**Assignment 1**

**Roll no. – A-44**

**Subject – DAP**

**Code-**

import numpy as np

import pandas as pd

data = pd.read\_csv('employee\_data.csv')

print("First few rows of the dataset:")

print(data.head())

print("\nMissing values in the dataset:")

print(data.isnull().sum())

data\_cleaned = data.dropna()

print("\nSummary statistics of numerical columns:")

print(data\_cleaned.describe())

avg\_salary\_by\_dept = data\_cleaned.groupby('Department')['Salary'].mean()

print("\nAverage salary by department:")

print(avg\_salary\_by\_dept)

import matplotlib.pyplot as plt

avg\_salary\_by\_dept.plot(kind='bar')

plt.title('Average Salary by Department')

plt.xlabel('Department')

plt.ylabel('Average Salary ($)')

plt.show()

**Output-**

First few rows of the dataset:

EmployeeID Name Department Age Salary

0 1 Employee1 HR 42 94223

1 2 Employee2 IT 33 41052

2 3 Employee3 Finance 26 45741

3 4 Employee4 HR 28 87368

4 5 Employee5 IT 26 73986

Missing values in the dataset:

EmployeeID 0

Name 0

Department 0

Age 0

Salary 0

dtype: int64

Summary statistics of numerical columns:

EmployeeID Age Salary

count 100.000000 100.000000 100.000000

mean 50.500000 39.520000 66601.760000

std 29.011492 11.311458 20142.788826

min 1.000000 22.000000 30894.000000

25% 25.750000 30.000000 50979.250000

50% 50.500000 38.000000 68735.000000

75% 75.250000 51.000000 84402.500000

max 100.000000 58.000000 99939.000000

Average salary by department:

Department

Finance 69954.291667

HR 68204.640000

IT 62359.031250

Marketing 67403.578947

Name: Salary, dtype: float64