    print('라운드 : ' + str(i + 1) + ', 순위 : ' + str(tr.text.split("\n")[0]))
    round.append(i + 1)
```

```
# 데이터 크롤링 실패시
if tr.text.split("\n")[0] == '':
    while(tr.text.split("\n")[0] == ''):
        print("재시도 . . .")
        driver.refresh()
        time.sleep(3)
        driver.get(url)

        time.sleep(1)
        driver.find_elements_by_class_name('current')[2].click()
        time.sleep(1)
        ul =
driver.find_elements_by_xpath('/html/body/main/div[2]/div[1]/div[1]/section/div[3]/ul/li'
)

    ul[i + 1].click()
    time.sleep(5)
    tr = driver.find_element_by_xpath("//tr[@data-filtered-table-row-name='Tottenham Hotspur']")

    print('라운드 : ' + str(i + 1) + ', 순위 : ' + str(tr.text.split("\n")[0]))

    rank.append(int(tr.text.split("\n")[0].strip()))
else:
    print('라운드 : ' + str(i + 1) + ', 순위 : ' + str(tr.text.split("\n")[0]))
else:
    rank.append(int(tr.text.split("\n")[0].strip()))
```

Crawling

크롤링

```
#라운드 순위
for i in range(len(ul) - 1):
    ul: [selenium.webdriver.remote.webelement.WebElement (session="2ee580367e0e7d1c...
    time.sleep(1)
    driver.find_elements_by_class_name('current')[2].click() driver: <selenium.webdriver.chrome.webdriver.WebDriver (session="2ee580367e0e7d1c...
    time.sleep(1)
    ul[i + 1].click() ul: [selenium.webdriver.remote.webelement.WebElement (session="2ee580367e0e7d1c...
    time.sleep(1)
    tr = driver.find_element_by_xpath("//tr[@data-filtered-table-row-name='Tottenham Hotspur']") driver: <selenium.webdriver.remote.webelement.WebElement (session="2ee580367e0e7d1c...

    print('라운드 : ' + str(i + 1) + ', 순위 : ' + str(tr.text.split("\n")[0])) i: 36 tr: <selenium.webdriver.remote.webelement.WebElement (session="2ee580367e0e7d1c...
    round.append(i + 1) round: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38]

#데이터 크롤링 실패시
if tr.text.split("\n")[0] == '':
    tr: <selenium.webdriver.remote.webelement.WebElement (session="2ee580367e0e7d1c...
    while(tr.text.split("\n")[0] == ''):
        tr: <selenium.webdriver.remote.webelement.WebElement (session="2ee580367e0e7d1c...
        print("재시도 . . .")
        driver.refresh() driver: <selenium.webdriver.chrome.webdriver.WebDriver (session="2ee580367e0e7d1c...
        time.sleep(3)
        driver.get(url) driver: <selenium.webdriver.chrome.webdriver.WebDriver (session="2ee580367e0e7d1c...

        time.sleep(1)
        driver.find_elements_by_class_name('current')[2].click() driver: <selenium.webdriver.chrome.webdriver.WebDriver (session="2ee580367e0e7d1c...
        time.sleep(1)
        ul = driver.find_elements_by_xpath('/html/body/main/div[2]/div[1]/div[1]/section/div[3]/ul/li') driver: <selenium.webdriver.chrome.webdriver.WebDriver (session="2ee580367e0e7d1c...
        ul[i + 1].click() ul: [selenium.webdriver.remote.webelement.WebElement (session="2ee580367e0e7d1c...
        time.sleep(5)
        tr = driver.find_element_by_xpath("//tr[@data-filtered-table-row-name='Tottenham Hotspur']") driver: <selenium.webdriver.remote.webelement.WebElement (session="2ee580367e0e7d1c...

        print('라운드 : ' + str(i + 1) + ', 순위 : ' + str(tr.text.split("\n")[0])) i: 36 tr: <selenium.webdriver.remote.webelement.WebElement (session="2ee580367e0e7d1c...

        rank.append(int(tr.text.split("\n")[0].strip())) rank: [11, 6, 8, 6, 7, 5, 3, 2, 1, 1, 1, 1, 2, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38]
    else:
```

해당 데이터

크롤링 실패 시,
재시도

```
38
라운드 : 1, 순위 : 11
라운드 : 2, 순위 : 6
라운드 : 3, 순위 : 8
라운드 : 4, 순위 : 6
라운드 : 5, 순위 : 7
라운드 : 6, 순위 : 5
라운드 : 7, 순위 : 3
라운드 : 8, 순위 : 2
라운드 : 9, 순위 : 1
라운드 : 10, 순위 : 1
라운드 : 11, 순위 : 1
라운드 : 12, 순위 : 1
라운드 : 13, 순위 : 2
라운드 : 14, 순위 : 6
라운드 : 15, 순위 : 5
라운드 : 16, 순위 : 7
라운드 : 17, 순위 : 4
라운드 : 18, 순위 : 6
라운드 : 19, 순위 : 5
라운드 : 20, 순위 : 6
라운드 : 21, 순위 : 6
라운드 : 22, 순위 : 8
라운드 : 23, 순위 : 8
라운드 : 24, 순위 : 9
라운드 : 25, 순위 : 9
라운드 : 26, 순위 : 8
라운드 : 27, 순위 :
재시도 . . .
라운드 : 27, 순위 : 7
라운드 : 27, 순위 : 7
라운드 : 28, 순위 : 8
라운드 : 29, 순위 : 6
라운드 : 30, 순위 : 6
라운드 : 31, 순위 : 7
라운드 : 32, 순위 : 6
라운드 : 33, 순위 : 7
라운드 : 34, 순위 : 6
라운드 : 35, 순위 : 7
라운드 : 36, 순위 : 6
라운드 : 37, 순위 : 7
```

```
692 print(round)
print(rank)
```

Crawling

크롤링

데이터 시각화

```
from matplotlib import pyplot as plt
import matplotlib.ticker as mticker

plt.figure(figsize=(20, 8))
# x축 눈금
plt.xticks([1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21,
22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38])
plt.ylim([int(20), int(0)]) # y축의 범위

plt.title('THS rankchart')
plt.gca().title.set_size(25)

plt.gca().yaxis.set_major_formatter(mticker.FormatStrFormatter('%i')) # 소수점 제거

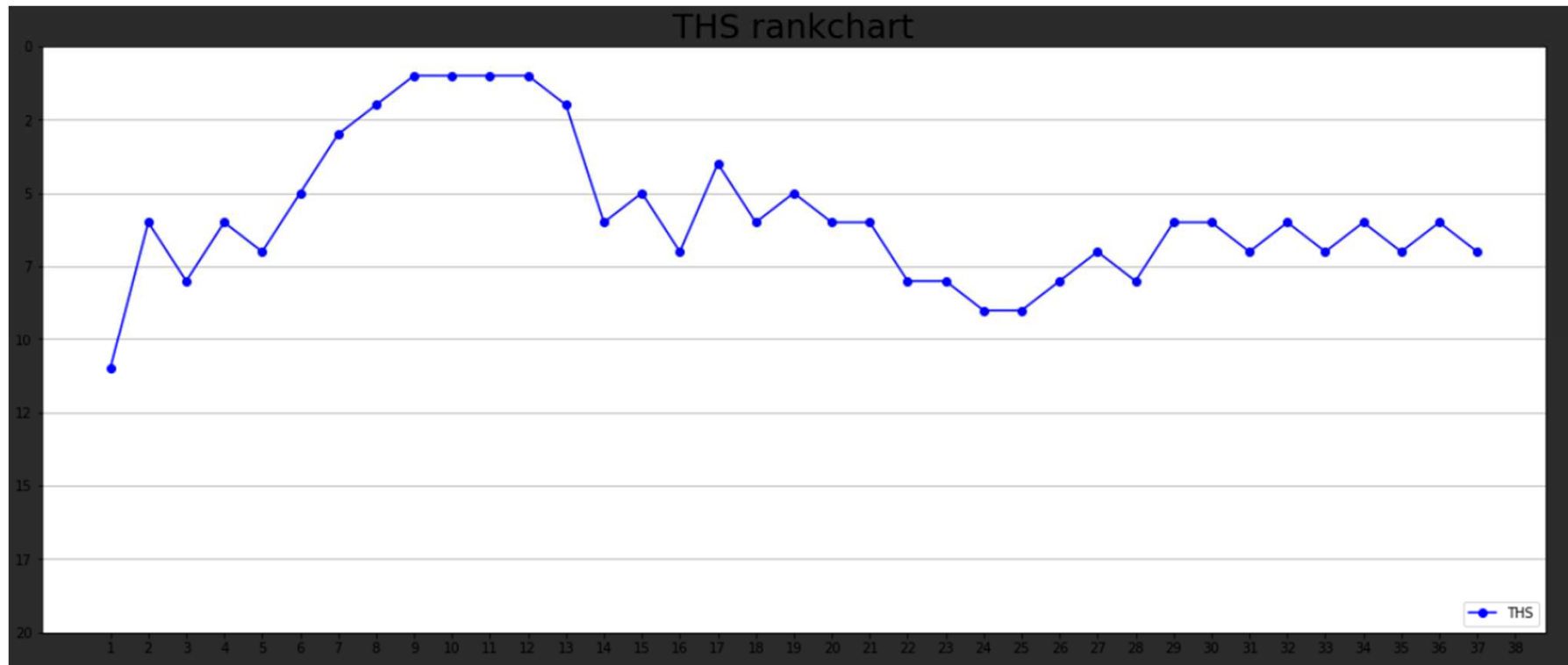
# 격자무늬
plt.grid(True, axis='y')

plt.plot(round, rank, color='blue', marker='o', label=('THS'))
plt.legend(loc=4)
plt.show()
```

Crawling

크롤링

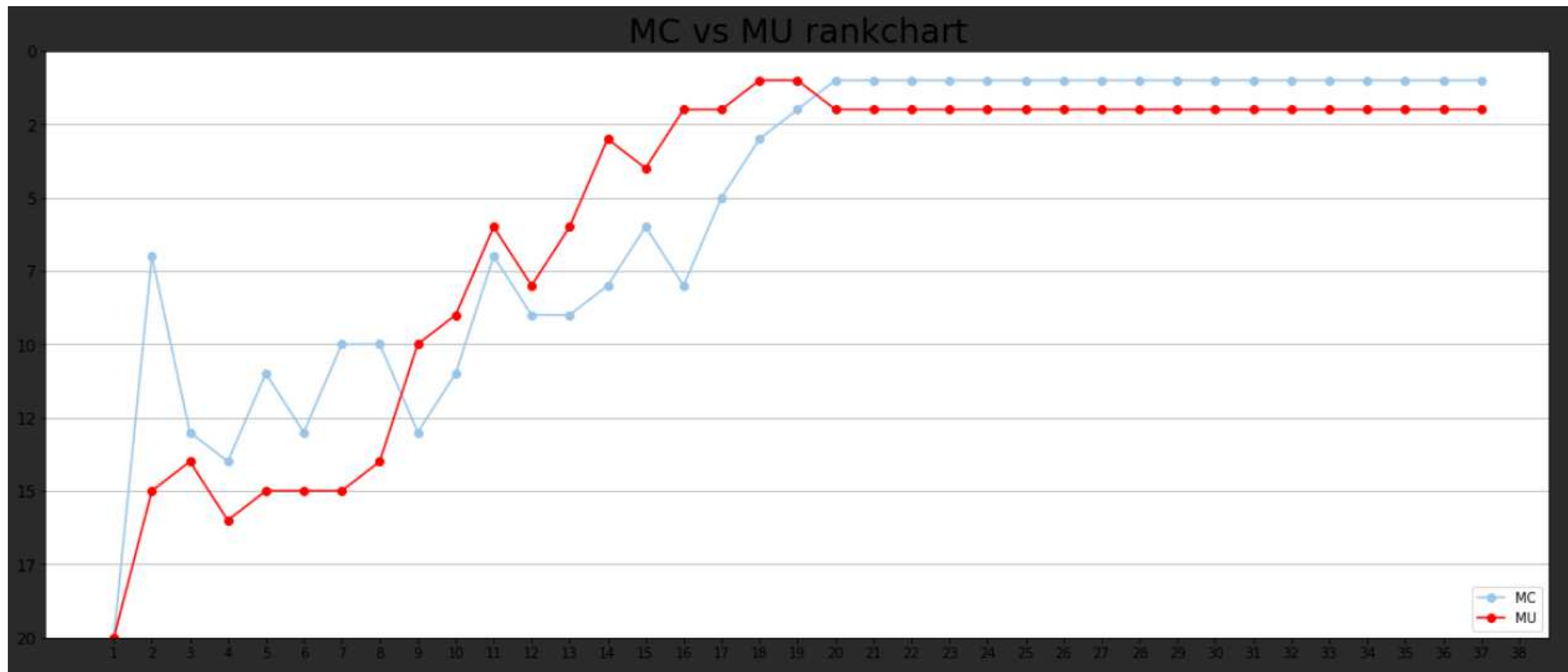
결과 그래프




Crawling

[추가] 맨시티vs 맨유 리그 순위

결과 그래프



A top-down view of four white ceramic coffee cups arranged in a square on a dark, textured wooden tray. Each cup contains a latte with intricate brown and white foam art. A semi-transparent, light-colored circle is centered over the cups, containing the text 'Thank You' and '감사합니다' in a dark, sans-serif font.

Thank You
감사합니다