In [12]:
```
import requests
import os
import shutil
```

In [6]:
```
years = list(range(1991,2022))
```

In [8]:
```
url_start = "https://www.basketball-reference.com/awards/awards_{}.html"

for year in years:
    url = url_start.format(year)

    data = requests.get(url)

    with open("mvp/{}.html".format(year), "w+") as f:
        f.write(data.text)
```

In [15]:
```
from bs4 import BeautifulSoup
```

In [46]:
```
with open("mvp/1991.html") as f:
    page = f.read()

soup = BeautifulSoup(page, 'html.parser')
soup.find('tr', class_="over_header").decompose()
```

In [47]:
```
mvp_table = soup.find_all(id="mvp")[0]
```

In [48]:
```
import pandas as pd
```

In [49]:
```
mvp_1991 = pd.read_html(str(mvp_table))[0]
```

In [50]:
```
mvp_1991.head(1)
```

Out[50]:

	Rank	Player	Age	Tm	First	Pts Won	Pts Max	Share	G	MP	PTS	TRB	AST	STL	BLI
0	1	Michael Jordan	27	CHI	77.0	891.0	960	0.928	82	37.0	31.5	6.0	5.5	2.7	1.(

In [51]:
```
mvp_1991["Year"] = 1991
```

In [52]:
```
mvp_1991.head()
```

Out[52]:

	Rank	Player	Age	Tm	First	Pts Won	Pts Max	Share	G	MP	...	TRB	AST	STL	BLK
0	1	Michael Jordan	27	CHI	77.0	891.0	960	0.928	82	37.0	...	6.0	5.5	2.7	1.0
1	2	Magic Johnson	31	LAL	10.0	497.0	960	0.518	79	37.1	...	7.0	12.5	1.3	0.2
2	3	David Robinson	25	SAS	6.0	476.0	960	0.496	82	37.7	...	13.0	2.5	1.5	3.9
3	4	Charles Barkley	27	PHI	2.0	222.0	960	0.231	67	37.3	...	10.1	4.2	1.6	0.5
4	5	Karl Malone	27	UTA	0.0	142.0	960	0.148	82	40.3	...	11.8	3.3	1.1	1.0

5 rows × 21 columns

In [55]:
```
dfs = []
for year in years:
    with open("mvp/{}.html".format(year)) as f:
        page = f.read()

    soup = BeautifulSoup(page, 'html.parser')
    soup.find('tr', class_="over_header").decompose()
    mvp_table = soup.find_all(id="mvp")[0]
    mvp_df = pd.read_html(str(mvp_table))[0]
    mvp_df["Year"] = year
    dfs.append(mvp_df)
```

In [57]:
```
mvps = pd.concat(dfs)

mvps.tail()
```

Out[57]:

	Rank	Player	Age	Tm	First	Pts Won	Pts Max	Share	G	MP	...	TRB	AST	STL	
10	11	Russell Westbrook	32	WAS	0.0	5.0	1010	0.005	65	36.4	...	11.5	11.7	1.4	
11	12	Ben Simmons	24	PHI	0.0	3.0	1010	0.003	58	32.4	...	7.2	6.9	1.6	
12	13T	James Harden	31	TOT	0.0	1.0	1010	0.001	44	36.6	...	7.9	10.8	1.2	
13	13T	LeBron James	36	LAL	0.0	1.0	1010	0.001	45	33.4	...	7.7	7.8	1.1	
14	13T	Kawhi Leonard	29	LAC	0.0	1.0	1010	0.001	52	34.1	...	6.5	5.2	1.6	

5 rows × 21 columns

In [91]:
```
mvps.to_csv("mvps.csv")
```

In [58]:
```
player_stats_url = "https://www.basketball-reference.com/leagues/NBA_{}_per_game.ht

for year in years:
    url = player_stats_url.format(year)

    data = requests.get(url)

    with open("player/{}.html".format(year), "w+") as f:
        f.write(data.text)
```

In [64]:
```
import os
from selenium import webdriver
from selenium.webdriver.common.keys import Keys
import time

# install selenium chrome driver from https://chromedriver.chromium.org/downloads
# xattr -d com.apple.quarantine chromedriver
```

In [68]:
```
driver = webdriver.Chrome(
    executable_path="/Users/aiden/chromedriver"
    )
```

```
/var/folders/xz/9z84c__j28g8tg28bmcthjj00000gn/T/ipykernel_76608/3988874453.py:1: De
precationWarning: executable_path has been deprecated, please pass in a Service obje
ct
  driver = webdriver.Chrome(
```

In [67]:
```
for year in years:
    url = player_stats_url.format(year)

    driver.get(url)
```

```
    driver.execute_script("window.scrollTo(1,10000)")
    time.sleep(2)

    with open("player/{}.html".format(year), "w+") as f:
        f.write(driver.page_source)
```

In [73]:
```
dfs = []
for year in years:
    with open("player/{}.html".format(year)) as f:
        page = f.read()

    soup = BeautifulSoup(page, 'html.parser')
    soup.find('tr', class_="thead").decompose()
    player_table = soup.find_all(id="per_game_stats")[0]
    player_df = pd.read_html(str(player_table))[0]
    player_df["Year"] = year
    dfs.append(player_df)
```

In [74]:
```
players = pd.concat(dfs)
```

Out[74]:

	Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	...	ORB	DRB	TRB	AST	S
725	536	Delon Wright	PG	28	SAC	27	8	25.8	3.9	8.3	...	1.0	2.9	3.9	3.6	
726	537	Thaddeus Young	PF	32	CHI	68	23	24.3	5.4	9.7	...	2.5	3.8	6.2	4.3	
727	538	Trae Young	PG	22	ATL	63	63	33.7	7.7	17.7	...	0.6	3.3	3.9	9.4	
728	539	Cody Zeller	C	28	CHO	48	21	20.9	3.8	6.8	...	2.5	4.4	6.8	1.8	
729	540	Ivica Zubac	C	23	LAC	72	33	22.3	3.6	5.5	...	2.6	4.6	7.2	1.3	

5 rows × 31 columns

In [76]:
```
players.head()
```

Out[76]:

	Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	...	ORB	DRB	TRB	AST	STL
0	1	Alaa Abdelnaby	PF	22	POR	43	0	6.7	1.3	2.7	...	0.6	1.4	2.1	0.3	0.1
1	2	Mahmoud Abdul-Rauf	PG	21	DEN	67	19	22.5	6.2	15.1	...	0.5	1.3	1.8	3.1	0.8
2	3	Mark Acres	C	28	ORL	68	0	19.3	1.6	3.1	...	2.1	3.2	5.3	0.4	0.4
3	4	Michael Adams	PG	28	DEN	66	66	35.5	8.5	21.5	...	0.9	3.0	3.9	10.5	2.2
4	5	Mark Aguirre	SF	31	DET	78	13	25.7	5.4	11.7	...	1.7	3.1	4.8	1.8	0.6

5 rows × 31 columns

In [92]:
```
players.to_csv("players.csv")
```

In [77]:
```
team_stats_url = "https://www.basketball-reference.com/leagues/NBA_{}_standings.htm
```

In [78]:
```
for year in years:
    url = team_stats_url.format(year)

    data = requests.get(url)

    with open("team/{}.html".format(year), "w+") as f:
        f.write(data.text)
```

In [86]:
```
dfs = []
for year in years:
    with open("team/{}.html".format(year)) as f:
        page = f.read()

    soup = BeautifulSoup(page, 'html.parser')
    soup.find('tr', class_="thead").decompose()
    e_table = soup.find_all(id="divs_standings_E")[0]
    e_df = pd.read_html(str(e_table))[0]
    e_df["Year"] = year
    e_df["Team"] = e_df["Eastern Conference"]
    del e_df["Eastern Conference"]
    dfs.append(e_df)

    w_table = soup.find_all(id="divs_standings_W")[0]
    w_df = pd.read_html(str(w_table))[0]
    w_df["Year"] = year
    w_df["Team"] = w_df["Western Conference"]
    del w_df["Western Conference"]
    dfs.append(w_df)
```

In [87]:
```
teams = pd.concat(dfs)
```

In [88]:
```
teams.tail()
```

Out[88]:

	W	L	W/L%	GB	PS/G	PA/G	SRS	Year	Team
13	42	30	.583	—	112.4	110.2	2.26	2021	Dallas Mavericks*
14	38	34	.528	4.0	113.3	112.3	1.07	2021	Memphis Grizzlies*
15	33	39	.458	9.0	111.1	112.8	-1.58	2021	San Antonio Spurs
16	31	41	.431	11.0	114.6	114.9	-0.20	2021	New Orleans Pelicans
17	17	55	.236	25.0	108.8	116.7	-7.50	2021	Houston Rockets

In [89]:
```
teams.head()
```

Out[89]:

	W	L	W/L%	GB	PS/G	PA/G	SRS	Year	Team
0	56	26	.683	—	111.5	105.7	5.22	1991	Boston Celtics*
1	44	38	.537	12.0	105.4	105.6	-0.39	1991	Philadelphia 76ers*
2	39	43	.476	17.0	103.1	103.3	-0.43	1991	New York Knicks*
3	30	52	.366	26.0	101.4	106.4	-4.84	1991	Washington Bullets
4	26	56	.317	30.0	102.9	107.5	-4.53	1991	New Jersey Nets

In [93]:
```
teams.to_csv("teams.csv")
```

In [ ]:
```

```