Experiment No:2

Course Outcome:CO1 **Blooms Level:L3**

Aim:To study basics of Python.

Abstract: The objective of this study is to understand the fundamental concepts of the Python programming language, which is widely used in various fields such as data science, web development, automation, and artificial intelligence. This exploration covers the basic syntax, data types, control structures, functions, and input/output operations in Python. The study also introduces key programming concepts such as loops, conditionals, and error handling.

Sample Input and Output:

Case 1:Numpy-Code

```
import numpy as np
a = np.array([[1, 2], [3, 4]])
transpose_a = a.T
sum_a = np.sum(a)
mean_a = np.mean(a)
dot_product = np.dot(a, transpose_a)
```

Output

```
# Original array
a =
[[1 2]
 [3 4]]

# Transpose of the array
transpose_a =
[[1 3]
 [2 4]]

# Sum of all elements
sum_a = 10

# Mean of all elements
mean_a = 2.5

# Dot product of a and its transpose
dot_product =
[[ 5 11]
[11 25]]
```

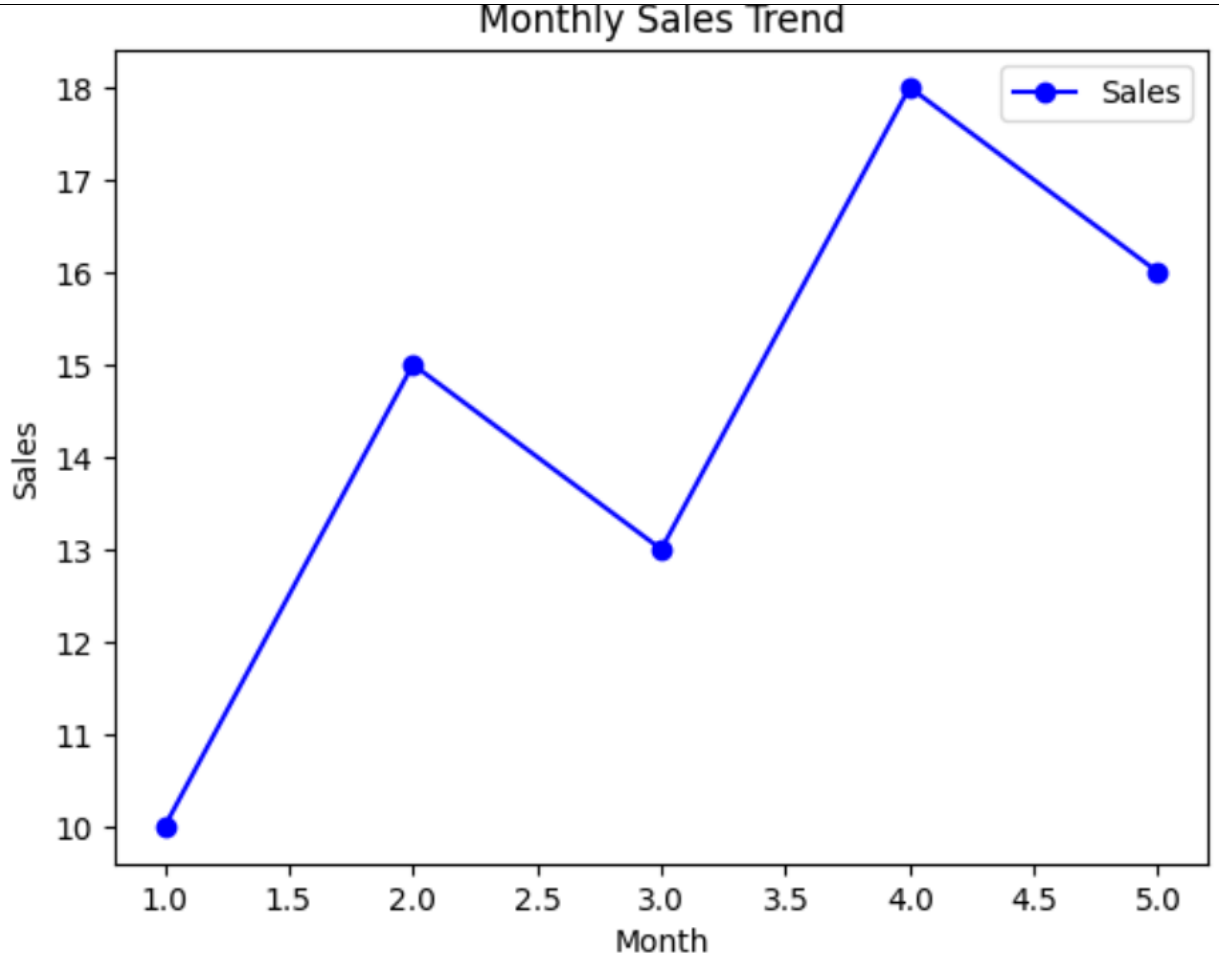
Case 2: Pandas –

```
import pandas as pd
data = {
```

```
    'Name': ['Alice', 'Bob', 'Charlie'],
    'Age': [25, 30, 35],
    'Salary': [50000, 60000, 70000]
  }
  df = pd.DataFrame(data)
  filtered_df = df[df['Age'] > 28]
  mean_salary = df['Salary'].mean()

  # Original DataFrame
  df =
      Name  Age  Salary
  0   Alice   25   50000
  1     Bob   30   60000
  2 Charlie   35   70000

  # Filtered DataFrame (Age > 28)
  filtered_df =
      Name  Age  Salary
  1     Bob   30   60000
  2 Charlie   35   70000

  # Mean salary
mean_salary = 60000.0
```

