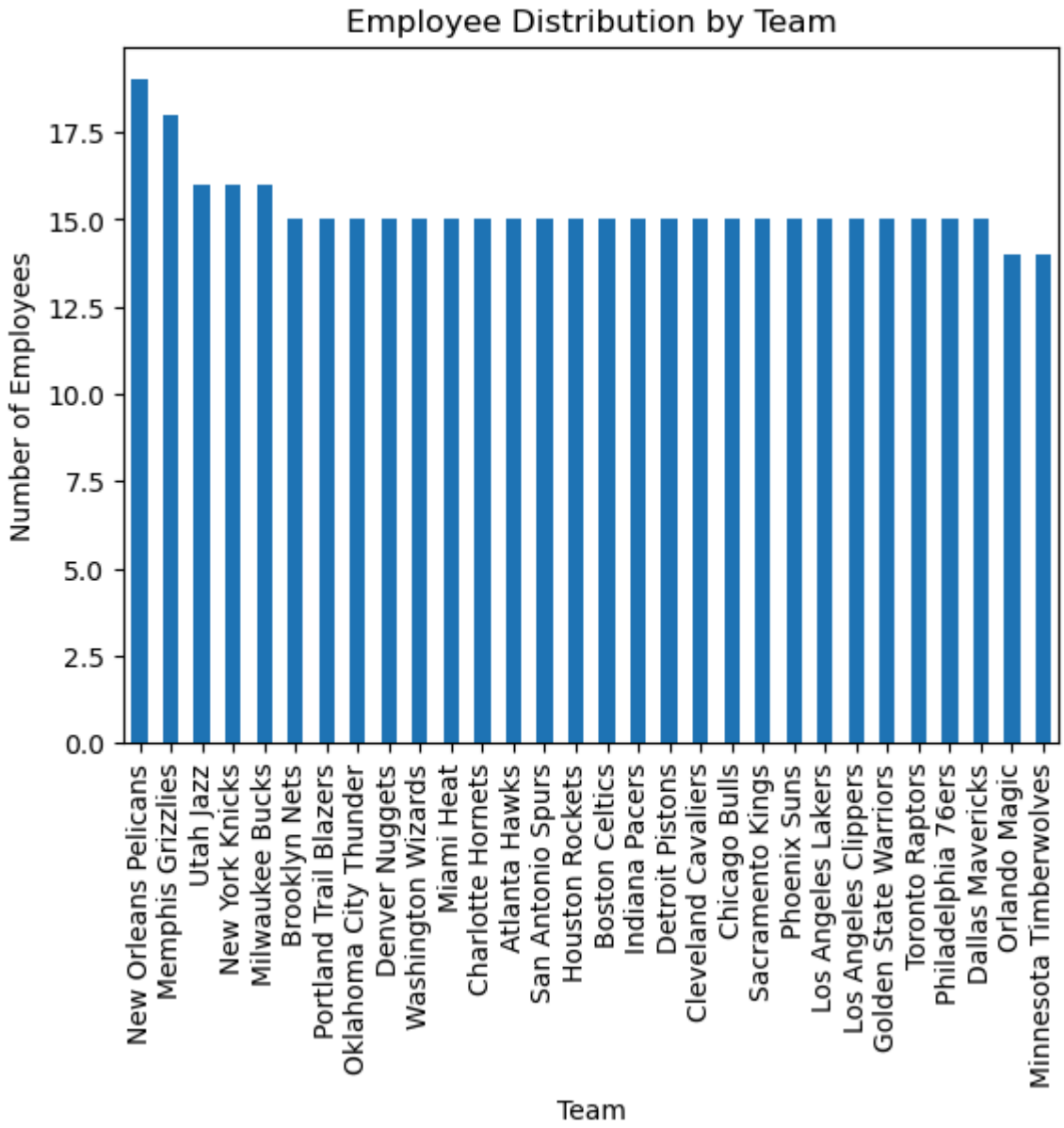
In [ ]:
```
# PYTHON PROJECT
```

In [29]:
```
# 1. Determine the distribution of employees across each team and calculate the
import pandas as pd
import random
df=pd.read_csv("myexcel.csv")
# Assuming your data is loaded into a pandas DataFrame called df
df['Height'] = df['Height'].apply(lambda x: random.randint(150, 180) if pd.isna(
```

In [35]:
```
team_distribution = df['Team'].value_counts()
total_employees = len(df)
team_percentage = (team_distribution / total_employees) * 100
```

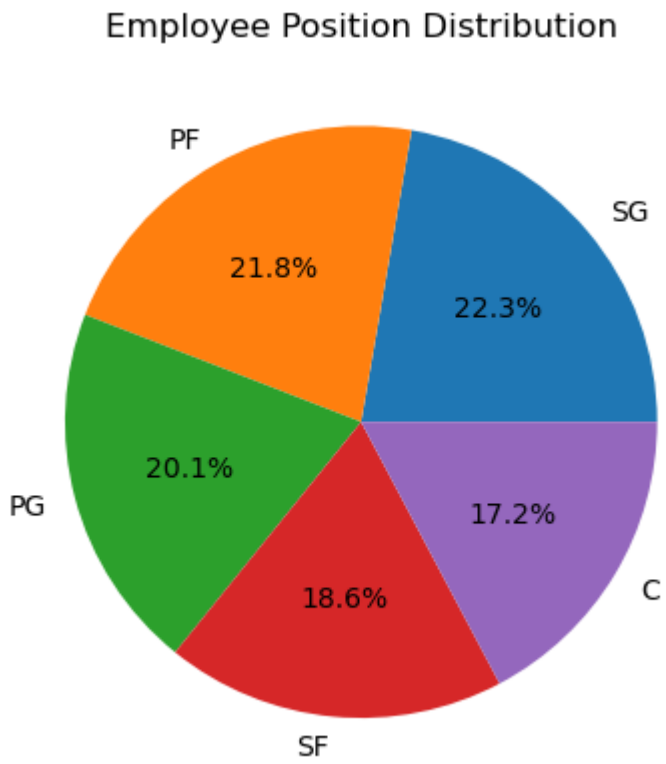
In [33]:
```
import matplotlib.pyplot as plt
team_distribution.plot(kind='bar', title='Employee Distribution by Team')
plt.xlabel('Team')
plt.ylabel('Number of Employees')
plt.show()
```



In [ ]:
```
```

In [37]:
```
# 2. Segregate employees based on their positions within the company.
position_distribution = df['Position'].value_counts()
```

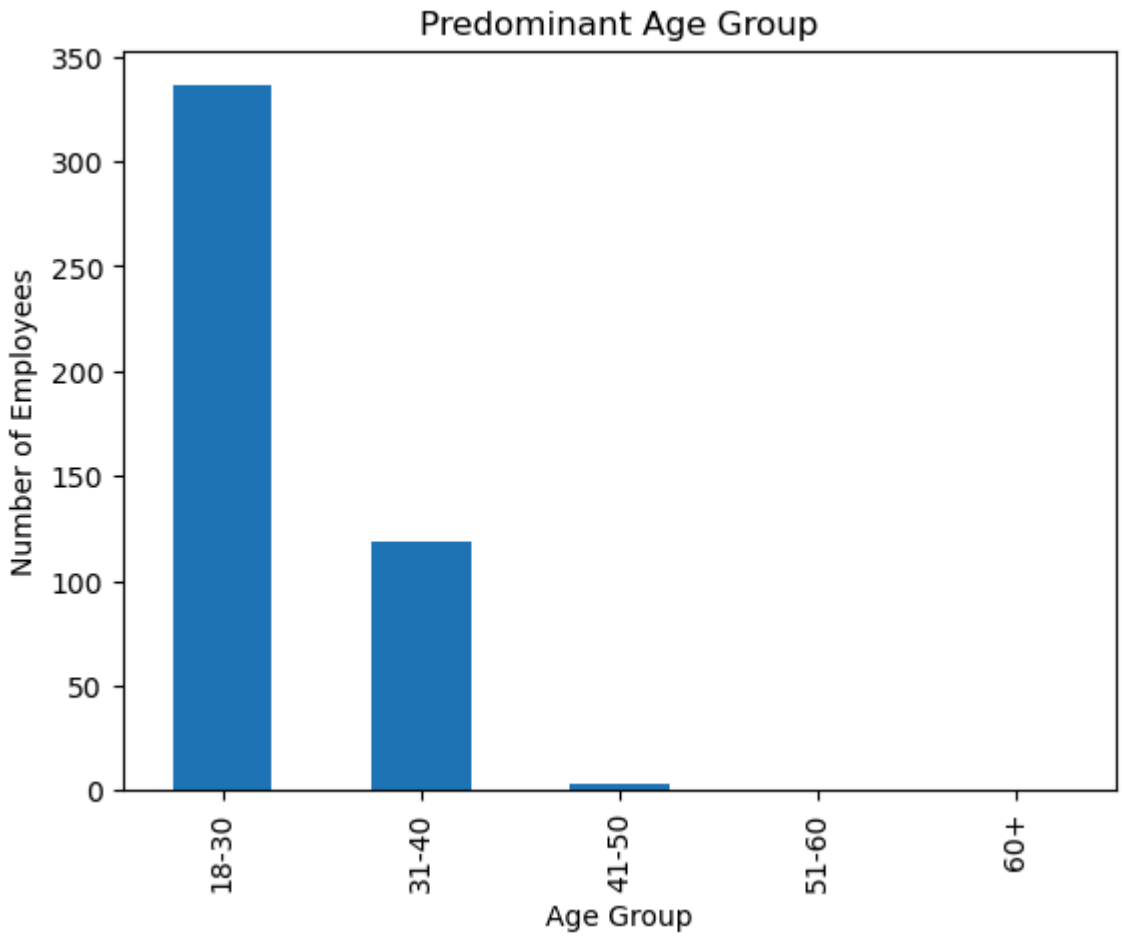
In [39]:
```
position_distribution.plot(kind='pie', title='Employee Position Distribution', a
plt.ylabel('')
plt.show()
```



In [ ]:
```
```

In [41]:
```
# 3. Identify the predominant age group among employees.
bins = [18, 30, 40, 50, 60, 100]
labels = ['18-30', '31-40', '41-50', '51-60', '60+']
df['Age_group'] = pd.cut(df['Age'], bins=bins, labels=labels, right=False)
age_group_distribution = df['Age_group'].value_counts()
```

In [43]:
```
age_group_distribution.plot(kind='bar', title='Predominant Age Group')
plt.xlabel('Age Group')
plt.ylabel('Number of Employees')
plt.show()
```

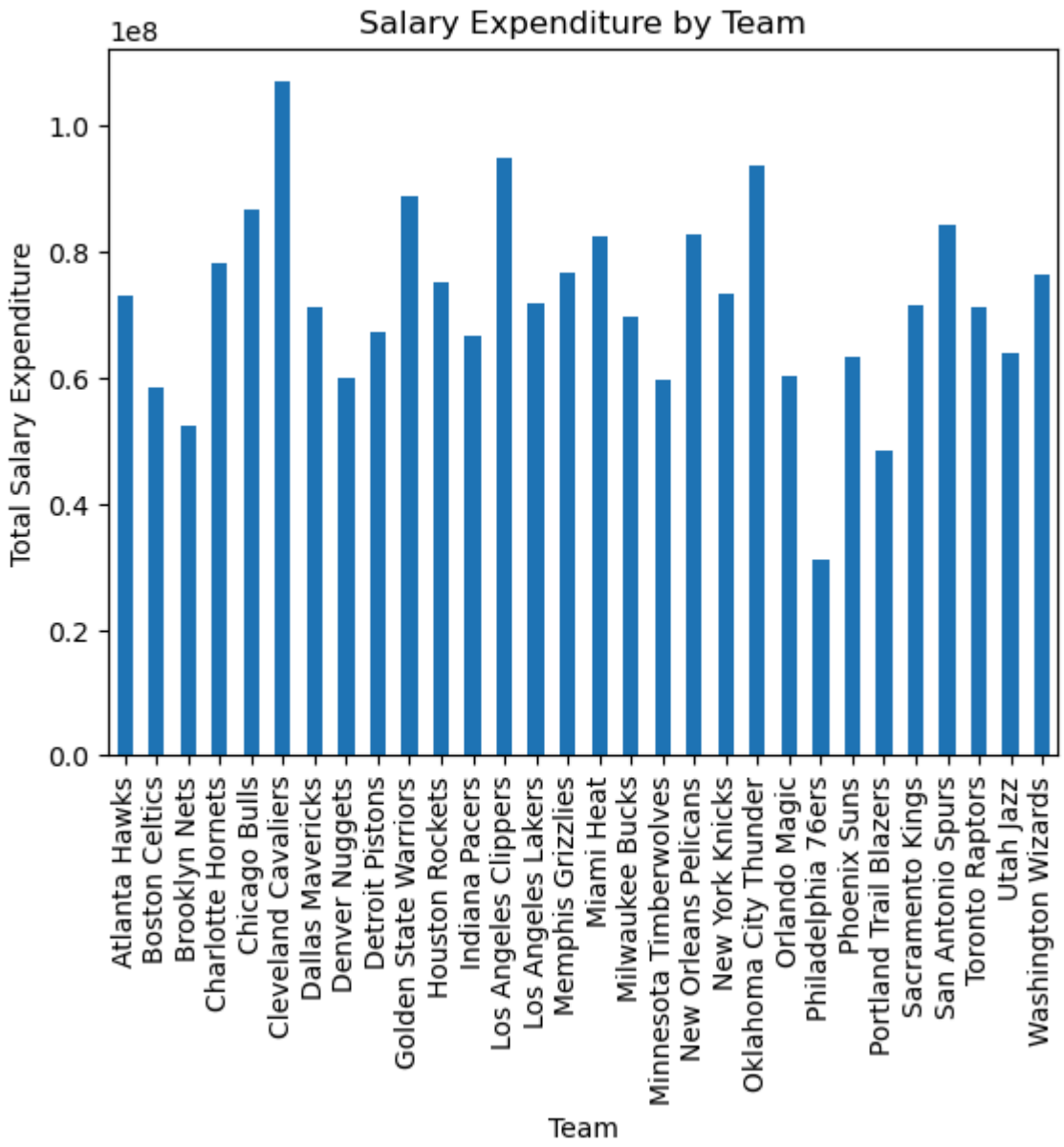


In [ ]:

In [45]:
```
# 4. Discover which team and position have the highest salary expenditure.
team_salary_expenditure = df.groupby('Team')['Salary'].sum()
position_salary_expenditure = df.groupby('Position')['Salary'].sum()

max_team = team_salary_expenditure.idxmax()
```

In [47]:
```
team_salary_expenditure.plot(kind='bar', title='Salary Expenditure by Team')
plt.xlabel('Team')
plt.ylabel('Total Salary Expenditure')
plt.show()
```



In [ ]:

In [49]:
```
# 5. Investigate if there's any correlation between age and salary, and represen
correlation = df['Age'].corr(df['Salary'])
```

In [51]:
```
df.plot(kind='scatter', x='Age', y='Salary', title='Correlation between Age and
plt.xlabel('Age')
plt.ylabel('Salary')
plt.show()
```


Correlation between Age and Salary
1e7
2.5
2.0
1.5
1.0
0.5
0.0
Salary
20
25
30
35
40
Age

In [ ]:

In [ ]:

In [ ]: