# INSIGHTS DERIVED FROM THE DATASET

Distribution of employees across each team

```
1   #  Determine the distribution of employees across each team
2   team_distribution = df['Team'].value_counts()
3   team_distribution
```

Found the percentage of distribution

```
1   total_employees =  len(df)
2   percentage_distribution = (team_distribution / total_employees)
3   percentage_distribution
```

Classified the employees based on each position

```
1    #print the employees in each position
2    position_distribution = df['Position'].value_counts()
3    position_distribution
```

|          | count |
|----------|-------|
| Position |       |
| SG       | 102   |
| PF       | 100   |
| PG       | 92    |
| SF       | 85    |
| C        | 79    |

Identified the age group among employees

```
1   # Identify age group among employees
2
3   bins = [18, 25, 35, 45, 55, 65]
4   labels = ['18-24', '25-34', '35-44', '45-54', '55-64']
5
6   # Create a new column
7   df['Age_Group'] = pd.cut(df['Age'], bins=bins, labels=labels,
    right=False)
8
9   age_group_distribution = df['Age_Group'].value_counts()
10  age_group_distribution
```

|            | count |
|------------|-------|
| **Age_Group** |    |
| 25-34      | 272   |
| 18-24      | 154   |
| 35-44      | 32    |
| 45-54      | 0     |
| 55-64      | 0     |

Identified which team and position has the highest salary

```
1   #Identify whcich team and position have highest salary
2   highest_salary_team = df.groupby('Team')['Salary'].mean().idxma
3   highest_salary_position = df.groupby('Position')['Salary'].mean
    ()
4   print(f"The higest  salaried team  is {highest_salary_team}")
5   print(f"The highest salaried position is { highest_salary_posit
```

```
The higest  salaried team  is Cleveland Cavaliers
The highest salaried position is C
```

Visually represented the correlation between Age and Salary

Age vs Salary