Case 3: MatplotLib-

```
  import matplotlib.pyplot as plt
  x = [1, 2, 3, 4, 5]
  y = [10, 15, 13, 18, 16]
  plt.plot(x, y, label='Sales', color='blue', marker='o')
  plt.title('Monthly Sales Trend')
  plt.xlabel('Month')
  plt.ylabel('Sales')
  plt.legend()
  plt.show()
```



Theory:

NUMPY NumPy is a library for the Python programming language, adding support for large, multidimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays

PANDAS Pandas is a column-oriented data analysis API. It's a great tool for handling and analyzing input data, and many ML frameworks support pandas data structures as inputs. The primary data structures in pandas are implemented as two classes: • DataFrame, which you can imagine as a relational data table, with rows and named columns. • Series, which is a single column. A DataFrame contains one or more Series and a name for each Series. The data frame is a commonly used abstraction for data manipulation.

MATPLOTLIB Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK.

SEABORN Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

Program and Output:

[Numpy Tutorial]
[Pandas Tutorial]
[Matplotlib Tutorial]
[Seaborn Tutorial]

Conclusion:
We have learnt the basics of NumPy and learnt how to access and manipulate data within a DataFrame and Series data structures of the pandas library. We have also performed visualization using matplotlib and explored the distribution and relationship between variables in the Tips dataset using various plots in Seaborn.

Exercise 1:

Abstract:
There is a College data set, which can be found in the file College.csv. It contains a number of variables for 777 different universities and colleges in the US. The are 18 columns in the dataset.
(a) Use the pd.read_csv() function to read the data into Python. Call the loaded data college. Make sure that you have the directory set to the correct location for the data.
(b) Use the describe() method of to produce a numerical summary of the variables in the data set.
(c) Use the pd.plotting.scatter_matrix() function to produce a scatterplot matrix of the first columns [Top10perc, Apps, Enroll].
(d) Use the boxplot() method of college to produce side-by-side boxplots of Outstate versus Private.
(e) Create a new qualitative variable, called Elite,bybinning the Top10perc variable into two groups based on whether or not the
proportion of students coming from the top 10% of their high school classes exceeds 50%.

Students shall draw flowchart of exercise question in the writeup and submit.

Exercise 2:

Abstract:
This exercise involves the Auto data set studied in the lab. Make sure
that the missing values have been removed from the data.
(a) Which of the predictors are quantitative, and which are qualitative?
(b) What is the range of each quantitative predictor? You can answer this using the min() and max() methods in numpy.
(c) What is the mean and standard deviation of each quantitative predictor?
(d) Now remove the 10th through 85th observations. What is the range, mean, and standard deviation of each predictor in the subset of the data that remains?
(e) Using the full data set, investigate the predictors graphically, using scatterplots . Create some plots highlighting the relationships among the predictors.Comment
on your findings.

Students shall draw flowchart of exercise question in the writeup and submit.