

```
In [1]: import pandas as pd

In [2]: df = pd.read_csv("all_seasons.csv")

In [3]: df.rename(columns = {'player_name':'name', 'team_abbreviation':'team', 'player_height':'height', 'player_weight':'weight', 'college':'college', 'country':'country', 'draft_year':'draft_year', 'draft_round':'draft_round', 'draft_number':'draft_number', 'avg_pts':'avg_pts', 'reb':'avg_reb', 'ast':'avg_ast'}, inplace = True)

In [4]: df['year'] = df['season'].apply(lambda x: int(x.split('-')[0]))

In [5]: df.head()

Out[5]:
```

	Unnamed: 0	name	team	age	height	weight	college	country	draft_year	draft_round	...	avg_reb	avg_ast	net
0	0	Randy Livingston	HOU	22.0	193.04	94.800728	Louisiana State	USA	1996	2	...	1.5	2.4	
1	1	Gaylon Nickerson	WAS	28.0	190.50	86.182480	Northwestern Oklahoma	USA	1994	2	...	1.3	0.3	
2	2	George Lynch	VAN	26.0	203.20	103.418976	North Carolina	USA	1993	1	...	6.4	1.9	
3	3	George McCloud	LAL	30.0	203.20	102.058200	Florida State	USA	1989	1	...	2.8	1.7	
4	4	George Zidek	DEN	23.0	213.36	119.748288	UCLA	USA	1995	1	...	1.7	0.3	

5 rows × 23 columns

```
In [6]: df.groupby(['name']).count()

Out[6]:
```

	Unnamed: 0	team	age	height	weight	college	country	draft_year	draft_round	draft_number	...	avg_reb	avg_ast
name													
A.C. Green	5	5	5	5	5	5	5	5	5	5	...	5	5
A.J. Bramlett	1	1	1	1	1	1	1	1	1	1	...	1	1
A.J. Guyton	3	3	3	3	3	3	3	3	3	3	...	3	3
A.J. Lawson	1	1	1	1	1	1	1	1	1	1	...	1	1
AJ Green	1	1	1	1	1	1	1	1	1	1	...	1	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...
Zion Williamson	3	3	3	3	3	3	3	3	3	3	...	3	3
Zoran Dragic	1	1	1	1	1	1	1	1	1	1	...	1	1
Zoran Planinic	3	3	3	3	3	3	3	3	3	3	...	3	3
Zydrunas Ilgauskas	13	13	13	13	13	13	13	13	13	13	...	13	13
Zylan Cheatham	2	2	2	2	2	2	2	2	2	2	...	2	2

2551 rows × 22 columns

```
In [8]: df['pts'] = (df['gp']*df['avg_pts']).astype(int)
df['reb'] = (df['gp']*df['avg_reb']).astype(int)
df['ast'] = (df['gp']*df['avg_ast']).astype(int)

In [9]: df.head()

Out[9]:
```

	Unnamed: 0	name	team	age	height	weight	college	country	draft_year	draft_round	...	oreb_pct	dreb_pct	usg_pct
0	0	Randy Livingston	HOU	22.0	193.04	94.800728	Louisiana State	USA	1996	2	...	0.042	0.071	
1	1	Gaylon Nickerson	WAS	28.0	190.50	86.182480	Northwestern Oklahoma	USA	1994	2	...	0.030	0.111	
2	2	George Lynch	VAN	26.0	203.20	103.418976	North Carolina	USA	1993	1	...	0.106	0.185	
3	3	George McCloud	LAL	30.0	203.20	102.058200	Florida State	USA	1989	1	...	0.027	0.111	
4	4	George Zidek	DEN	23.0	213.36	119.748288	UCLA	USA	1995	1	...	0.102	0.169	

5 rows × 26 columns

```
In [35]: df['total_pts'] = df.groupby('name')['pts'].cumsum()
df['total_reb'] = df.groupby('name')['reb'].cumsum()
df['total_ast'] = df.groupby('name')['ast'].cumsum()

In [36]: df[df['name']=='Kyrie Irving'].head()

Out[36]:
```

	Unnamed: 0	name	team	age	height	weight	college	country	draft_year	draft_round	...	ts_pct	ast_pct	season	year
6734	6734	Kyrie Irving	CLE	20.0	190.5	86.636072	Duke	Australia	2011	1	...	0.566	0.324	2011-12	201
7594	7594	Kyrie Irving	CLE	21.0	190.5	86.636072	Duke	Australia	2011	1	...	0.553	0.315	2012-13	201
8027	8027	Kyrie Irving	CLE	22.0	190.5	87.543256	Duke	Australia	2011	1	...	0.533	0.311	2013-14	201
8508	8508	Kyrie Irving	CLE	23.0	190.5	87.543256	Duke	Australia	2011	1	...	0.583	0.236	2014-15	201
8984	8984	Kyrie Irving	CLE	24.0	190.5	87.543256	Duke	Australia	2011	1	...	0.540	0.253	2015-16	201

5 rows × 29 columns

```
In [15]: !pip install bubbly

Collecting bubbly
  Downloading bubbly-1.0.2.tar.gz (5.1 kB)
  Preparing metadata (setup.py) ... done
Requirement already satisfied: plotly in /Users/sameerbaloch/anaconda3/lib/python3.11/site-packages (from bubbly) (5.9.0)
Requirement already satisfied: pandas in /Users/sameerbaloch/anaconda3/lib/python3.11/site-packages (from bubbly) (1.5.3)
Requirement already satisfied: python-dateutil>=2.8.1 in /Users/sameerbaloch/anaconda3/lib/python3.11/site-packages (from pandas->bubbly) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /Users/sameerbaloch/anaconda3/lib/python3.11/site-packages (from pandas->bubbly) (2022.7)
Requirement already satisfied: numpy>=1.21.0 in /Users/sameerbaloch/anaconda3/lib/python3.11/site-packages (from pandas->bubbly) (1.24.3)
Requirement already satisfied: tenacity>=6.2.0 in /Users/sameerbaloch/anaconda3/lib/python3.11/site-packages (from plotly->bubbly) (8.2.2)
Requirement already satisfied: six>=1.5 in /Users/sameerbaloch/anaconda3/lib/python3.11/site-packages (from python-dateutil>=2.8.1->pandas->bubbly) (1.16.0)
Building wheels for collected packages: bubbly
  Building wheel for bubbly (setup.py) ... done
  Created wheel for bubbly: filename=bubbly-1.0.2-py3-none-any.whl size=5415 sha256=afe1c4c64f25a5c6d26fff8aafb8a0ce8850b6f64b675a296097a6b671f319b5
  Stored in directory: /Users/sameerbaloch/Library/Caches/pip/wheels/83/9a/47/387700a4f2acbd0b6b06f3a389913db17982d52e25ee5f9187
Successfully built bubbly
Installing collected packages: bubbly
Successfully installed bubbly-1.0.2

In [37]: import bubbly
from bubbly.bubbly import bubbleplot
from plotly.offline import init_notebook_mode, iplot

In [38]: import warnings
warnings.filterwarnings('ignore')

In [39]: NumTopScorers = 10
top_scorers_names = df.groupby(['name'])['pts'].sum().sort_values(ascending=False).index[:NumTopScorers]
top_scorers_names

Out[39]: Index(['LeBron James', 'Kobe Bryant', 'Dirk Nowitzki', 'Carmelo Anthony', 'Kevin Durant', 'Tim Duncan', 'Paul Pierce', 'Vince Carter', 'Kevin Garnett', 'James Harden'], dtype='object', name='name')

In [40]: topp = df[df['name'].isin(top_scorers_names)][['name', 'team', 'age', 'draft_year', 'year', 'total_pts', 'total_ast', 'total_reb']]

In [41]: for p in top_scorers_names:
    first_year = topp[topp['name']==p]['year'].min()
    fill_missing_years = []
    last_row = topp[(topp['name']==) & (topp['year']==first_year)].to_dict('records')[0]
    for y in range(1996+1, 2021+1):
        if y<first_year: continue
        query = topp[(topp['name']==p) & (topp['year']==y)]
        if query.empty: # no data in year 'y' is found
            row = last_row.copy()
            row['year'] = y
            fill_missing_years.append(row)
        else:
            last_row = query.to_dict('records')[0]
    topp = pd.concat([topp, pd.DataFrame(fill_missing_years)])

In [42]: topp[topp['name']=='Kobe Bryant'].tail(10)

Out[42]:
```

	name	team	age	draft_year	year	total_pts	total_ast	total_reb
7576	Kobe Bryant	LAL	34.0	1996	2012	31603	5894	6575
8021	Kobe Bryant	LAL	35.0	1996	2013	31685	5931	6600
8516	Kobe Bryant	LAL	36.0	1996	2014	32465	6127	6799
8961	Kobe Bryant	LAL	37.0	1996	2015	33626	6311	7043
0	Kobe Bryant	LAL	37.0	1996	2016	33626	6311	7043
1	Kobe Bryant	LAL	37.0	1996	2017	33626	6311	7043
2	Kobe Bryant	LAL	37.0	1996	2018	33626	6311	7043
3	Kobe Bryant	LAL	37.0	1996	2019	33626	6311	7043
4	Kobe Bryant	LAL	37.0	1996	2020	33626	6311	7043
5	Kobe Bryant	LAL	37.0	1996	2021	33626	6311	7043

```
In [43]: topp[topp['name']=='Kevin Durant'].tail()

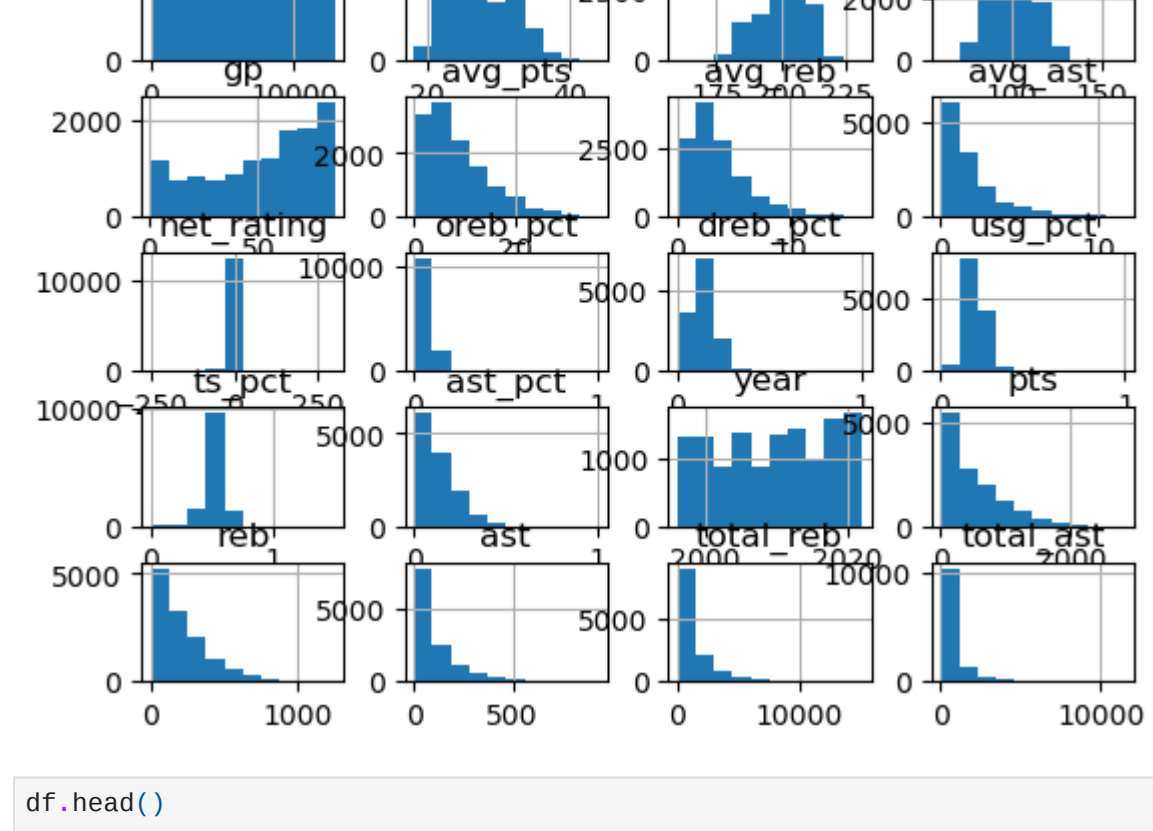
Out[43]:
```

	name	team	age	draft_year	year	total_pts	total_ast	total_reb
10505	Kevin Durant	GSW	30.0	2007	2018	22934	3480	5995
11568	Kevin Durant	BKN	32.0	2007	2020	23875	3676	6243
12189	Kevin Durant	BKN	33.0	2007	2021	25519	4028	6650
12663	Kevin Durant	PHX	34.0	2007	2022	26886	4263	6964
0	Kevin Durant	GSW	30.0	2007	2019	22934	3480	5995

```
In [26]: topp.sort_values(['draft_year', 'name', 'year'], inplace=True)

In [31]: df.hist()

Out[31]: array([[<Axes: title={'center': 'Unnamed: 0'}>,
<Axes: title={'center': 'age'}>,
<Axes: title={'center': 'height'}>,
<Axes: title={'center': 'weight'}>],
[<Axes: title={'center': 'gp'}>,
<Axes: title={'center': 'avg_pts'}>,
<Axes: title={'center': 'avg_reb'}>,
<Axes: title={'center': 'avg_ast'}>],
[<Axes: title={'center': 'net_rating'}>,
<Axes: title={'center': 'oreb_pct'}>,
<Axes: title={'center': 'dreb_pct'}>,
<Axes: title={'center': 'usg_pct'}>],
[<Axes: title={'center': 'ts_pct'}>,
<Axes: title={'center': 'ast_pct'}>,
<Axes: title={'center': 'year'}>,
<Axes: title={'center': 'pts'}>],
[<Axes: title={'center': 'reb'}>, <Axes: title={'center': 'ast'}>],
[<Axes: title={'center': 'total_reb'}>,
<Axes: title={'center': 'total_ast'}>]], dtype=object)
```



```
In [62]: df.head()

Out[62]:
```

	Unnamed: 0	name	team	age	height	weight	college	country	draft_year	draft_round	...	ts_pct	ast_pct	season
0	0	Randy Livingston	HOU	22.0	193.04	94.800728	Louisiana State	USA	1996	2	...	0.487	0.248	1996-97
1	1	Gaylon Nickerson	WAS	28.0	190.50	86.182480	Northwestern Oklahoma	USA	1994	2	...	0.497	0.043	1996-97
2	2	George Lynch	VAN	26.0	203.20	103.418976	North Carolina	USA	1993	1	...	0.512	0.125	1996-97
3	3	George McCloud	LAL	30.0	203.20	102.058200	Florida State	USA	1989	1	...	0.527	0.125	1996-97
4	4	George Zidek	DEN	23.0	213.36	119.748288	UCLA	USA	1995	1	...	0.500	0.064	1996-97

5 rows × 29 columns

```
In [56]: import matplotlib.pyplot as plt
import seaborn as sns

In [63]: pts_data = pd.DataFrame(df.groupby('name')['pts'].mean().sort_values(ascending = False).reset_index())

In [64]: pts_data

Out[64]:
```

	name	pts
0	Michael Jordan	1948.500000
1	LeBron James	1932.050000
2	Luka Doncic	1818.800000
3	Trae Young	1797.200000
4	Kevin Durant	1792.400000
5	James Harden	1763.285714
6	Damian Lillard	1762.090909
7	Allen Iverson	1741.142857
8	Karl Malone	1697.875000
9	Donovan Mitchell	1694.166667

```
In [79]: plt.figure(figsize = (11,9) , dpi = 100)
sns.barplot(data = pts_data , x = 'name' , y = 'pts')
plt.xticks(rotation = 45);
```

