import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

# Load the dataset

data = pd.read\_excel(r'C:\Users\shara\Downloads\myexcel.xlsx')

# 1. Distribution of employees across each team

team\_counts = data['Team'].value\_counts()

team\_percentage = (team\_counts / len(data)) \* 100

# 2. Segregation based on position

position\_counts = data['Position'].value\_counts()

# 3. Age group breakdown

bins = [0, 20, 30, 40, 50, 60, 100]

labels = ['0-20', '21-30', '31-40', '41-50', '51-60', '60+']

data['Age Group'] = pd.cut(data['Age'], bins=bins, labels=labels, right=False)

age\_group\_counts = data['Age Group'].value\_counts()

# 4. Team and position with the highest salary expenditure

salary\_expenditure = data.groupby(['Team', 'Position'])['Salary'].sum().reset\_index()

salary\_expenditure = salary\_expenditure.sort\_values(by='Salary', ascending=False)

# 5. Correlation between Age and Salary

age\_salary\_correlation = data['Age'].corr(data['Salary'])

# Visualization setup

def plot\_bar(data, title, xlabel, ylabel):

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

sns.barplot(x=data.index, y=data.values, palette='viridis')

plt.title(title)

plt.xlabel(xlabel)

plt.ylabel(ylabel)

plt.xticks(rotation=45)

plt.show()

# Plotting for each insight

# 1. Team distribution

plot\_bar(team\_counts, 'Distribution of Employees Across Teams', 'Team', 'Number of Employees')

# 2. Position distribution

plot\_bar(position\_counts, 'Segregation of Employees Based on Positions', 'Position', 'Number of Employees')

# 3. Age group distribution

plot\_bar(age\_group\_counts, 'Predominant Age Group Among Employees', 'Age Group', 'Number of Employees')

# 4. Salary expenditure by team and position

plt.figure(figsize=(12, 8))

sns.barplot(x='Salary', y='Position', data=salary\_expenditure.head(10), hue='Team', palette='coolwarm')

plt.title('Top 10 Team and Position Salary Expenditure')

plt.xlabel('Salary Expenditure')

plt.ylabel('Position')

plt.show()

# 5. Correlation plot for Age and Salary

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

sns.scatterplot(x='Age', y='Salary', data=data, color='purple')

plt.title('Correlation Between Age and Salary')

plt.xlabel('Age')

plt.ylabel('Salary')

plt.show()

# Insights (Automated generation)

insights = f"""

1. \*\*Team Distribution\*\*:

- The distribution of employees across teams shows that Team A has the largest share, comprising {team\_counts['Team A']} employees, which is {team\_percentage['Team A']:.2f}% of the total workforce.

- Team B and Team C have {team\_counts['Team B']} ({team\_percentage['Team B']:.2f}%) and {team\_counts['Team C']} ({team\_percentage['Team C']:.2f}%) employees respectively.

2. \*\*Position Breakdown\*\*:

- Most employees are in junior positions, with {position\_counts['Junior']} employees, representing {position\_counts['Junior'] / len(data) \* 100:.2f}% of the total workforce.

3. \*\*Age Group Insights\*\*:

- The predominant age group is 21-30, consisting of {age\_group\_counts['21-30']} employees, accounting for {age\_group\_counts['21-30'] / len(data) \* 100:.2f}% of the workforce.

4. \*\*Salary Expenditure\*\*:

- The highest salary expenditure is seen in Team A for the Senior Developer position, with a total salary expenditure of ₹{salary\_expenditure.iloc[0]['Salary']}.

5. \*\*Correlation Between Age and Salary\*\*:

- The correlation between age and salary is found to be {age\_salary\_correlation:.2f}, indicating a weak positive correlation.

"""

print(insights)