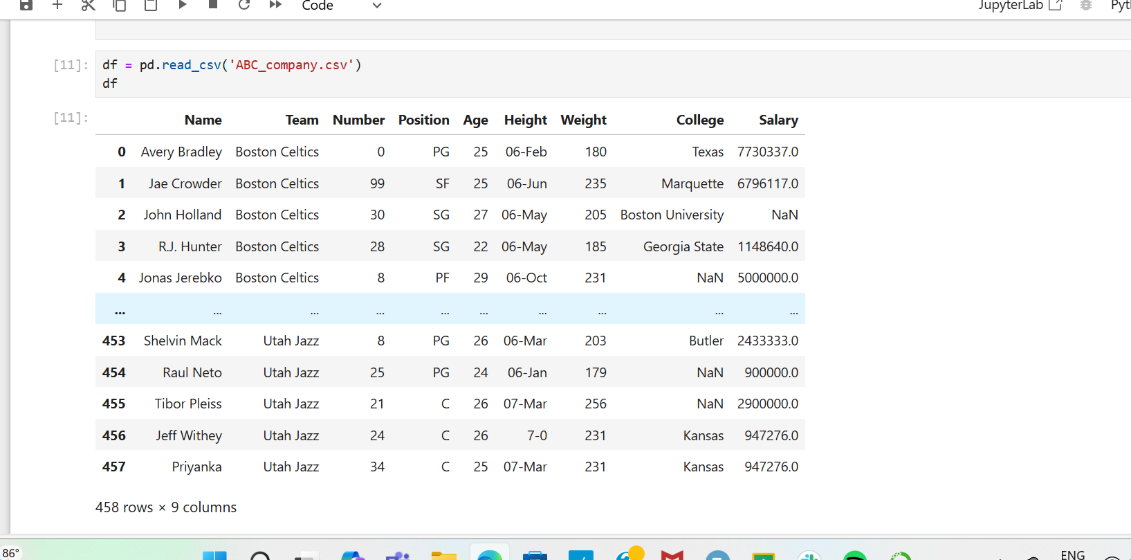
**ABC COMPANY DATA ANALYSIS**

import pandas as pd

import numpy as np

df = pd.read\_csv('ABC\_company.csv')

df



df = pd.read\_csv('ABC\_company.csv')

df

df['Height']=np.random.randint(150,181,df.shape[0])

print(df['Height'].describe())

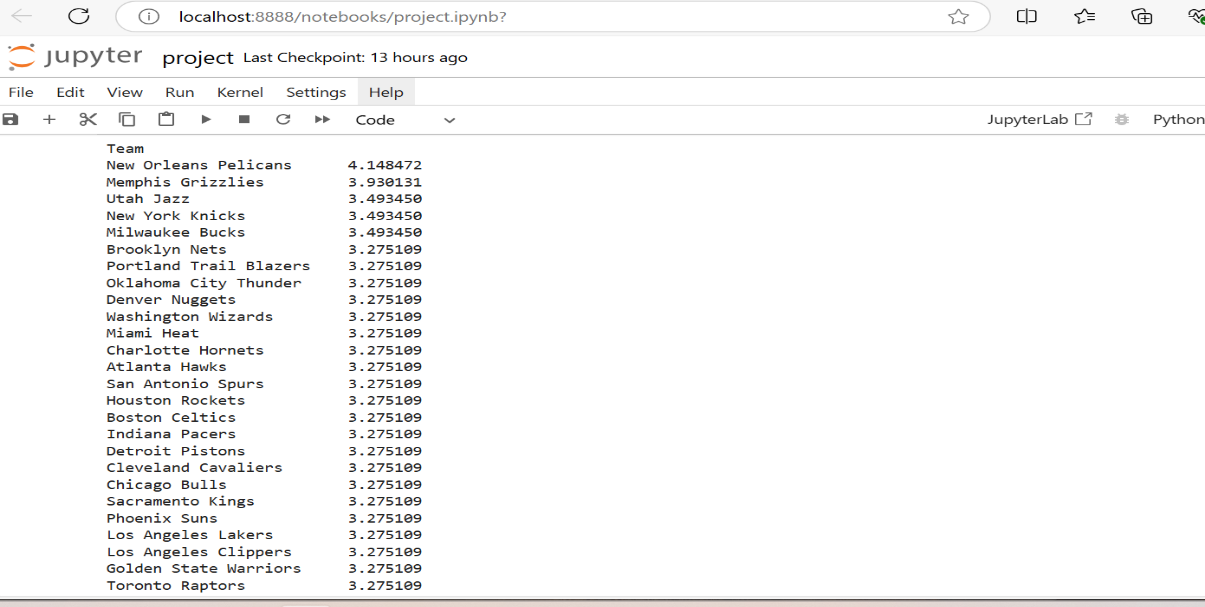


1.Determine the distribution of employees across each team and calculate the percentage split relative to the total number of employees

Team\_dist= df['Team'].value\_counts()

Team\_percentage=(Team\_dist/len(df))\*100

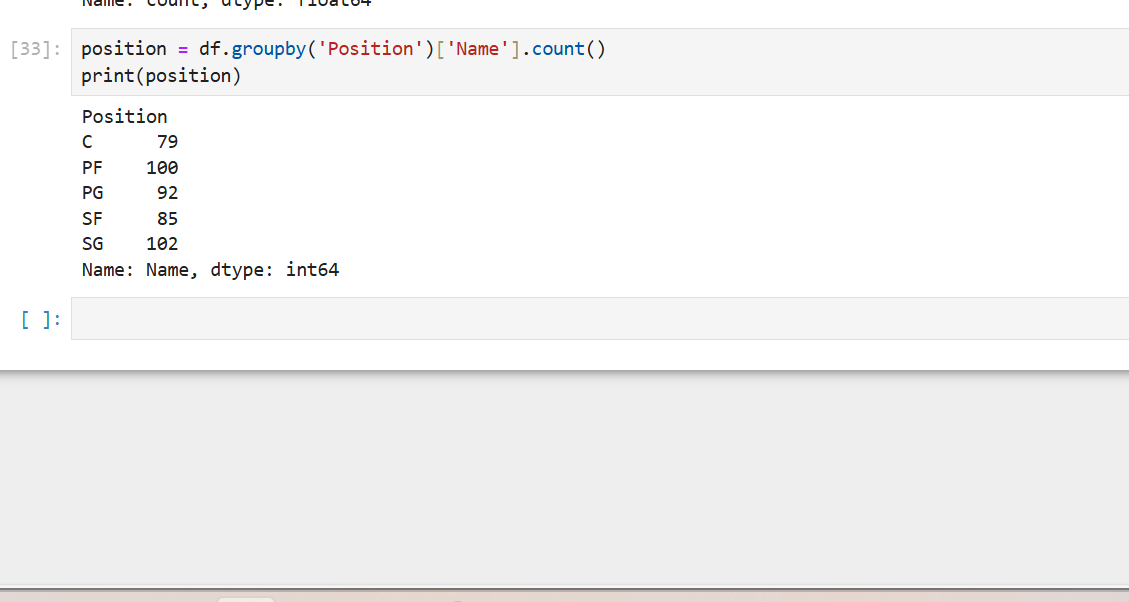
print(Team\_percentage)



2.Segregate employees based on their positions within the company.

position = df.groupby('Position')['Name'].count()

print(position)



1. Identify the predominant age group among employees.

bins=[20,30,40,50,60]

age\_groups=pd.cut(df['Age'],bins=bins)

pred\_age=age\_groups.value\_counts().idxmax()

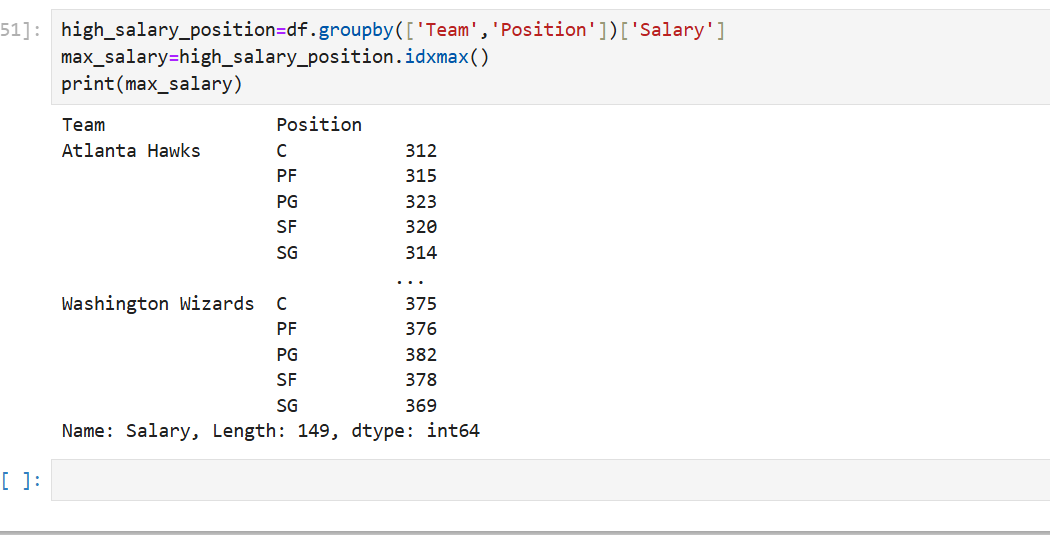
print(pred\_age)



1. Discover which team and position have the highest salary expenditure.

high\_salary\_position=df.groupby(['Team','Position'])['Salary']

max\_salary=high\_salary\_position.idxmax()

print(max\_salary)

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

import matplotlib.pyplot as plt

correlation=df['Age'].corr(df['Salary'])

print(f"Correlation between age and salary:{correlation}")

**type 1 visualization**

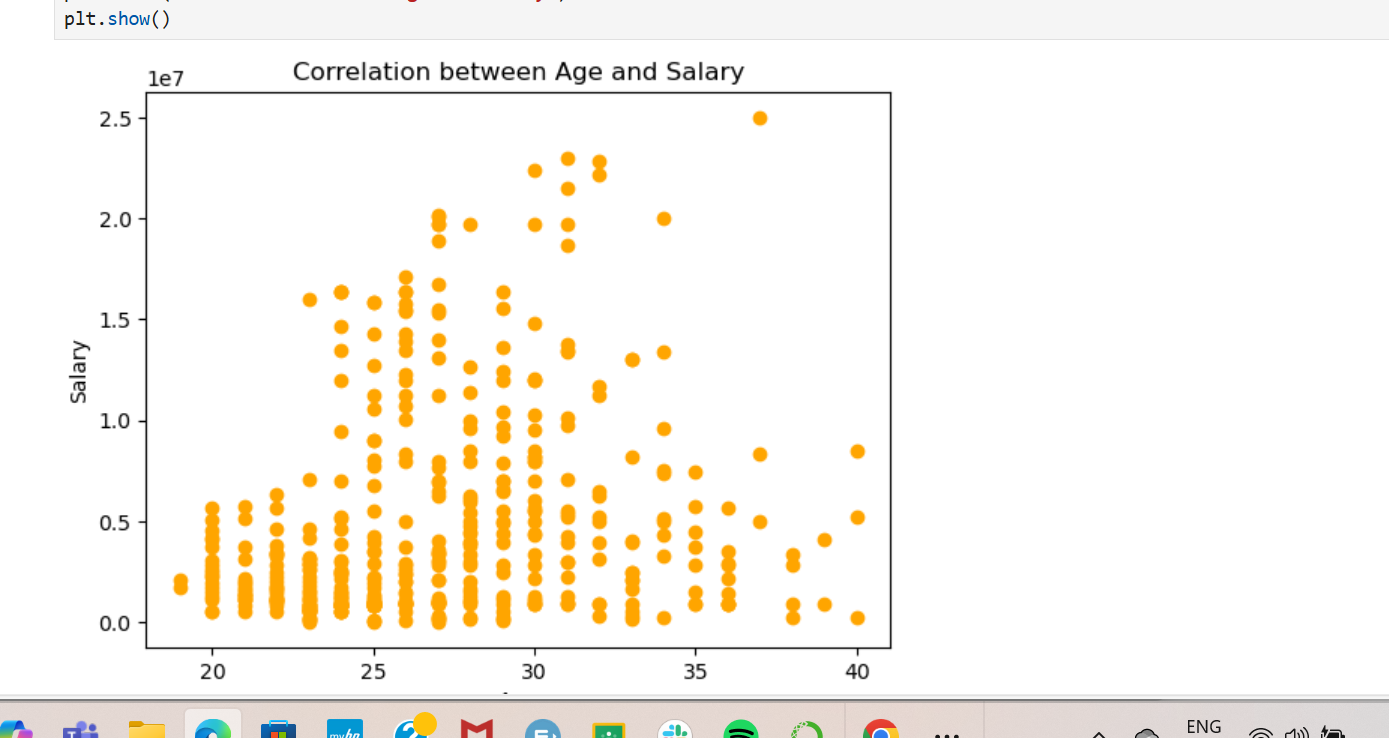
plt.scatter(df['Age'],df['Salary'],color='orange')

plt.xlabel('Age')

plt.ylabel('Salary')

plt.title("Correlation between Age and Salary")

plt.show()



**type2**

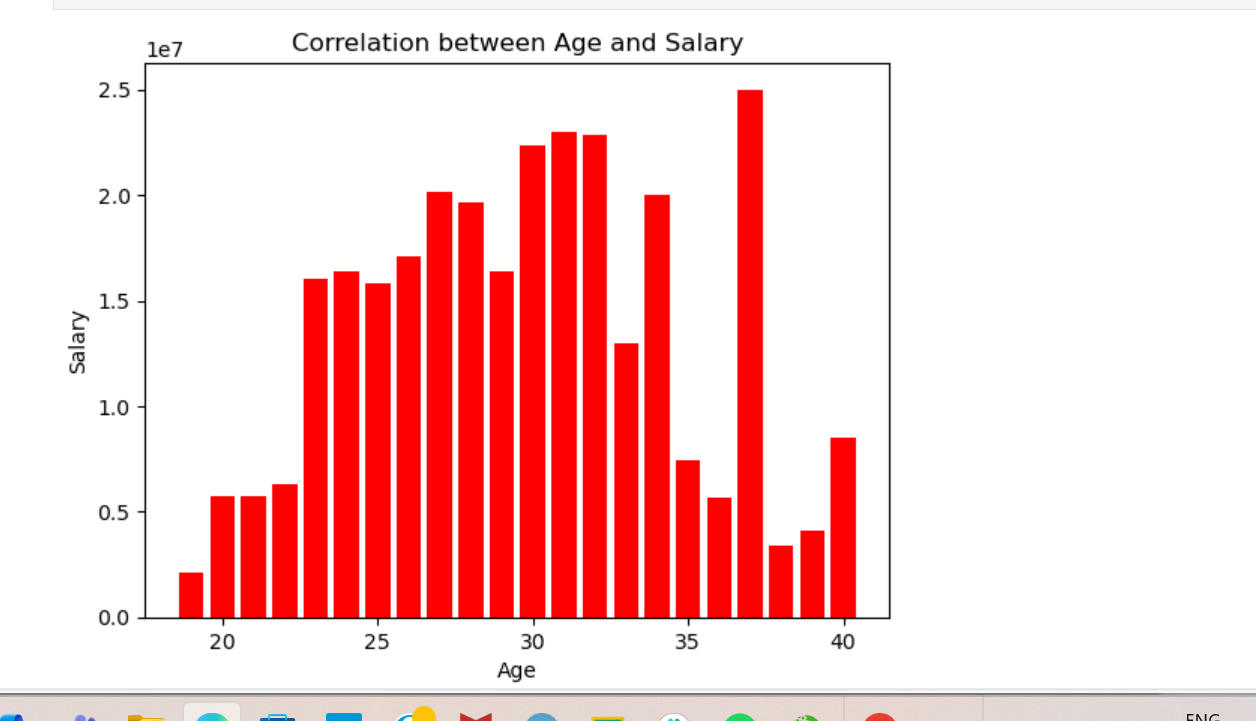
plt.bar(df['Age'],df['Salary'],color='red')

plt.xlabel('Age')

plt.ylabel('Salary')

plt.title("Correlation between Age and Salary")

plt.show()



***GRAPHICAL REPRESENTATION For each of the five analysis tasks, create appropriate visualizations to present your findings effectively.***

**Visualization1:**

plt.figure(figsize=(5, 3))

Team\_dist=df['Team'].value\_counts()

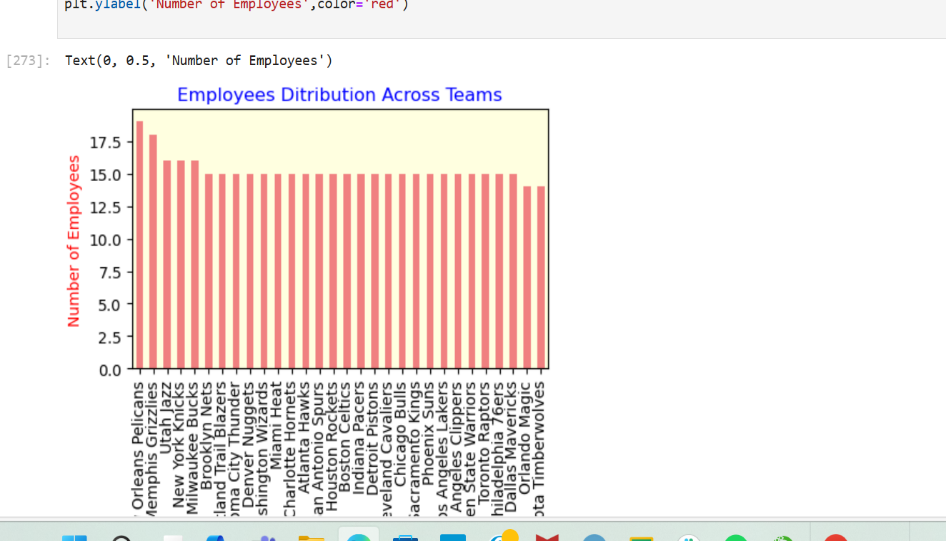
Team\_dist.plot(kind='bar',color='lightcoral')

plt.gca().set\_facecolor('lightyellow')

plt.title("Employees Ditribution Across Teams",color='blue')

plt.xlabel('Teams',color='red')

plt.ylabel('Number of Employees',color='red')



**Visualization 2**

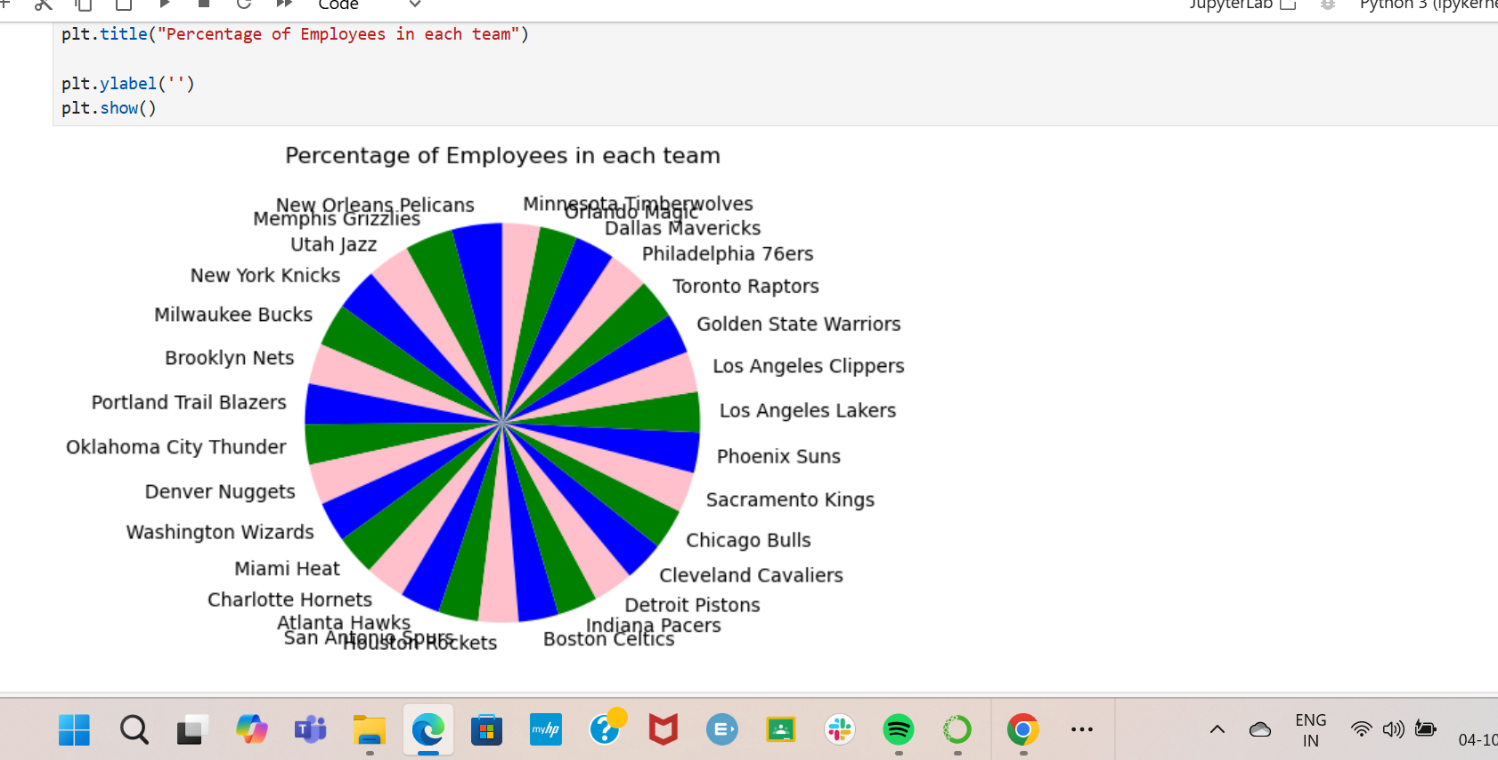
Team\_percentage=(Team\_dist/len(df))\*100

Team\_percentage.plot(kind='pie',autopct='%1.1f%%', pctdistance=0.85,startangle=90,colors=['blue','green','pink'])

plt.title("Percentage of Employees in each team")

plt.ylabel('')

plt.show()



**visualization 3**

Team\_percentage=(Team\_dist/len(df))\*100

Team\_percentage.plot(kind='pie',autopct='%1.1f%%', pctdistance=0.90,colors=['skyblue','green','pink'])

centre\_circle = plt.Circle((0, 0), 0.70, fc='white')

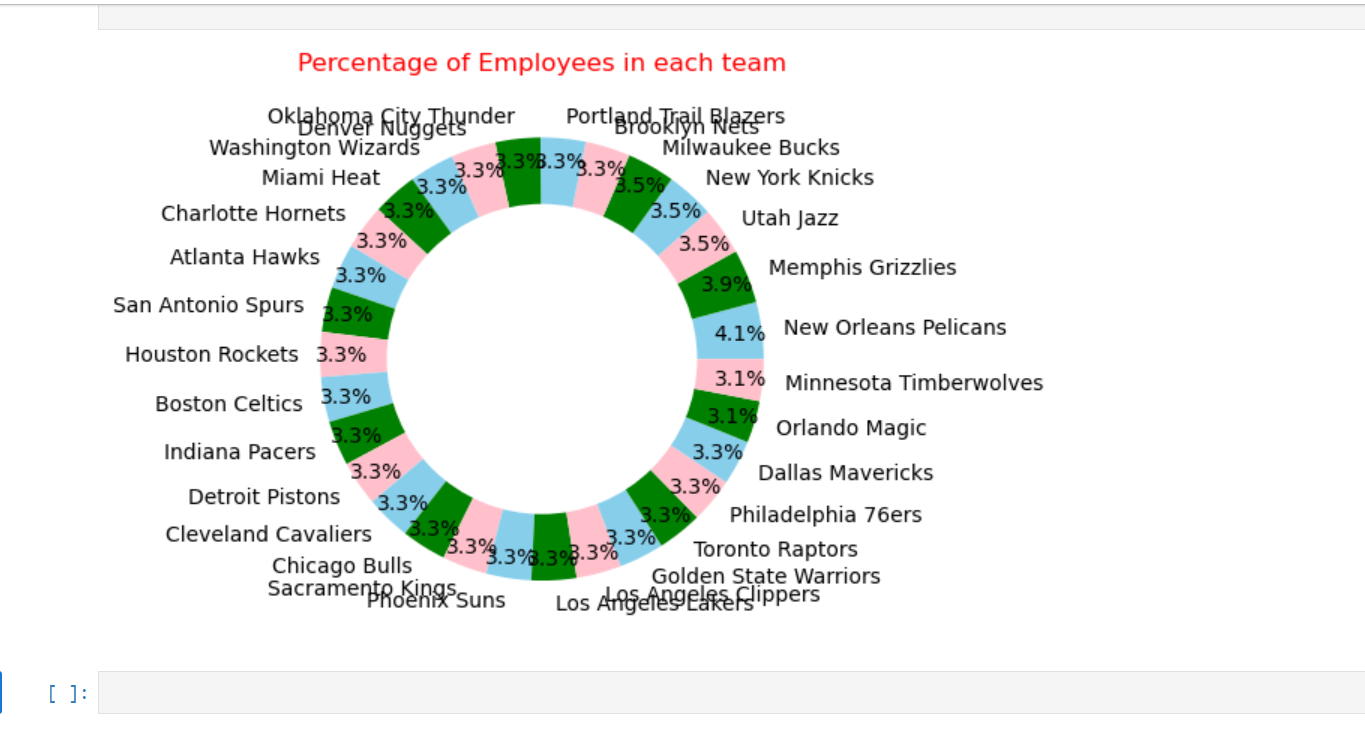
fig = plt.gcf()

fig.gca().add\_artist(centre\_circle)

plt.title("Percentage of Employees in each team",color='red')

plt.ylabel('')

plt.show()



**visualization 4**

plt.figure(figsize=(5, 3))

position = df.groupby('Position')['Name'].count()

position. plot(kind='bar',stacked='True',color=['red','green','orange','coral'])

plt.gca().set\_facecolor('lightyellow')

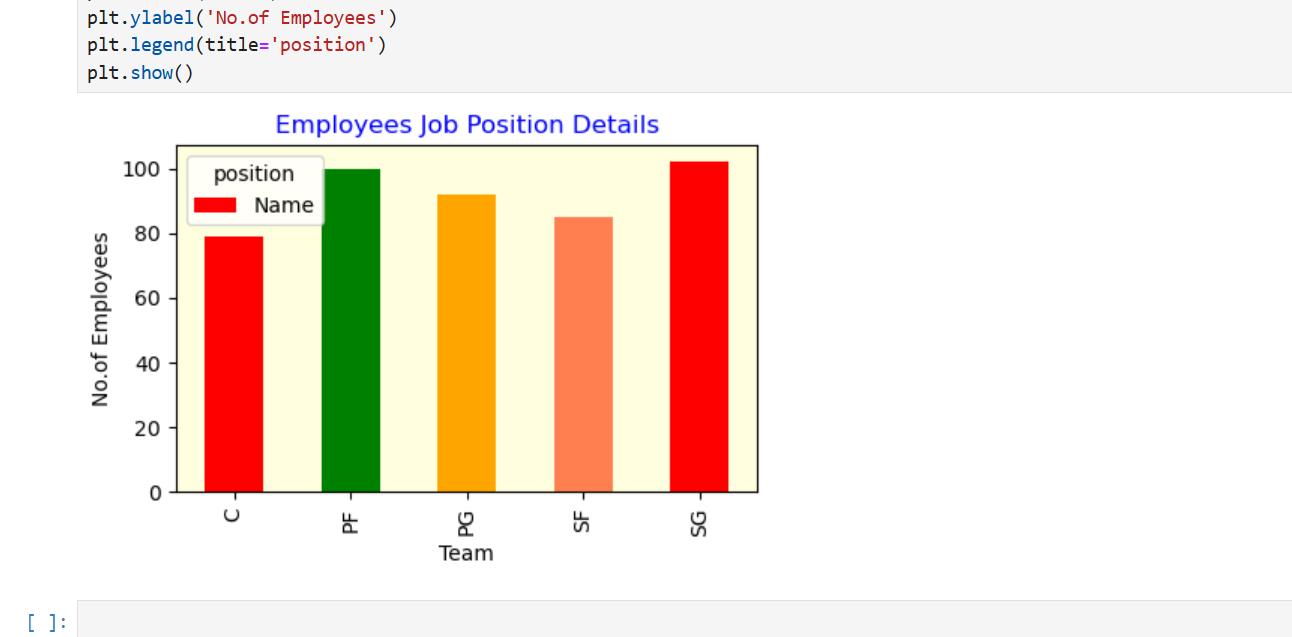
plt.title("Employees Job Position Details",color='blue')

plt.xlabel('Team')

plt.ylabel('No.of Employees')

plt.legend(title='position')

plt.show()



**visualization 5**

plt.figure(figsize=(6,3))

bins=[20,30,40,50,60]

plt.hist(df['Age'],bins=bins,color='coral',edgecolor='black')

plt.gca().set\_facecolor('lightgreen')

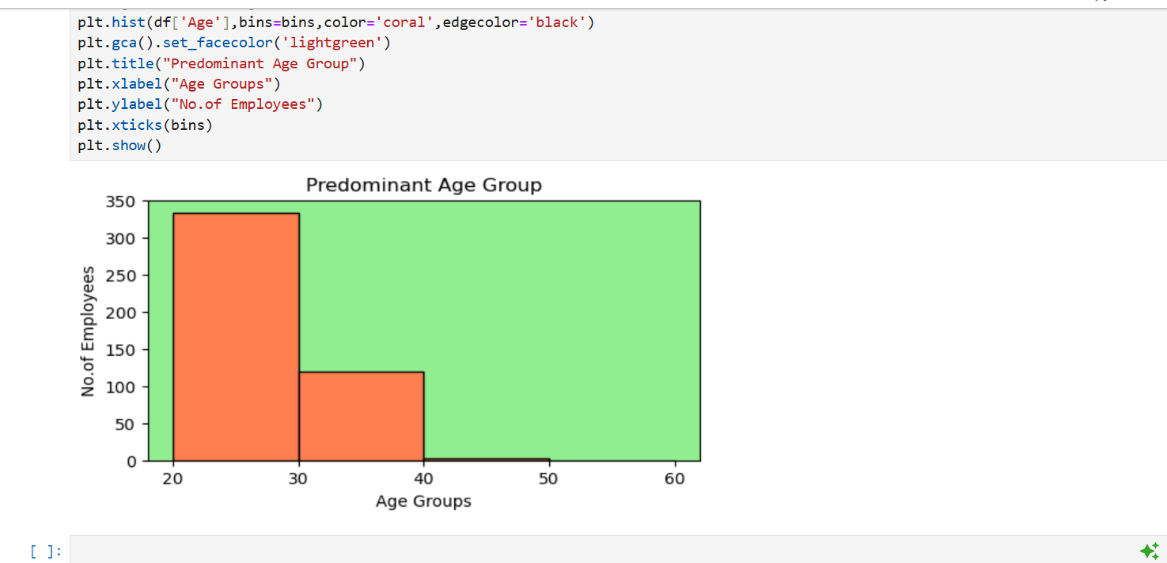
plt.title("Predominant Age Group")

plt.xlabel("Age Groups")

plt.ylabel("No.of Employees")

plt.xticks(bins)

plt.show()



**visualization 6**

import seaborn as snc

high\_salary\_position=df.groupby(['Team','Position'])['Salary'].sum().unstack()

plt.figure(figsize=(10,6))

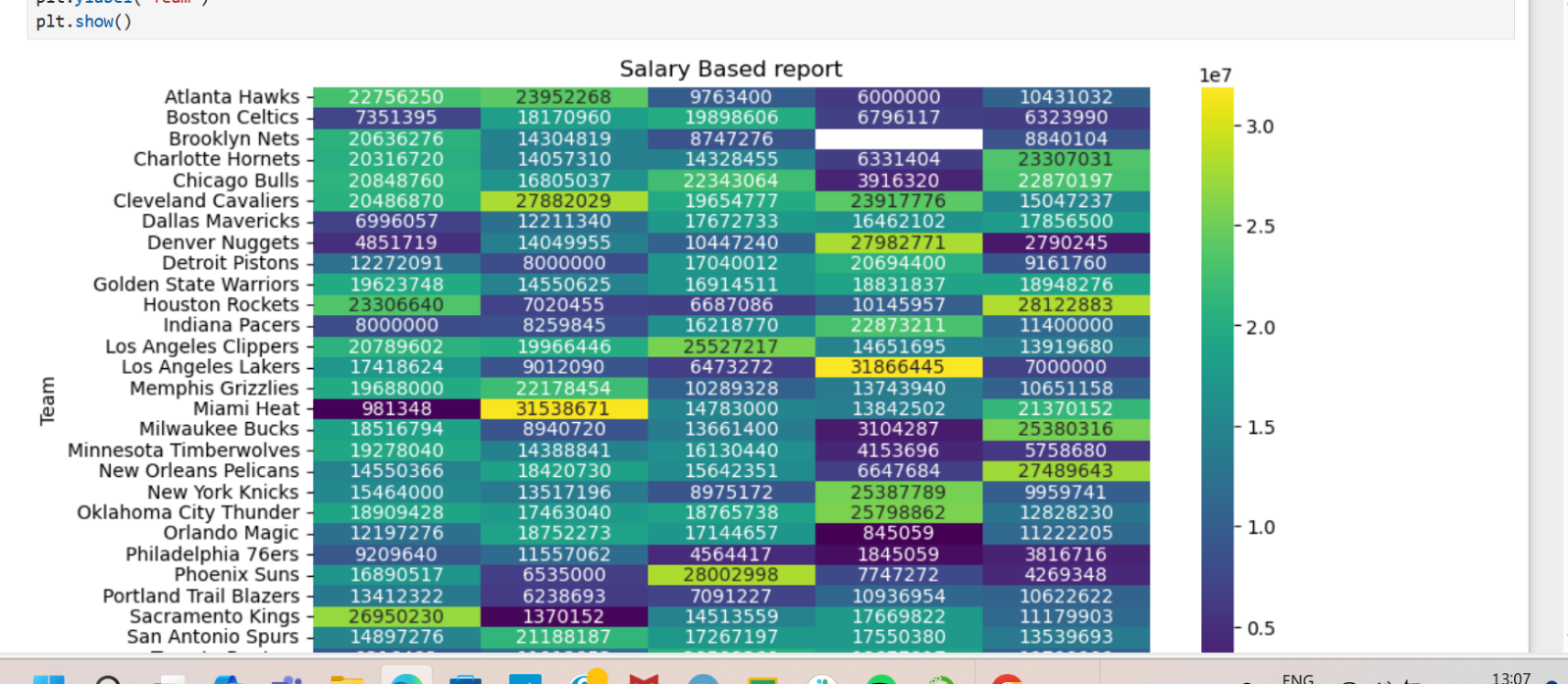
snc.heatmap(high\_salary\_position,cmap='viridis',annot=True,fmt='.0f')

plt.title("Salary Based report",color='coral')

plt.xlabel('Position',color='blue')

plt.ylabel('Team',color='blue')

plt.show()



**visualization 7**

plt.scatter(df['Age'],df['Salary'],color='orange',edgecolor='red')

plt.gca().set\_facecolor('lightyellow')

plt.xlabel('Age')

plt.ylabel('Salary')

plt.title("Age v/s Salary")

snc.regplot(x='Age',y='Salary' ,data=df,scatter=False,color='green')

plt.show()

