

```
In [1]: # Import necessary libraries
import pandas as pd # Pandas for data manipulation
import matplotlib.pyplot as plt # Matplotlib for basic plotting
import seaborn as sns # Seaborn for statistical data visualization
```

```
In [2]: # Read the CSV file into a DataFrame
df = pd.read_csv("all_seasons.csv")
```

```
In [3]: # Rename selected columns for better readability
df.rename(columns = {'player_name':'name', 'team_abbreviation':'team', 'player_height':'height', 'player_weight':'weight',
                    'pts':'avg_pts', 'reb':'avg_reb', 'ast':'avg_ast'}, inplace = True)
```

```
In [4]: # Display the first few rows of the DataFrame
df.head()
```

	Unnamed: 0	name	team	age	height	weight	college	country	draft_year	draft_round	...	avg_pts	avg_reb	avg_ast	net_rating	oreb_pct	dreb_pct	usg_pct	ts_pct	ast_pct	season
0	0	Randy Livingston	HOU	22.0	193.04	94.800728	Louisiana State	USA	1996	2	...	3.9	1.5	2.4	0.3	0.042	0.071	0.169	0.487	0.248	1996-97
1	1	Gaylon Nickerson	WAS	28.0	190.50	86.182480	Northwestern Oklahoma	USA	1994	2	...	3.8	1.3	0.3	8.9	0.030	0.111	0.174	0.497	0.043	1996-97
2	2	George Lynch	VAN	26.0	203.20	103.418976	North Carolina	USA	1993	1	...	8.3	6.4	1.9	-8.2	0.106	0.185	0.175	0.512	0.125	1996-97
3	3	George McCloud	LAL	30.0	203.20	102.058200	Florida State	USA	1989	1	...	10.2	2.8	1.7	-2.7	0.027	0.111	0.206	0.527	0.125	1996-97
4	4	George Zidek	DEN	23.0	213.36	119.748288	UCLA	USA	1995	1	...	2.8	1.7	0.3	-14.1	0.102	0.169	0.195	0.500	0.064	1996-97

5 rows × 22 columns

```
In [15]: # Top Scorers
NumTopScorers = 10
top_scorers_names = df.groupby(['name'])['pts'].sum().sort_values(ascending=False).index[:NumTopScorers]
top_scorers_names
```

```
Out[15]: 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 [25]: # Display the first 10 rows of data
df.head(11)
```

	Unnamed: 0	name	team	age	height	weight	college	country	draft_year	draft_round	...	usg_pct	ts_pct	ast_pct	season	pts	reb	ast	total_pts	total_reb	total_ast
0	0	Randy Livingston	HOU	22.0	193.04	94.800728	Louisiana State	USA	1996	2	...	0.169	0.487	0.248	1996-97	249	96	153	249	96	153
1	1	Gaylon Nickerson	WAS	28.0	190.50	86.182480	Northwestern Oklahoma	USA	1994	2	...	0.174	0.497	0.043	1996-97	15	5	1	15	5	1
2	2	George Lynch	VAN	26.0	203.20	103.418976	North Carolina	USA	1993	1	...	0.175	0.512	0.125	1996-97	340	262	77	340	262	77
3	3	George McCloud	LAL	30.0	203.20	102.058200	Florida State	USA	1989	1	...	0.206	0.527	0.125	1996-97	652	179	108	652	179	108
4	4	George Zidek	DEN	23.0	213.36	119.748288	UCLA	USA	1995	1	...	0.195	0.500	0.064	1996-97	145	88	15	145	88	15
5	5	Gerald Wilkins	ORL	33.0	198.12	102.058200	Tennessee-Chattanooga	USA	1985	2	...	0.203	0.503	0.143	1996-97	848	176	176	848	176	176
6	6	Gheorghe Muresan	WAS	26.0	231.14	137.438376	None	USA	1993	2	...	0.185	0.618	0.024	1996-97	773	481	29	773	481	29
7	7	Glen Rice	CHH	30.0	203.20	99.790240	Michigan	USA	1989	1	...	0.272	0.605	0.088	1996-97	2117	316	158	2117	316	158
8	8	Glenn Robinson	MIL	24.0	200.66	106.594120	Purdue	USA	1994	1	...	0.278	0.528	0.146	1996-97	1688	504	248	1688	504	248
9	9	Grant Hill	DET	24.0	203.20	102.058200	Duke	USA	1994	1	...	0.283	0.556	0.356	1996-97	1712	720	584	1712	720	584
10	10	Gary Trent	POR	22.0	203.20	113.398000	Ohio	USA	1995	1	...	0.212	0.569	0.077	1996-97	885	426	90	885	426	90

11 rows × 28 columns

```
In [39]: # Display the last few rows of the data
df.tail()
```

	Unnamed: 0	name	team	age	height	weight	college	country	draft_year	draft_round	...	usg_pct	ts_pct	ast_pct	season	pts	reb	ast	total_pts	total_reb	total_ast
12839	12839	Joel Embiid	PHI	29.0	213.36	127.005760	Kansas	Cameroon	2014	1	...	0.370	0.655	0.233	2022-23	2184	673	277	10718	4403	1359
12840	12840	John Butler Jr.	POR	20.0	213.36	86.182480	Florida State	USA	Undrafted	Undrafted	...	0.102	0.411	0.066	2022-23	45	17	11	45	17	11
12841	12841	John Collins	ATL	25.0	205.74	102.511792	Wake Forest	USA	2017	1	...	0.168	0.593	0.052	2022-23	930	461	85	5763	2899	536
12842	12842	Jericho Sims	NYK	24.0	208.28	113.398000	Texas	USA	2021	2	...	0.074	0.780	0.044	2022-23	176	244	26	266	412	46
12843	12843	JaMychal Green	GSW	33.0	205.74	102.965384	Alabama	USA	Undrafted	Undrafted	...	0.169	0.650	0.094	2022-23	364	205	51	4190	2990	500

5 rows × 28 columns

```
In [6]: # Calculate cumulative sums for points, rebounds, and assists
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 [7]: # Display the first few rows of the DataFrame for Kyrie Irving
df[df['name']=='Kyrie Irving'].head()
```

	Unnamed: 0	name	team	age	height	weight	college	country	draft_year	draft_round	...	usg_pct	ts_pct	ast_pct	season	pts	reb	ast	total_pts	total_reb	total_ast
6734	6734	Kyrie Irving	CLE	20.0	190.5	86.636072	Duke	Australia	2011	1	...	0.281	0.566	0.324	2011-12	943	188	275	943	188	275
7594	7594	Kyrie Irving	CLE	21.0	190.5	86.636072	Duke	Australia	2011	1	...	0.298	0.553	0.315	2012-13	1327	218	348	2270	406	623
8027	8027	Kyrie Irving	CLE	22.0	190.5	87.543256	Duke	Australia	2011	1	...	0.280	0.533	0.311	2013-14	1476	255	433	3746	661	1056
8508	8508	Kyrie Irving	CLE	23.0	190.5	87.543256	Duke	Australia	2011	1	...	0.260	0.583	0.236	2014-15	1627	240	390	5373	901	1446
8984	8984	Kyrie Irving	CLE	24.0	190.5	87.543256	Duke	Australia	2011	1	...	0.293	0.540	0.253	2015-16	1038	159	249	6411	1060	1695

5 rows × 28 columns

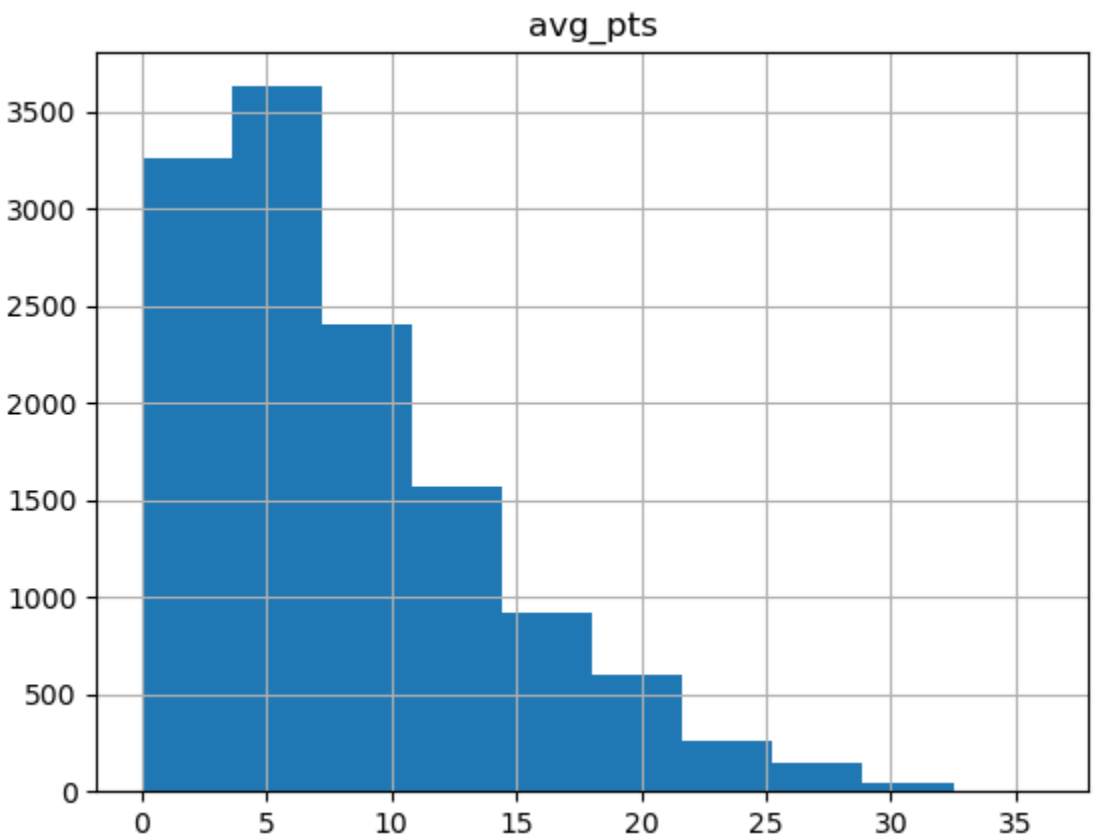
```
In [30]: df['pts'] = (df['gp']*df['avg_pts']).astype(int)
df['reb'] = (df['gp']*df['avg_reb']).astype(int)
df['ast'] = (df['gp']*df['avg_ast']).astype(int)
```

```
In [31]: # Display first few rows
df.head()
```

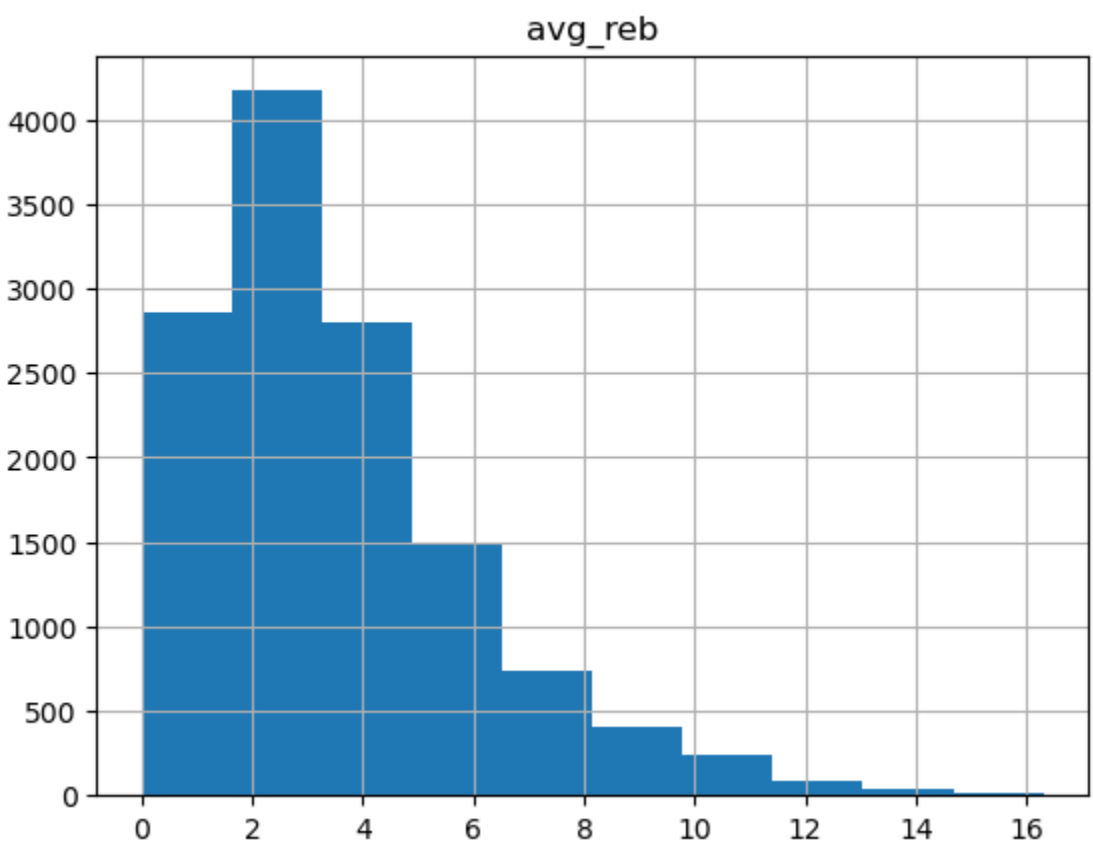
	Unnamed: 0	name	team	age	height	weight	college	country	draft_year	draft_round	...	usg_pct	ts_pct	ast_pct	season	pts	reb	ast	total_pts	total_reb	total_ast
0	0	Randy Livingston	HOU	22.0	193.04	94.800728	Louisiana State	USA	1996	2	...	0.169	0.487	0.248	1996-97	249	96	153	249	96	153
1	1	Gaylon Nickerson	WAS	28.0	190.50	86.182480	Northwestern Oklahoma	USA	1994	2	...	0.174	0.497	0.043	1996-97	15	5	1	15	5	1
2	2	George Lynch	VAN	26.0	203.20	103.418976	North Carolina	USA	1993	1	...	0.175	0.512	0.125	1996-97	340	262	77	340	262	77
3	3	George McCloud	LAL	30.0	203.20	102.058200	Florida State	USA	1989	1	...	0.206	0.527	0.125	1996-97	652	179	108	652	179	108
4	4	George Zidek	DEN	23.0	213.36	119.748288	UCLA	USA	1995	1	...	0.195	0.500	0.064	1996-97	145	88	15	145	88	15

5 rows × 28 columns

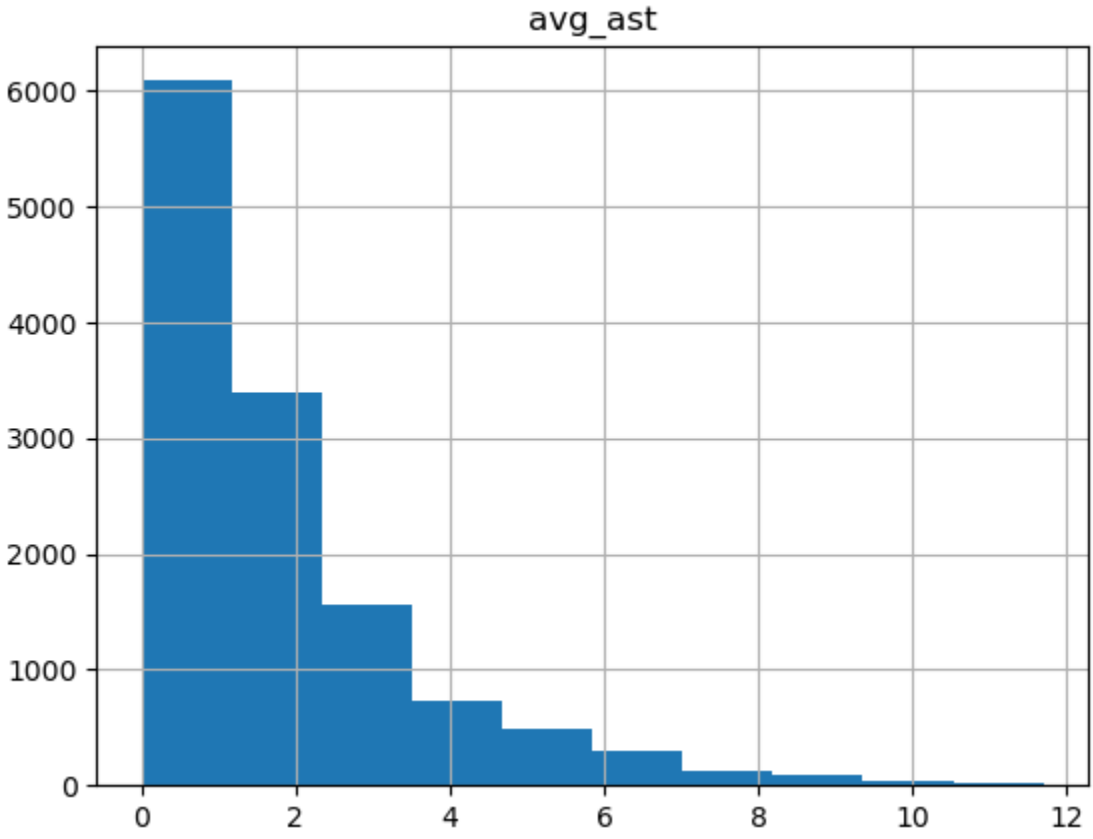
```
In [35]: df.hist(['avg_pts']); # Histograms for Average Points
```



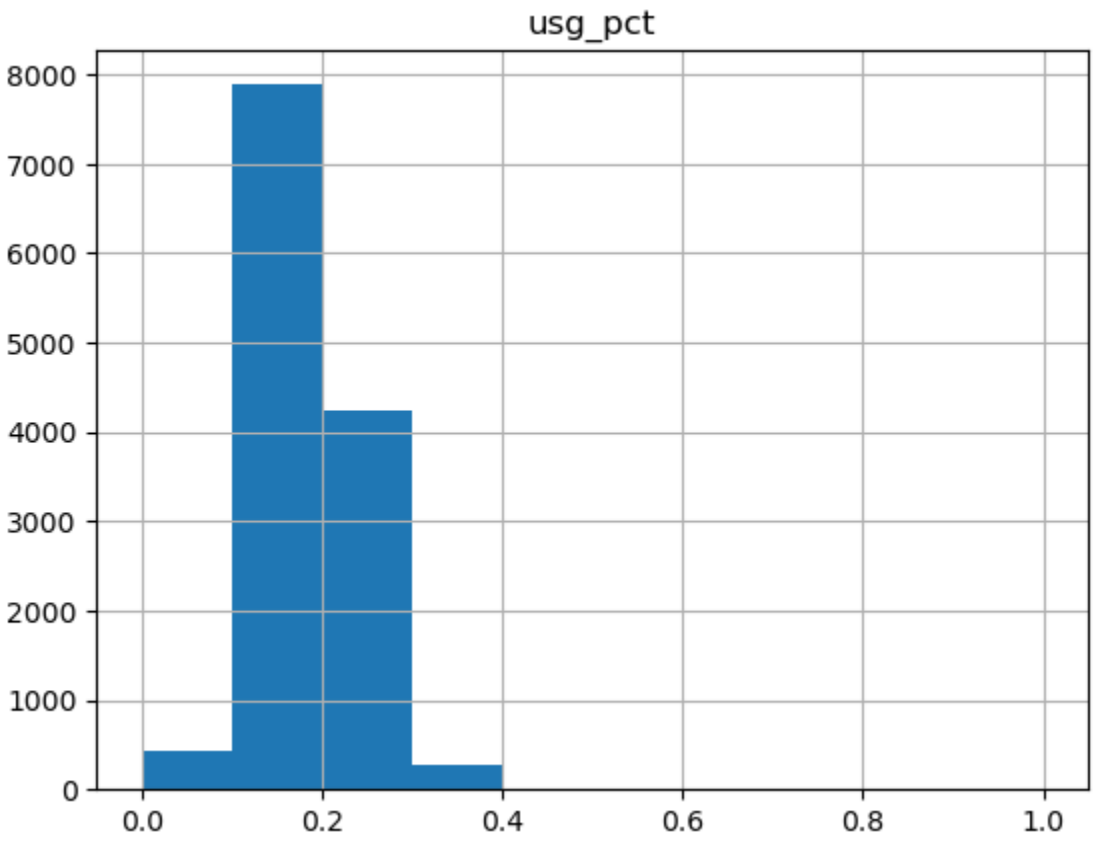
```
In [32]: # Average Rebounds
df.hist(['avg_reb']);
```



```
In [36]: # Average Assists
df.hist(['avg_ast']);
```



```
In [38]: # Usage Percentage
df.hist(['usg_pct']);
```

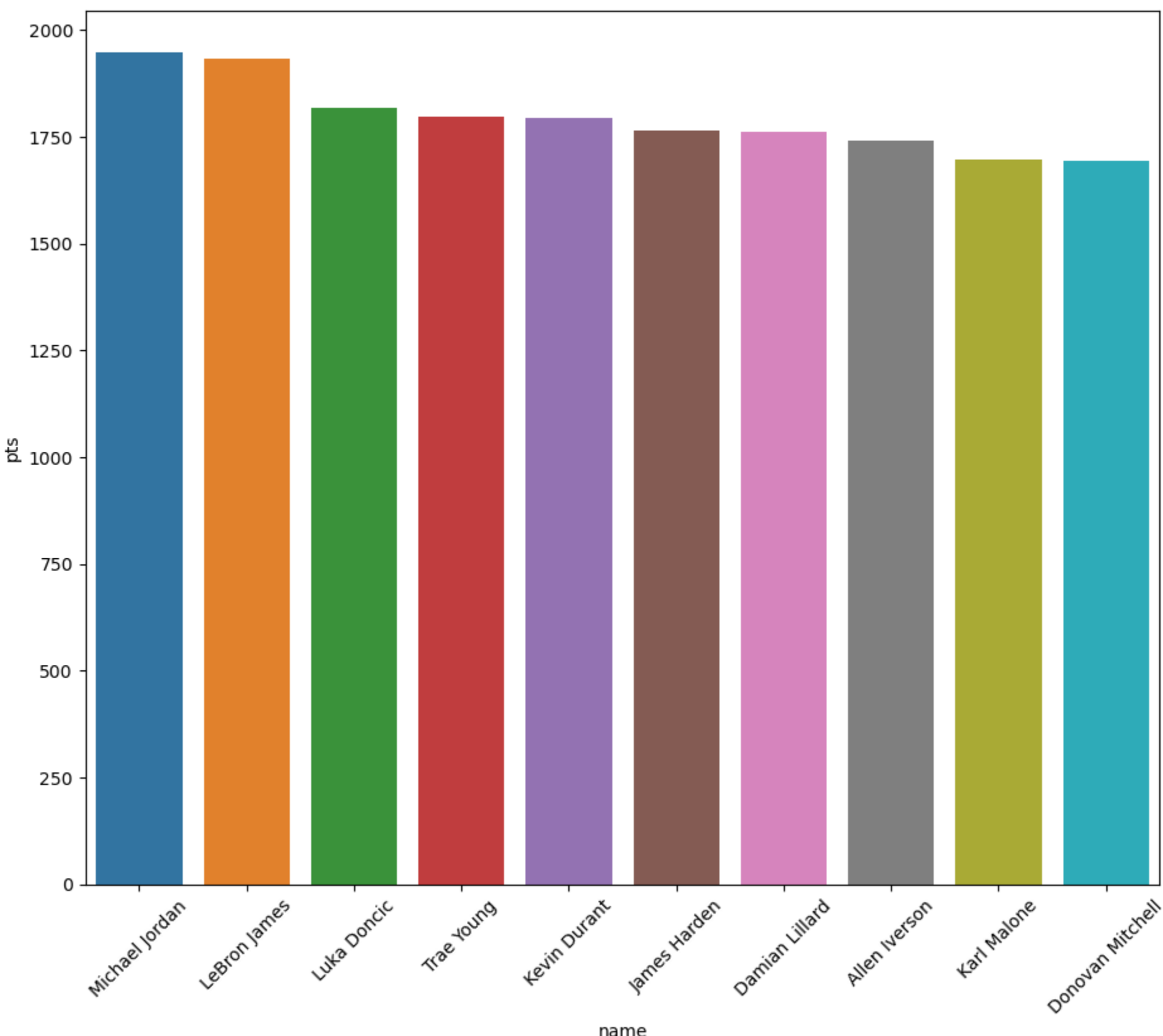


```
In [8]: # Create a DataFrame containing mean points for the top 10 players
pts_data = pd.DataFrame(df.groupby('name')['pts'].mean().sort_values(ascending = False).reset_index())[:10]
```

```
In [9]: # Display the resulting DataFrame
pts_data
```

	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 [11]: # Create a bar plot for the top 10 players with the highest mean points
plt.figure(figsize = (11,9) , dpi = 100)
sns.barplot(data = pts_data , x = 'name' , y = 'pts')
plt.xticks(rotation = 45);
```



```
In [ ]:
```