

```
In [ ]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import matplotlib.ticker as mtick
import sqlite3
import seaborn as sns
from matplotlib.offsetbox import OffsetImage, AnnotationBbox
from selenium import webdriver
from selenium.webdriver.common.keys import Keys
from bs4 import BeautifulSoup
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
import time
import requests
import shutil
import datetime
from scipy.stats import norm
from sklearn.preprocessing import OneHotEncoder, OrdinalEncoder
import os
import winsound
os.chdir('C:\\\\Users\\\\Travis\\\\OneDrive\\\\Data Science\\\\Personal_Projects\\\\Sports\\\\NBA')
import warnings
warnings.filterwarnings('ignore')

from selenium.common.exceptions import WebDriverException
```

```
In [ ]: url= 'https://www.nba.com/stats/teams'
driver = webdriver.Chrome()
```

```
In [ ]: driver.get(url)
# # find by class StatsTeamsList_divContent__JvxSY
soup = BeautifulSoup(driver.page_source, 'html.parser')

# get all links
links = soup.find_all('a', href=True)
# keep only links and link text
links = [[link['href'], link.text] for link in links]
# keep only links with /stats/team/ in them
links = [link for link in links if '/stats/team/' in link[0]]
# save to dataframe
links = pd.DataFrame(links, columns=['link', 'team'])
links
```

Out[ ]:

	link	team
0	/stats/team/1610612738/	Boston Celtics
1	/stats/team/1610612758/	Sacramento Kings
2	/stats/team/1610612740/	New Orleans Pelicans
3	/stats/team/1610612744/	Golden State Warriors
4	/stats/team/1610612762/	Utah Jazz
...	...	...
225	/stats/team/1610612738/	Boston Celtics
226	/stats/team/1610612744/	Golden State Warriors
227	/stats/team/1610612738/	Boston Celtics
228	/stats/team/1610612742/	Dallas Mavericks
229	/stats/team/1610612744/	Golden State Warriors

230 rows × 2 columns

In [ ]:

```
links['url2'] = links['link'].str[6:]
links
```

Out[ ]:

	link	team	url2
0	/stats/team/1610612738/	Boston Celtics	/team/1610612738/
1	/stats/team/1610612758/	Sacramento Kings	/team/1610612758/
2	/stats/team/1610612740/	New Orleans Pelicans	/team/1610612740/
3	/stats/team/1610612744/	Golden State Warriors	/team/1610612744/
4	/stats/team/1610612762/	Utah Jazz	/team/1610612762/
...	...	...	...
225	/stats/team/1610612738/	Boston Celtics	/team/1610612738/
226	/stats/team/1610612744/	Golden State Warriors	/team/1610612744/
227	/stats/team/1610612738/	Boston Celtics	/team/1610612738/
228	/stats/team/1610612742/	Dallas Mavericks	/team/1610612742/
229	/stats/team/1610612744/	Golden State Warriors	/team/1610612744/

230 rows × 3 columns

```
In [ ]: links['url'] = 'https://www.nba.com' + links['url2']
# if link ends with /, remove it
links['url'] = links['url'].apply(lambda x: x[:-1] if x[-1] == '/' else x)

# remove duplicates
links = links.drop_duplicates(subset='url')
links = links.sort_values('team')
links

# find by class StatsTeamsList_divContent__JvxSY
```

Out[ ]:	link	team	url2	url
42	/stats /team/1610612737/	Atlanta Hawks	/team/1610612737/	<a href="https://www.nba.com/team/1610612737">https://www.nba.com/team/1610612737</a>
0	/stats /team/1610612738/	Boston Celtics	/team/1610612738/	<a href="https://www.nba.com/team/1610612738">https://www.nba.com/team/1610612738</a>
15	/stats /team/1610612751/	Brooklyn Nets	/team/1610612751/	<a href="https://www.nba.com/team/1610612751">https://www.nba.com/team/1610612751</a>
9	/stats /team/1610612766/	Charlotte Hornets	/team/1610612766/	<a href="https://www.nba.com/team/1610612766">https://www.nba.com/team/1610612766</a>
29	/stats /team/1610612741/	Chicago Bulls	/team/1610612741/	<a href="https://www.nba.com/team/1610612741">https://www.nba.com/team/1610612741</a>
51	/stats /team/1610612739	Cleveland Cavaliers	/team/1610612739	<a href="https://www.nba.com/team/1610612739">https://www.nba.com/team/1610612739</a>
34	/stats /team/1610612742/	Dallas Mavericks	/team/1610612742/	<a href="https://www.nba.com/team/1610612742">https://www.nba.com/team/1610612742</a>
11	/stats /team/1610612743/	Denver Nuggets	/team/1610612743/	<a href="https://www.nba.com/team/1610612743">https://www.nba.com/team/1610612743</a>
52	/stats /team/1610612765	Detroit Pistons	/team/1610612765	<a href="https://www.nba.com/team/1610612765">https://www.nba.com/team/1610612765</a>
3	/stats /team/1610612744/	Golden State Warriors	/team/1610612744/	<a href="https://www.nba.com/team/1610612744">https://www.nba.com/team/1610612744</a>
8	/stats /team/1610612745/	Houston Rockets	/team/1610612745/	<a href="https://www.nba.com/team/1610612745">https://www.nba.com/team/1610612745</a>
12	/stats /team/1610612754/	Indiana Pacers	/team/1610612754/	<a href="https://www.nba.com/team/1610612754">https://www.nba.com/team/1610612754</a>
66	/stats /team/1610612746	LA Clippers	/team/1610612746	<a href="https://www.nba.com/team/1610612746">https://www.nba.com/team/1610612746</a>
67	/stats /team/1610612747	Los Angeles Lakers	/team/1610612747	<a href="https://www.nba.com/team/1610612747">https://www.nba.com/team/1610612747</a>
6	/stats /team/1610612763/	Memphis Grizzlies	/team/1610612763/	<a href="https://www.nba.com/team/1610612763">https://www.nba.com/team/1610612763</a>
22	/stats /team/1610612748/	Miami Heat	/team/1610612748/	<a href="https://www.nba.com/team/1610612748">https://www.nba.com/team/1610612748</a>
5	/stats /team/1610612749/	Milwaukee Bucks	/team/1610612749/	<a href="https://www.nba.com/team/1610612749">https://www.nba.com/team/1610612749</a>
23	/stats /team/1610612750/	Minnesota Timberwolves	/team/1610612750/	<a href="https://www.nba.com/team/1610612750">https://www.nba.com/team/1610612750</a>
2	/stats /team/1610612740/	New Orleans Pelicans	/team/1610612740/	<a href="https://www.nba.com/team/1610612740">https://www.nba.com/team/1610612740</a>
7	/stats /team/1610612752/	New York Knicks	/team/1610612752/	<a href="https://www.nba.com/team/1610612752">https://www.nba.com/team/1610612752</a>
24	/stats /team/1610612760/	Oklahoma City Thunder	/team/1610612760/	<a href="https://www.nba.com/team/1610612760">https://www.nba.com/team/1610612760</a>

	<b>link</b>	<b>team</b>	<b>url2</b>	<b>url</b>
<b>58</b>	/stats /team/1610612753	Orlando Magic	/team/1610612753	<a href="https://www.nba.com/team/1610612753">https://www.nba.com/team/1610612753</a>
<b>39</b>	/stats /team/1610612755/	Philadelphia 76ers	/team/1610612755/	<a href="https://www.nba.com/team/1610612755">https://www.nba.com/team/1610612755</a>
<b>14</b>	/stats /team/1610612756/	Phoenix Suns	/team/1610612756/	<a href="https://www.nba.com/team/1610612756">https://www.nba.com/team/1610612756</a>
<b>38</b>	/stats /team/1610612757/	Portland Trail Blazers	/team/1610612757/	<a href="https://www.nba.com/team/1610612757">https://www.nba.com/team/1610612757</a>
<b>1</b>	/stats /team/1610612758/	Sacramento Kings	/team/1610612758/	<a href="https://www.nba.com/team/1610612758">https://www.nba.com/team/1610612758</a>
<b>13</b>	/stats /team/1610612759/	San Antonio Spurs	/team/1610612759/	<a href="https://www.nba.com/team/1610612759">https://www.nba.com/team/1610612759</a>
<b>20</b>	/stats /team/1610612761/	Toronto Raptors	/team/1610612761/	<a href="https://www.nba.com/team/1610612761">https://www.nba.com/team/1610612761</a>
<b>4</b>	/stats /team/1610612762/	Utah Jazz	/team/1610612762/	<a href="https://www.nba.com/team/1610612762">https://www.nba.com/team/1610612762</a>
<b>18</b>	/stats /team/1610612764/	Washington Wizards	/team/1610612764/	<a href="https://www.nba.com/team/1610612764">https://www.nba.com/team/1610612764</a>

```
In [ ]: links['team_name'] = links['team'].apply(lambda x: x.split(' ')[-1])
links
```

Out[ ]:		link	team	url2	url	team_name
	42	/stats /team/1610612737/	Atlanta Hawks	/team/1610612737/	https://www.nba.com /team/1610612737	Hawks
	0	/stats /team/1610612738/	Boston Celtics	/team/1610612738/	https://www.nba.com /team/1610612738	Celtics
	15	/stats /team/1610612751/	Brooklyn Nets	/team/1610612751/	https://www.nba.com /team/1610612751	Nets
	9	/stats /team/1610612766/	Charlotte Hornets	/team/1610612766/	https://www.nba.com /team/1610612766	Hornets
	29	/stats /team/1610612741/	Chicago Bulls	/team/1610612741/	https://www.nba.com /team/1610612741	Bulls
	51	/stats /team/1610612739	Cleveland Cavaliers	/team/1610612739	https://www.nba.com /team/1610612739	Cavaliers
	34	/stats /team/1610612742/	Dallas Mavericks	/team/1610612742/	https://www.nba.com /team/1610612742	Mavericks
	11	/stats /team/1610612743/	Denver Nuggets	/team/1610612743/	https://www.nba.com /team/1610612743	Nuggets
	52	/stats /team/1610612765	Detroit Pistons	/team/1610612765	https://www.nba.com /team/1610612765	Pistons
	3	/stats /team/1610612744/	Golden State Warriors	/team/1610612744/	https://www.nba.com /team/1610612744	Warriors
	8	/stats /team/1610612745/	Houston Rockets	/team/1610612745/	https://www.nba.com /team/1610612745	Rockets
	12	/stats /team/1610612754/	Indiana Pacers	/team/1610612754/	https://www.nba.com /team/1610612754	Pacers
	66	/stats /team/1610612746	LA Clippers	/team/1610612746	https://www.nba.com /team/1610612746	Clippers
	67	/stats /team/1610612747	Los Angeles Lakers	/team/1610612747	https://www.nba.com /team/1610612747	Lakers
	6	/stats /team/1610612763/	Memphis Grizzlies	/team/1610612763/	https://www.nba.com /team/1610612763	Grizzlies
	22	/stats /team/1610612748/	Miami Heat	/team/1610612748/	https://www.nba.com /team/1610612748	Heat
	5	/stats /team/1610612749/	Milwaukee Bucks	/team/1610612749/	https://www.nba.com /team/1610612749	Bucks
	23	/stats /team/1610612750/	Minnesota Timberwolves	/team/1610612750/	https://www.nba.com /team/1610612750	Timberwolves
	2	/stats /team/1610612740/	New Orleans Pelicans	/team/1610612740/	https://www.nba.com /team/1610612740	Pelicans
	7	/stats /team/1610612752/	New York Knicks	/team/1610612752/	https://www.nba.com /team/1610612752	Knicks
	24	/stats /team/1610612760/	Oklahoma City Thunder	/team/1610612760/	https://www.nba.com /team/1610612760	Thunder

	<b>link</b>	<b>team</b>	<b>url2</b>	<b>url</b>	<b>team_name</b>
<b>58</b>	/stats /team/1610612753	Orlando Magic	/team/1610612753	https://www.nba.com /team/1610612753	Magic
<b>39</b>	/stats /team/1610612755/	Philadelphia 76ers	/team/1610612755/	https://www.nba.com /team/1610612755	76ers
<b>14</b>	/stats /team/1610612756/	Phoenix Suns	/team/1610612756/	https://www.nba.com /team/1610612756	Suns
<b>38</b>	/stats /team/1610612757/	Portland Trail Blazers	/team/1610612757/	https://www.nba.com /team/1610612757	Blazers
<b>1</b>	/stats /team/1610612758/	Sacramento Kings	/team/1610612758/	https://www.nba.com /team/1610612758	Kings
<b>13</b>	/stats /team/1610612759/	San Antonio Spurs	/team/1610612759/	https://www.nba.com /team/1610612759	Spurs
<b>20</b>	/stats /team/1610612761/	Toronto Raptors	/team/1610612761/	https://www.nba.com /team/1610612761	Raptors
<b>4</b>	/stats /team/1610612762/	Utah Jazz	/team/1610612762/	https://www.nba.com /team/1610612762	Jazz
<b>18</b>	/stats /team/1610612764/	Washington Wizards	/team/1610612764/	https://www.nba.com /team/1610612764	Wizards

```
In [ ]: links['fixed_url'] = links['url']+ '/' + links['team_name'].str.lower()
```

```
In [ ]: all_roster_links = links['fixed_url'].to_list()
all_roster_links
```

```
Out[ ]: ['https://www.nba.com/team/1610612737/hawks',
 'https://www.nba.com/team/1610612738/celtics',
 'https://www.nba.com/team/1610612751/nets',
 'https://www.nba.com/team/1610612766/hornets',
 'https://www.nba.com/team/1610612741/bulls',
 'https://www.nba.com/team/1610612739/cavaliers',
 'https://www.nba.com/team/1610612742/mavericks',
 'https://www.nba.com/team/1610612743/nuggets',
 'https://www.nba.com/team/1610612765/pistons',
 'https://www.nba.com/team/1610612744/warriors',
 'https://www.nba.com/team/1610612745/rockets',
 'https://www.nba.com/team/1610612754/pacers',
 'https://www.nba.com/team/1610612746/clippers',
 'https://www.nba.com/team/1610612747/lakers',
 'https://www.nba.com/team/1610612763/grizzlies',
 'https://www.nba.com/team/1610612748/heat',
 'https://www.nba.com/team/1610612749/bucks',
 'https://www.nba.com/team/1610612750/timberwolves',
 'https://www.nba.com/team/1610612740/pelicans',
 'https://www.nba.com/team/1610612752/knicks',
 'https://www.nba.com/team/1610612760/thunder',
 'https://www.nba.com/team/1610612753/magic',
 'https://www.nba.com/team/1610612755/76ers',
 'https://www.nba.com/team/1610612756/suns',
 'https://www.nba.com/team/1610612757/blazers',
 'https://www.nba.com/team/1610612758/kings',
 'https://www.nba.com/team/1610612759/spurs',
 'https://www.nba.com/team/1610612761/raptors',
 'https://www.nba.com/team/1610612762/jazz',
 'https://www.nba.com/team/1610612764/wizards']
```

```
In [ ]: pd.read_html('https://www.nba.com/team/1610612737/hawks')[0]
```

Out[ ]:

	Player	#	Pos	Height	Weight	Birthdate	Age	Exp	School
0	Frank Kaminsky	0	F-C	7-0	240 lbs	APR 04, 1993	29	7	Wisconsin
1	Jalen Johnson	1	F	6-8	219 lbs	DEC 18, 2001	21	1	Duke
2	Trent Forrest	2	G	6-4	210 lbs	JUN 12, 1998	24	2	Florida State
3	Aaron Holiday	3	G	6-0	185 lbs	SEP 30, 1996	26	4	UCLA
4	Dejounte Murray	5	G	6-5	180 lbs	SEP 19, 1996	26	5	Washington
5	Jarrett Culver	7	G-F	6-6	195 lbs	FEB 20, 1999	23	3	Texas Tech
6	Justin Holiday	8	F-G	6-6	180 lbs	APR 05, 1989	33	9	Washington
7	Trae Young	11	G	6-1	164 lbs	SEP 19, 1998	24	4	Oklahoma
8	De'Andre Hunter	12	F-G	6-8	221 lbs	DEC 02, 1997	25	3	Virginia
9	Bogdan Bogdanovic	13	G	6-5	225 lbs	AUG 18, 1992	30	5	Fenerbahce
10	AJ Griffin	14	F	6-6	220 lbs	AUG 25, 2003	19	R	Duke
11	Clint Capela	15	C	6-10	256 lbs	MAY 18, 1994	28	8	Elan Chalon
12	Onyeka Okongwu	17	F-C	6-9	240 lbs	DEC 11, 2000	22	2	Southern California
13	John Collins	20	F-C	6-9	226 lbs	SEP 23, 1997	25	5	Wake Forest
14	Tyrese Martin	22	G	6-6	215 lbs	MAR 07, 1999	23	R	Connecticut
15	Vit Krejci	27	G	6-8	195 lbs	JUN 19, 2000	22	1	Zaragoza

In [ ]: `all_roster_links[0]`

Out[ ]: `'https://www.nba.com/team/1610612737/hawks'`

In [ ]: `rosters = pd.DataFrame()`

```
for f in all_roster_links:
    # wait for page to load
    time.sleep(3)
    try:
        data = pd.read_html(f)[0]
    except:
        time.sleep(5)
        data = pd.read_html(f)[0]
    team = f.split('/')[-1]
    data['team'] = team
    data.to_csv('data/team/Current_Rosters/'+team+'.csv', index=False)
    rosters = rosters.append(data)
    print(team)
```

`rosters`

hawks  
celtics  
nets  
hornets  
bulls  
cavaliers  
mavericks  
nuggets  
pistons  
warriors  
rockets  
pacers  
clippers  
lakers  
grizzlies  
heat  
bucks  
timberwolves  
pelicans  
knicks  
thunder  
magic  
76ers  
suns  
blazers  
kings  
spurs  
raptors  
jazz  
wizards

Out[ ]:

	Player	#	Pos	Height	Weight	Birthdate	Age	Exp	School	team
0	Frank Kaminsky	0.0	F-C	7-0	240 lbs	APR 04, 1993	29	7	Wisconsin	hawks
1	Jalen Johnson	1.0	F	6-8	219 lbs	DEC 18, 2001	21	1	Duke	hawks
2	Trent Forrest	2.0	G	6-4	210 lbs	JUN 12, 1998	24	2	Florida State	hawks
3	Aaron Holiday	3.0	G	6-0	185 lbs	SEP 30, 1996	26	4	UCLA	hawks
4	Dejounte Murray	5.0	G	6-5	180 lbs	SEP 19, 1996	26	5	Washington	hawks
...	...	...	...	...	...	...	...	...	...	...
12	Monte Morris	22.0	G	6-2	183 lbs	JUN 27, 1995	27	5	Iowa State	wizards
13	Corey Kispert	24.0	F	6-6	224 lbs	MAR 03, 1999	23	1	Gonzaga	wizards
14	Kyle Kuzma	33.0	F	6-9	221 lbs	JUL 24, 1995	27	5	Utah	wizards
15	Delon Wright	55.0	G	6-5	185 lbs	APR 26, 1992	30	7	Utah	wizards
16	Taj Gibson	67.0	F	6-9	232 lbs	JUN 24, 1985	37	13	Southern California	wizards

499 rows × 10 columns

In [ ]: `# save roster data  
rosters.to_csv('data/team/Current_Rosters/0_All_rosters.csv', index=False)`

In [ ]: