In [1]:

```
# Import necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset
url = "https://docs.google.com/spreadsheets/d/1VP9BE_eI2yl6uUHSm4mGiiwjRdoqCqnkcIjsv5Q2ex4/export?fo
df = pd.read_csv(url)

# Display the first few rows of the dataset
df.head()
```

Out[1]:

	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 [2]:

```
# Replace height values with random numbers between 150 and 180
np.random.seed(0)  # For reproducibility
df['height'] = np.random.randint(150, 181, size=df.shape[0])

# Ensure data consistency and integrity
df.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    object 
 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(1), int64(3), object(5)
memory usage: 34.1+ KB
```

In [4]:

```
total_employees= len(df)
total_employees
```

Out[4]:

```
458
```

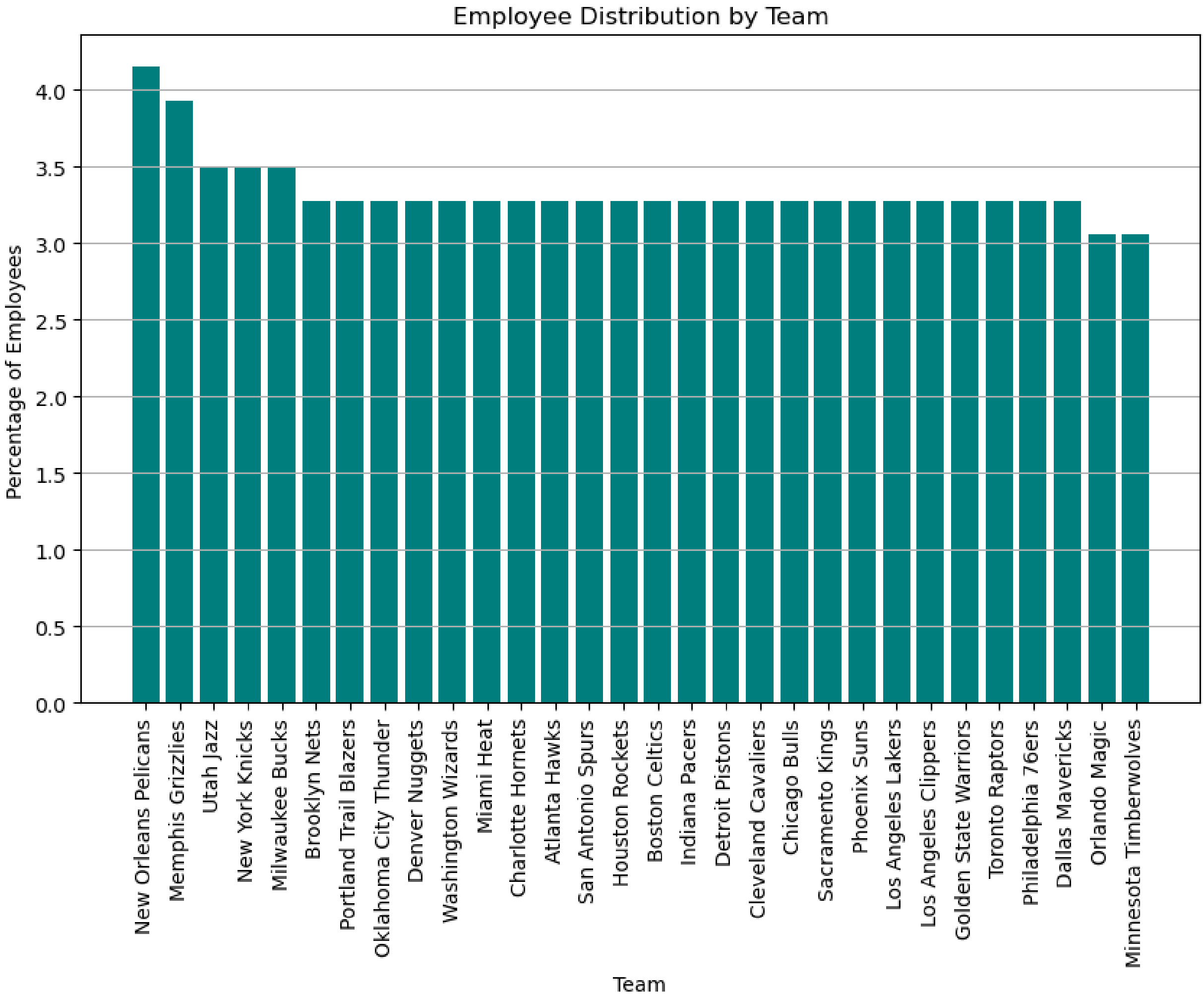
```
In [5]: team_counts= df['Team'].value_counts()
        team_counts
```

```
Out[5]: 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 [6]: team_percentages = (team_counts/total_employees)*100
        team_percentages
```

```
Out[6]: 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 [7]:
```
plt.figure(figsize=(10, 6))
plt.bar(team_percentages.index, team_percentages.values, color='teal')
plt.xlabel('Team')
plt.ylabel('Percentage of Employees')
plt.title('Employee Distribution by Team')
plt.xticks(rotation=90)
plt.grid(axis='y')
plt.show()
```



In [9]:
```
position_segregation=df.groupby('Position').size().reset_index(name='Number of employees')
position_segregation.index += 1
position_segregation
```

Out[9]:
```
   Position  Number of employees
1         C                   79
2        PF                  100
3        PG                   92
4        SF                   85
5        SG                  102
```

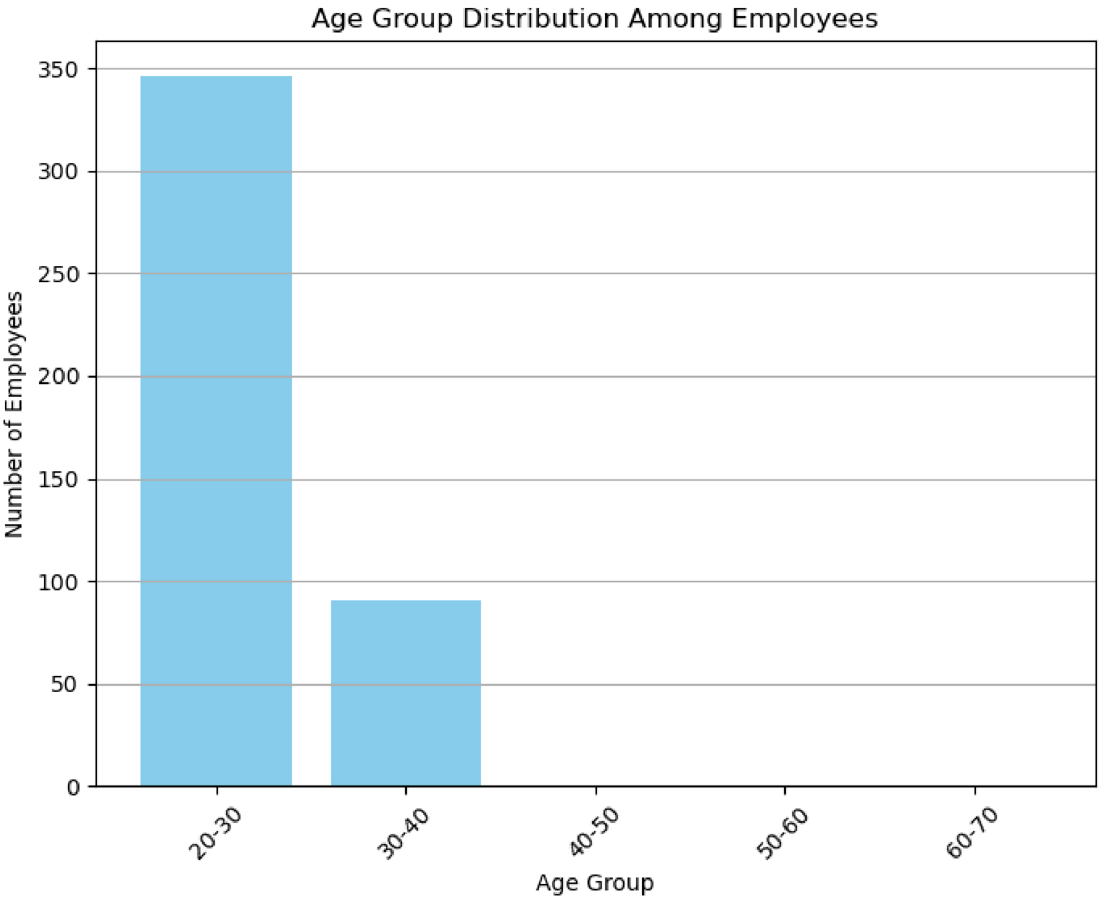
```
In [10]: age_group = pd.cut(df['Age'], bins=[20, 30, 40, 50, 60, 100],labels=['20-30', '30-40', '40-50', '50-
         age_group_counts = age_group.value_counts()
         age_group_counts
```

```
Out[10]: Age
         20-30    346
         30-40     91
         40-50      0
         50-60      0
         60-70      0
         Name: count, dtype: int64
```

```
In [11]: most_common_age_group = age_group_counts.idxmax()
         count=age_group_counts.max()
         print("Age group with the most employees is", most_common_age_group, "with count of employees",count
```

```
         Age group with the most employees is 20-30 with count of employees 346
```

```
In [12]: plt.figure(figsize=(8, 6))
         plt.bar(age_group_counts.index, age_group_counts.values, color='skyblue')
         plt.xlabel('Age Group')
         plt.ylabel('Number of Employees')
         plt.title('Age Group Distribution Among Employees')
         plt.xticks(rotation=45)
         plt.grid(axis='y')
         plt.show()
```

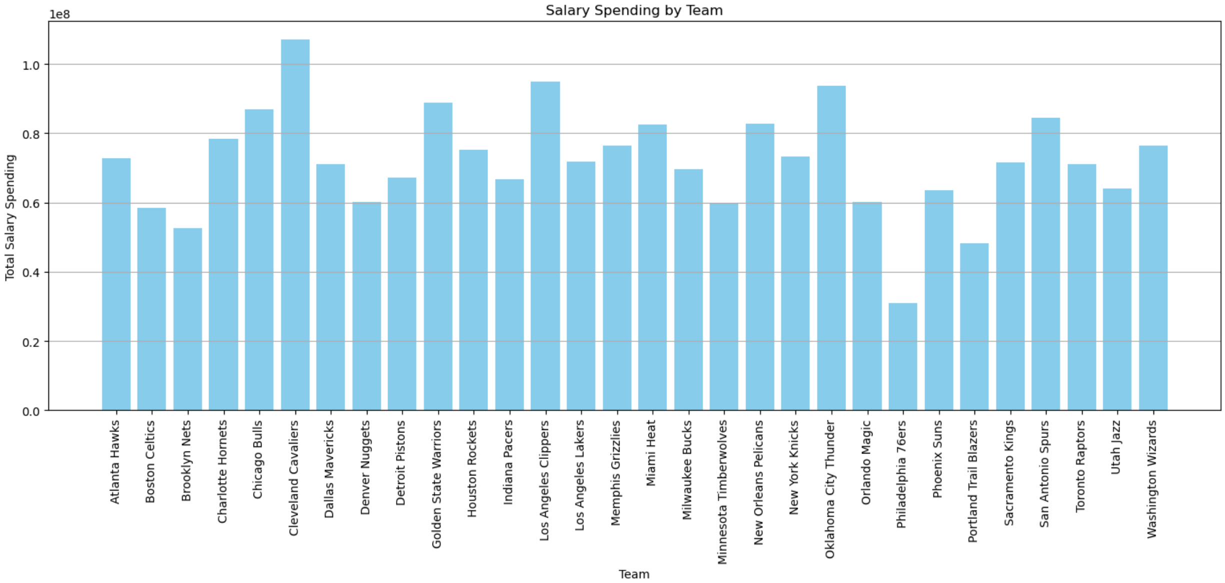


```
In [13]: team_salary = df.groupby('Team')['Salary'].sum()
         highest_team = team_salary.idxmax()
         print ("Highest Salary Team:",highest_team)
```

```
         Highest Salary Team: Cleveland Cavaliers
```

In [14]:

```
plt.figure(figsize=(18, 6))
plt.bar(team_salary.index, team_salary.values, color='skyblue')
plt.xlabel('Team')
plt.ylabel('Total Salary Spending')
plt.title('Salary Spending by Team')
plt.xticks(rotation=90)
plt.grid(axis='y')
plt.show()
```



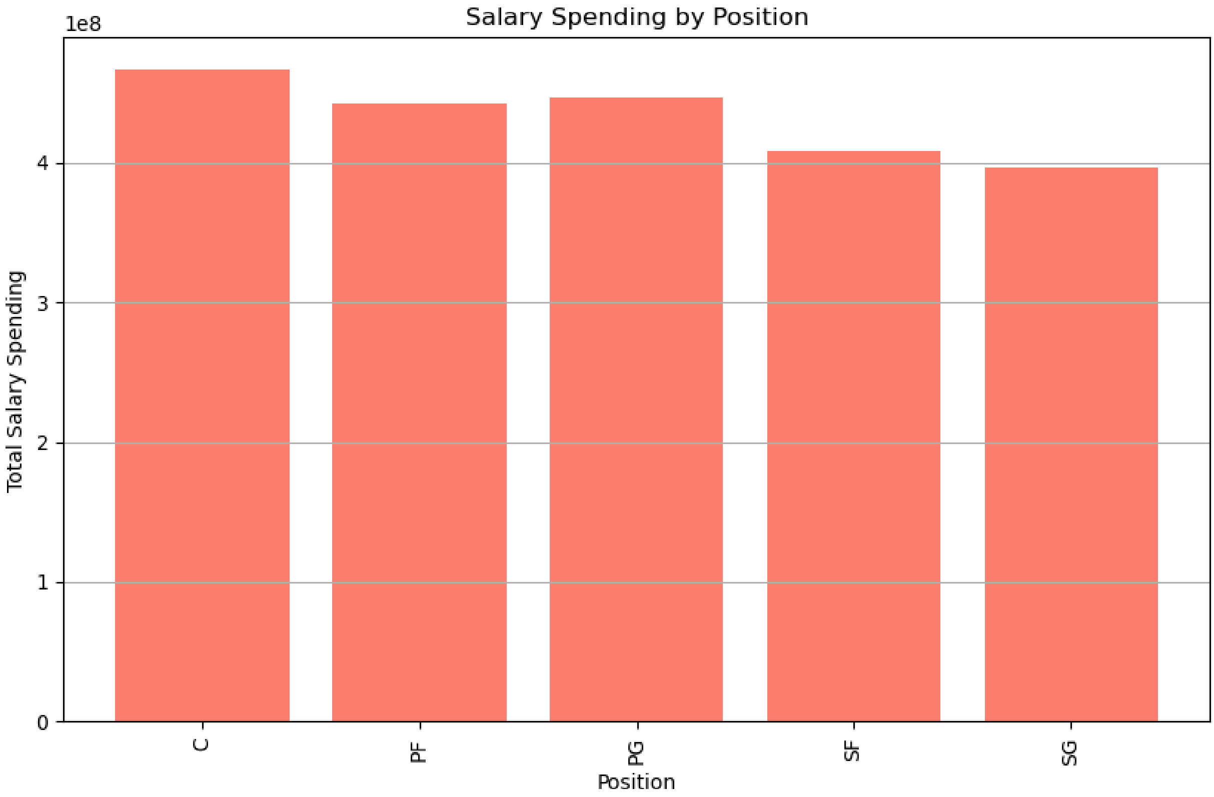
In [15]:

```
highest_position = df.groupby('Position')['Salary'].sum().idxmax()
print ("Highest Salary Position:",highest_position)
```

```
Highest Salary Position: C
```

In [16]:

```
position_salary = df.groupby('Position')['Salary'].sum()
plt.figure(figsize=(10, 6))
plt.bar(position_salary.index, position_salary.values, color='salmon')
plt.xlabel('Position')
plt.ylabel('Total Salary Spending')
plt.title('Salary Spending by Position')
plt.xticks(rotation=90)
plt.grid(axis='y')
plt.show()
```



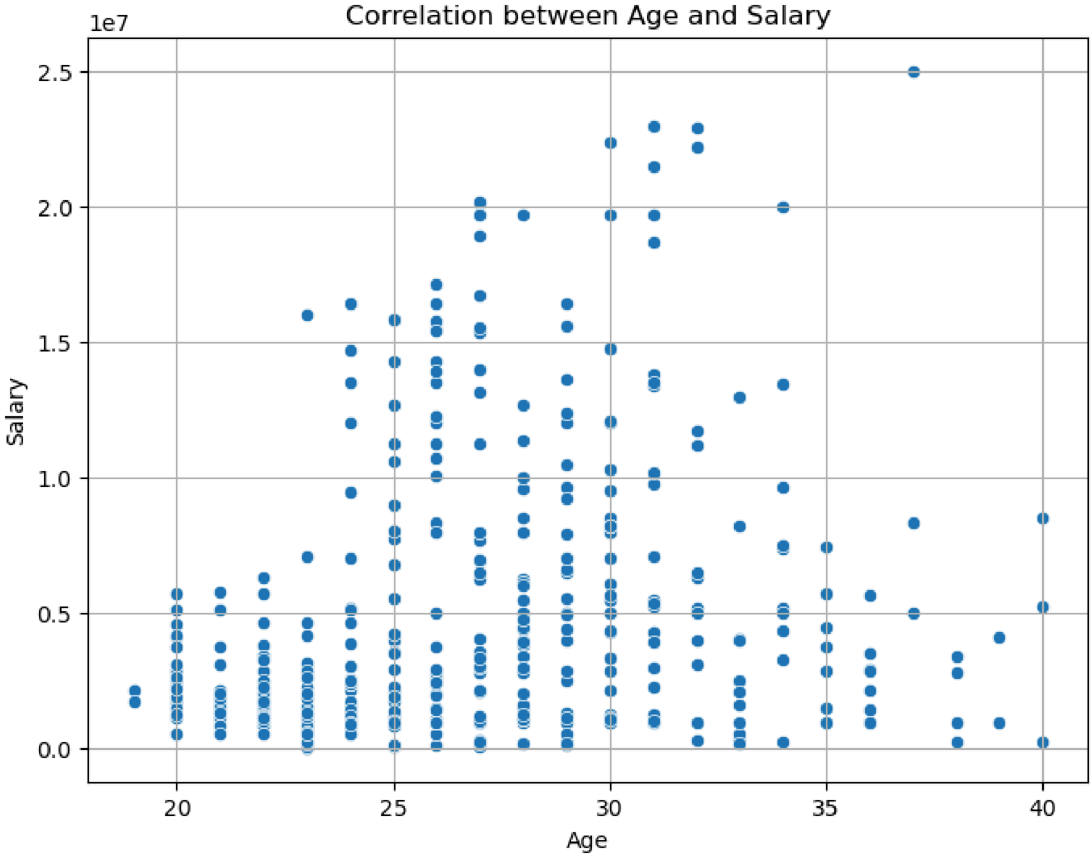
In [17]:

```
# Correlation between age and salary
correlation=df['Age'].corr(df['Salary'])
print("Correlation between Age and Salary:", correlation)
```

```
Correlation between Age and Salary: 0.21400941226570974
```

In [18]:

```
plt.figure(figsize=(8, 6))
sns.scatterplot(x='Age', y='Salary', data=df)
plt.xlabel('Age')
plt.ylabel('Salary')
plt.title('Correlation between Age and Salary')
plt.grid(True)
plt.show()
```



In [ ]: