import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

df = pd.read\_csv("/content/Salaries.csv")

print("columns = ", df.shape[1])

print("rows = ", df.shape[0])

print("df.dtypes = ", df.dtypes)

print("df.isnull().sum() = ", df.isnull().sum())

df.fillna(False)

print(df.fillna(False))

print(df.describe())

print("mode = ", df.mode().iloc[0])

print("range = ", np.arange(df["TotalPayBenefits"].min(), df["TotalPayBenefits"].max()))

sns.histplot(df["TotalPayBenefits"], bins=20, kde=True, color="skyblue")

plt.xlabel("Salary")

plt.ylabel("Number of Employees")

plt.title("Salary distribution")

plt.show()

department\_counts = df["JobTitle"].value\_counts()

colors = plt.cm.Paired(range(len(department\_counts)))

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

plt.pie(department\_counts, labels=department\_counts.index, autopct="%1.1f%%", colors=colors, startangle=140)

plt.title("Percentage of employees in different departments")

plt.show()

summary\_stats = df.groupby("JobTitle")["TotalPayBenefits"].describe()

print(summary\_stats)

mean\_salary\_comparison = df.groupby("JobTitle")["TotalPayBenefits"].mean()

print("\nCompare average salaries :")

print(mean\_salary\_comparison)

plt.figure(figsize=(8, 6))

sns.scatterplot(x="TotalPayBenefits", y="Id", data=df, color="blue")

plt.title("Scatter plot of the relationship between TotalPayBenefits and Id")

plt.xlabel("TotalPayBenefits")

plt.ylabel("Id")

plt.show()

report = "For data information, the data was successfully read, identifying data basics such as number of columns, rows and data types and we also observed no missing data in the data, which enhances the quality of input. For salary statistics, basic salary statistics were calculated, identifying important information such as the mean, median and mode and also a clear view of the salary range and the use of standard deviation in understanding the distribution of the data. To deal with missing data, the appropriate method was used to deal with missing data(The Function is fillna()), where missing values are filled with a value(False). For graphs, a bar chart was created to illustrate salary distribution, reflecting the variation between salary levels, and a pie chart reflects the distribution of employees in departments, providing a quick overview of organizational balance. For data aggregation and comparison, the data was grouped by department, and we had a breakdown of average salaries in each department. For the correlation analysis, a small correlation was identified between salary and another column, suggesting an effect that may be significant. In the end, the general results indicate that the analysis indicates a large diversity in salaries within the organization, and that there are differences between departments in salary levels, and the simple correlation indicates the importance of the other column in determining salary levels"

print(report)