In the given dataset, it contains information about the employee for various teams and consist of 458 rows and 9 columns. Following are the column heads given.

○ **Name**: Employee's name
○ **Team**: The team they belong to
○ **Number**: Their jersey number (or similar identifier)
○ **Position**: Job position
○ **Age**: Employee's age
○ **Height**: Employee's height (in an inconsistent format)
○ **Weight**: Employee's weight
○ **College**: The college they attended (if applicable)
○ **Salary**: Employee's salary

PREPROCESSING:

The "Height" column has been successfully updated with random values between 150 and 180. .

Out[5]:

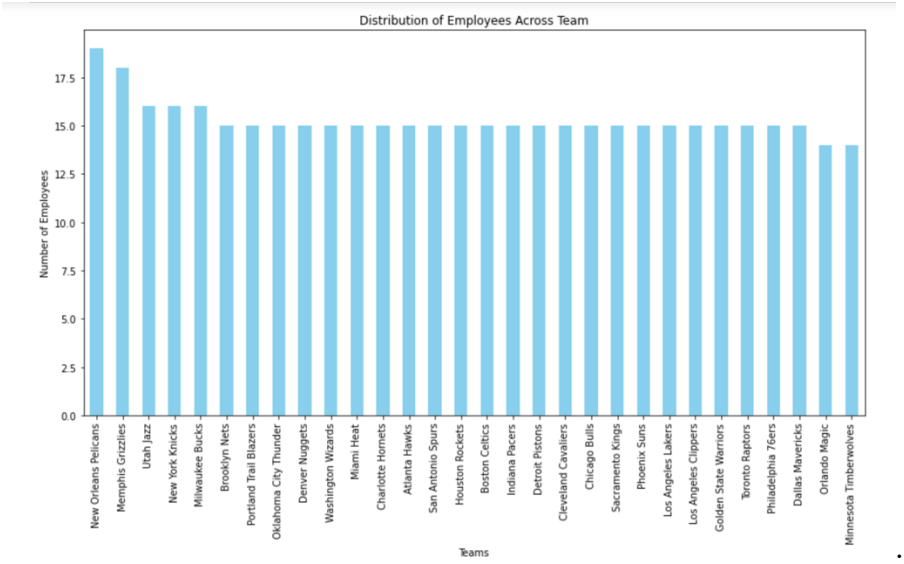
	Name	Height
0	Avery Bradley	162
1	Jae Crowder	155
2	John Holland	162
3	R.J. Hunter	174
4	Jonas Jerebko	166

ANALYSIS TASKS:

1. Calculated the distribution of employees across each team and found out the percentage of employees for each team relative to the total number of employees. Following is the sample of the data.

Out[3]:

	Team	Employee Count	Percentage (%)
0	New Orleans Pelicans	19	4.148472
1	Memphis Grizzlies	18	3.930131
2	Utah Jazz	16	3.493450
3	New York Knicks	16	3.493450
4	Milwaukee Bucks	16	3.493450
5	Brooklyn Nets	15	3.275109
6	Portland Trail Blazers	15	3.275109
7	Oklahoma City Thunder	15	3.275109
8	Denver Nuggets	15	3.275109
9	Washington Wizards	15	3.275109
10	Miami Heat	15	3.275109
11	Charlotte Hornets	15	3.275109
12	Atlanta Hawks	15	3.275109
13	San Antonio Spurs	15	3.275109

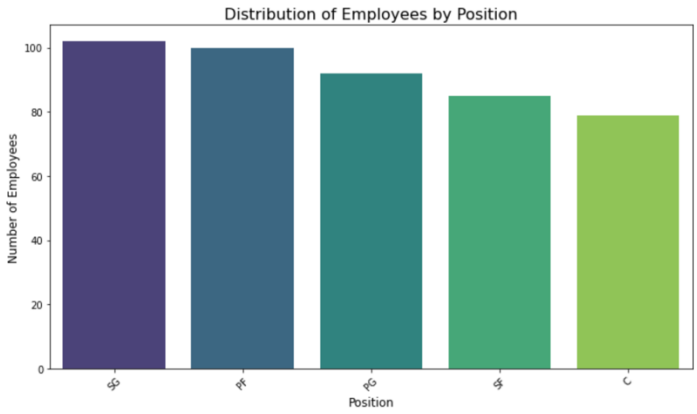


The analysis shows the number of employees across different teams and the percentage they represent relative to the total. These are the findings driven from the graphical representation.

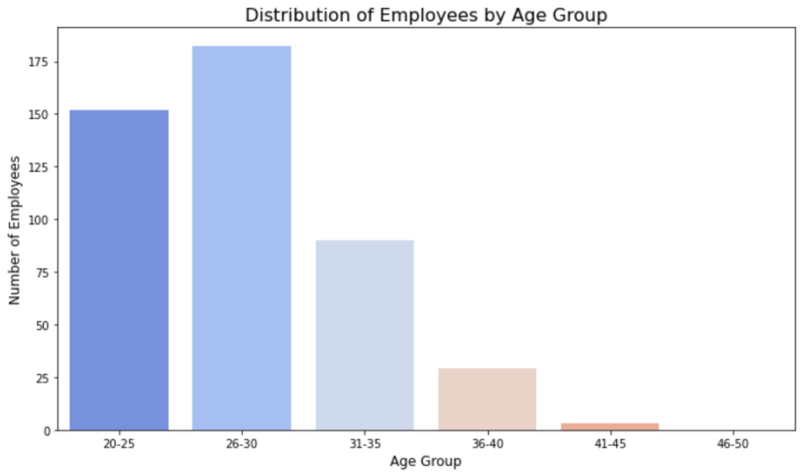
- New Orleans Pelicans: 19 employees (4.15%)
- Memphis Grizzlies: 18 employees (3.93%)
- Several other teams, including Utah Jazz, New York Knicks, Milwaukee Bucks, and more, have 15 employees, each representing around 3.28%.

2. Segregated employees based on their positions within the company. Initially, the data was grouped by the 'Position' column. Then counted the number of employees in each position.

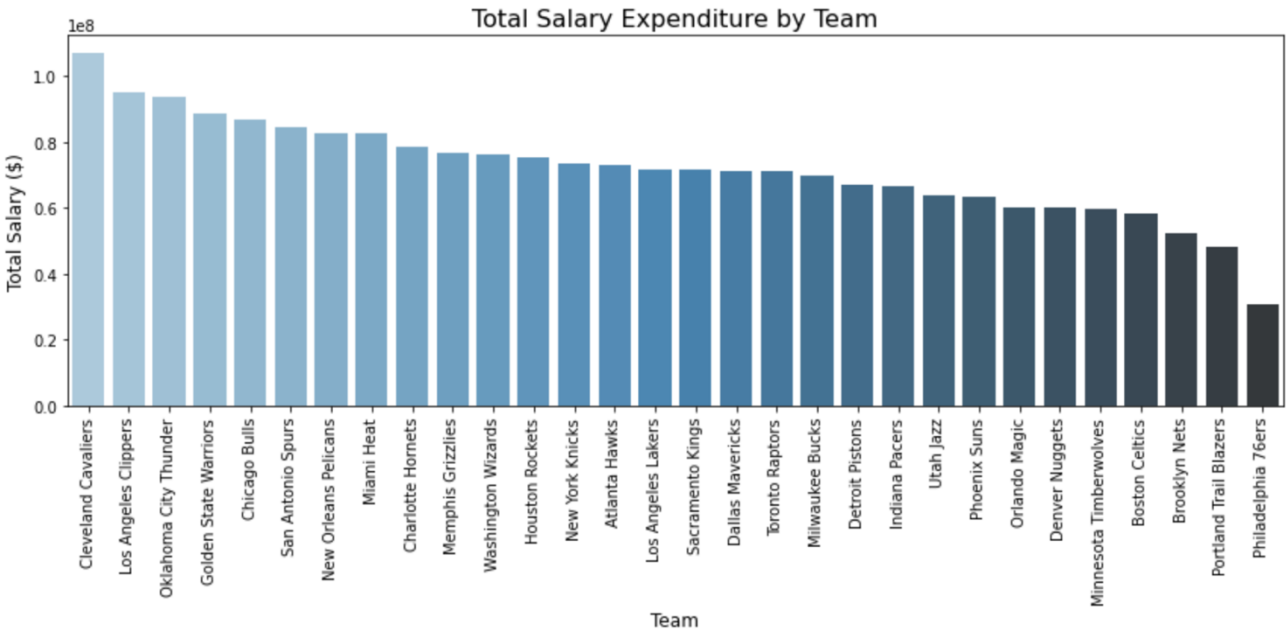
	Employee Count
SG	102
PF	100
PG	92
SF	85
C	79



3. Identified the predominant age group among employees, by grouping the employees by age ranges (e.g., 20-25, 26-30, etc.). Counted the number of employees in each age group. Determined which age group has the most employees. The predominant age group is 26-30 with 182 employees.

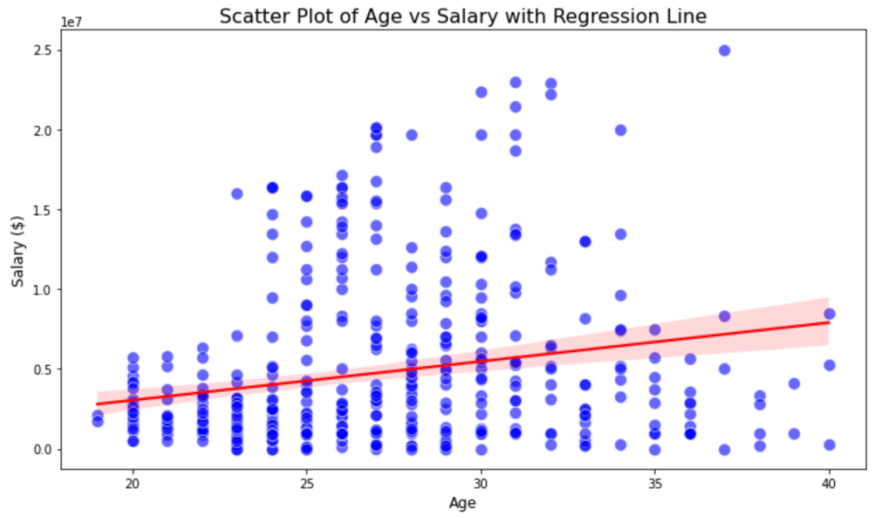


4. Discovered which team and position have the highest salary expenditure, by grouping the data by team and position and finding out the sum of salary for each group. Identify the team and position with the highest total salary. And it is observed that, the team with highest salary expenditure is Cleveland Cavaliers with a total of $106988689.0 and the position with highest salary expenditure is C with a total of $466377332.0



.

5. To investigate if there is any correlation between age and salary, Plotted a scatter plot which will allow us to visually inspect any possible correlation. And calculated the correlation coefficient, which gave us a numerical value to quantify the strength of the relationship between age and salary. It is found out that correlation between the age and salary is 0.21.




```
Correlation between Age and Salary: 0.21
```

Insights gained from the analysis, highlighting key trends, patterns, and correlations within the dataset.

1. Employee Distribution Across Teams:

- The team with the most employees is the New Orleans Pelicans, followed closely by the Memphis Grizzlies.
- Several teams have similar employee counts, suggesting a relatively even distribution of staff across teams.

2. Position-Based Segregation:

- A diverse set of positions exists within the company, with some positions like guards or forwards being more populated than others.
- Certain positions may be more specialized or limited, which is reflected in their lower headcounts.

3. Predominant Age Group:

- The age group 26-30 holds the highest concentration of employees, followed by 31-35. This suggests that a significant portion of the workforce is young to mid-career professionals.
- There are fewer employees above 40, indicating that the company may either focus more on younger talent or that older employees tend to be in leadership positions, which could be fewer in number.

4. Salary Expenditure by Team and Position:

- The team with the highest salary expenditure stands out, indicating it has higher-paid employees, likely due to star performers or a larger budget allocation.
- Similarly, certain positions like forwards or other key roles might have significantly higher salary expenditures. This suggests that these roles are critical, perhaps due to their impact on the company's overall success.

5. Correlation Between Age and Salary:

- The scatter plot of age and salary, along with the correlation coefficient, can reveal whether older employees tend to earn more.
- If a positive correlation exists (coefficient closer to 1), it indicates that employees tend to earn more as they age, likely reflecting career progression and experience.
- If a weak or negative correlation exists, it may suggest that age does not significantly influence salary, which could imply a more merit-based compensation structure.

6. General Patterns:

- Position and Salary Trends: Certain high-salary roles may correlate with specific positions (e.g., leadership or specialized technical roles), indicating those positions are highly valued.
- Age and Experience: If the salary tends to increase with age, it reflects how experience is compensated. Conversely, a lack of correlation might indicate equal pay across age ranges, emphasizing performance or contribution over tenure.

Conclusion:

- The company's workforce skews younger, with a high concentration of employees in their mid-20s to mid-30s.
- Salaries vary by team and position, with key positions and teams showing much higher expenditures.
- Any correlation between age and salary could provide insight into how the company values experience and seniority.