In [5]:
```python
#PREPROCESSING

import pandas as pd
import numpy as np

# Load the dataset
df = pd.read_csv('ABC.csv')

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

print(df.info())
```

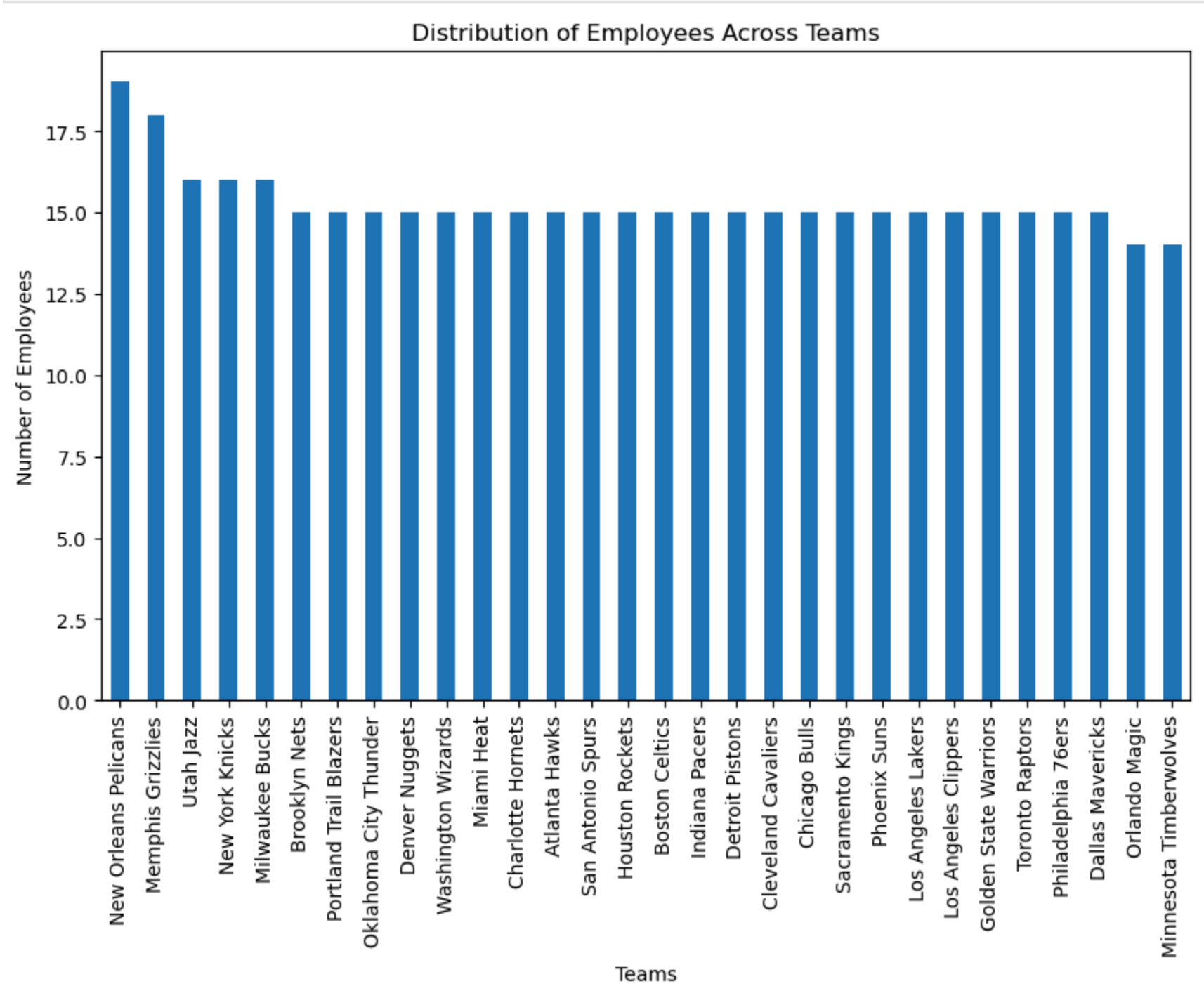
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 458 entries, 0 to 457
Data columns (total 10 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Name      458 non-null    object
 1   Team      458 non-null    object
 2   Number    458 non-null    int64
 3   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
None
```

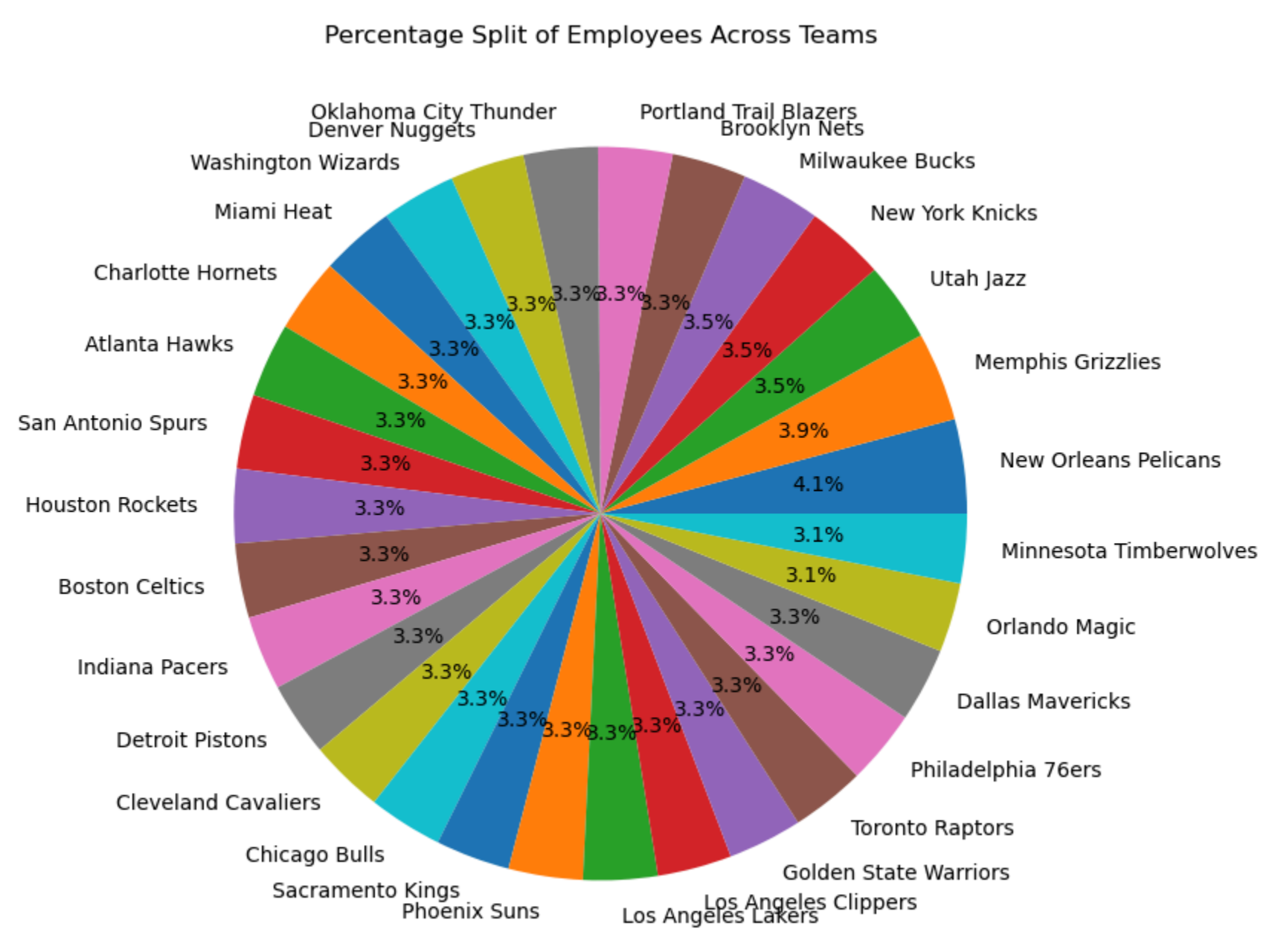
In [ ]:

In [10]:
```python
#TASK1
team_distribution = df['Team'].value_counts()
team_percentage = (team_distribution / df.shape[0]) * 100
print(team_distribution)
print(team_percentage)
```

```
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
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 [11]:
```python
#representation
import matplotlib.pyplot as plt

# Bar plot for team distribution
plt.figure(figsize=(10, 6))
team_distribution.plot(kind='bar')
plt.title('Distribution of Employees Across Teams')
plt.xlabel('Teams')
plt.ylabel('Number of Employees')
plt.show()

# Pie chart for team percentage
plt.figure(figsize=(8, 8))
team_percentage.plot(kind='pie', autopct='%1.1f%%')
plt.title('Percentage Split of Employees Across Teams')
plt.ylabel('')
plt.show()
```

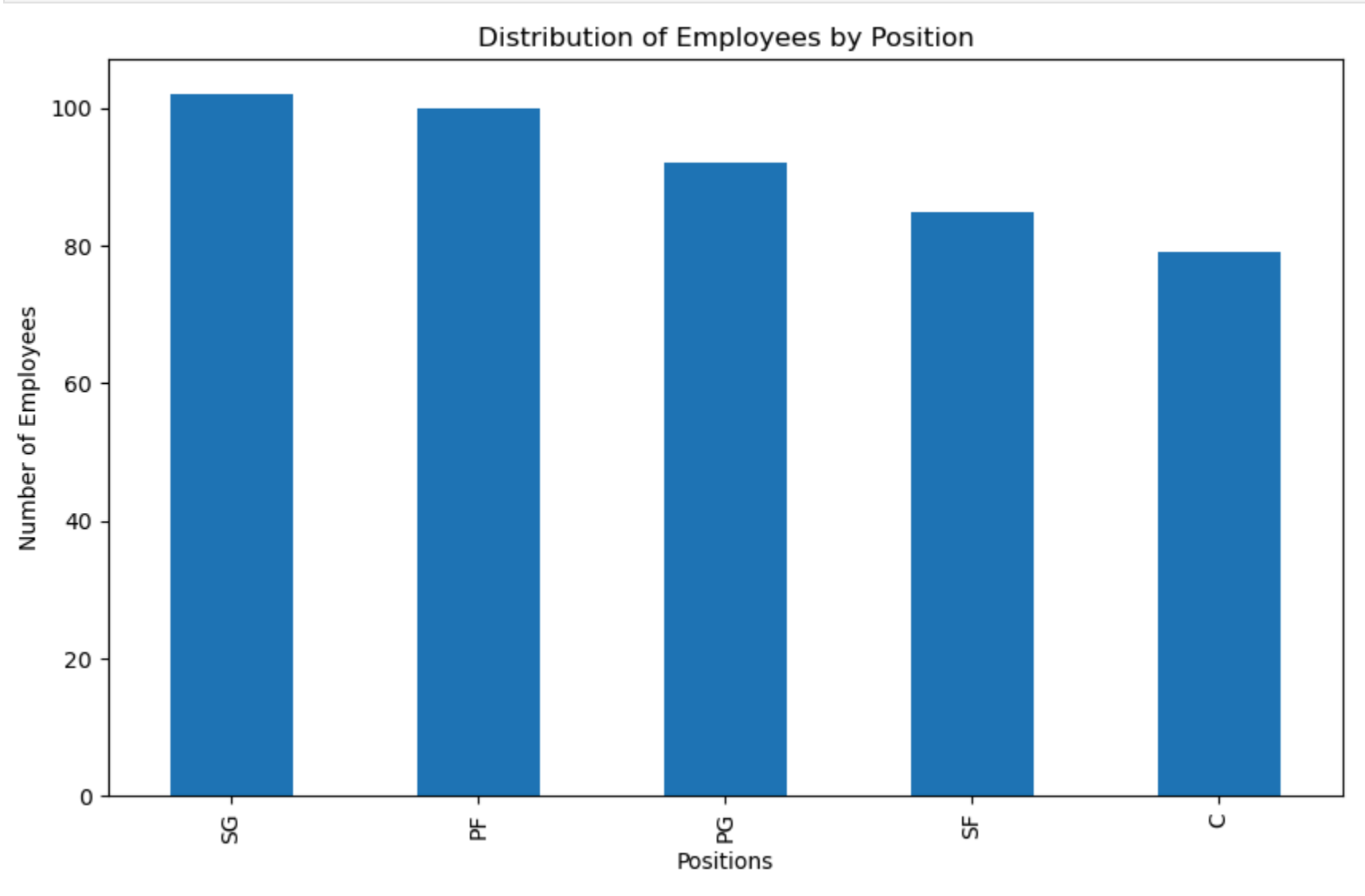




In [13]:
```python
#TASK2
position_distribution = df['Position'].value_counts()
print(position_distribution)
```

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

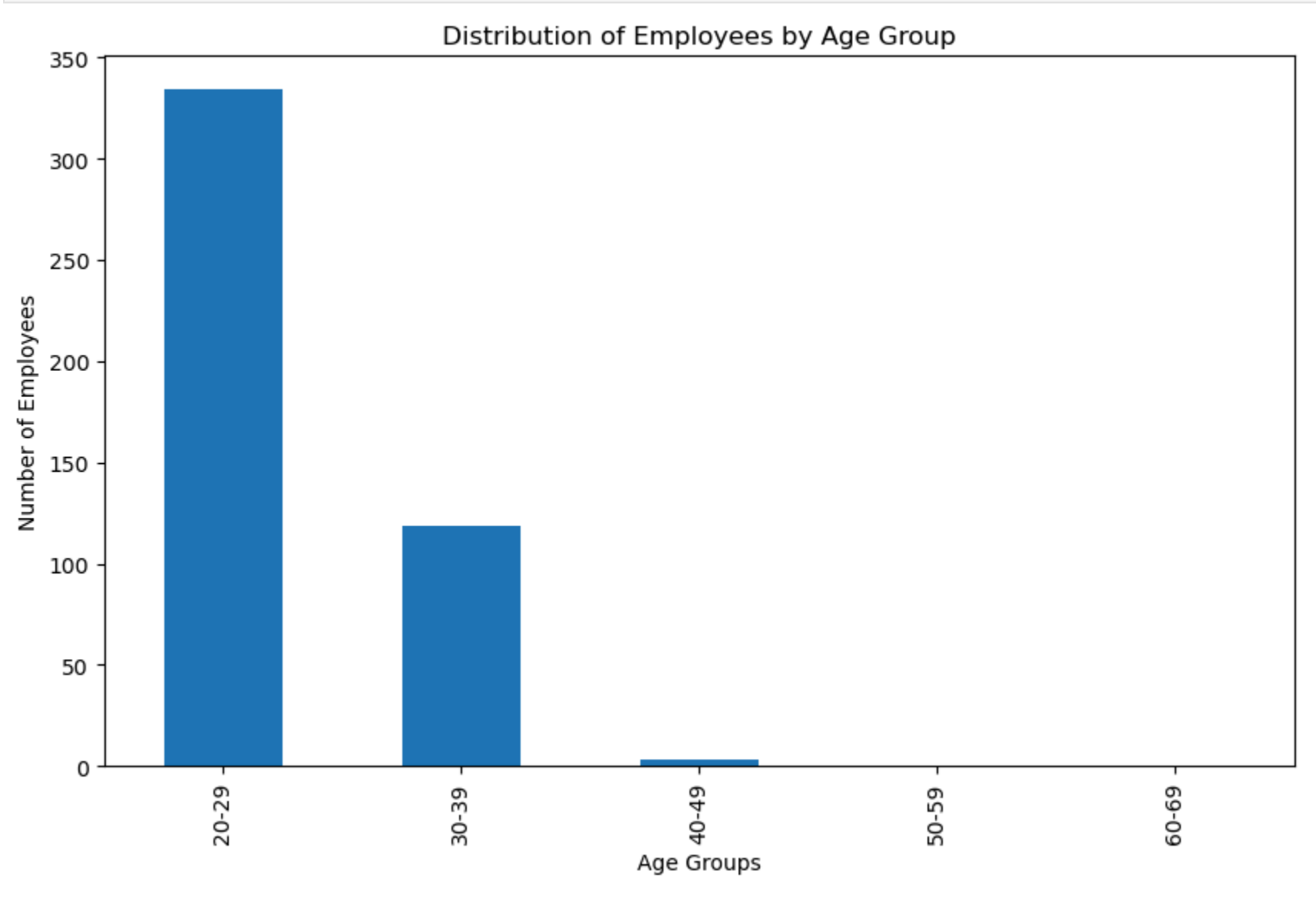
In [14]:
```python
#representation
plt.figure(figsize=(10, 6))
position_distribution.plot(kind='bar')
plt.title('Distribution of Employees by Position')
plt.xlabel('Positions')
plt.ylabel('Number of Employees')
plt.show()
```



In [16]:
```python
#TASK3
bins = [20, 30, 40, 50, 60, 70]
labels = ['20-29', '30-39', '40-49', '50-59', '60-69']
df['age_group'] = pd.cut(df['Age'], bins=bins, labels=labels, right=False)
age_group_distribution = df['age_group'].value_counts().sort_index()
print(age_group_distribution)
```

```
age_group
20-29    334
30-39    119
40-49      3
50-59      0
60-69      0
Name: count, dtype: int64
```

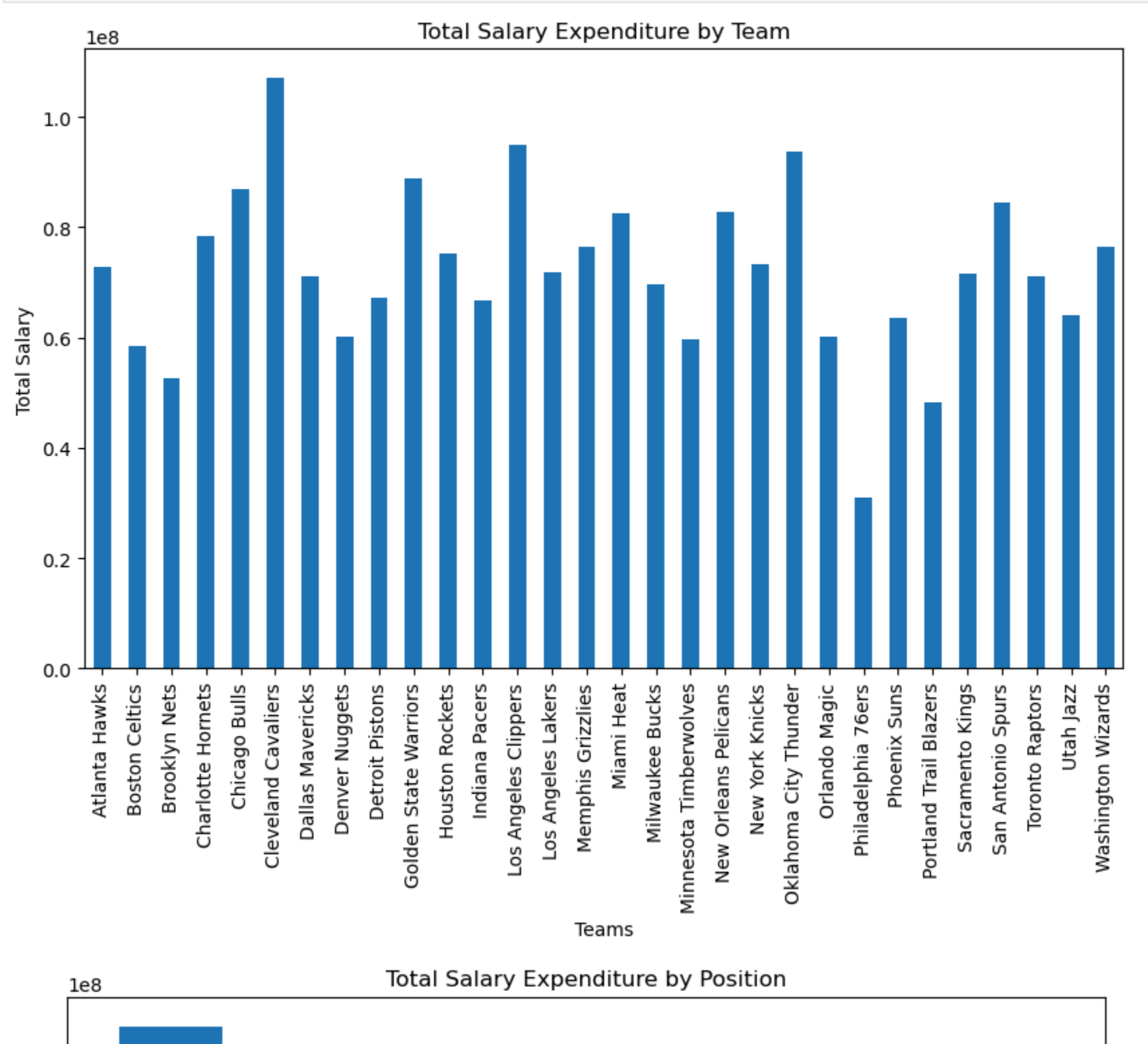
In [17]:
```python
#representation
plt.figure(figsize=(10, 6))
age_group_distribution.plot(kind='bar')
plt.title('Distribution of Employees by Age Group')
plt.xlabel('Age Groups')
plt.ylabel('Number of Employees')
plt.show()
```

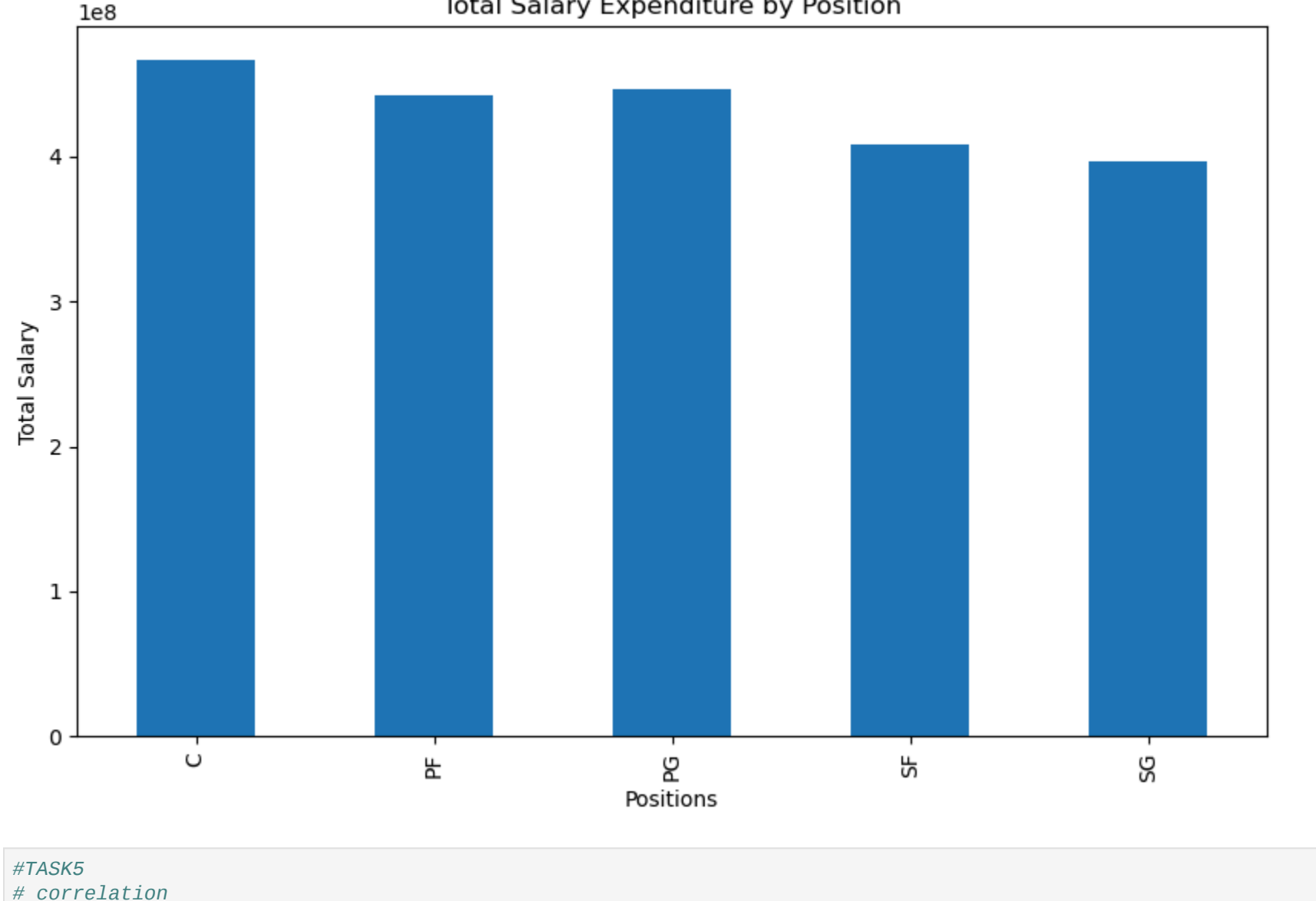


In [18]:
```python
#TASK4
team_salary = df.groupby('Team')['Salary'].sum()
position_salary = df.groupby('Position')['Salary'].sum()
print('Total salary by team:\n', team_salary)
print('Total salary by position:\n', position_salary)
```

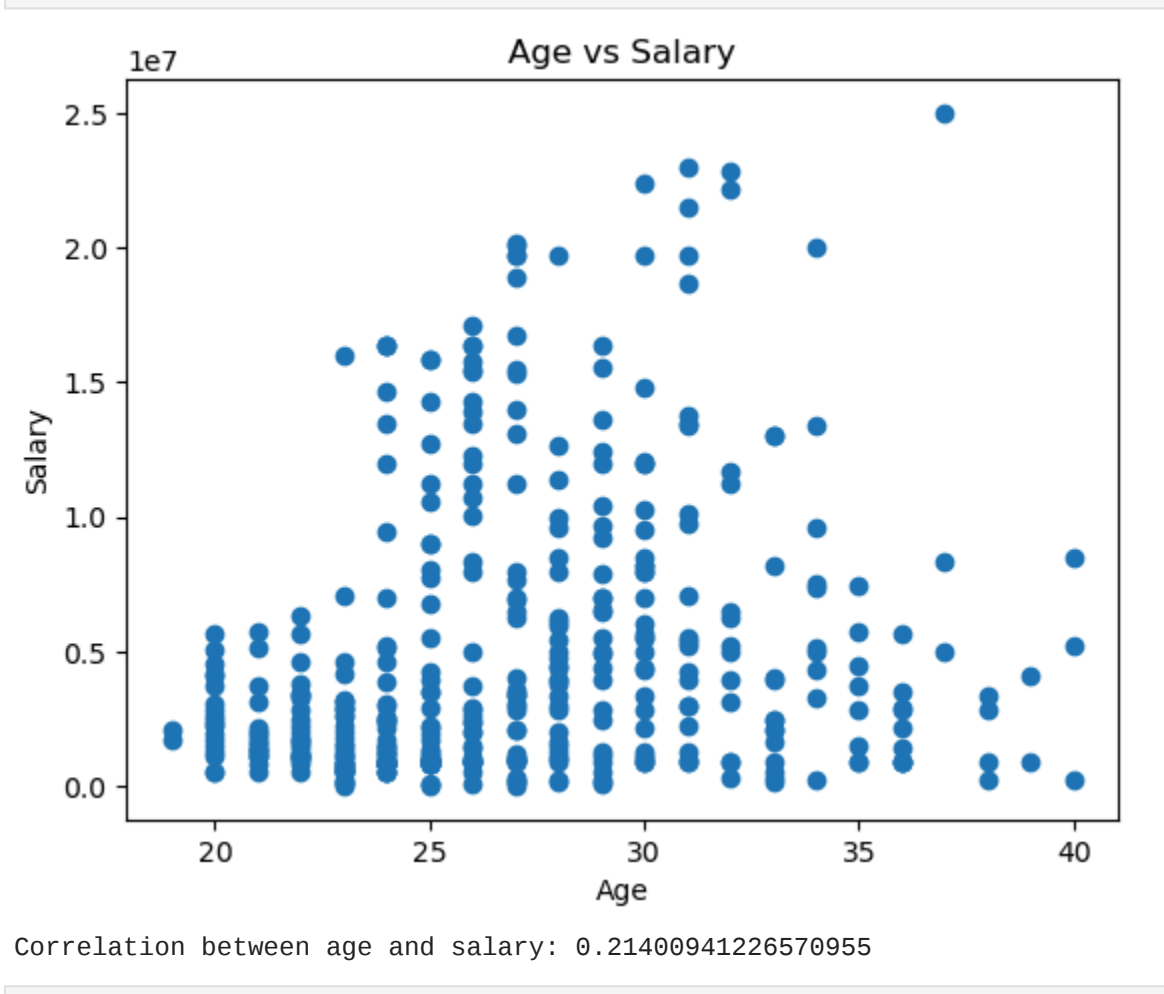
```
Total salary by team:
 Team
Atlanta Hawks              72902950.0
Boston Celtics             58541068.0
Brooklyn Nets              52528475.0
Charlotte Hornets          78340920.0
Chicago Bulls              86783378.0
Cleveland Cavaliers       106988689.0
Dallas Mavericks           71198732.0
Denver Nuggets             60121930.0
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          76550880.0
Miami Heat                 82515673.0
Milwaukee Bucks            69603517.0
Minnesota Timberwolves     59709697.0
New Orleans Pelicans       82750774.0
New York Knicks            73303898.0
Oklahoma City Thunder      93765298.0
Orlando Magic              60161470.0
Philadelphia 76ers         30992894.0
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
Total salary by position:
 Position
C     466377332.0
PF    442560850.0
PG    446848971.0
SF    408020976.0
SG    396976258.0
Name: Salary, dtype: float64
```

In [19]:
```python
#representation
# Bar plot for salary expenditure by team
plt.figure(figsize=(10, 6))
team_salary.plot(kind='bar')
plt.title('Total Salary Expenditure by Team')
plt.xlabel('Teams')
plt.ylabel('Total Salary')
plt.show()

# Bar plot for salary expenditure by position
plt.figure(figsize=(10, 6))
position_salary.plot(kind='bar')
plt.title('Total Salary Expenditure by Position')
plt.xlabel('Positions')
plt.ylabel('Total Salary')
plt.show()
```





In [22]:
```python
#TASK5
# correlation
plt.scatter(df['Age'], df['Salary'])
plt.title('Age vs Salary')
plt.xlabel('Age')
plt.ylabel('Salary')
plt.show()

# correlation coefficient
correlation = df[['Age', 'Salary']].corr().iloc[0, 1]
print('Correlation between age and salary:', correlation)
```




```
Correlation between age and salary: 0.21400941226570955
```

In [ ]: