In [3]:
```python
import pandas as pd
import matplotlib.pyplot as plt
import os
```

In [4]:
```python
df = pd.read_csv('all_seasons.csv')
```

In [5]:
```python
df.head()
```

Out[5]:

	Unnamed: 0	player_name	team_abbreviation	age	player_height	player_weight	college	country	draft_year	draft_round	...
0	0	Randy Livingston	HOU	22.0	193.04	94.800728	Louisiana State	USA	1996	2	...
1	1	Gaylon Nickerson	WAS	28.0	190.50	86.182480	Northwestern Oklahoma	USA	1994	2	...
2	2	George Lynch	VAN	26.0	203.20	103.418976	North Carolina	USA	1993	1	...
3	3	George McCloud	LAL	30.0	203.20	102.058200	Florida State	USA	1989	1	...
4	4	George Zidek	DEN	23.0	213.36	119.748288	UCLA	USA	1995	1	...

5 rows × 22 columns

In [6]:
```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12844 entries, 0 to 12843
Data columns (total 22 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Unnamed: 0         12844 non-null  int64
 1   player_name        12844 non-null  object
 2   team_abbreviation  12844 non-null  object
 3   age                12844 non-null  float64
 4   player_height      12844 non-null  float64
 5   player_weight      12844 non-null  float64
 6   college            12844 non-null  object
 7   country            12844 non-null  object
 8   draft_year         12844 non-null  object
 9   draft_round        12844 non-null  object
 10  draft_number       12844 non-null  object
 11  gp                 12844 non-null  int64
 12  pts                12844 non-null  float64
 13  reb                12844 non-null  float64
 14  ast                12844 non-null  float64
 15  net_rating         12844 non-null  float64
 16  oreb_pct           12844 non-null  float64
 17  dreb_pct           12844 non-null  float64
 18  usg_pct            12844 non-null  float64
 19  ts_pct             12844 non-null  float64
 20  ast_pct            12844 non-null  float64
 21  season             12844 non-null  object
dtypes: float64(12), int64(2), object(8)
memory usage: 2.2+ MB
```

In [6]:
```python
df.shape
```

Out[6]:
```
(12844, 22)
```

In [7]:
```python
df
```

Out[7]:

	Unnamed: 0	player_name	team_abbreviation	age	player_height	player_weight	college	country	draft_year	draft_ro
0	0	Randy Livingston	HOU	22.0	193.04	94.800728	Louisiana State	USA	1996	
1	1	Gaylon Nickerson	WAS	28.0	190.50	86.182480	Northwestern Oklahoma	USA	1994	
2	2	George Lynch	VAN	26.0	203.20	103.418976	North Carolina	USA	1993	
3	3	George McCloud	LAL	30.0	203.20	102.058200	Florida State	USA	1989	
4	4	George Zidek	DEN	23.0	213.36	119.748288	UCLA	USA	1995	
...	...	...	...	...	...	...	...	...	...	
12839	12839	Joel Embiid	PHI	29.0	213.36	127.005760	Kansas	Cameroon	2014	
12840	12840	John Butler Jr.	POR	20.0	213.36	86.182480	Florida State	USA	Undrafted	Undra
12841	12841	John Collins	ATL	25.0	205.74	102.511792	Wake Forest	USA	2017	
12842	12842	Jericho Sims	NYK	24.0	208.28	113.398000	Texas	USA	2021	
12843	12843	JaMychal Green	GSW	33.0	205.74	102.965384	Alabama	USA	Undrafted	Undra

12844 rows × 22 columns

In [8]:
```python
df.pts.value_counts()
```

Out[8]:
```
2.0     174
0.0     161
3.0     149
4.0     149
5.0     139
       ...
29.8      1
29.3      1
33.0      1
35.4      1
33.1      1
Name: pts, Length: 319, dtype: int64
```

In [11]:
```python
df.team_abbreviation.value_counts()
```

Out[11]:
```
CLE    450
TOR    446
MIA    443
DAL    443
WAS    442
LAC    442
ATL    439
PHI    438
SAS    434
HOU    433
LAL    429
ORL    428
IND    428
DEN    428
POR    428
MIL    427
GSW    426
NYK    426
BOS    425
CHI    423
DET    419
SAC    418
UTA    417
MIN    417
PHX    415
MEM    370
CHA    305
NJN    257
OKC    255
BKN    200
SEA    182
NOP    175
NOH    143
CHH     89
VAN     72
NOK     32
Name: team_abbreviation, dtype: int64
```

In [13]:
```python
new_cols_dict ={'player_name':'Player'}
```

In [14]:
```python
new_cols_dict
```

Out[14]:
```
{'player_name': 'Player'}
```

In [14]:
```python
plt.figure(figsize = (10,5))
```

Out[14]:
```
<Figure size 1000x500 with 0 Axes>

<Figure size 1000x500 with 0 Axes>
```

In [26]:
```python
data = {
    'player_name': ['Randy Livingston', 'Gaylon Nickerson', 'George Lynch', 'George McCloud', 'George Z
    'college': ['Louisiana State', 'Northwestern Oklahoma', 'North Carolina', 'Florida State', 'UCLA']
}
```

In [27]:
```python
data
```

Out[27]:
```
{'player_name': ['Randy Livingston',
  'Gaylon Nickerson',
  'George Lynch',
  'George McCloud',
  'George Zidek'],
 'college': ['Louisiana State',
  'Northwestern Oklahoma',
  'North Carolina',
  'Florida State',
  'UCLA']}
```

In [8]:
```python
print(df.mean(numeric_only=True))
```

```
Unnamed: 0       6421.500000
age                27.045313
player_height     200.555097
player_weight     100.263279
gp                 51.154158
pts                 8.212582
reb                 3.558486
ast                 1.824681
net_rating         -2.226339
oreb_pct            0.054073
dreb_pct            0.140646
usg_pct             0.184641
ts_pct              0.513138
ast_pct             0.131595
dtype: float64
```

In [9]:
```python
print(df.median(numeric_only=True))
```

```
Unnamed: 0       6421.50000
age                26.00000
player_height     200.66000
player_weight      99.79024
gp                 57.00000
pts                 6.70000
reb                 3.00000
ast                 1.20000
net_rating         -1.30000
oreb_pct            0.04000
dreb_pct            0.13050
usg_pct             0.18100
ts_pct              0.52500
ast_pct             0.10300
dtype: float64
```

In [10]:
```python
print(df.mode(numeric_only=True))
```

```
       Unnamed: 0   age  player_height  player_weight    gp  pts  reb  ast  \
0               0  24.0         205.74       99.79024  82.0  2.0  2.0  0.3
1               1   NaN            NaN            NaN   NaN  NaN  NaN  NaN
2               2   NaN            NaN            NaN   NaN  NaN  NaN  NaN
3               3   NaN            NaN            NaN   NaN  NaN  NaN  NaN
4               4   NaN            NaN            NaN   NaN  NaN  NaN  NaN
...           ...   ...            ...            ...   ...  ...  ...  ...
12839       12839   NaN            NaN            NaN   NaN  NaN  NaN  NaN
12840       12840   NaN            NaN            NaN   NaN  NaN  NaN  NaN
12841       12841   NaN            NaN            NaN   NaN  NaN  NaN  NaN
12842       12842   NaN            NaN            NaN   NaN  NaN  NaN  NaN
12843       12843   NaN            NaN            NaN   NaN  NaN  NaN  NaN

       net_rating  oreb_pct  dreb_pct  usg_pct  ts_pct  ast_pct
0            -0.3       0.0       0.0    0.183     0.0      0.0
1             0.8       NaN       NaN      NaN     NaN      NaN
2             NaN       NaN       NaN      NaN     NaN      NaN
3             NaN       NaN       NaN      NaN     NaN      NaN
4             NaN       NaN       NaN      NaN     NaN      NaN
...           ...       ...       ...      ...     ...      ...
12839         NaN       NaN       NaN      NaN     NaN      NaN
12840         NaN       NaN       NaN      NaN     NaN      NaN
12841         NaN       NaN       NaN      NaN     NaN      NaN
12842         NaN       NaN       NaN      NaN     NaN      NaN
12843         NaN       NaN       NaN      NaN     NaN      NaN

[12844 rows x 14 columns]
```

In [29]:
```python
df
```

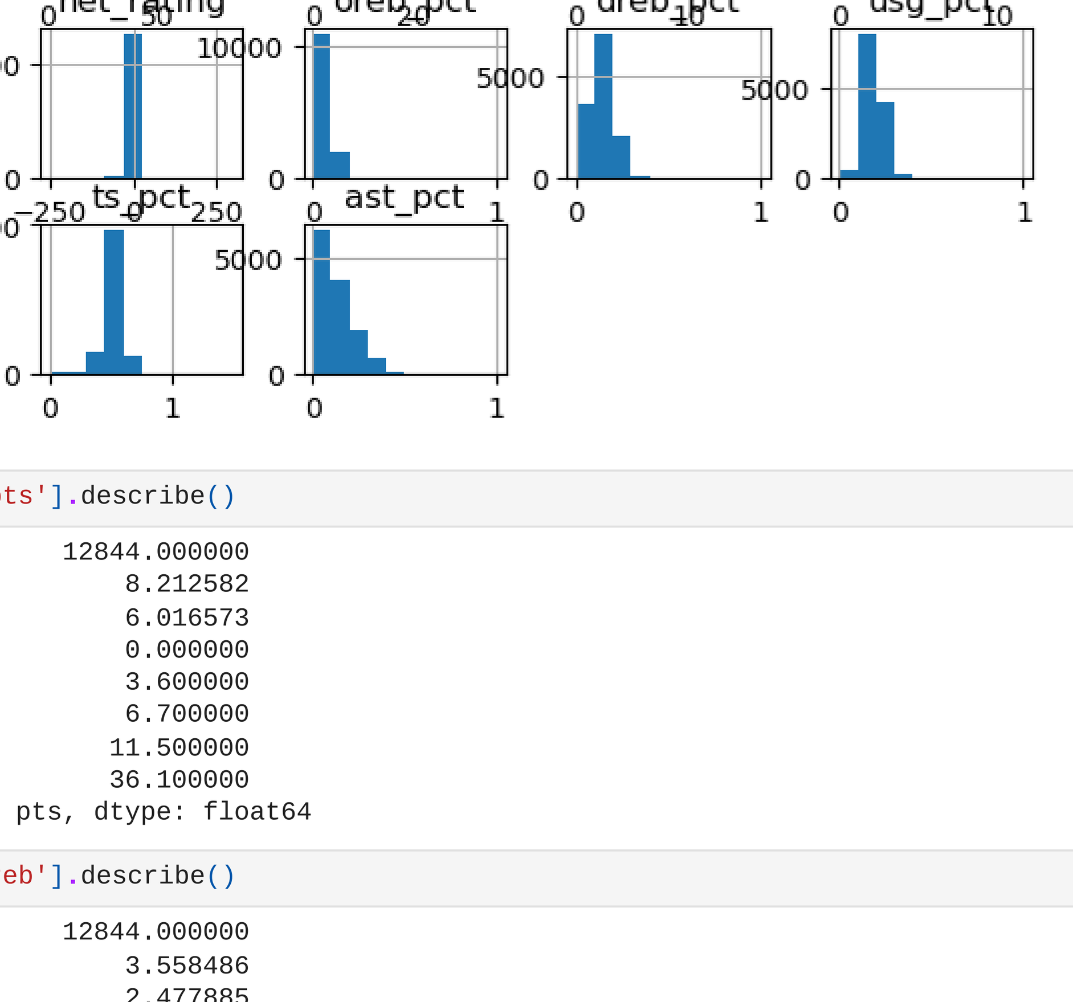
Out[29]:

	player_name	college
0	Randy Livingston	Louisiana State
1	Gaylon Nickerson	Northwestern Oklahoma
2	George Lynch	North Carolina
3	George McCloud	Florida State
4	George Zidek	UCLA

In [31]:
```python
country_counts = df['college'].value_counts()
```

In [7]:
```python
df.hist()
```

Out[7]:
```
array([[<Axes: title={'center': 'Unnamed: 0'}>,
        <Axes: title={'center': 'age'}>,
        <Axes: title={'center': 'player_height'}>,
        <Axes: title={'center': 'player_weight'}>],
       [<Axes: title={'center': 'gp'}>, <Axes: title={'center': 'pts'}>,
        <Axes: title={'center': 'reb'}>, <Axes: title={'center': '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: >, <Axes: >]],
      dtype=object)
```



In [11]:
```python
df['pts'].describe()
```

Out[11]:
```
count    12844.000000
mean         8.212582
std          6.016573
min          0.000000
25%          3.600000
50%          6.700000
75%         11.500000
max         36.100000
Name: pts, dtype: float64
```

In [12]:
```python
df['reb'].describe()
```

Out[12]:
```
count    12844.000000
mean         3.558486
std          2.477885
min          0.000000
25%          1.800000
50%          3.000000
75%          4.700000
max         16.300000
Name: reb, dtype: float64
```

In [14]:
```python
df['ast'].describe()
```

Out[14]:
```
count    12844.000000
mean         1.824681
std          1.800840
min          0.000000
25%          0.600000
50%          1.200000
75%          2.400000
max         11.700000
Name: ast, dtype: float64
```

In [15]:
```python
df.describe()
```

Out[15]:

	Unnamed: 0	age	player_height	player_weight	gp	pts	reb	ast	net_
count	12844.000000	12844.000000	12844.000000	12844.000000	12844.000000	12844.000000	12844.000000	12844.000000	12844.0
mean	6421.500000	27.045313	200.555097	100.263279	51.154158	8.212582	3.558486	1.824681	-2.2
std	3707.887763	4.339211	9.111090	12.426628	25.084904	6.016573	2.477885	1.800840	12.6
min	0.000000	18.000000	160.020000	60.327736	1.000000	0.000000	0.000000	0.000000	-250.0
25%	3210.750000	24.000000	193.040000	90.718400	31.000000	3.600000	1.800000	0.600000	-6.4
50%	6421.500000	26.000000	200.660000	99.790240	57.000000	6.700000	3.000000	1.200000	-1.3
75%	9632.250000	30.000000	208.280000	108.862080	73.000000	11.500000	4.700000	2.400000	3.2
max	12843.000000	44.000000	231.140000	163.293120	85.000000	36.100000	16.300000	11.700000	300.0

In [7]:
```python
df.head(11)
```

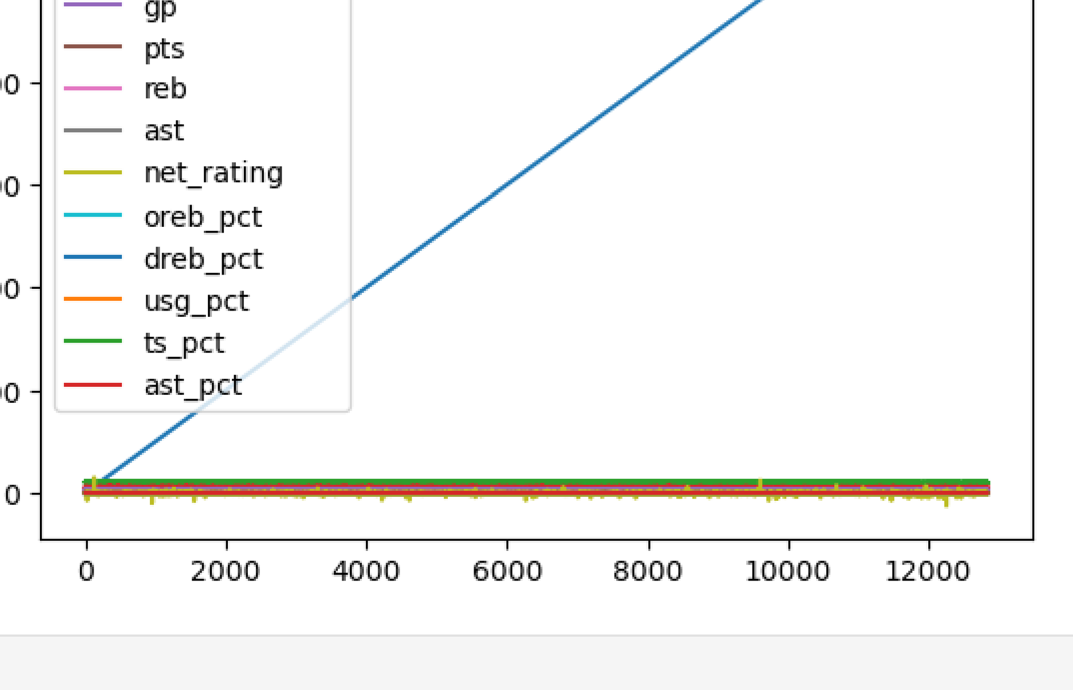
Out[7]:

	Unnamed: 0	player_name	team_abbreviation	age	player_height	player_weight	college	country	draft_year	draft_round
0	0	Randy Livingston	HOU	22.0	193.04	94.800728	Louisiana State	USA	1996	2
1	1	Gaylon Nickerson	WAS	28.0	190.50	86.182480	Northwestern Oklahoma	USA	1994	2
2	2	George Lynch	VAN	26.0	203.20	103.418976	North Carolina	USA	1993	1
3	3	George McCloud	LAL	30.0	203.20	102.058200	Florida State	USA	1989	1
4	4	George Zidek	DEN	23.0	213.36	119.748288	UCLA	USA	1995	1
5	5	Gerald Wilkins	ORL	33.0	198.12	102.058200	Tennessee-Chattanooga	USA	1985	2
6	6	Gheorghe Muresan	WAS	26.0	231.14	137.438376	None	USA	1993	2
7	7	Glen Rice	CHH	30.0	203.20	99.790240	Michigan	USA	1989	1
8	8	Glenn Robinson	MIL	24.0	200.66	106.594120	Purdue	USA	1994	1
9	9	Grant Hill	DET	24.0	203.20	102.058200	Duke	USA	1994	1
10	10	Gary Trent	POR	22.0	203.20	113.398000	Ohio	USA	1995	1

11 rows × 22 columns

In [8]:
```python
df.plot()
```

Out[8]:
```
<Axes: >
```



In [ ]: