

```
In [6]: import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
import seaborn as sns
data=pd.read_excel("C:/Users/user/Downloads/myexcel.xlsx")
data
```

```
Out[6]:
```

	Name	Team	Number	Position	Age	Height	Weight	College	Salary
0	Avery Bradley	Boston Celtics	0	PG	25	2023-02-06 00:00:00	180	Texas	7730337.0
1	Jae Crowder	Boston Celtics	99	SF	25	2023-06-06 00:00:00	235	Marquette	6796117.0
2	John Holland	Boston Celtics	30	SG	27	2023-05-06 00:00:00	205	Boston University	NaN
3	R.J. Hunter	Boston Celtics	28	SG	22	2023-05-06 00:00:00	185	Georgia State	1148640.0
4	Jonas Jerebko	Boston Celtics	8	PF	29	2023-10-06 00:00:00	231	NaN	5000000.0
...
453	Shelvin Mack	Utah Jazz	8	PG	26	2023-03-06 00:00:00	203	Butler	2433333.0
454	Raul Neto	Utah Jazz	25	PG	24	2023-01-06 00:00:00	179	NaN	900000.0
455	Tibor Pleiss	Utah Jazz	21	C	26	2023-03-07 00:00:00	256	NaN	2900000.0
456	Jeff Withey	Utah Jazz	24	C	26	7-0	231	Kansas	947276.0
457	Priyanka	Utah Jazz	34	C	25	2023-03-07 00:00:00	231	Kansas	947276.0

458 rows × 9 columns

```
In [7]: data.head(5)
```

```
Out[7]:
```

	Name	Team	Number	Position	Age	Height	Weight	College	Salary
0	Avery Bradley	Boston Celtics	0	PG	25	2023-02-06 00:00:00	180	Texas	7730337.0
1	Jae Crowder	Boston Celtics	99	SF	25	2023-06-06 00:00:00	235	Marquette	6796117.0
2	John Holland	Boston Celtics	30	SG	27	2023-05-06 00:00:00	205	Boston University	NaN
3	R.J. Hunter	Boston Celtics	28	SG	22	2023-05-06 00:00:00	185	Georgia State	1148640.0
4	Jonas Jerebko	Boston Celtics	8	PF	29	2023-10-06 00:00:00	231	NaN	5000000.0

```
In [11]: data.columns
```

```
Out[11]: Index(['Name', 'Team', 'Number', 'Position', 'Age', 'Height', 'Weight',  
              'College', 'Salary'],  
              dtype='object')
```

```
In [15]: data["Height"] = np.random.randint(150, 180, size=len(data))  
data
```

```
Out[15]:
```

	Name	Team	Number	Position	Age	Height	Weight	College	Salary	height
0	Avery Bradley	Boston Celtics	0	PG	25	174	180	Texas	7730337.0	177
1	Jae Crowder	Boston Celtics	99	SF	25	158	235	Marquette	6796117.0	165
2	John Holland	Boston Celtics	30	SG	27	177	205	Boston University	NaN	152
3	R.J. Hunter	Boston Celtics	28	SG	22	178	185	Georgia State	1148640.0	173
4	Jonas Jerebko	Boston Celtics	8	PF	29	170	231	NaN	5000000.0	164
...
453	Shelvin Mack	Utah Jazz	8	PG	26	158	203	Butler	2433333.0	167
454	Raul Neto	Utah Jazz	25	PG	24	160	179	NaN	900000.0	170
455	Tibor Pleiss	Utah Jazz	21	C	26	164	256	NaN	2900000.0	178
456	Jeff Withey	Utah Jazz	24	C	26	174	231	Kansas	947276.0	153
457	Priyanka	Utah Jazz	34	C	25	178	231	Kansas	947276.0	154

458 rows × 10 columns

```
In [16]: data.head(3)
```

```
Out[16]:
```

	Name	Team	Number	Position	Age	Height	Weight	College	Salary	height
0	Avery Bradley	Boston Celtics	0	PG	25	174	180	Texas	7730337.0	177
1	Jae Crowder	Boston Celtics	99	SF	25	158	235	Marquette	6796117.0	165
2	John Holland	Boston Celtics	30	SG	27	177	205	Boston University	NaN	152

```
In [17]: data.describe()
```

```
Out[17]:
```

	Number	Age	Height	Weight	Salary	height
count	458.000000	458.000000	458.000000	458.000000	4.470000e+02	458.000000
mean	17.713974	26.934498	164.310044	221.543668	4.833970e+06	164.790393
std	15.966837	4.400128	8.840920	26.343200	5.226620e+06	8.480106
min	0.000000	19.000000	150.000000	161.000000	3.088800e+04	150.000000
25%	5.000000	24.000000	156.000000	200.000000	1.025210e+06	158.000000
50%	13.000000	26.000000	164.000000	220.000000	2.836186e+06	166.000000
75%	25.000000	30.000000	172.000000	240.000000	6.500000e+06	171.750000
max	99.000000	40.000000	179.000000	307.000000	2.500000e+07	179.000000

```
In [18]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 458 entries, 0 to 457
Data columns (total 10 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Name         458 non-null    object
1   Team         458 non-null    object
2   Number       458 non-null    int64
3   Position     458 non-null    object
4   Age          458 non-null    int64
5   Height       458 non-null    int32
6   Weight       458 non-null    int64
7   College      374 non-null    object
8   Salary       447 non-null    float64
9   height       458 non-null    int32
dtypes: float64(1), int32(2), int64(3), object(4)
memory usage: 32.3+ KB
```

```
In [21]: data.isnull().sum()
```

```
Out[21]: Name         0
Team         0
Number       0
Position     0
Age          0
Height       0
Weight       0
College      84
Salary       11
height       0
dtype: int64
```

```
In [23]: data.shape
```

```
Out[23]: (458, 10)
```

```
In [26]: data["College"]
```

```
Out[26]: 0          Texas
1      Marquette
2  Boston University
3      Georgia State
4          NaN
...
453      Butler
454          NaN
455          NaN
456      Kansas
457      Kansas
Name: College, Length: 458, dtype: object
```

```
In [31]: data["College"].fillna("Not Specified",inplace=True)
data
```

```
Out[31]:
```

	Name	Team	Number	Position	Age	Height	Weight	College	Salary	height
0	Avery Bradley	Boston Celtics	0	PG	25	174	180	Texas	7730337.0	177
1	Jae Crowder	Boston Celtics	99	SF	25	158	235	Marquette	6796117.0	165
2	John Holland	Boston Celtics	30	SG	27	177	205	Boston University	NaN	152
3	R.J. Hunter	Boston Celtics	28	SG	22	178	185	Georgia State	1148640.0	173
4	Jonas Jerebko	Boston Celtics	8	PF	29	170	231	Not Specified	5000000.0	164
...
453	Shelvin Mack	Utah Jazz	8	PG	26	158	203	Butler	2433333.0	167
454	Raul Neto	Utah Jazz	25	PG	24	160	179	Not Specified	900000.0	170
455	Tibor Pleiss	Utah Jazz	21	C	26	164	256	Not Specified	2900000.0	178
456	Jeff Withey	Utah Jazz	24	C	26	174	231	Kansas	947276.0	153
457	Priyanka	Utah Jazz	34	C	25	178	231	Kansas	947276.0	154

458 rows × 10 columns

```
In [34]: data["Salary"].fillna(method="pad",inplace=True)
data
```

```
Out[34]:
```

	Name	Team	Number	Position	Age	Height	Weight	College	Salary	height
0	Avery Bradley	Boston Celtics	0	PG	25	174	180	Texas	7730337.0	177
1	Jae Crowder	Boston Celtics	99	SF	25	158	235	Marquette	6796117.0	165
2	John Holland	Boston Celtics	30	SG	27	177	205	Boston University	6796117.0	152
3	R.J. Hunter	Boston Celtics	28	SG	22	178	185	Georgia State	1148640.0	173
4	Jonas Jerebko	Boston Celtics	8	PF	29	170	231	Not Specified	5000000.0	164
...
453	Shelvin Mack	Utah Jazz	8	PG	26	158	203	Butler	2433333.0	167
454	Raul Neto	Utah Jazz	25	PG	24	160	179	Not Specified	900000.0	170
455	Tibor Pleiss	Utah Jazz	21	C	26	164	256	Not Specified	2900000.0	178
456	Jeff Withey	Utah Jazz	24	C	26	174	231	Kansas	947276.0	153
457	Priyanka	Utah Jazz	34	C	25	178	231	Kansas	947276.0	154

458 rows × 10 columns

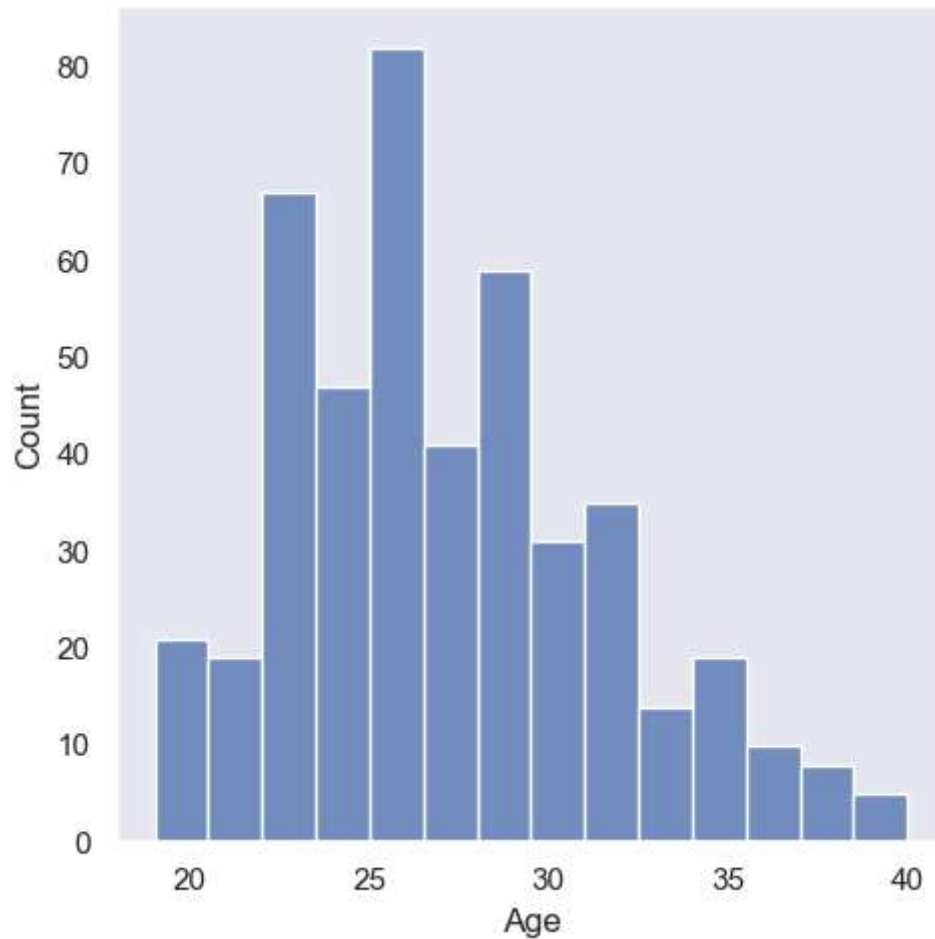
```
In [107]: data["Age"].value_counts().idxmax()
```

```
Out[107]: 24
```

```
In [59]: sns.set(style="dark")
sns.displot(data['Age'])
```

C:\Users\user\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserWarning: The figure layout has changed to tight
self._figure.tight_layout(*args, **kwargs)

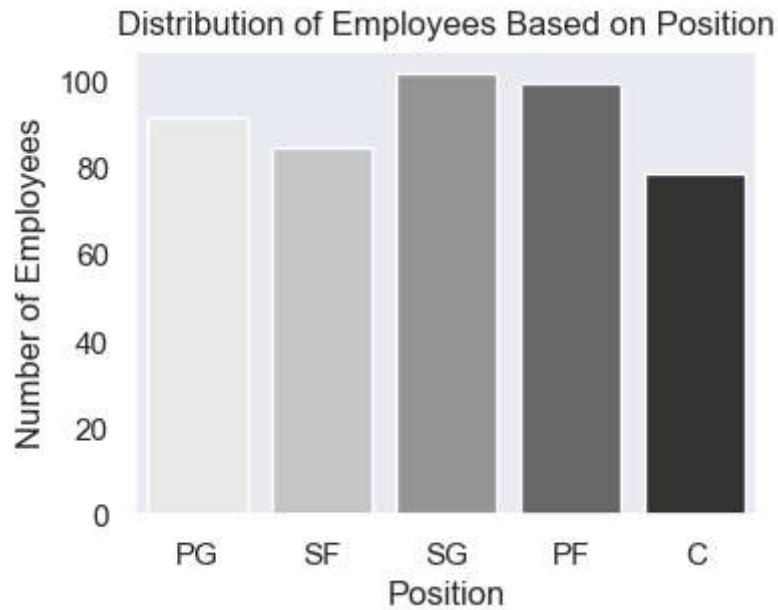
```
Out[59]: <seaborn.axisgrid.FacetGrid at 0x1f708563ad0>
```



```
In [61]: data['Position'].value_counts()
```

```
Out[61]: Position
SG      102
PF      100
PG       92
SF       85
C        79
Name: count, dtype: int64
```

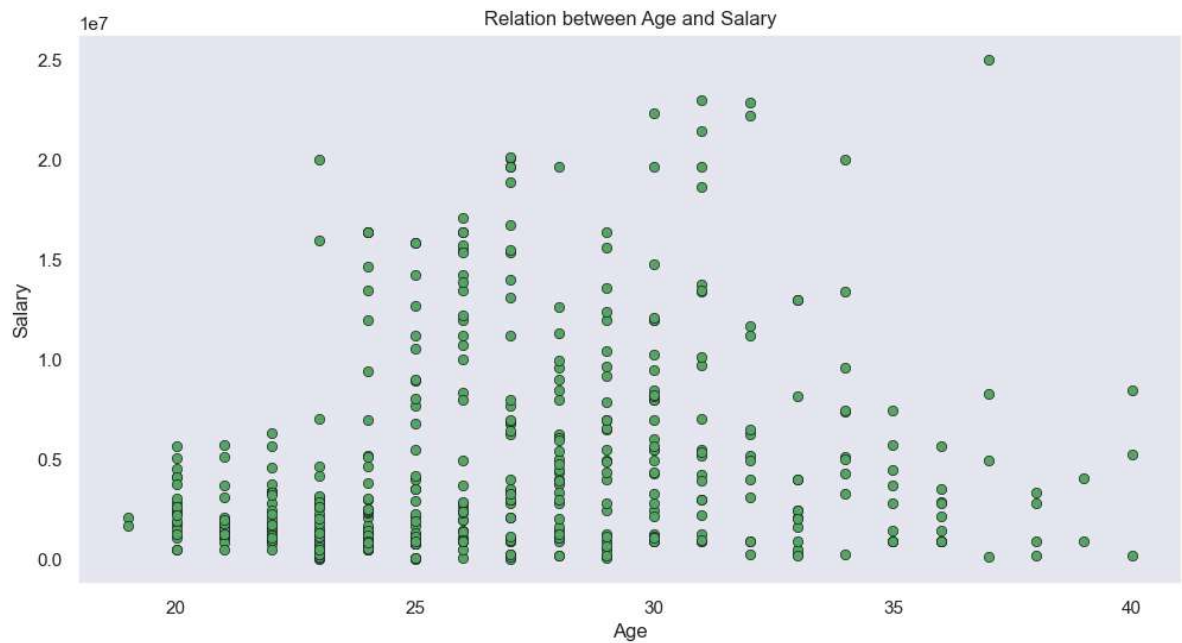
```
In [81]: plt.figure(figsize=(4, 3))
sns.countplot(x='Position',data=data ,palette='Greys')
plt.title('Distribution of Employees Based on Position')
plt.xlabel('Position')
plt.ylabel('Number of Employees')
plt.show()
```



```
In [87]: correl = data['Age'].corr(data['Salary'])
print("Extend of corelation between age and salary is",correl)
```

Extend of corelation between age and salary is 0.19826344340402224

```
In [78]: plt.figure(figsize=(12, 6))
sns.scatterplot(x='Age', y='Salary', data=data, edgecolor="k", marker="o", c="g")
plt.title('Relation between Age and Salary')
plt.xlabel('Age')
plt.ylabel('Salary')
plt.show()
```



Type *Markdown* and LaTeX: α^2


```
In [110]: team_expense=data.groupby("Team")["Salary"].sum()  
team_expense
```

```
Out[110]: Team  
Atlanta Hawks          72902950.0  
Boston Celtics         65337185.0  
Brooklyn Nets          52528475.0  
Charlotte Hornets      78340920.0  
Chicago Bulls          86783378.0  
Cleveland Cavaliers    107935965.0  
Dallas Mavericks       71198732.0  
Denver Nuggets         60380419.0  
Detroit Pistons        67168263.0  
Golden State Warriors  88868997.0  
Houston Rockets        75283021.0  
Indiana Pacers         66751826.0  
Los Angeles Clippers   94854640.0  
Los Angeles Lakers     71770431.0  
Memphis Grizzlies      88713462.0  
Miami Heat             104997393.0  
Milwaukee Bucks        69603517.0  
Minnesota Timberwolves 60859197.0  
New Orleans Pelicans   82750774.0  
New York Knicks        73303898.0  
Oklahoma City Thunder  93765298.0  
Orlando Magic          60161470.0  
Philadelphia 76ers     31160300.0  
Phoenix Suns           63445135.0  
Portland Trail Blazers 48301818.0  
Sacramento Kings       71683666.0  
San Antonio Spurs      84442733.0  
Toronto Raptors        71117611.0  
Utah Jazz              64007367.0  
Washington Wizards     76328636.0  
Name: Salary, dtype: float64
```

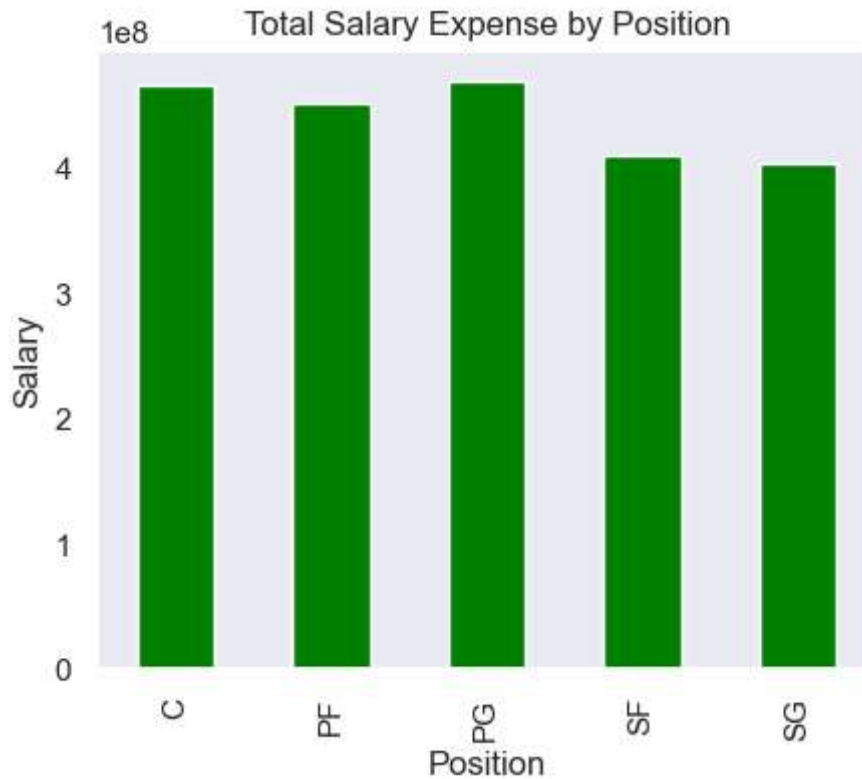
```
In [111]: max_expense=team_expense.idxmax()  
max_expense
```

```
Out[111]: 'Cleveland Cavaliers'
```

```
In [120]: position_expense=data.groupby("Position")["Salary"].sum()  
position_expense
```

```
Out[120]: Position  
C      466377332.0  
PF     452877756.0  
PG     470011553.0  
SF     410502696.0  
SG     404978140.0  
Name: Salary, dtype: float64
```

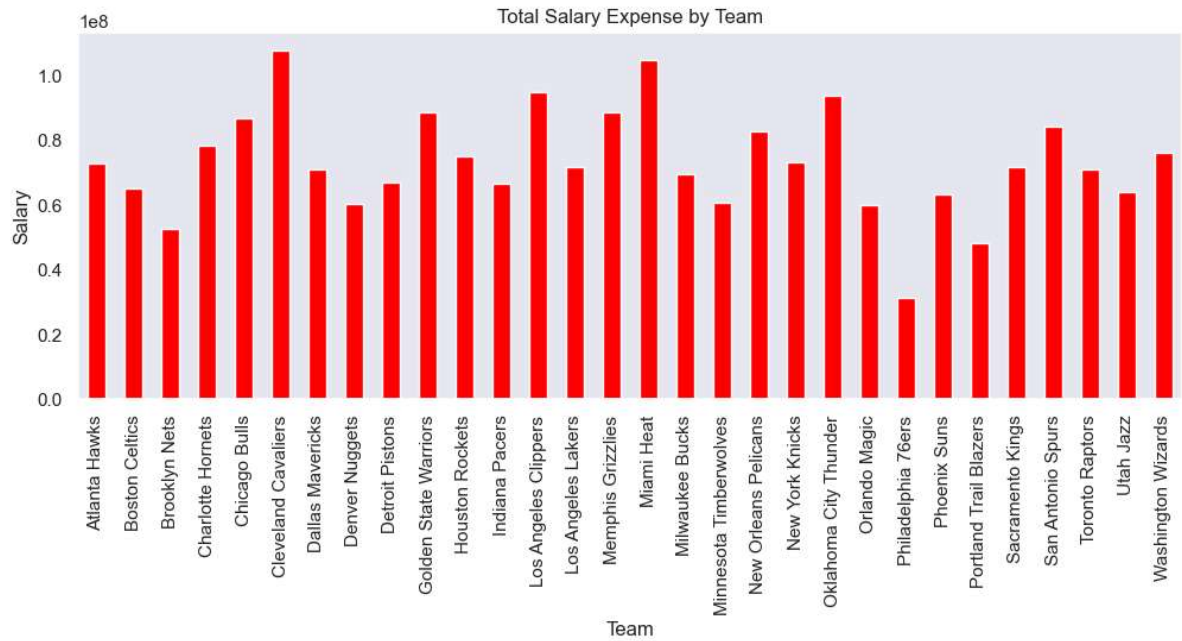
```
In [125]: plt.figure(figsize=(5, 4))
position_expense.plot(kind='bar', color='Green')
plt.title('Total Salary Expense by Position')
plt.xlabel('Position')
plt.ylabel('Salary')
plt.show()
```



```
In [121]: max_expense_position=position_expense.idxmax()
max_expense_position
```

```
Out[121]: 'PG'
```

```
In [145]: plt.figure(figsize=(12, 4))
team_expense.plot(kind='bar', color='red')
plt.title('Total Salary Expense by Team')
plt.xlabel('Team')
plt.ylabel('Salary')
plt.show()
```



```
In [131]: total_players=len(data)
total_players
```

Out[131]: 458

```
In [132]: team_distribution = data['Team'].value_counts()  
team_distribution
```

```
Out[132]: Team  
New Orleans Pelicans    19  
Memphis Grizzlies      18  
Utah Jazz               16  
New York Knicks        16  
Milwaukee Bucks        16  
Brooklyn Nets          15  
Portland Trail Blazers  15  
Oklahoma City Thunder  15  
Denver Nuggets         15  
Washington Wizards     15  
Miami Heat             15  
Charlotte Hornets      15  
Atlanta Hawks          15  
San Antonio Spurs      15  
Houston Rockets        15  
Boston Celtics         15  
Indiana Pacers         15  
Detroit Pistons        15  
Cleveland Cavaliers    15  
Chicago Bulls          15  
Sacramento Kings       15  
Phoenix Suns           15  
Los Angeles Lakers     15  
Los Angeles Clippers   15  
Golden State Warriors  15  
Toronto Raptors        15  
Philadelphia 76ers      15  
Dallas Mavericks       15  
Orlando Magic          14  
Minnesota Timberwolves 14  
Name: count, dtype: int64
```

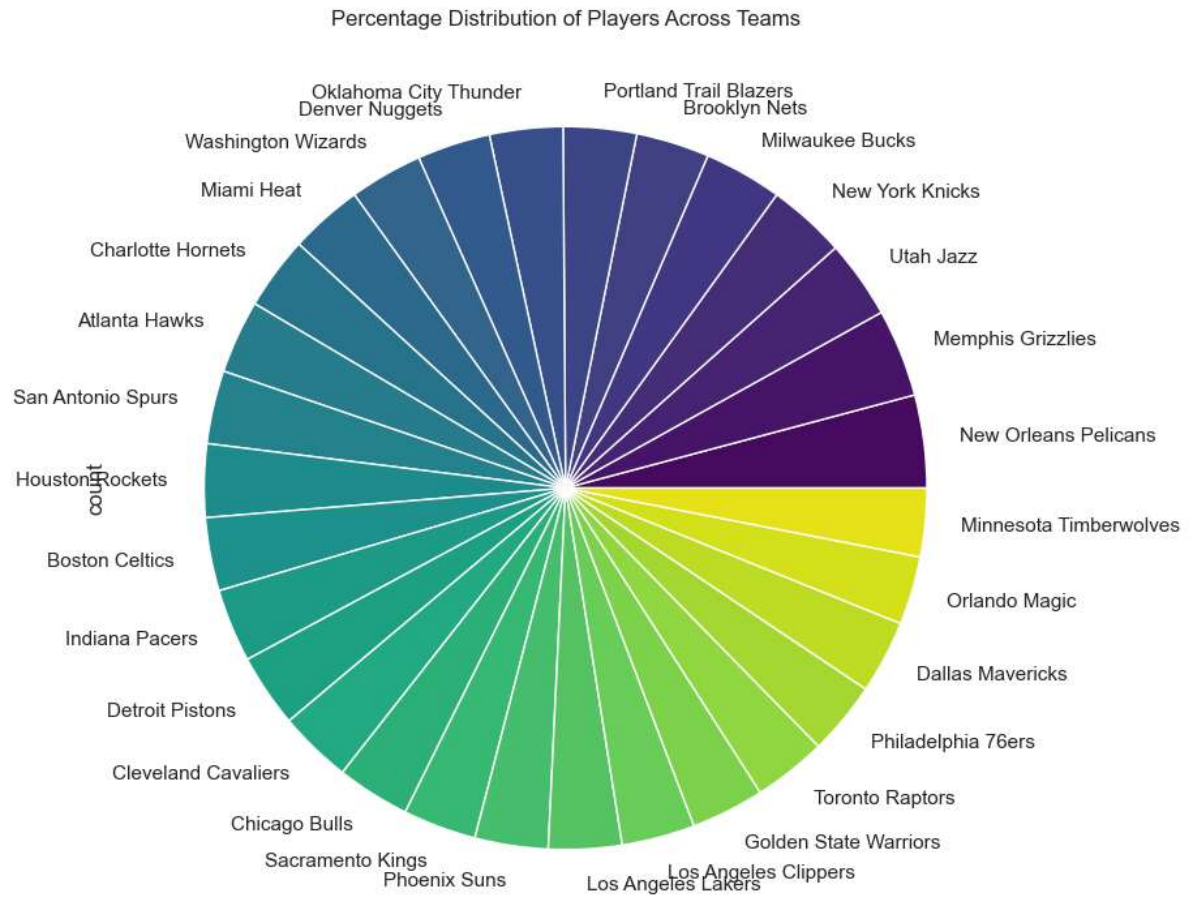
```
In [134]: team_percentage=(team_distribution/total_players)*100  
print("The percentage of the given exploration is",team_percentage)
```

The percentage of the given exploration is Team

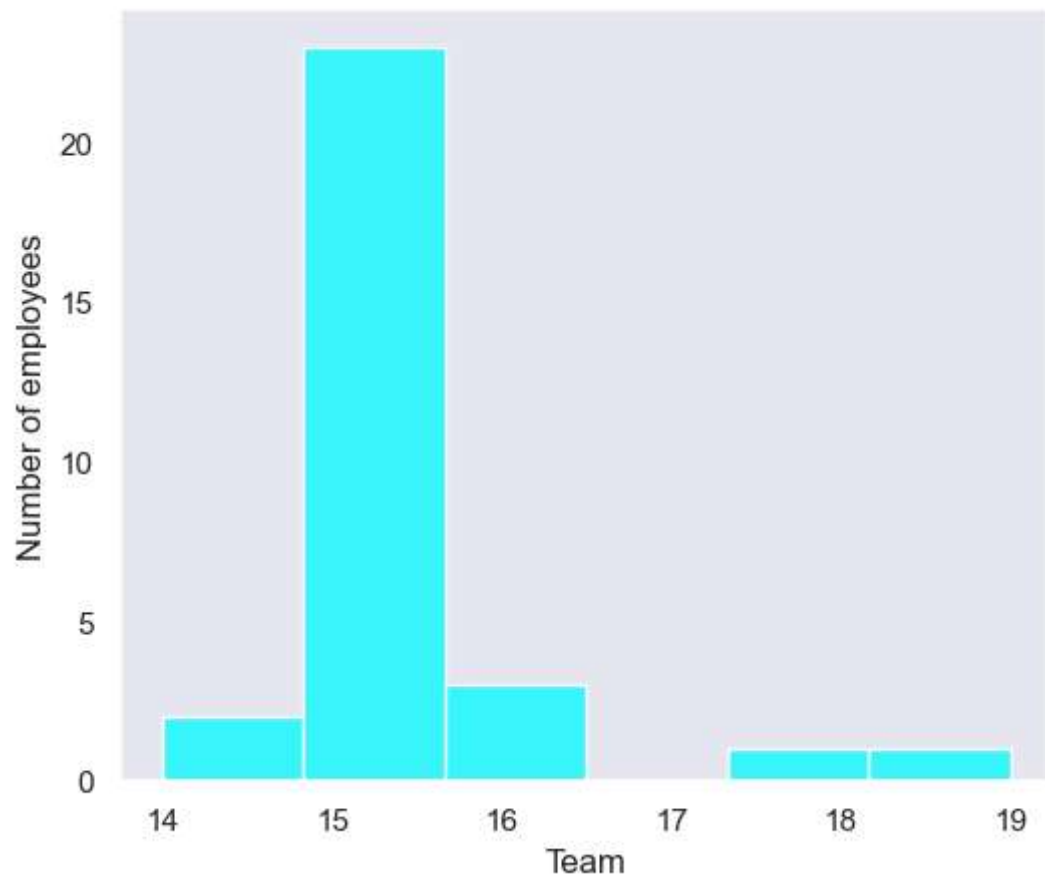
New Orleans Pelicans	4.148472
Memphis Grizzlies	3.930131
Utah Jazz	3.493450
New York Knicks	3.493450
Milwaukee Bucks	3.493450
Brooklyn Nets	3.275109
Portland Trail Blazers	3.275109
Oklahoma City Thunder	3.275109
Denver Nuggets	3.275109
Washington Wizards	3.275109
Miami Heat	3.275109
Charlotte Hornets	3.275109
Atlanta Hawks	3.275109
San Antonio Spurs	3.275109
Houston Rockets	3.275109
Boston Celtics	3.275109
Indiana Pacers	3.275109
Detroit Pistons	3.275109
Cleveland Cavaliers	3.275109
Chicago Bulls	3.275109
Sacramento Kings	3.275109
Phoenix Suns	3.275109
Los Angeles Lakers	3.275109
Los Angeles Clippers	3.275109
Golden State Warriors	3.275109
Toronto Raptors	3.275109
Philadelphia 76ers	3.275109
Dallas Mavericks	3.275109
Orlando Magic	3.056769
Minnesota Timberwolves	3.056769

Name: count, dtype: float64

```
In [140]: plt.figure(figsize=(16, 9))  
team_percentage.plot(kind='pie', colors=sns.color_palette('viridis', len(team_)),  
plt.title('Percentage Distribution of Players Across Teams')  
plt.show()
```



```
In [158]: plt.figure(figsize=(6, 5))  
sns.histplot(x=team_distribution,data=data,color="cyan")  
plt.xlabel("Team")  
plt.ylabel("Number of employees")  
plt.show()
```



```
In [ ]:
```