

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

```
In [3]: import numpy as np
```

```
In [5]: import matplotlib.pyplot as plt
```

```
In [13]: data = pd.read_csv('myexcel.csv')
```

```
In [15]: df = pd.read_csv('myexcel.csv')
df.head(5)
```

```
Out[15]:
```

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

```
In [21]: data['Height'] = np.random.randint(150, 181, size=data.shape[0])
```

```
In [23]: data.head(5)
```

```
Out[23]:
```

	Name	Team	Number	Position	Age	Height	Weight	College	Salary
0	Avery Bradley	Boston Celtics	0	PG	25	159	180	Texas	7730337.0
1	Jae Crowder	Boston Celtics	99	SF	25	162	235	Marquette	6796117.0
2	John Holland	Boston Celtics	30	SG	27	175	205	Boston University	NaN
3	R.J. Hunter	Boston Celtics	28	SG	22	155	185	Georgia State	1148640.0
4	Jonas Jerebko	Boston Celtics	8	PF	29	177	231	NaN	5000000.0

```
In [25]: # Q1
```

```
In [27]: team_counts = data['Team'].value_counts()
team_percentage = (team_counts / len(data)) * 100
print("Team distribution:\n", team_counts)
print("Team percentage distribution:\n", team_percentage)
```

## Team distribution:

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

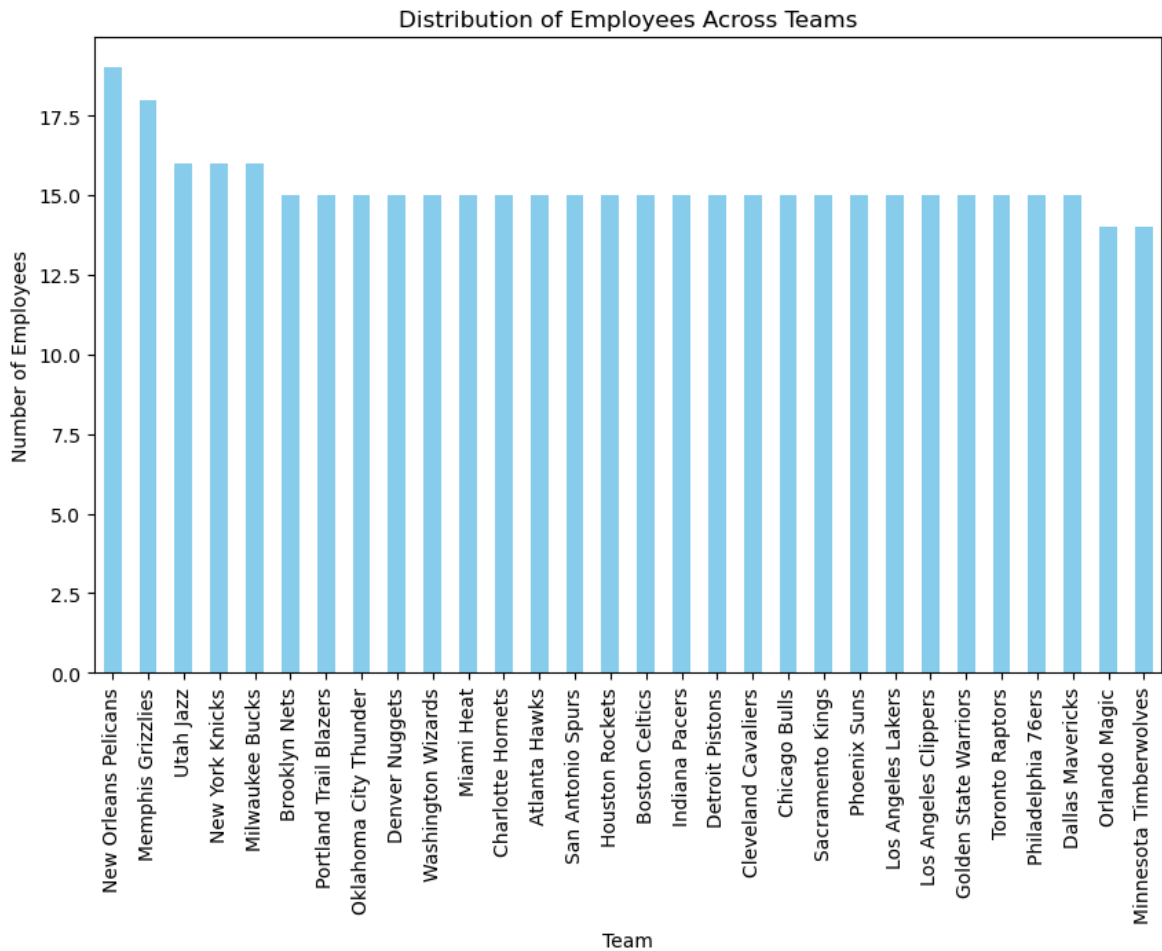
Name: count, dtype: int64

## Team percentage distribution:

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 [31]: plt.figure(figsize=(10, 6))
team_counts.plot(kind='bar', color='skyblue')
plt.title('Distribution of Employees Across Teams')
plt.xlabel('Team')
plt.ylabel('Number of Employees')
plt.show()
```



```
In [35]: # Q2
```

```
In [39]: position_counts = data['Position'].value_counts()
print("Position counts:\n", position_counts)
```

```
Position counts:
Position
SG      102
PF      100
PG       92
SF       85
C        79
Name: count, dtype: int64
```

```
In [41]: plt.figure(figsize=(10, 6))
position_counts.plot(kind='bar', color='coral')
plt.title('Distribution of Employees by Position')
```

```
plt.xlabel('Position')
plt.ylabel('Number of Employees')
plt.show()
```



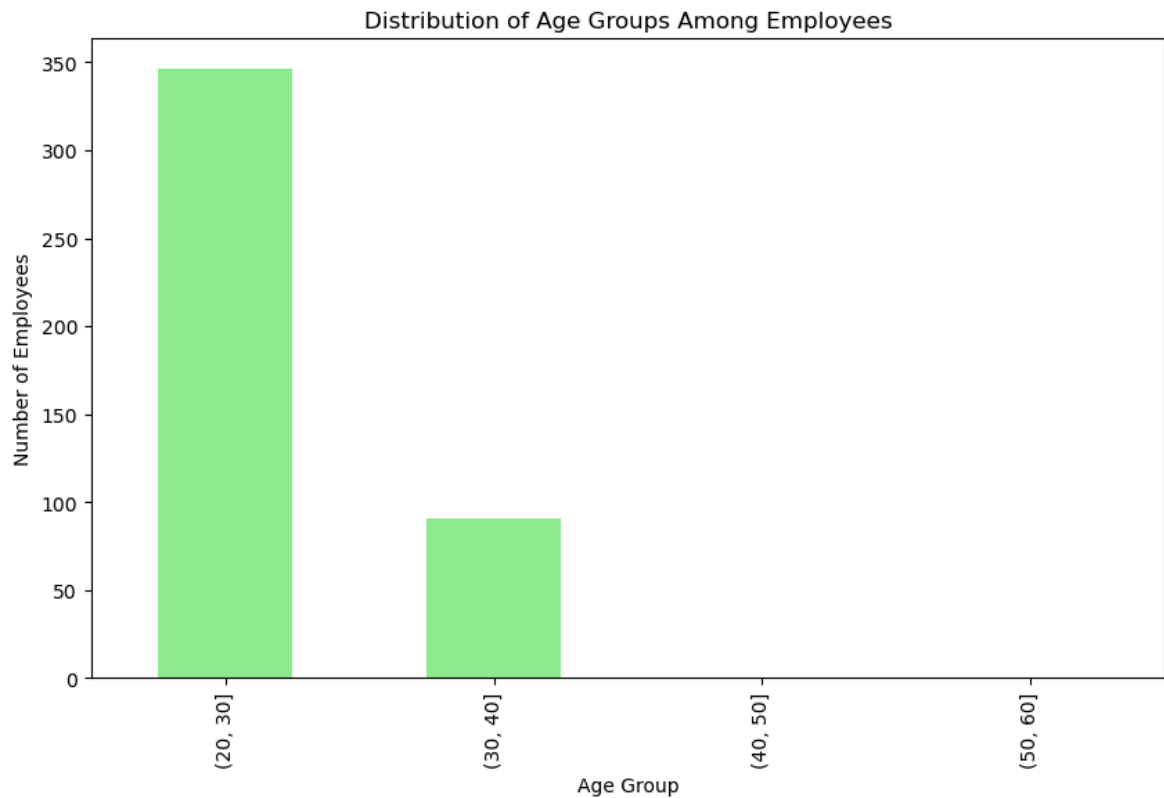
In [43]: # Q3

```
In [45]: age_bins = [20, 30, 40, 50, 60]
age_groups = pd.cut(data['Age'], bins=age_bins).value_counts()
print("Predominant age group:\n", age_groups)
```

Predominant age group:

```
Age
(20, 30]    346
(30, 40]     91
(40, 50]      0
(50, 60]      0
Name: count, dtype: int64
```

```
In [47]: plt.figure(figsize=(10, 6))
age_groups.plot(kind='bar', color='lightgreen')
plt.title('Distribution of Age Groups Among Employees')
plt.xlabel('Age Group')
plt.ylabel('Number of Employees')
plt.show()
```



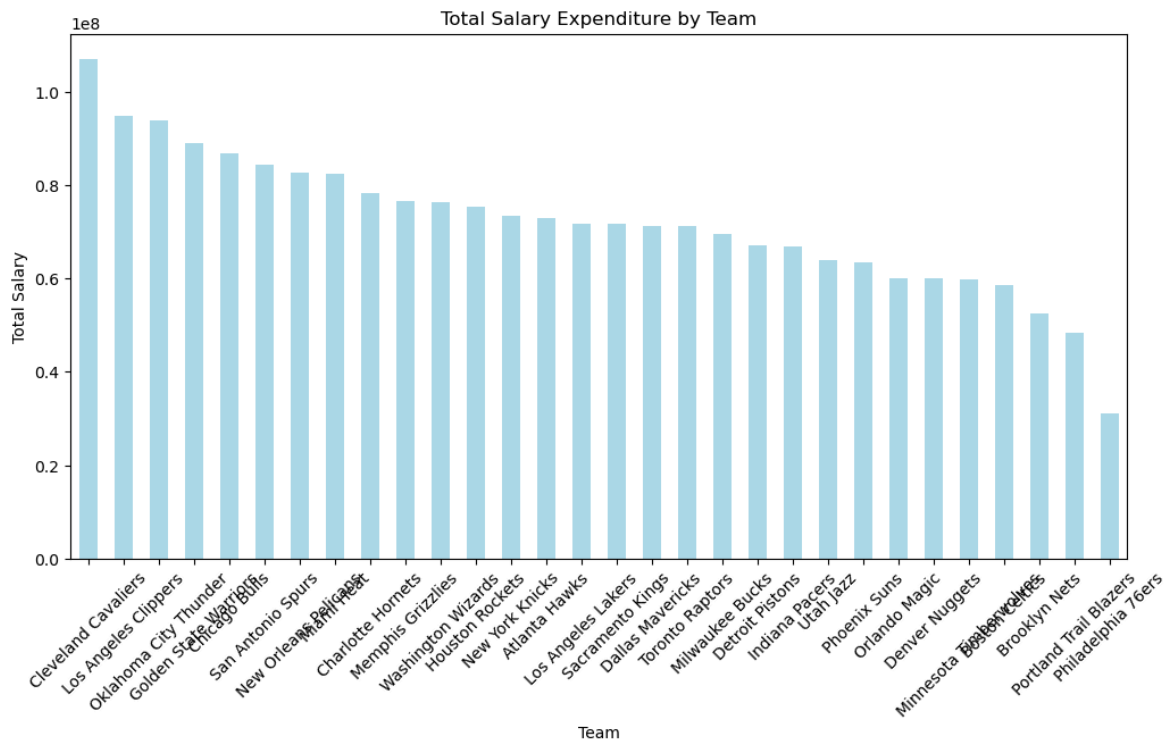
In [49]: # Q4

```
In [51]: highest_salary_team = data.groupby('Team')['Salary'].sum().idxmax()
highest_salary_position = data.groupby('Position')['Salary'].sum().idxmax()
print("Team with highest salary expenditure:", highest_salary_team)
print("Position with highest salary expenditure:", highest_salary_position)
```

Team with highest salary expenditure: Cleveland Cavaliers  
Position with highest salary expenditure: C

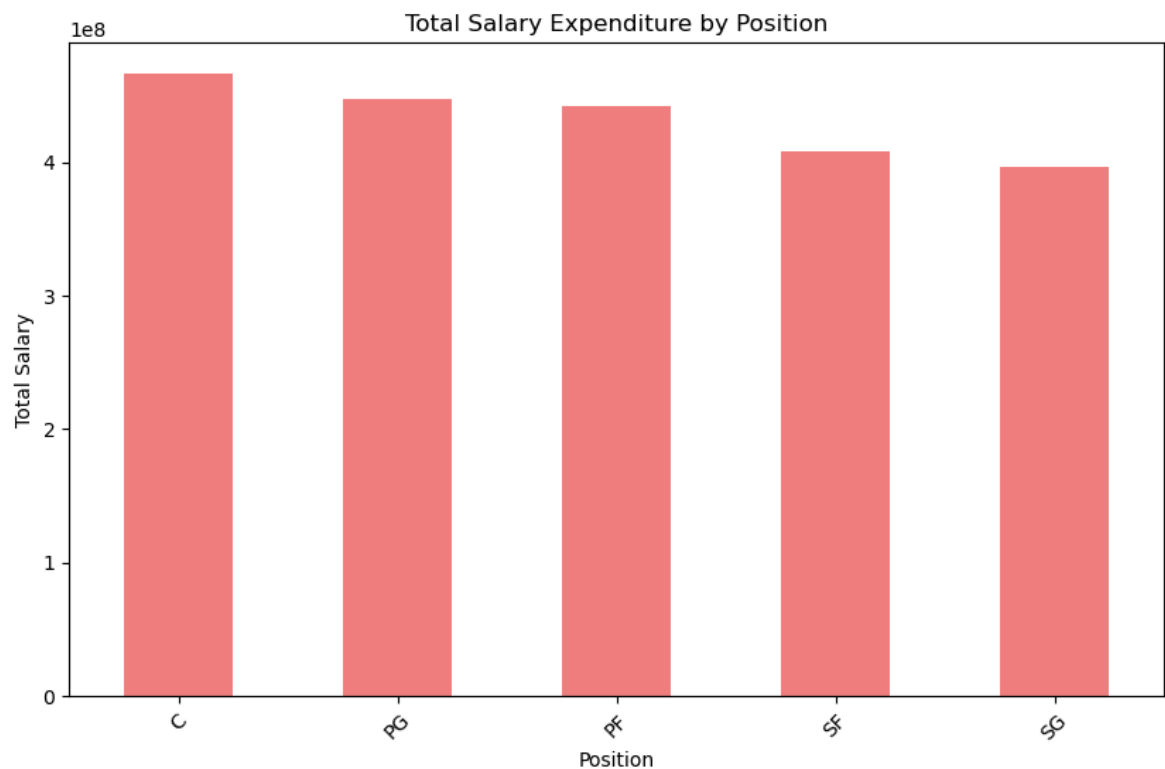
```
In [63]: team_salary = data.groupby('Team')['Salary'].sum().sort_values(ascending=False)
```

```
In [65]: plt.figure(figsize=(12, 6))
team_salary.plot(kind='bar', color='lightblue')
plt.title('Total Salary Expenditure by Team')
plt.xlabel('Team')
plt.ylabel('Total Salary')
plt.xticks(rotation=45)
plt.show()
```



```
In [67]: position_salary = data.groupby('Position')['Salary'].sum().sort_values(ascending
```

```
In [69]: plt.figure(figsize=(10, 6))
position_salary.plot(kind='bar', color='lightcoral')
plt.title('Total Salary Expenditure by Position')
plt.xlabel('Position')
plt.ylabel('Total Salary')
plt.xticks(rotation=45)
plt.show()
```



```
In [53]: # Q5
```

```
In [57]: correlation = data['Age'].corr(data['Salary'])  
print("Correlation between Age and Salary:", correlation)
```

Correlation between Age and Salary: 0.21400941226570985

```
In [61]: plt.figure(figsize=(10, 6))  
plt.scatter(data['Age'], data['Salary'], color='purple', alpha=0.5)  
plt.title('Correlation between Age and Salary')  
plt.xlabel('Age')  
plt.ylabel('Salary')  
plt.show()
```



```
In [ ]:
```