In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sn
```

In [3]:
```
df=pd.read_csv(r"C:\Users\neeth\Downloads\myexcel - myexcel.csv (1).csv")
df
```

Out[3]:

	Name	Team	Number	Position	Age	Height	Weight	College	Salary
0	Avery Bradley	Boston Celtics	0	PG	25	06-Feb	180	Texas	7730337.0
1	Jae Crowder	Boston Celtics	99	SF	25	06-Jun	235	Marquette	6796117.0
2	John Holland	Boston Celtics	30	SG	27	06-May	205	Boston University	NaN
3	R.J. Hunter	Boston Celtics	28	SG	22	06-May	185	Georgia State	1148640.0
4	Jonas Jerebko	Boston Celtics	8	PF	29	06-Oct	231	NaN	5000000.0
...	...	...	...	...	...	...	...	...	...
453	Shelvin Mack	Utah Jazz	8	PG	26	06-Mar	203	Butler	2433333.0
454	Raul Neto	Utah Jazz	25	PG	24	06-Jan	179	NaN	900000.0
455	Tibor Pleiss	Utah Jazz	21	C	26	07-Mar	256	NaN	2900000.0
456	Jeff Withey	Utah Jazz	24	C	26	7-0	231	Kansas	947276.0
457	Priyanka	Utah Jazz	34	C	25	07-Mar	231	Kansas	947276.0

458 rows × 9 columns

Replace the 'Height' column with random numbers between 150 and 180

In [12]:
```
df['Height']=np.random.randint(150,181,size=len(df))#by using the random randint function through
print(df)# displayed the updated data
```

```
              Name            Team  Number Position  Age  Height  Weight  \
0    Avery Bradley  Boston Celtics       0       PG   25     175     180
1      Jae Crowder  Boston Celtics      99       SF   25     160     235
2     John Holland  Boston Celtics      30       SG   27     169     205
3      R.J. Hunter  Boston Celtics      28       SG   22     153     185
4    Jonas Jerebko  Boston Celtics       8       PF   29     155     231
..             ...             ...     ...      ...  ...     ...     ...
453   Shelvin Mack       Utah Jazz       8       PG   26     158     203
454      Raul Neto       Utah Jazz      25       PG   24     167     179
455   Tibor Pleiss       Utah Jazz      21        C   26     179     256
456    Jeff Withey       Utah Jazz      24        C   26     153     231
457       Priyanka       Utah Jazz      34        C   25     172     231

               College     Salary
0                Texas  7730337.0
1            Marquette  6796117.0
2    Boston University        NaN
3        Georgia State  1148640.0
4                  NaN  5000000.0
..                 ...        ...
453            Butler  2433333.0
454               NaN   900000.0
455               NaN  2900000.0
456            Kansas   947276.0
457            Kansas   947276.0

