In [1]:
```python
import pandas as pd
import numpy as np
```

In [2]:
```python
import pandas as pd
import numpy as np

# Load the dataset
data = pd.read_csv('C:/Users/15196/Downloads/myexcel - myexcel.csv.csv')

# Replace the "height" column with random numbers between 150 and 180
np.random.seed(42)  # For reproducibility
data['height'] = np.random.randint(150, 181, size=len(data))

print(data.head())
data.to_csv('modified_dataset.csv',index=False)
```

```
            Name            Team  Number Position  Age  Height  Weight  \
0  Avery Bradley  Boston Celtics       0       PG   25  06-Feb     180
1    Jae Crowder  Boston Celtics      99       SF   25  06-Jun     235
2    John Holland  Boston Celtics      30       SG   27  06-May     205
3    R.J. Hunter  Boston Celtics      28       SG   22  06-May     185
4   Jonas Jerebko  Boston Celtics       8       PF   29  06-Oct     231

              College     Salary  height
0              Texas  7730337.0     156
1          Marquette  6796117.0     169
2  Boston University        NaN     178
3      Georgia State  1148640.0     164
4                NaN  5000000.0     160
```

# Analysis Tasks

1. Distribution of Employees Across Each Team

In [3]:
```python
import matplotlib.pyplot as plt
import seaborn as sns
```

In [4]:
```python
Team_distribution = data['Team'].value_counts()
Team_percentage = (Team_distribution / len(data)) * 100

plt.figure(figsize=(10,6))
Team_distribution.plot(kind='bar',color='blue')
plt.title('Distribution of Employees Across Teams')
plt.xlabel('Team')
plt.ylabel('Number of Employees')
plt.xticks(rotation=90)
plt.show()

print(Team_distribution)
print(Team_percentage)
```
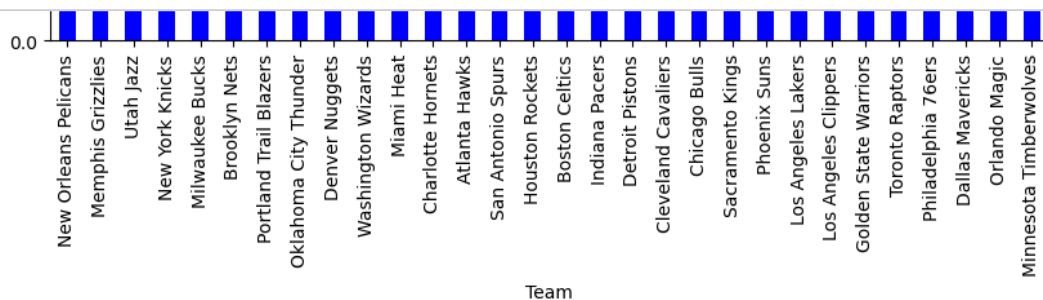


```
Team
New Orleans Pelicans      19
Memphis Grizzlies         18
Utah Jazz                 16
New York Knicks           16
Milwaukee Bucks           16
Brooklyn Nets             15
Portland Trail Blazers    15
Oklahoma City Thunder     15
Denver Nuggets            15
```

2.Segregate Employees Based on Their Positions

In [5]:
```python
position_distribution = data['Position'].value_counts()
```

In [6]:
```python
plt.figure(figsize=(10,6))
position_distribution.plot(kind='bar',color='blue')
plt.title('Distribution of Emolyees by Position')
plt.xlabel('Position')
plt.ylabel('Number of Employees')
plt.xticks(rotation=0)
plt.show()

print(position_distribution)
```



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

3. Predominant Age Group Among Employees

```python
In [7]: import pandas as pd
        bins = [20,25,30,35,40,45,50]
        labels=['20-25','25-30','30-35','35-40','40-45','45-50']
        data['Age Group']= pd.cut(data['Age'], bins=bins,labels=labels, right=False
        age_group_distribution = data['Age Group'].value_counts().sort_index()

        plt.figure(figsize=(10,6))
        age_group_distribution.plot(kind='bar',color='red')
        plt.title('Distribution of Employees by Age Group')
        plt.xlabel('Age Group')
        plt.ylabel('Number of Employees')
        plt.xticks(rotation=0)
        plt.show()

        print(age_group_distribution)
```
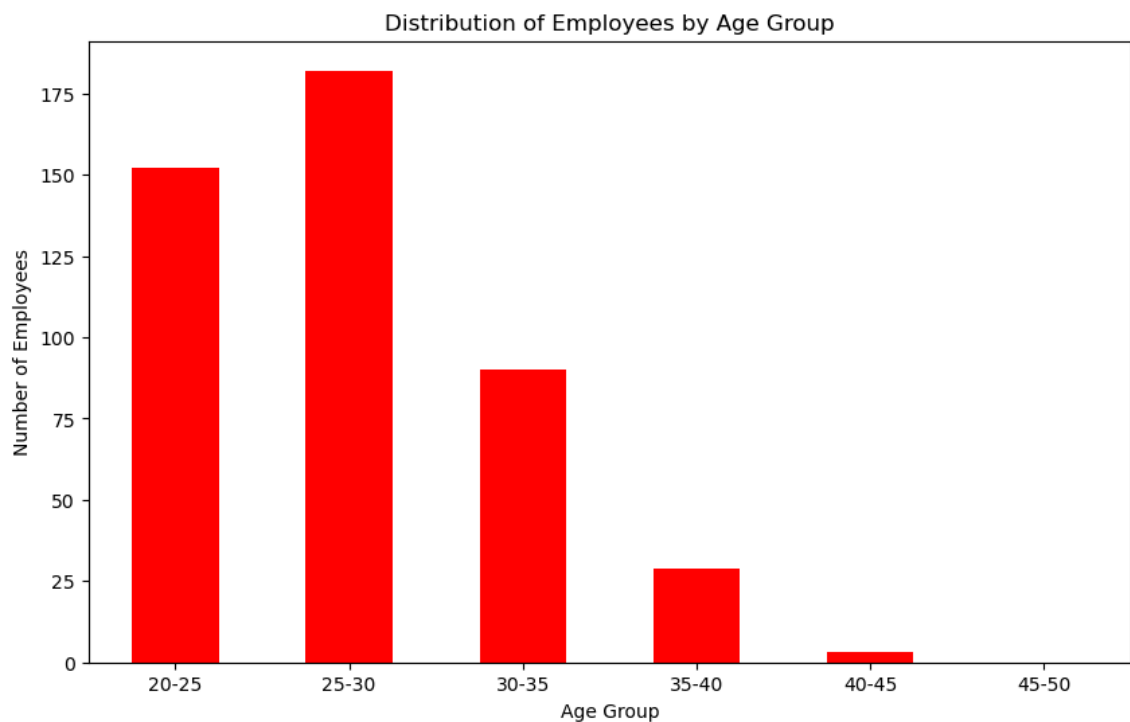


```
Age Group
20-25    152
25-30    182
30-35     90
35-40     29
40-45      3
45-50      0
Name: count, dtype: int64
```

4. Team and Position with the Highest Salary Expenditure

In [8]:
```python
team_salary = data.groupby('Team')['Salary'].sum().sort_values(ascending=Fa
position_salary = data.groupby('Position')['Salary'].sum().sort_values(asce
plt.figure(figsize=(10,6))
team_salary.plot(kind='bar',color='green')
plt.title('Total Salary by Team')
plt.xlabel('Team')
plt.ylabel('Total Salary')
plt.xticks(rotation=90)
plt.show()


plt.figure(figsize=(10,6))
position_salary.plot(kind='bar',color='green')
plt.title('Total Salary by Position')
plt.xlabel('Position')
plt.ylabel('Total Salary')
plt.xticks(rotation=0)
plt.show()

print (team_salary)
print(position_salary)
```
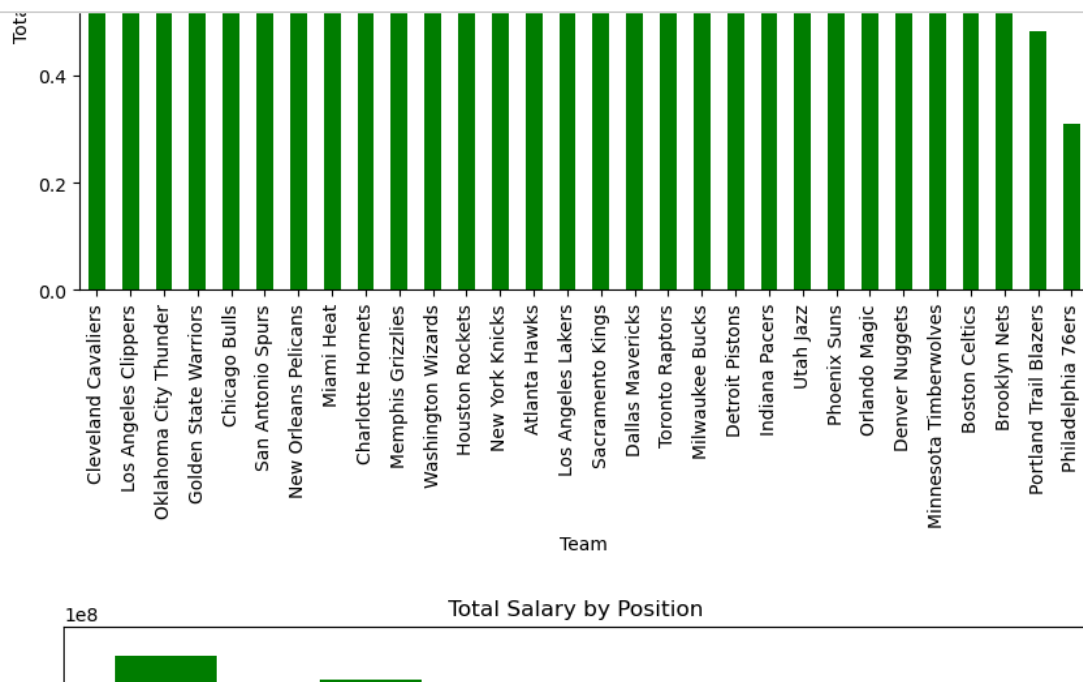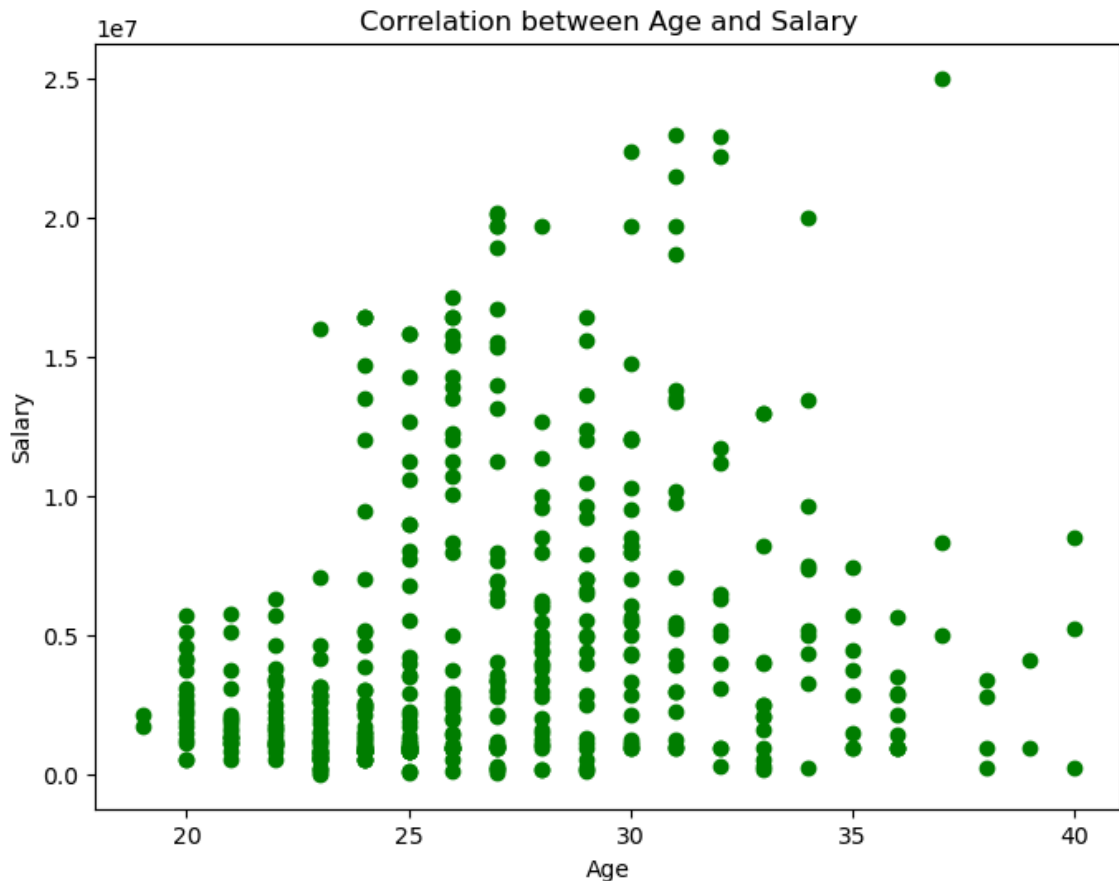

5. Correlation Between Age and Salary

In [9]:
```python
correlation = data[['Age','Salary']].corr()
plt.figure(figsize=(8,6))
plt.scatter(data['Age'],data['Salary'],color='green')
plt.title('Correlation between Age and Salary')
plt.xlabel('Age')
plt.ylabel('Salary')
plt.show()

print(correlation)
```



```
              Age     Salary
Age      1.000000   0.214009
Salary   0.214009   1.000000
```

DATA STORY

Insights From the Analysis

```
 1.Team Distribution:

a.The distribution of employees across different teams shows which teams
are more heavily staffed.
b.The percentage split provides a clear picture of the company's focus
areas.

##2.Position Segregation:

a.Understanding the number of employees in various positions helps in
identifying roles with potential overstaffing or understaffing.
```

3.Age Group Analysis:

a.Identifying the predominant age group aids in workforce planning and predicting retirement waves.

4.Salary Expenditure:

 a.The analysis of salary expenditure by team and position reveals where the company is investing most of its payroll budget.
 b.This helps in budget allocation and identifying high-cost areas.

5.Age-Salary Correlation:

a.Understanding the correlation between age and salary can inform salary structure and career progression policies.

In [ ]: