

In [1]:

```
1 import pandas as pd
2 import numpy as np
3 import requests
4 import time
```

In [4]:

```
1 test_url = 'https://stats.nba.com/stats/leagueLeaders?LeagueID=00&PerMode=Totals&Scope=S&Season=2012-13&SeasonType'
```

In [8]:

```
1 r = requests.get(url= test_url).json()
```

In [11]:

```
1 table_headers = r['resultSet']['headers']
```

In [17]:

```
1 pd.DataFrame(r['resultSet']['rowSet'],columns=table_headers)
```

Out[17]:

	PLAYER_ID	RANK	PLAYER	TEAM_ID	TEAM	GP	MIN	FGM	FGA	FG_PCT	...	REB	AST	STL	BLK	TOV	PF	PTS	EFF	AST_T
0	201142	1	Kevin Durant	1610612760	OKC	81	3119	731	1433	0.510	...	640	374	116	105	280	143	2280	2462	1
1	977	2	Kobe Bryant	1610612747	LAL	78	3013	738	1595	0.463	...	433	469	106	25	287	173	2133	1921	1
2	2544	3	LeBron James	1610612748	MIA	76	2877	765	1354	0.565	...	610	551	129	67	226	110	2036	2446	2
3	201935	4	James Harden	1610612745	HOU	78	2985	585	1337	0.438	...	379	455	142	38	295	178	2023	1872	1
4	2546	5	Carmelo Anthony	1610612752	NYK	67	2482	669	1489	0.449	...	460	171	52	32	175	205	1920	1553	C
...
463	203130	463	Darius Johnson-Odom	1610612747	LAL	4	6	0	4	0.000	...	4	1	0	0	0	0	0	1	C
464	2545	463	Darko Milicic	1610612738	BOS	1	5	0	1	0.000	...	1	0	0	0	2	1	0	-2	C
465	202458	463	Justin Dentmon	1610612742	DAL	2	4	0	2	0.000	...	0	0	0	0	0	0	0	-2	C
466	2679	463	Matt Carroll	1610612766	CHA	1	6	0	0	0.000	...	0	1	0	0	0	1	0	1	C
467	200780	463	Solomon Jones	1610612752	NYK	2	26	0	1	0.000	...	3	0	0	1	3	3	0	0	C

468 rows × 28 columns

In [44]:

```
1 temp_df1 = pd.DataFrame(r['resultSet']['rowSet'],columns=table_headers)
2 temp_df2 = pd.DataFrame({'Year': ['2012-13' for i in range(len(temp_df1))],
3                           'Season_type': ['Regular%20Season' for i in range(len(temp_df1))])
4 temp_df3 = pd.concat([temp_df2, temp_df1], axis = 1)
```

```
In [45]: 1 temp_df3
```

Out[45]:

	Year	Season_type	PLAYER_ID	RANK	PLAYER	TEAM_ID	TEAM	GP	MIN	FGM	...	REB	AST	STL	BLK	TOV	PF	PTS	E
0	2012-13	Regular%20Season	201142	1	Kevin Durant	1610612760	OKC	81	3119	731	...	640	374	116	105	280	143	2280	24
1	2012-13	Regular%20Season	977	2	Kobe Bryant	1610612747	LAL	78	3013	738	...	433	469	106	25	287	173	2133	19
2	2012-13	Regular%20Season	2544	3	LeBron James	1610612748	MIA	76	2877	765	...	610	551	129	67	226	110	2036	24
3	2012-13	Regular%20Season	201935	4	James Harden	1610612745	HOU	78	2985	585	...	379	455	142	38	295	178	2023	18
4	2012-13	Regular%20Season	2546	5	Carmelo Anthony	1610612752	NYK	67	2482	669	...	460	171	52	32	175	205	1920	15
...
463	2012-13	Regular%20Season	203130	463	Darius Johnson-Odom	1610612747	LAL	4	6	0	...	4	1	0	0	0	0	0	0
464	2012-13	Regular%20Season	2545	463	Darko Milicic	1610612738	BOS	1	5	0	...	1	0	0	0	2	1	0	0
465	2012-13	Regular%20Season	202458	463	Justin Dentmon	1610612742	DAL	2	4	0	...	0	0	0	0	0	0	0	0
466	2012-13	Regular%20Season	2679	463	Matt Carroll	1610612766	CHA	1	6	0	...	0	1	0	0	0	1	0	0
467	2012-13	Regular%20Season	200780	463	Solomon Jones	1610612752	NYK	2	26	0	...	3	0	0	1	3	3	0	0

468 rows × 30 columns

```
In [46]: 1 del temp_df1, temp_df2, temp_df3, df
```

```
In [33]: 1 df_cols = ['Year', 'Season_type'] + table_headers
```

```
In [47]: 1 pd.DataFrame(columns=df_cols)
```

Out[47]:

Year	Season_type	PLAYER_ID	RANK	PLAYER	TEAM_ID	TEAM	GP	MIN	FGM	...	REB	AST	STL	BLK	TOV	PF	PTS	EFF	AST_TOV
------	-------------	-----------	------	--------	---------	------	----	-----	-----	-----	-----	-----	-----	-----	-----	----	-----	-----	---------

0 rows × 30 columns

```
In [49]: 1 df = pd.DataFrame(columns=df_cols)
2
3 years = ['2012-13', '2013-14', '2014-15', '2015-16', '2016-17', '2017-18', '2018-19', '2019-20', '2020-21', '2021-22']
4 season_types = ['Regular%20Season', 'Playoffs']
5
6 begin_time = time.time()
7
8 for y in years:
9     for s in season_types:
10         url = 'https://stats.nba.com/stats/leagueLeaders?LeagueID=00&PerMode=Totals&Scope=S&Season='+y+'&SeasonTy
11         r = requests.get(url=url).json()
12
13         temp_df1 = pd.DataFrame(r['resultSet']['rowSet'], columns=table_headers)
14         temp_df2 = pd.DataFrame({'Year': [y for i in range(len(temp_df1))],
15                                 'Season_type': [s for i in range(len(temp_df1))]}))
16         temp_df3 = pd.concat([temp_df2, temp_df1], axis = 1)
17         df =pd.concat([df,temp_df3], axis = 0)
18         print(f'Scraped data for {y} {s}')
19         lag = np.random.uniform(5,20)
20         print(f'waiting for {round(lag,1)} seconds')
21         time.sleep(lag)
22
23
24 df.to_excel('nba_players_data.xlsx', headers = False)
```

Scraped data for 2012-13 Regular%20Season
waiting for 19.8 seconds
Scraped data for 2012-13 Playoffs
waiting for 18.9 seconds
Scraped data for 2013-14 Regular%20Season
waiting for 5.6 seconds
Scraped data for 2013-14 Playoffs
waiting for 18.5 seconds
Scraped data for 2014-15 Regular%20Season
waiting for 15.7 seconds
Scraped data for 2014-15 Playoffs
waiting for 9.6 seconds
Scraped data for 2015-16 Regular%20Season
waiting for 10.2 seconds
Scraped data for 2015-16 Playoffs
waiting for 18.3 seconds
Scraped data for 2016-17 Regular%20Season
waiting for 13.2 seconds
Scraped data for 2016-17 Playoffs
waiting for 19.8 seconds
Scraped data for 2017-18 Regular%20Season
waiting for 19.0 seconds
Scraped data for 2017-18 Playoffs
waiting for 13.7 seconds
Scraped data for 2018-19 Regular%20Season
waiting for 14.8 seconds
Scraped data for 2018-19 Playoffs
waiting for 13.6 seconds
Scraped data for 2019-20 Regular%20Season
waiting for 16.7 seconds
Scraped data for 2019-20 Playoffs
waiting for 11.9 seconds
Scraped data for 2020-21 Regular%20Season
waiting for 6.3 seconds
Scraped data for 2020-21 Playoffs
waiting for 13.7 seconds
Scraped data for 2021-22 Regular%20Season
waiting for 12.4 seconds
Scraped data for 2021-22 Playoffs
waiting for 12.5 seconds

NameError Traceback (most recent call last)
Input In [49], in <module>
 20 print(f'waiting for {round(lag,1)} seconds')
 21 time.sleep(lag)
----> 23 print(f'Process completed in {round((time.time() - begin)/60,2)} minutes')
 24 df.to_excel('nba_players_data.xlsx', headers = False)

NameError: name 'begin' is not defined

```
In [56]: 1 df.to_excel('nba_players_data.xlsx', index = False)
```

In [57]:

1df.head()

Out[57]:

	Year	Season_type	PLAYER_ID	RANK	PLAYER	TEAM_ID	TEAM	GP	MIN	FGM	...	REB	AST	STL	BLK	TOV	PF	PTS	EFF
0	2012-13	Regular%20Season	201142	1	Kevin Durant	1610612760	OKC	81	3119	731	...	640	374	116	105	280	143	2280	2462
1	2012-13	Regular%20Season	977	2	Kobe Bryant	1610612747	LAL	78	3013	738	...	433	469	106	25	287	173	2133	1921
2	2012-13	Regular%20Season	2544	3	LeBron James	1610612748	MIA	76	2877	765	...	610	551	129	67	226	110	2036	2446
3	2012-13	Regular%20Season	201935	4	James Harden	1610612745	HOU	78	2985	585	...	379	455	142	38	295	178	2023	1872
4	2012-13	Regular%20Season	2546	5	Carmelo Anthony	1610612752	NYK	67	2482	669	...	460	171	52	32	175	205	1920	1553

5 rows × 30 columns

In []:

1