[458 rows x 9 columns]
```

In [14]:
```
df
```

Out[14]:

	Name	Team	Number	Position	Age	Height	Weight	College	Salary
0	Avery Bradley	Boston Celtics	0	PG	25	175	180	Texas	7730337.0
1	Jae Crowder	Boston Celtics	99	SF	25	160	235	Marquette	6796117.0
2	John Holland	Boston Celtics	30	SG	27	169	205	Boston University	NaN
3	R.J. Hunter	Boston Celtics	28	SG	22	153	185	Georgia State	1148640.0
4	Jonas Jerebko	Boston Celtics	8	PF	29	155	231	NaN	5000000.0
...	...	...	...	...	...	...	...	...	...
453	Shelvin Mack	Utah Jazz	8	PG	26	158	203	Butler	2433333.0
454	Raul Neto	Utah Jazz	25	PG	24	167	179	NaN	900000.0
455	Tibor Pleiss	Utah Jazz	21	C	26	179	256	NaN	2900000.0
456	Jeff Withey	Utah Jazz	24	C	26	153	231	Kansas	947276.0
457	Priyanka	Utah Jazz	34	C	25	172	231	Kansas	947276.0

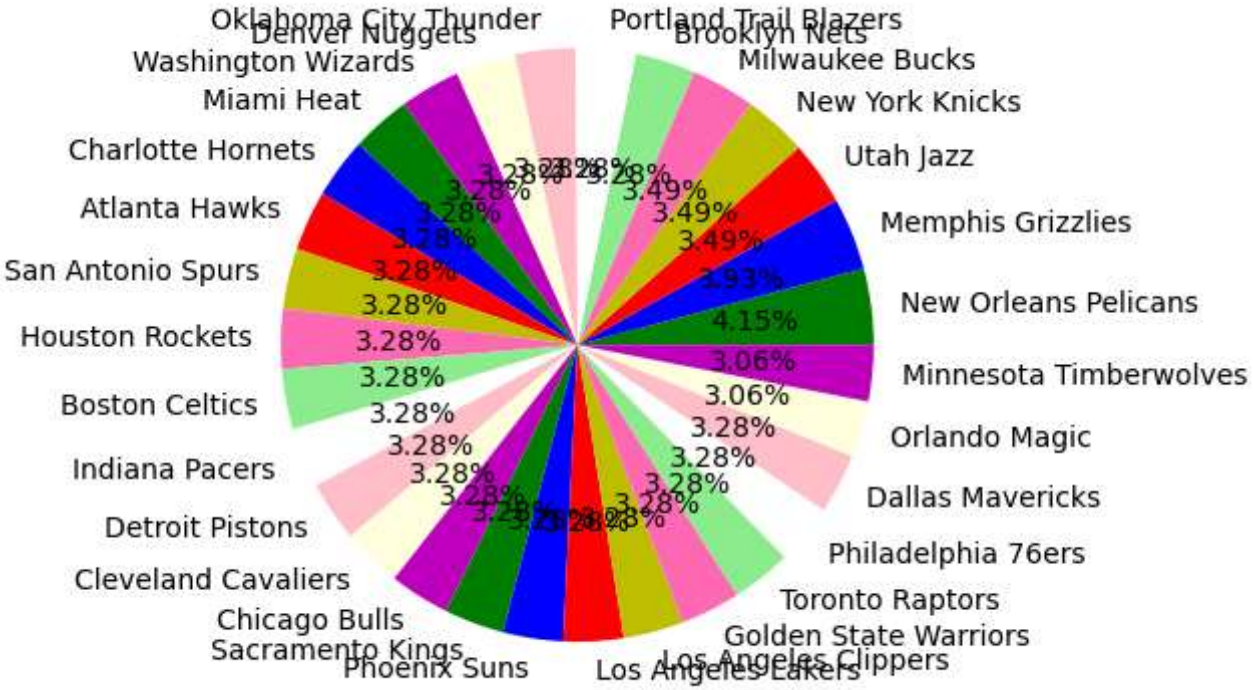
458 rows × 9 columns

1.Determine the distribution of employees across each team and calculate the percentage split relative to the total number of employees

In [17]:
```
emp_distri_team = df['Team'].value_counts() #to get the number of employees across each team.
emp_distri_team

total_no_emp=len(df) #to get the total no_of_employees
total_no_emp

percentage_split=(emp_distri_team/len(df))* 100 # to get the percentage split of employees in ea
percentage_split

Team_data= pd.DataFrame({
    'Team':emp_distri_team.index,
    'Number of Employees':emp_distri_team.values,
    'Percentage':percentage_split.values
})
print(Team_data)
```

	Team	Number of Employees	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
14	Houston Rockets	15	3.275109
15	Boston Celtics	15	3.275109
16	Indiana Pacers	15	3.275109
17	Detroit Pistons	15	3.275109
18	Cleveland Cavaliers	15	3.275109
19	Chicago Bulls	15	3.275109
20	Sacramento Kings	15	3.275109
21	Phoenix Suns	15	3.275109
22	Los Angeles Lakers	15	3.275109
23	Los Angeles Clippers	15	3.275109
24	Golden State Warriors	15	3.275109
25	Toronto Raptors	15	3.275109
26	Philadelphia 76ers	15	3.275109
27	Dallas Mavericks	15	3.275109
28	Orlando Magic	14	3.056769
29	Minnesota Timberwolves	14	3.056769

In [19]:
```
# pie chart visualization of percentage split of employees across each team
My_colors=('g','b','r','y','hotpink','lightgreen','white','pink','lightyellow','m')
plt.pie(Team_data['Percentage'],labels=Team_data['Team'],autopct='%1.2f%%',colors=My_colors)
plt.show()
##INSIGHT= The percentage split of employees across teams shows that "New Orleans Pelicans" has
```



2. Segregate employees based on their positions within the company

In [22]:
```
Positions=df.groupby('Position') #to get the complete employees segregated list based on their po
for x in Positions:
    print(x)
```

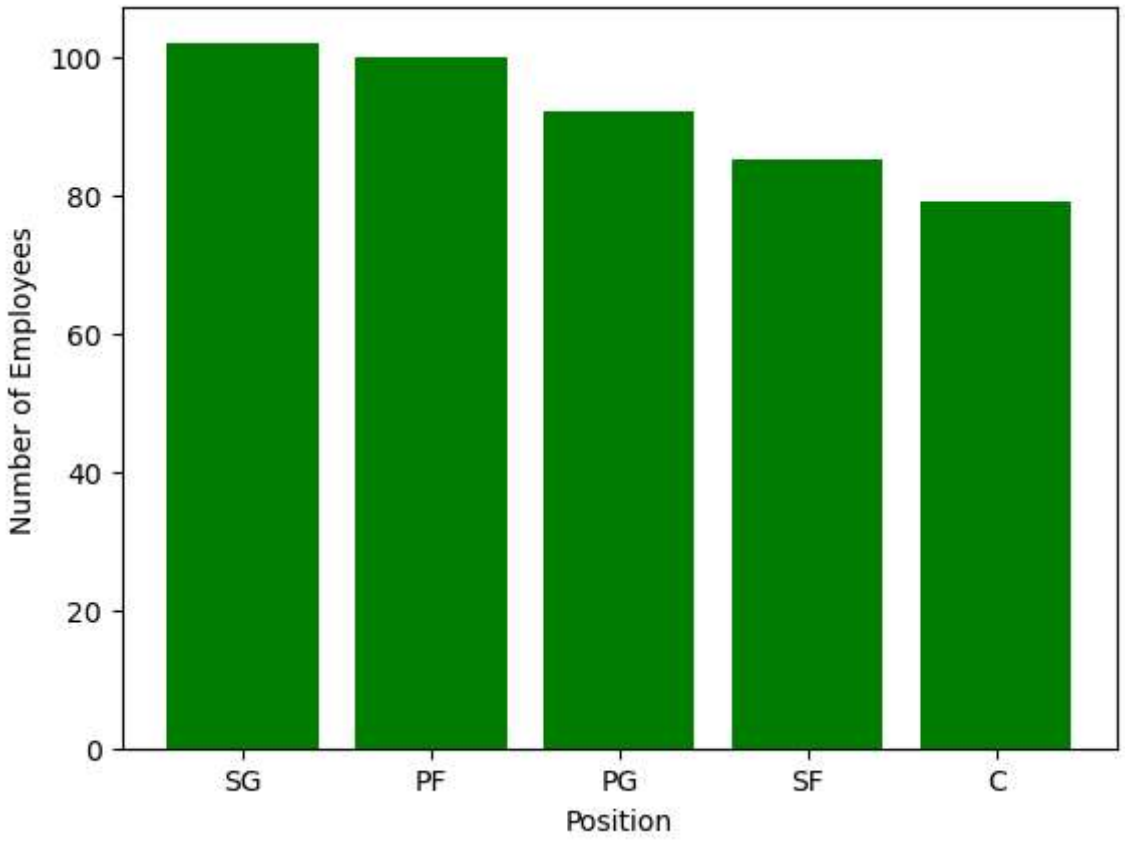
```
('C',                Name                   Team  Number Position  Age  Height  \
7       Kelly Olynyk          Boston Celtics      41        C   25     165
10   Jared Sullinger          Boston Celtics       7        C   24     161
14      Tyler Zeller          Boston Celtics      44        C   26     174
23       Brook Lopez           Brooklyn Nets      11        C   28     167
27        Henry Sims           Brooklyn Nets      14        C   26     154
..               ...                     ...     ...      ...  ...     ...
439    Mason Plumlee  Portland Trail Blazers      24        C   26     164
447      Rudy Gobert               Utah Jazz      27        C   23     164
455     Tibor Pleiss               Utah Jazz      21        C   26     179
456      Jeff Withey               Utah Jazz      24        C   26     153
457         Priyanka               Utah Jazz      34        C   25     172

     Weight         College      Salary
7       238         Gonzaga   2165160.0
10      260      Ohio State   2569260.0
14      253  North Carolina   2616975.0
23      275        Stanford  19689000.0
27      248      Georgetown    947276.0
..      ...             ...         ...
439     235            Duke   1415520.0
447     245             NaN   1175880.0
455     256             NaN   2900000.0
456     231          Kansas    947276.0
457     231          Kansas    947276.0

[79 rows x 9 columns])
('PF',                 Name                   Team  Number Position  Age  Height  \
4       Jonas Jerebko          Boston Celtics       8       PF   29     155
5        Amir Johnson          Boston Celtics      90       PF   29     159
6       Jordan Mickey          Boston Celtics      55       PF   21     165
24   Chris McCullough           Brooklyn Nets       1       PF   21     156
25       Willie Reed           Brooklyn Nets      33       PF   26     180
..               ...                     ...     ...      ...  ...     ...
435   Meyers Leonard  Portland Trail Blazers      11       PF   24     151
441      Noah Vonleh  Portland Trail Blazers      21       PF   20     154
442     Trevor Booker               Utah Jazz      33       PF   28     157
446    Derrick Favors               Utah Jazz      15       PF   24     155
452        Trey Lyles               Utah Jazz      41       PF   20     153

     Weight       College      Salary
4       231           NaN   5000000.0
5       240           NaN  12000000.0
6       235           LSU   1170960.0
24      200      Syracuse   1140240.0
25      220   Saint Louis    947276.0
..      ...           ...         ...
435     245      Illinois   3075880.0
441     240       Indiana   2637720.0
442     228       Clemson   4775000.0
446     265  Georgia Tech  12000000.0
452     234      Kentucky   2239800.0

[100 rows x 9 columns])
('PG',               Name                   Team  Number Position  Age  Height  \
0    Avery Bradley          Boston Celtics       0       PG   25     175
8     Terry Rozier          Boston Celtics      12       PG   22     160
9     Marcus Smart          Boston Celtics      36       PG   22     173
11   Isaiah Thomas          Boston Celtics       4       PG   27     180
19    Jarrett Jack           Brooklyn Nets       2       PG   32     169
..             ...                     ...     ...      ...  ...     ...
440  Brian Roberts  Portland Trail Blazers       2       PG   30     172
```

```
443     Trey Burke              Utah Jazz       3       PG   23     180
445     Dante Exum              Utah Jazz      11       PG   20     162
453   Shelvin Mack              Utah Jazz       8       PG   26     158
454      Raul Neto              Utah Jazz      25       PG   24     167

     Weight         College     Salary
0       180           Texas  7730337.0
8       190      Louisville  1824360.0
9       220  Oklahoma State  3431040.0
11      185      Washington  6912869.0
19      200    Georgia Tech  6300000.0
..      ...             ...        ...
440     173          Dayton  2854940.0
443     191        Michigan  2658240.0
445     190             NaN  3777720.0
453     203          Butler  2433333.0
454     179             NaN   900000.0

[92 rows x 9 columns])
('SF',                         Name                    Team  Number Position  Age  \
1               Jae Crowder          Boston Celtics      99       SF   25
32   Thanasis Antetokounmpo         New York Knicks      43       SF   23
33          Carmelo Anthony         New York Knicks       7       SF   32
35         Cleanthony Early         New York Knicks      11       SF   25
42             Lance Thomas         New York Knicks      42       SF   28
..                      ...                     ...     ...      ...  ...
428         Al-Farouq Aminu  Portland Trail Blazers       8       SF   25
432        Maurice Harkless  Portland Trail Blazers       4       SF   23
448          Gordon Hayward              Utah Jazz      20       SF   26
450              Joe Ingles              Utah Jazz       2       SF   28
451           Chris Johnson              Utah Jazz      23       SF   26

     Height  Weight        College      Salary
1       160     235      Marquette   6796117.0
32      152     205            NaN     30888.0
33      156     240       Syracuse  22875000.0
35      156     210  Wichita State    845059.0
42      159     235           Duke   1636842.0
..      ...     ...            ...         ...
428     168     215    Wake Forest   8042895.0
432     179     215     St. John's   2894059.0
448     158     226         Butler  15409570.0
450     178     226            NaN   2050000.0
451     158     206         Dayton    981348.0

[85 rows x 9 columns])
('SG',                  Name                    Team  Number Position  Age  Height  \
2        John Holland          Boston Celtics      30       SG   27     169
3         R.J. Hunter          Boston Celtics      28       SG   22     153
12        Evan Turner          Boston Celtics      11       SG   27     179
13        James Young          Boston Celtics      13       SG   20     154
15   Bojan Bogdanovic           Brooklyn Nets      44       SG   27     156
..                ...                     ...     ...      ...  ...     ...
433  Gerald Henderson  Portland Trail Blazers       9       SG   28     161
437     C.J. McCollum  Portland Trail Blazers       3       SG   24     164
438      Luis Montero  Portland Trail Blazers      44       SG   23     167
444        Alec Burks              Utah Jazz      10       SG   24     177
449       Rodney Hood              Utah Jazz       5       SG   23     159

     Weight            College     Salary
2       205  Boston University        NaN
3       185      Georgia State  1148640.0
```

```
12      220         Ohio State  3425510.0
13      215           Kentucky  1749840.0
15      216                NaN  3425510.0
..      ...                ...        ...
433     215               Duke  6000000.0
437     200             Lehigh  2525160.0
438     185     Westchester CC   525093.0
444     214           Colorado  9463484.0
449     206               Duke  1348440.0

[102 rows x 9 columns])
```

visualization and insight

In [25]:
```
#Bar plot visualization of 'Employee Distribution by Position'
Team_Positions=df['Position'].value_counts() # get the totalcount of respected positions
Team_Positions

plt.xlabel('Position')
plt.ylabel('Number of Employees')

plt.bar(Team_Positions.index,Team_Positions.values,color='green')
plt.show()

##INSIGHT:The position-wise segregation shows that "SG" is the most favourable position,followed
```



3. Identify the predominant age group among employees.¶

In [28]:
```
# defining bins and labels
bins = [10,20,30,40,50]
labels = ['<20','20-30','30-40','40-50']
## creating an Age_Group column
```

```
df['Age_Group'] = pd.cut(df['Age'], bins=[10,20,30,40,50],labels=['<20','20-30','30-40','40-50']
(df['Age_Group'])
```

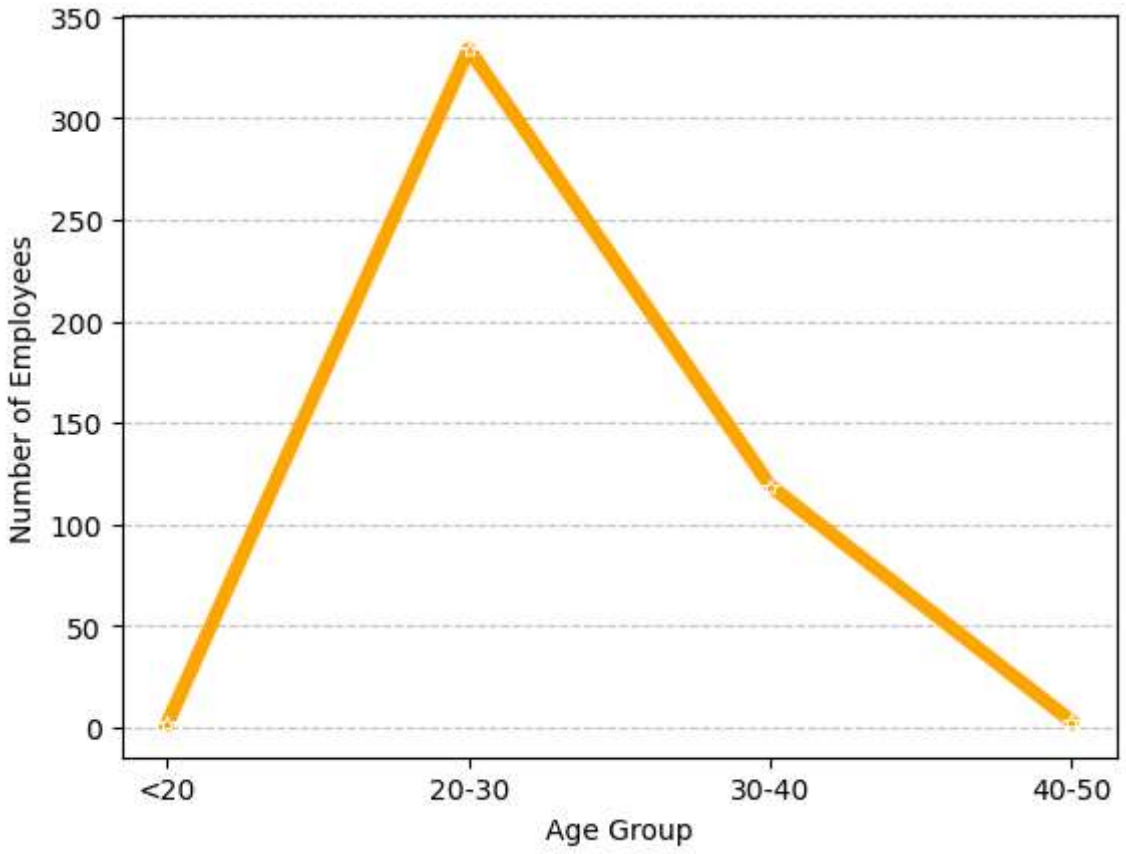
Out[28]:
```
0      20-30
1      20-30
2      20-30
3      20-30
4      20-30
       ...
453    20-30
454    20-30
455    20-30
456    20-30
457    20-30
Name: Age_Group, Length: 458, dtype: category
Categories (4, object): ['<20' < '20-30' < '30-40' < '40-50']
```

In [32]:
```
#find the total count correponding to each age group
Age_Group_distribution = df['Age_Group'].value_counts()
Age_Group_distribution
 ##predominant age group among employees
predominant_age_group=Age_Group_distribution.idxmax()
predominant_age_group
print(f"The predominant_age_group is:{predominant_age_group}")
```

```
The predominant_age_group is:20-30
```

visualization and insight

In [52]:
```
Age_Group_distribution = df['Age_Group'].value_counts()
Age_Group_distribution
## line plot visualization of age group distribution

plt.xlabel('Age Group')
plt.ylabel('Number of Employees')

sn.lineplot(x=Age_Group_distribution.index, y=Age_Group_distribution.values, marker='*', linewid

plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.show()
## INSIGHT:The predominant age group among employees is 26-30, followed by 18-25, showing that th
```



Discover which team and position have the highest salary expenditure.

In [155…

```
## Grouping data by 'Team'  to calculate total salary
Team_salary = df.groupby('Team')['Salary'].sum().sort_values(ascending=False)
Team_salary
Highest_Team_salary=Team_salary.idxmax()
Highest_Team_salary_value=Team_salary.max()
Highest_Team_salary_value
#Grouping data by 'Position' to calculate total salary
Position_salary = df.groupby('Position')['Salary'].sum().sort_values(ascending=False)
Position_salary
Highest_Position_salary=Position_salary.idxmax()
Highest_Position_salary_value=Position_salary.max()
Highest_Position_salary_value
#combine the above 2 to get the result
Result={
    'Highest_Team_salary':(Highest_Team_salary,Highest_Team_salary_value),
    'Highest_Position_salary':(Highest_Position_salary,Highest_Position_salary_value)
}
Result
```

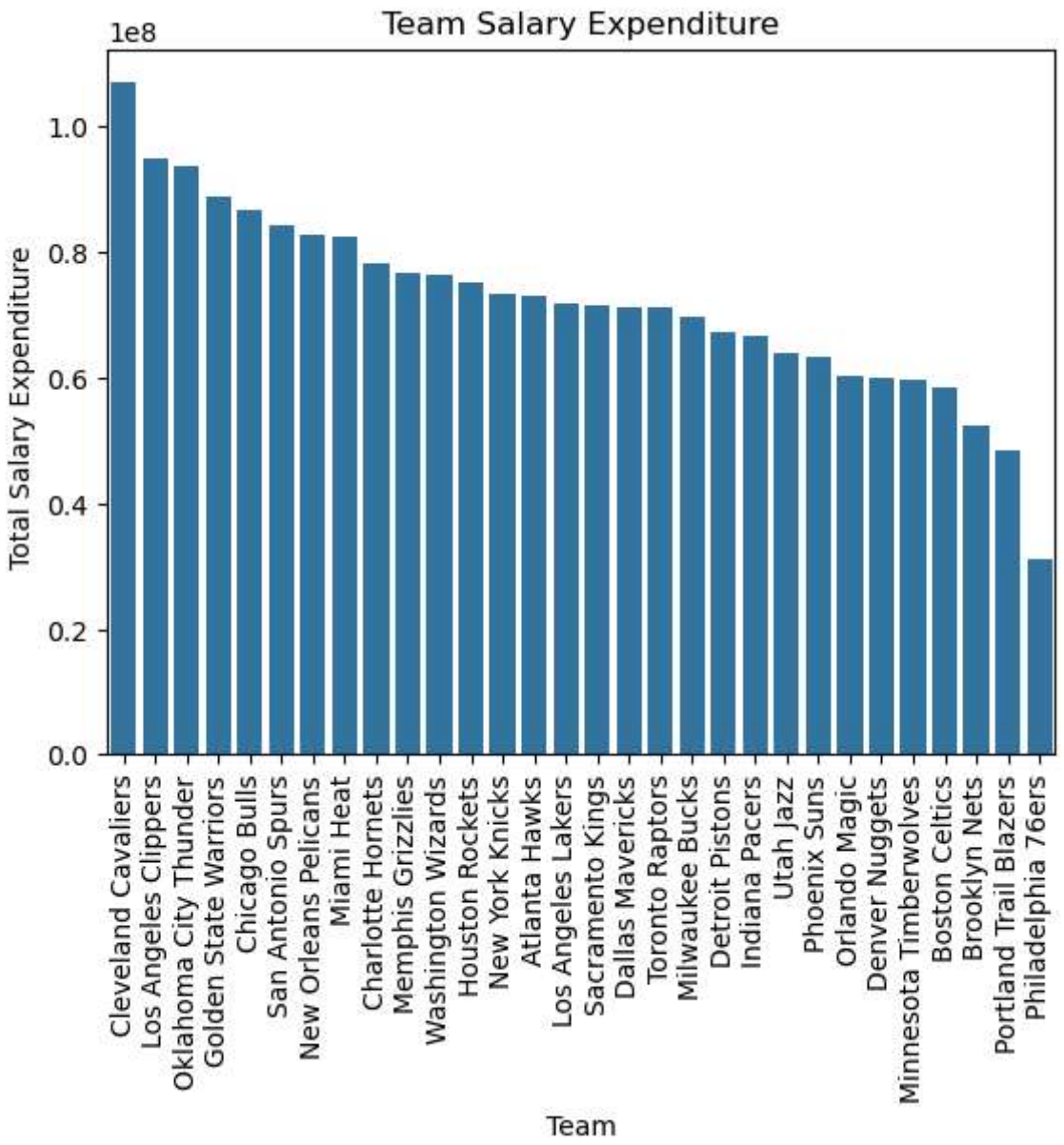
Out[155…

```
{'Highest_Team_salary': ('Cleveland Cavaliers', 106988689.0),
 'Highest_Position_salary': ('C', 466377332.0)}
```

In [149…

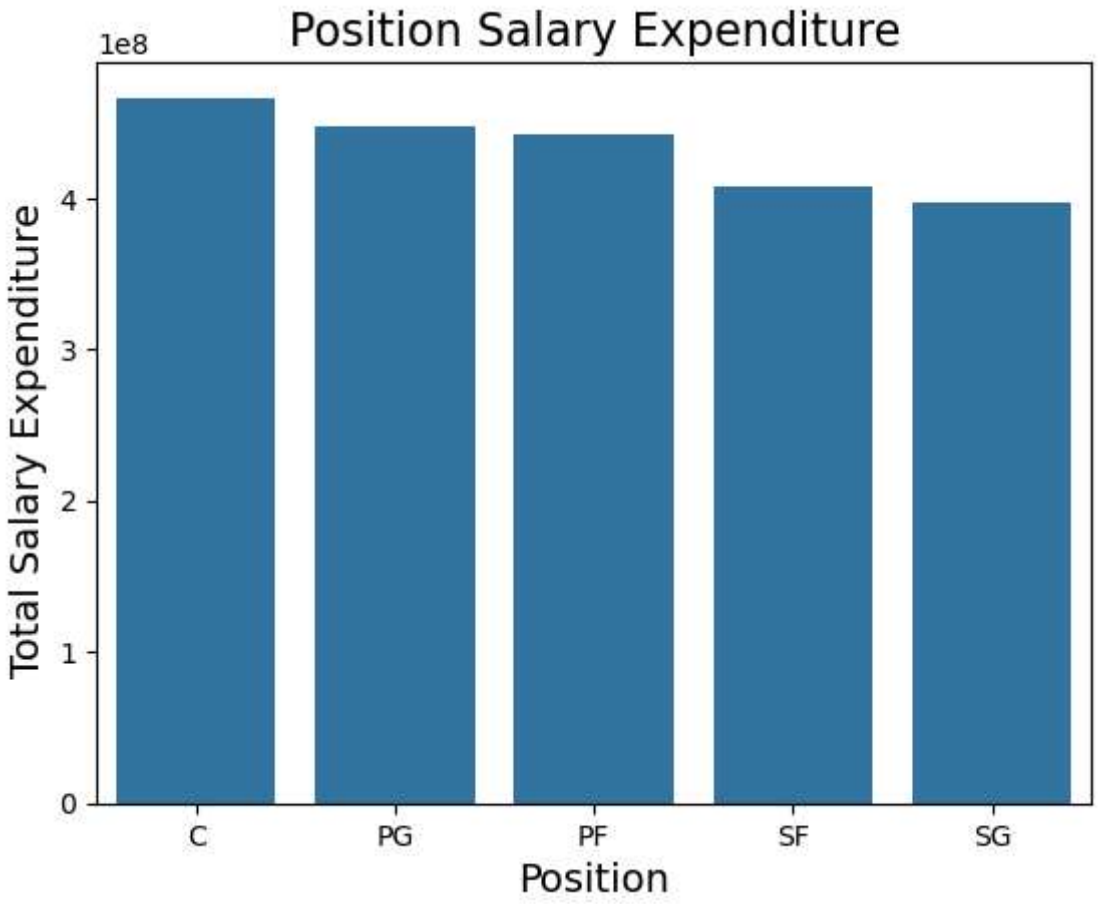
```
##Team Salary Expenditure: A bar chart highlights the distribution of total salary expenditure ac
sn.barplot(
    x=Team_salary.index,
    y=Team_salary.values,
)
plt.title('Team Salary Expenditure')
plt.xlabel('Team')
plt.ylabel('Total Salary Expenditure ')
```

```
plt.xticks(rotation=90)

plt.show()

##INSIGHT:Team with the highest salary expenditure is the Cleveland Cavaliers lead with a total s
```



In [151…

```
## bar chart shows the distribution of salary expenditure by position, with (C position leading.
sn.barplot(
    x=Position_salary.index,
    y=Position_salary.values,
)
plt.title('Position Salary Expenditure', fontsize=16)
plt.xlabel('Position', fontsize=14)
plt.ylabel('Total Salary Expenditure ', fontsize=14)

plt.show()
##INSIGHT: C position has the highest salary expenditure at 466,377,332.
```

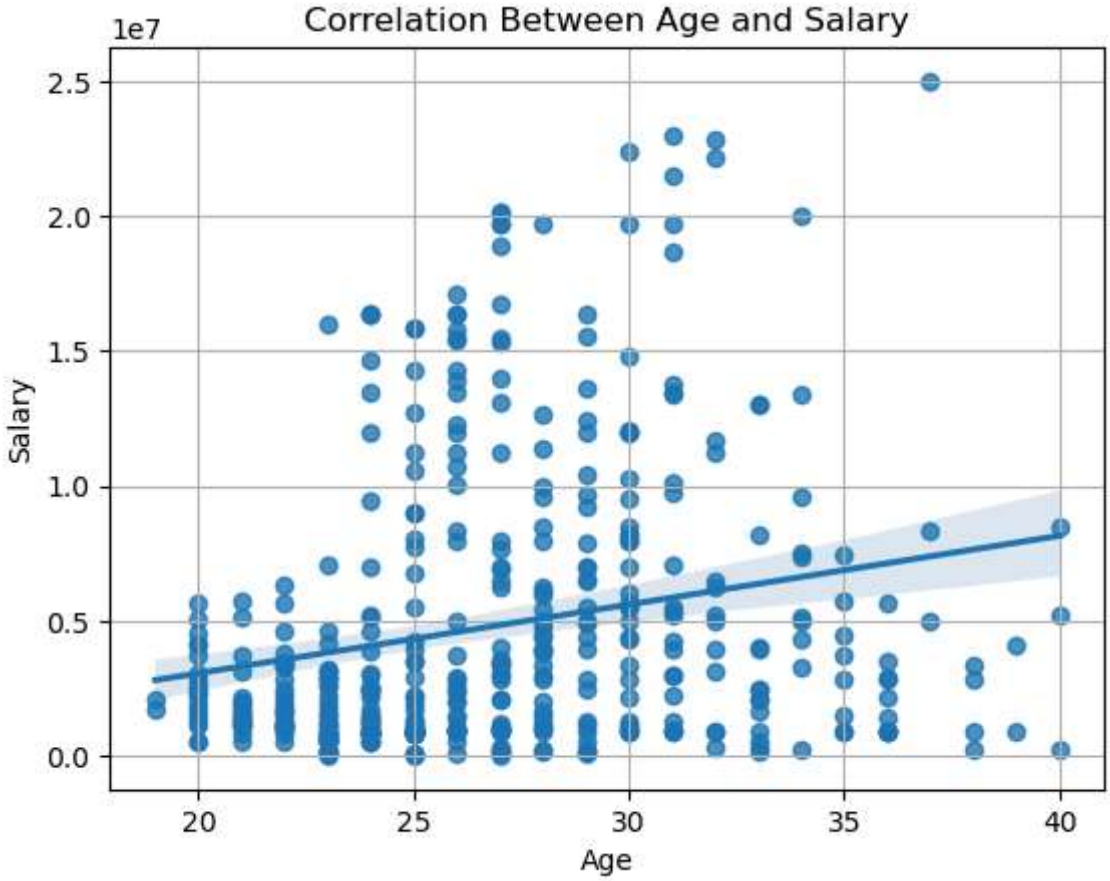


5. Investigate if there's any correlation between age and salary, and represent it visually.

In [187…
```
#check for correlation between 'Age' and 'Salary'
correlation=df[[ 'Age','Salary']].corr().iloc[0, 1]
correlation
```

Out[187…
```
0.21400941226570955
```

In [191…
```
## Scatter plot with trendline

sn.regplot(x='Age', y='Salary', data=df)
plt.xlabel('Age')
plt.ylabel('Salary')

plt.title('Correlation Between Age and Salary')
plt.grid(True)
plt.show()
##INSIGHT: A weak positive correlation 0.21 between age and salary. This indicates that salary te
```


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 [ ]: