

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

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

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

# Display the updated DataFrame
print(df[['height']].head())
```

	height
0	156
1	169
2	178
3	164
4	160

```
In [4]: ► team_distribution = df['Team'].value_counts()
team_percentage = (team_distribution / len(df)) * 100

# Display results
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
```

```

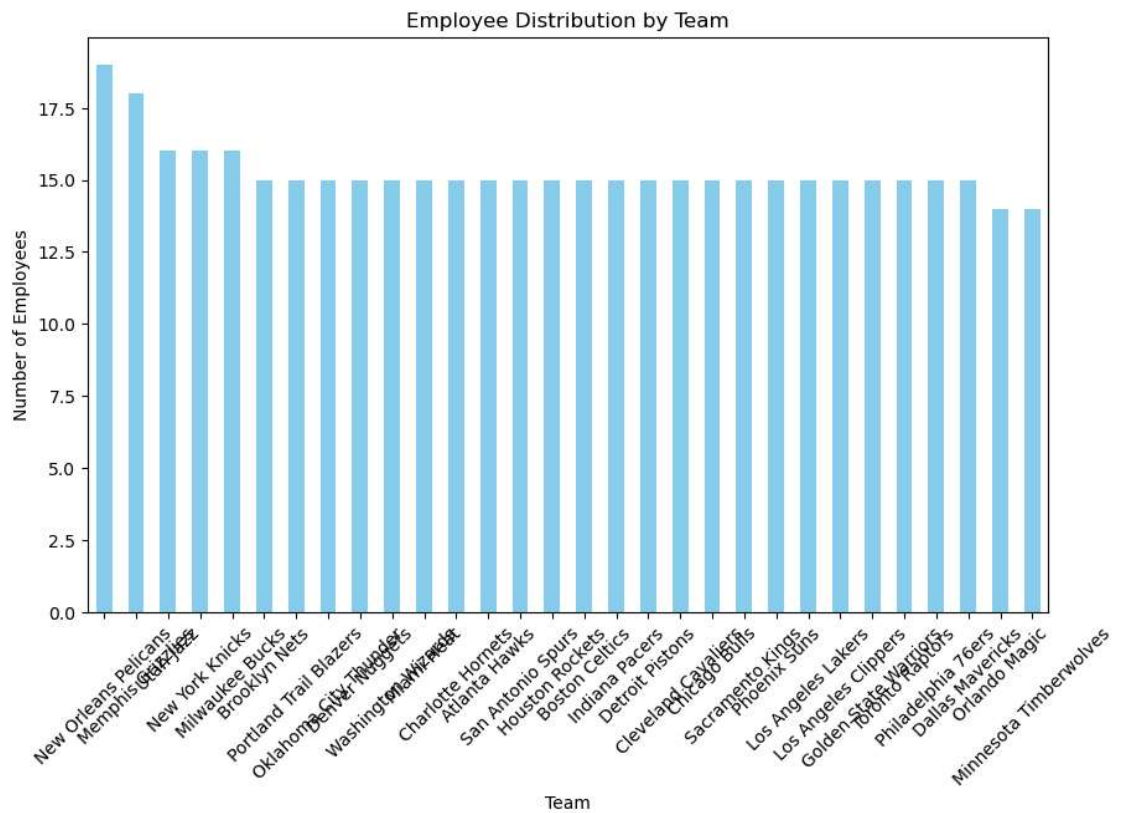
Team
New Orleans Pelicans    4.148472
Memphis Grizzlies       3.930131
Utah Jazz               3.493450
New York Knicks         3.493450
Milwaukee Bucks         3.493450
Brooklyn Nets           3.275109
Portland Trail Blazers   3.275109
Oklahoma City Thunder   3.275109
Denver Nuggets           3.275109
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 [5]: #visualization
import matplotlib.pyplot as plt

plt.figure(figsize=(10, 6))
team_distribution.plot(kind='bar', color='skyblue')
plt.title('Employee Distribution by Team')
plt.xlabel('Team')
plt.ylabel('Number of Employees')
plt.xticks(rotation=45)
plt.show()
```

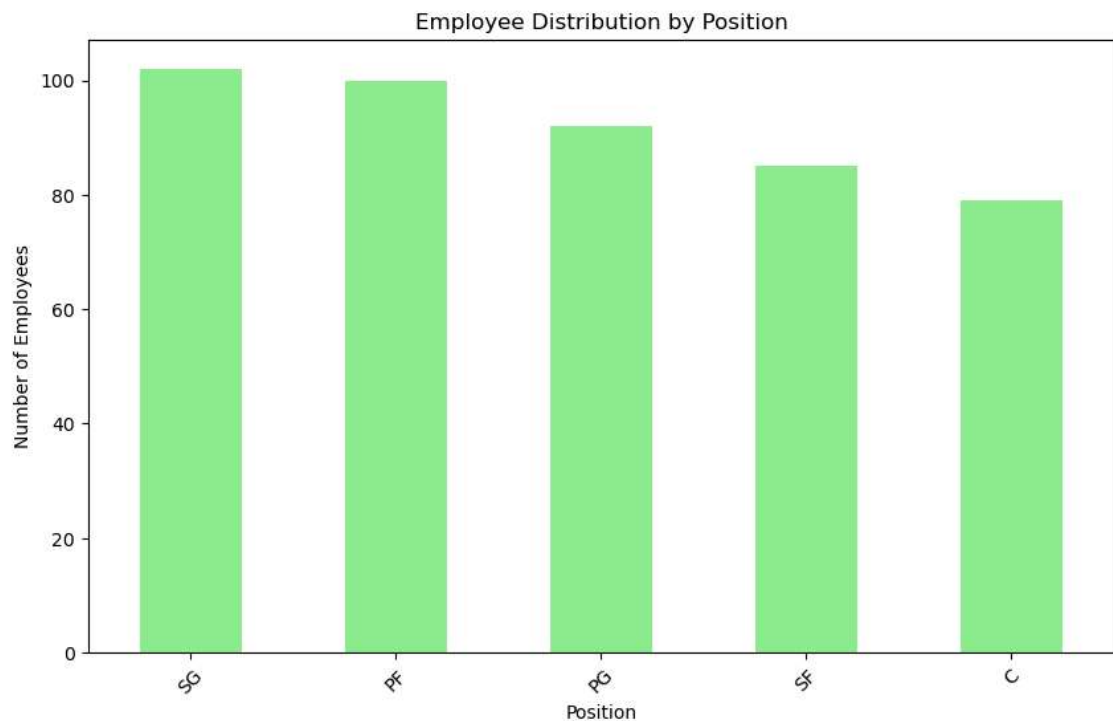


```
In [7]: ▶ #segregation by position
position_distribution = df['Position'].value_counts()

# Display results
print(position_distribution)
```

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

```
In [8]: ▶ #visualisation
plt.figure(figsize=(10, 6))
position_distribution.plot(kind='bar', color='lightgreen')
plt.title('Employee Distribution by Position')
plt.xlabel('Position')
plt.ylabel('Number of Employees')
plt.xticks(rotation=45)
plt.show()
```



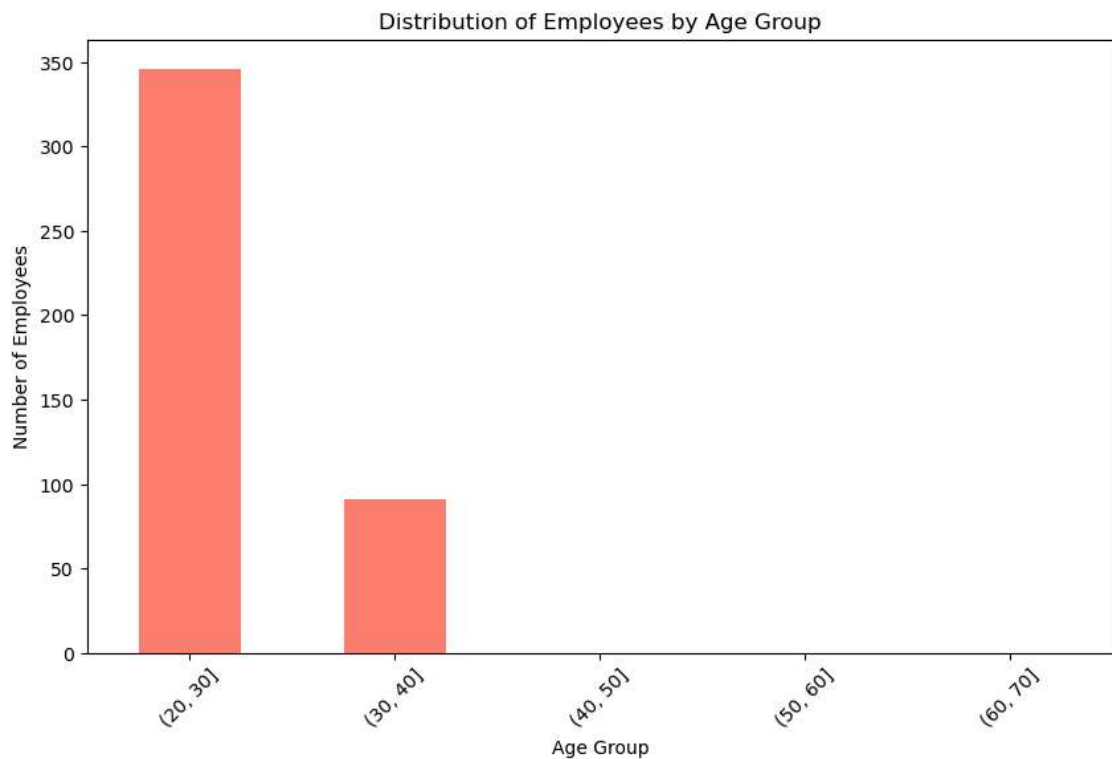
```
In [9]: ▶ #predominant age group
age_bins = [20, 30, 40, 50, 60, 70]
df['Age_group'] = pd.cut(df['Age'], bins=age_bins)

age_group_distribution = df['Age_group'].value_counts()

# Display results
print(age_group_distribution)
```

```
Age_group
(20, 30]    346
(30, 40]     91
(40, 50]      0
(50, 60]      0
(60, 70]      0
Name: count, dtype: int64
```

```
In [10]: ▶ #visualization
plt.figure(figsize=(10, 6))
age_group_distribution.plot(kind='bar', color='salmon')
plt.title('Distribution of Employees by Age Group')
plt.xlabel('Age Group')
plt.ylabel('Number of Employees')
plt.xticks(rotation=45)
plt.show()
```



```
In [11]: #highest salary expenditure
highest_salary_team = df.groupby('Team')['Salary'].sum().idxmax()
highest_salary_position = df.groupby('Position')['Salary'].sum().idxmax()

# Display results
print(f'Highest Salary Team: {highest_salary_team}')
print(f'Highest Salary Position: {highest_salary_position}')
```

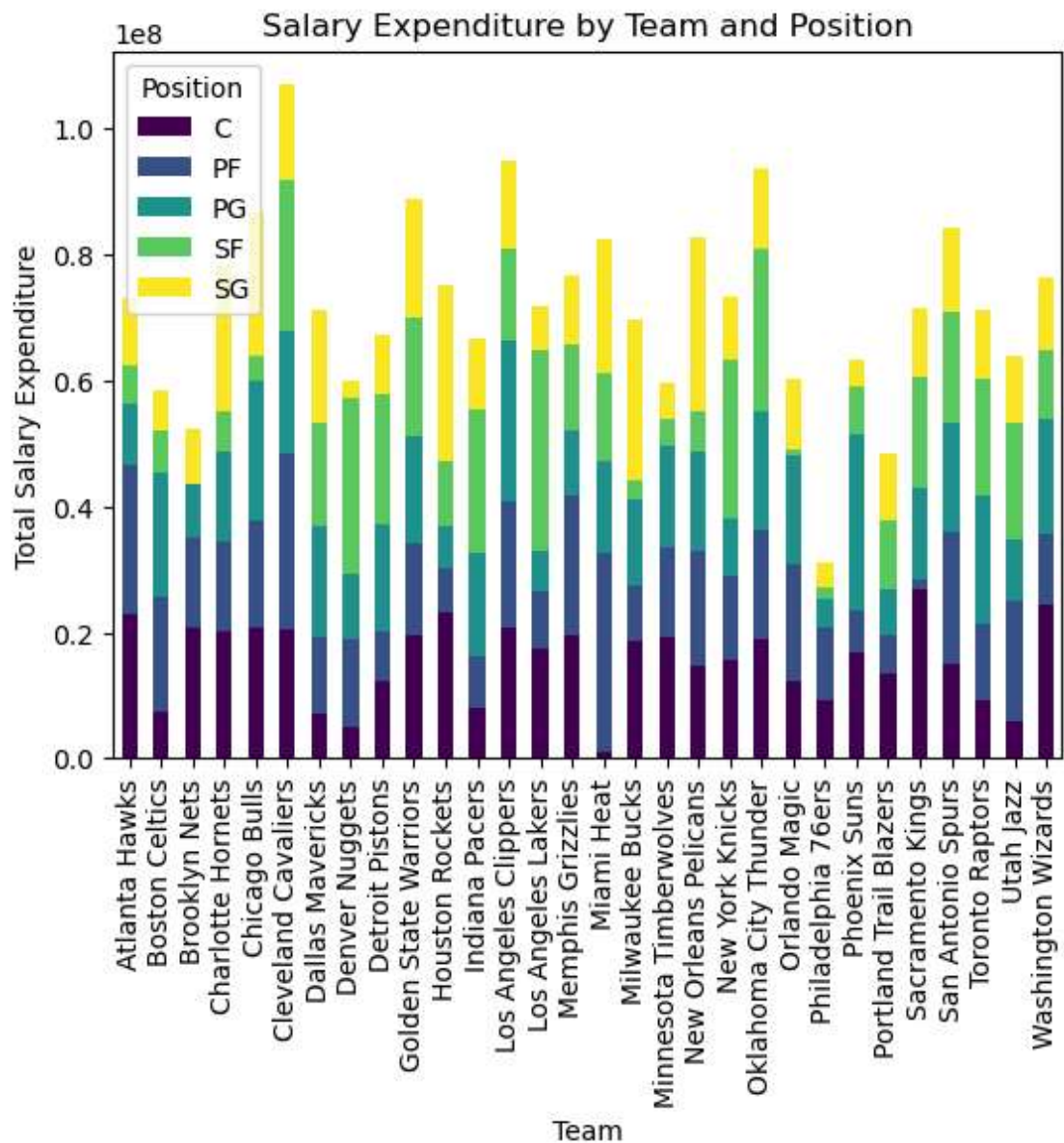
Highest Salary Team: Cleveland Cavaliers

Highest Salary Position: C

```
In [12]: #visualisation
salary_expenditure = df.groupby(['Team', 'Position'])['Salary'].sum().unstack()

plt.figure(figsize=(12, 8))
salary_expenditure.plot(kind='bar', stacked=True, colormap='viridis')
plt.title('Salary Expenditure by Team and Position')
plt.ylabel('Total Salary Expenditure')
plt.xlabel('Team')
plt.legend(title='Position')
plt.show()
```

<Figure size 1200x800 with 0 Axes>

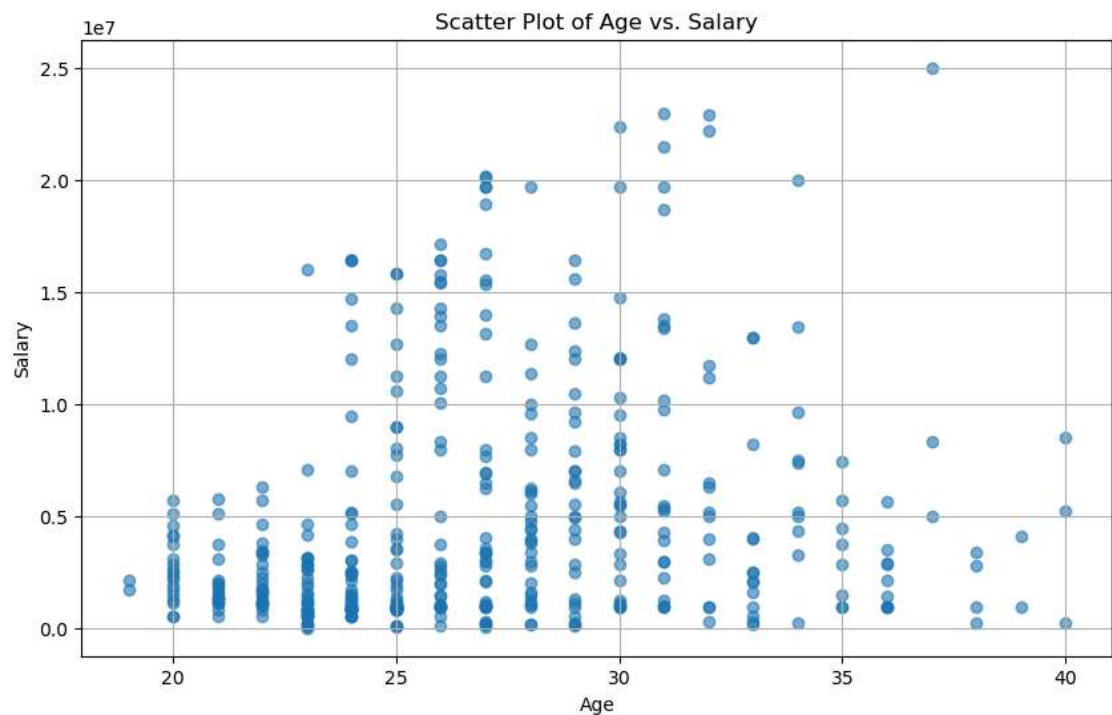



```
In [13]: ▶ #Correlation Between Age and Salary
correlation = df[['Age', 'Salary']].corr().iloc[0, 1]

# Display results
print(f'Correlation between Age and Salary: {correlation}')
```

Correlation between Age and Salary: 0.21400941226570955

```
In [14]: ▶ #visualization
plt.figure(figsize=(10, 6))
plt.scatter(df['Age'], df['Salary'], alpha=0.6)
plt.title('Scatter Plot of Age vs. Salary')
plt.xlabel('Age')
plt.ylabel('Salary')
plt.grid()
plt.show()
```



```
In [15]: ▶ # Insights based on analysis
team_distribution_insight = f"Total employees: {len(df)}. The largest team
position_distribution_insight = f"The most common position is {position_di
predominant_age_group_insight = f"The predominant age group is {age_group_
highest_salary_insight = f"The team with the highest salary expenditure is
correlation_insight = f"The correlation between age and salary is {correla
data_story = f""
Data Story Insights:
1. {team_distribution_insight}
2. {position_distribution_insight}
3. {predominant_age_group_insight}
4. {highest_salary_insight}
5. {correlation_insight}
"""
print(data_story)
```

Data Story Insights:

1. Total employees: 458. The largest team is New Orleans Pelicans with 19 members, constituting 4.15% of the workforce.
2. The most common position is SG with 102 employees.
3. The predominant age group is (20, 30] with 346 employees.
4. The team with the highest salary expenditure is Cleveland Cavaliers and the position with the highest salary is C.
5. The correlation between age and salary is 0.21, indicating a moderate positive relationship.

In []: ▶