**PRACTICAL-2**

**AIM:-** Create Two-dimensional data with the help of data frames and perform ifferent operations on it.

**THEORY**

A 2D array is an array of arrays that can be represented in matrix form, like rows and columns. In this array, the position of data elements is defined with two indices instead of a single index. In Python, we can access two-dimensional array elements using two indices.

Data Frame is a 2-dimensional labelled data structure with columns of potentially different types. You can think of it like a spread sheet or SQL table, or a dict of Series objects. It is generally the most commonly used pandas object.

**SOURCE CODE AND OUTPUTS**

import numpy as np

import pandas as pd

Employee = {

'Name':['Raj','Omkar','Aryan','Devam','Riya','Davi','Abhishake','Ankita','Rinku','Bhushan'],

'Age':[21,32,24,35,42,32,36,52,46,30],

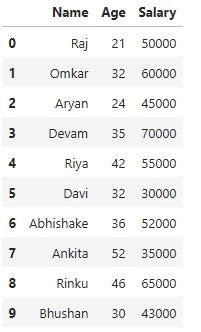
'Salary':[50000, 60000, 45000, 70000, 55000,30000, 52000, 35000, 65000, 43000]

}

**# Data frame**

df =pd.DataFrame(Employee)

df



**# Accessing columns**

print(df['Name'])

0 Raj

1 Omkar

2 Aryan

3 Devam

4 Riya

5 Davi

6 Abhishake

7 Ankita

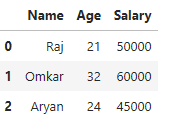
8 Rinku

9 Bhushan

Name: Name, dtype: object

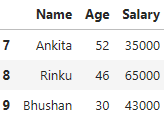
**# Slicing rows head first (3 value)**

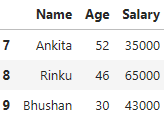
df.head(3)



**# Slicing rows tail last (3 value)**

df.tail(3)





**# Filtering (conditions)**

print(df[df['Salary']> 50000])

print(' ')

print( df[df['Age'] < 30])

Name Age Salary

1 Omkar 32 60000

3 Devam 35 70000

4 Riya 42 55000

6 Abhishake 36 52000

8 Rinku 46 65000

Name Age Salary

0 Raj 21 50000

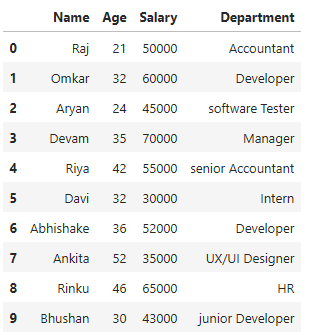
2 Aryan 24 45000

**# Adding a new column**

df['Department']=['Accountant','Developer','softwareTester','Manager','senior ccountant','Intern','Developer','UX/UI Designer','HR','junior Developer']

# Show DataFrame

df



**# Information**

df.info()

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 10 entries, 0 to 9

Data columns (total 4 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Name 10 non-null object

1 Age 10 non-null int64

2 Salary 10 non-null int64

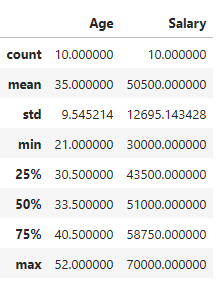
3 Department 10 non-null object

dtypes: int64(2), object(2)

memory usage: 448.0+ bytes

**# Describe**

df.describe()



**# Removing**

df = df.drop(columns=['Department'])



**# Creating Data Visualization**

import pandas as pd

import matplotlib.pyplot as plt

**# Load the data into a Series object**

series = pd.Series([1, 2, 3, 4, 5])

# Explore the data

print(series.head())

0 1

1 2

2 3

3 4

4 5

dtype: int64

print(series.dtype)

int64

print(len(series))

5

print(series.unique())

[1 2 3 4 5]

print(series.isna())

0 False

1 False

2 False

3 False

4 False

dtype: bool

**# Analyze the data**

print(series.describe())

count 5.000000

mean 3.000000

std 1.581139

min 1.000000

25% 2.000000

50% 3.000000

75% 4.000000

max 5.000000

dtype: float64

**# Plot the data**

plt.hist(series)

plt.xlabel("Value")

plt.ylabel("Frequency")

plt.title("Distribution of Series")

plt.show()

