

St. Francis Institute of Technology, Mumbai-400 103

A.Y. 2022-23

Class: SE-ITA/ITB, Semester: IV

Subject: **Python Lab.**

Experiment – 13: Python program to implement Data Visualization techniques using advanced python libraries.

1. **Aim:** To implement a python program for the following:
 - a) Read an existing .csv file
 - b) Create a dataframe using .csv file and display the frame.
 - c) Visualize the dataframe using Bar chart,Histogram,Pie chart,Line Graph.
2. **Prerequisite:** Knowledge of advanced python libraries used in data science.
3. **Objectives:** To demonstrate the usage of data frames and different types of visualization techniques.
4. **Requirements:** Personal Computer (PC), Windows /Linux Operating System, IDLE 3.6 for Python3, sqlite3, MySQL server5.5.

5. **Pre-Experiment Exercise:**

Theory:

1. **Bar Graph:**

A **bar chart** or **bar graph** is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally.

- **Modules Required:**

```
import matplotlib.pyplot as plt
import pandas as pd
```

- **Extracting x and y columns:**

```
x=df[“roll_no”]
y=df[“name”]
```

- **Labeling x and y columns:**

```
plt.xlabel(“roll”)
plt.ylabel(“stud_name”)
```

- **Giving Title to the Graph:**

```
plt.title(“Name of the Title”)
```

- **Giving different colors and labels to data:**

```
plt.legend()
```

2. **Histogram:**

Shows the distribution of values in terms of intervals.

histtype=‘bar’ to show the histogram in terms of bars.The other types are barstacked,step,stepfilled etc.

Syntax:

```
plt.hist(stud_age,bins,histtype='stepfilled',rwidth=2,color='cyan')
```

rwidth indicates the bar’s width.

3. **Pie Chart:**

Shows circle that is divided into sectors and each sector represents a proportion of the

whole.

Syntax:

```
plt.pie(slices,labels=depts,colors=color,autopct="%.1f%%",startangle=90,shadow=True)
```

The **shadow** option indicates that the pie chart should be displayed with a shadow or not.

The **autopct** indicates how to display the percentages on the slices.

4. **Line Graph:**

A simple graph that shows results in the form of lines.

Syntax:

```
plt.plot(x,y,'colorname')
```

6.Laboratory Exercise:

A. Procedure:

- i. Open Idle for python
- ii. Open editor in Idle from menu file-new
- iii. Type python code with proper syntax
- iv. Save file with .py extension
- v. Execute the code inside the saved file using shortcut key F5 or using menu: Run-Run module

B. Program code with comments:

Write and execute your program code to achieve the given aim and attach it **with your own comments with neat indentation**.

7.Post-Experiments Exercise

A. Extended Theory:

1. Explain different data visualization techniques with their syntax's.
2. Explain Data Frames and different techniques to create Data Frames.

B. Questions/Programs:

1. Write a Python program to create Data Frame using Dictionary, List of Tuples.
2. Write a Python program to implement any 5 operations on Data Frame.

C. Conclusion:

1. Write what was performed in the experiment/program.
2. What is the significance of experiment/program?

D. References

- [1] James Payne, "Beginning Python: Using Python 2.6 and Python 3.1", WroxPublication.
- [2] Dr.Nageswara Rao,"Core Python Programming",Wiley Publication.
- [2] <https://www.python.org/>
- [3] www.pythonforbeginners.com

In-Lab Exercise

To implement a python program for the following:

- a) Read an existing .csv file
- b) Create a dataframe using .csv file and display the frame.
- c) Visualize the dataframe using Bar chart,Histogram,Pie chart,Line Graph.

```
import matplotlib.pyplot as plt
import math
import numpy as np
```

Line Graph of Years from 2000 to 2012 vs tables of 2 and 3:

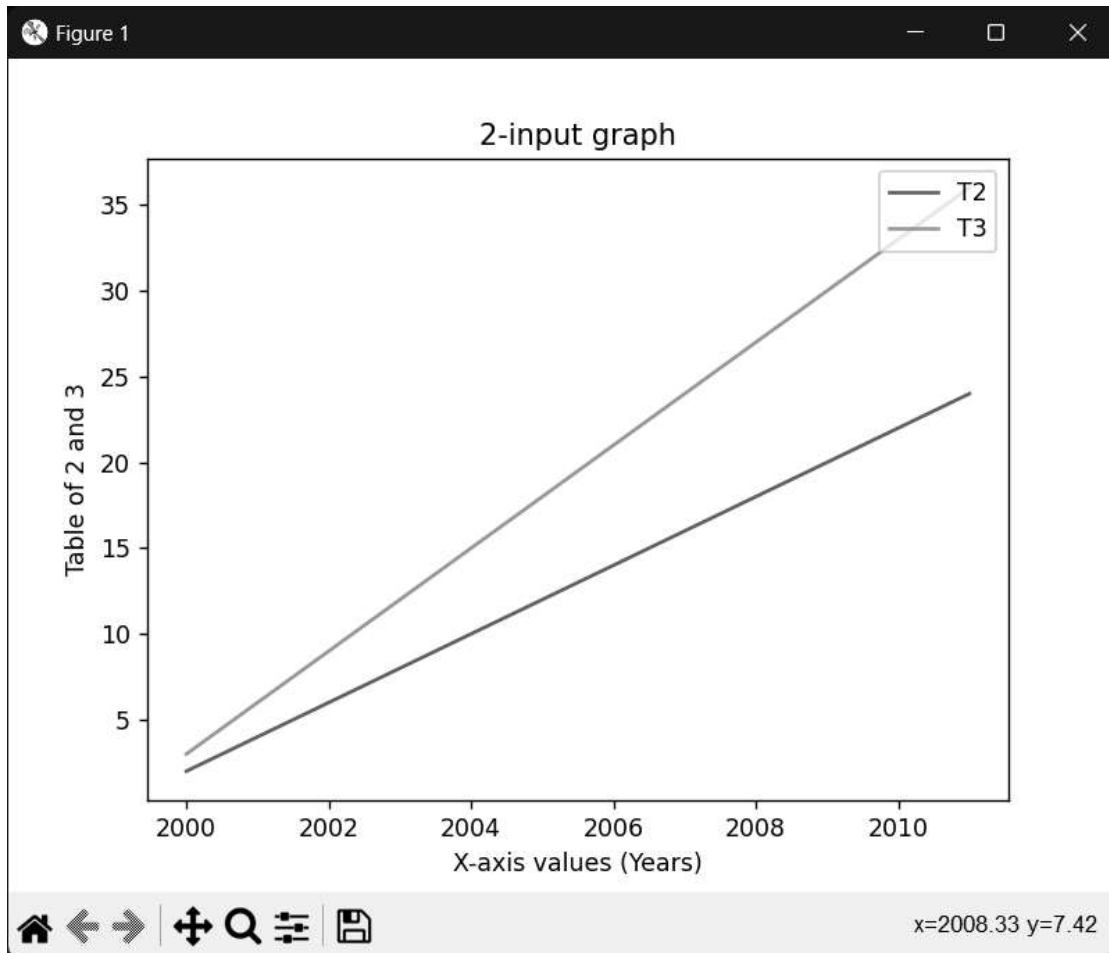
```
#X-axis data
years = range(2000,2012)

#Y-axis data
a = [2,4,6,8,10,12,14,16,18,20,22,24]
b = [3,6,9,12,15,18,21,24,27,30,33,36]

#Ploting the line graphs
plt.plot(years,a)
plt.plot(years,b)

#Adding Labels,Legend and Titles
plt.xlabel("X-axis values (Years)")
plt.ylabel("Table of 2 and 3")
plt.title("2-input graph")
plt.legend(["T2","T3"],loc = "upper right")
plt.show()
```

Plot of Line Graph of Years from 2000 to 2012 vs tables of 2 and 3:



2. Histogram

Code:

```
#Generating Random Data
data = np.random.randn(1000)

#Plotting Histogram
plt.hist(data, bins=10, color = "teal", edgecolor="black")

#Adding Labels and Titles
```

```
plt.xlabel("Values")
plt.ylabel("Frequency")
plt.title("Basic Histogram")

plt.show()

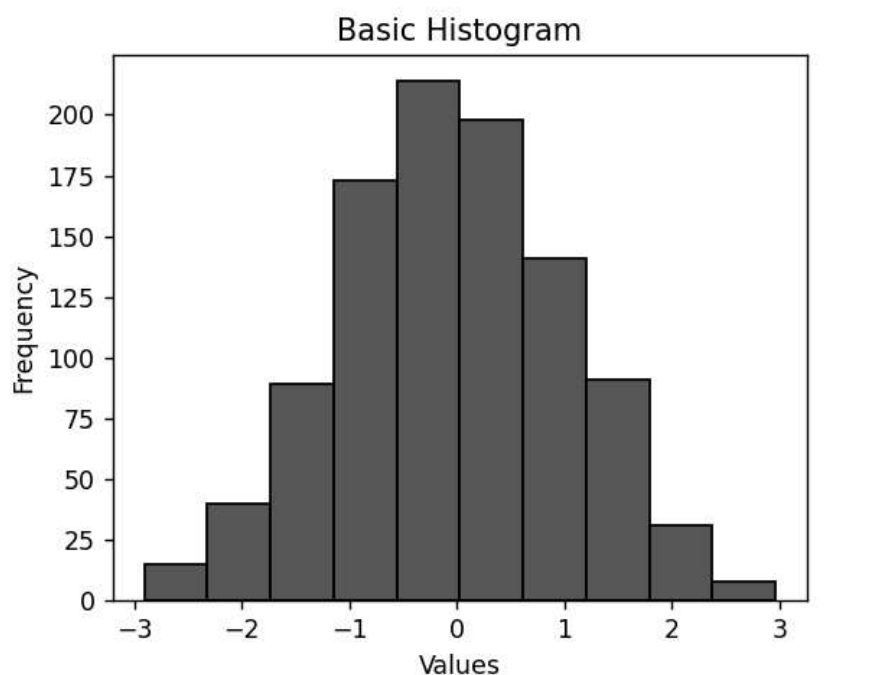
#Generating Random Data
data1 = np.random.randn(1000)
data1 = np.random.randn(1000)

#Ploting Histogram
plt.hist(data1, bins=20, color = "blue", edgecolor="black")
plt.hist(data, bins=20, color = "teal", edgecolor="black")

#Adding Labels and Titles
plt.xlabel("Values")
plt.ylabel("Frequency")
plt.title("Histogram")

plt.show()
```

Plot of Histogram:

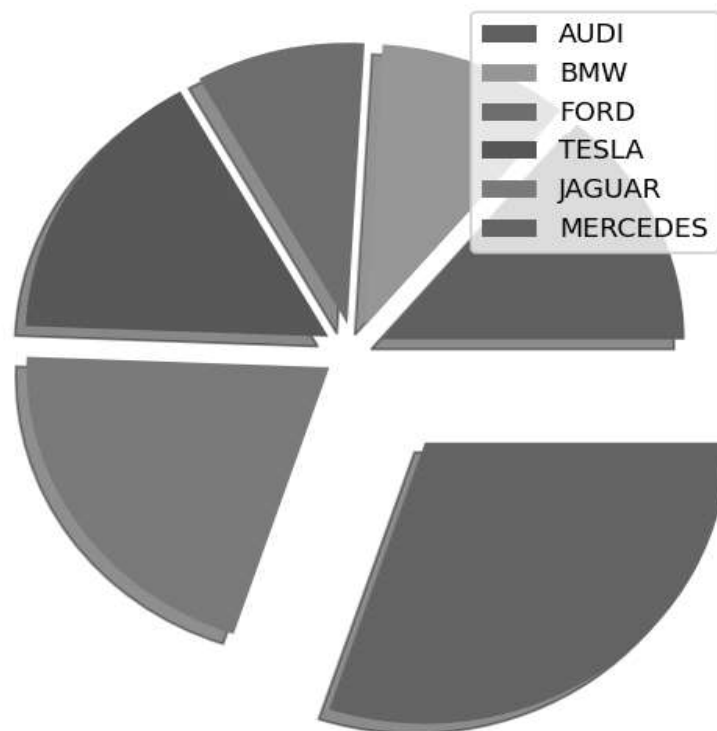


3. Pie Charts

Code:

```
cars = ["AUDI","BMW","FORD","TESLA", "JAGUAR", "MERCEDES"]  
data = [23,17,15,27,34,50]  
explodes = [0.1,0.1,0.1,0.1,0.1,0.4]  
fig = plt.figure(figsize=(5,7))  
plt.pie(data , shadow = True, explode = explodes)  
plt.legend(cars,loc = "upper right")  
plt.show()
```

Plot:



Post-Lab Exercise:

1. Write a Python program to create Data Frame using Dictionary, List of Tuples.

Code:

```
import pandas as pd

# Using dictionary
data_dict = {
    'Name': ['Vishal', 'Joseph', 'Kevin', 'Ajay'],
    'Address': ['Jalgoan', 'Ram Mandir', 'Malad', 'Borivali'],
    'Rollno': [63, 57, 5, 4],
}

df_dict = pd.DataFrame(data_dict)
print("DataFrame using Dictionary:")
print(df_dict)

# Using list of tuples
data_list = [
    ('Vishal', 'Jalgoan', 63),
    ('Joseph', 'Ram Mandir', 57),
    ('Kevin', 'Malad', 5),
    ('Ajay', 'Borivali', 4)
]

df_list = pd.DataFrame(data_list, columns=['Name', 'Address', 'Rollno'])
print("\nDataFrame using List of Tuples:")
print(df_list)
```


Output:

```
Command Prompt
F:\College Stuff\Vishal Mahajan SE IT SEM 4\Python Lab>python exp13.py
DataFrame using Dictionary:
   Name    Address  Rollno
0  Vishal   Jalgoan     63
1  Joseph  Ram Mandir    57
2   Kevin    Malad       5
3   Ajay    Borivali     4

DataFrame using List of Tuples:
   Name    Address  Rollno
0  Vishal   Jalgoan     63
1  Joseph  Ram Mandir    57
2   Kevin    Malad       5
3   Ajay    Borivali     4
```

2. Write a Python program to implement any 5 operations on Data Frame.

Code:

```
import pandas as pd

# Creating a DataFrame
data = {
    'Name': ['Vishal', 'Joseph', 'Kevin', 'Ajay'],
    'Address': ['Jalgoan', 'Ram Mandir', 'Malad', 'Borivali'],
    'Rollno': [63, 57, 5, 4],
    'Marks': [90, 65, 95, 75]
}

df = pd.DataFrame(data)
print("\nOriginal DataFrame:")
print(df)

# Accessing and displaying data
print("\nAccessing and displaying data:")
print("First row:")
print(df.iloc[0])
print("\nLast 2 rows:")
print(df.tail(2))
print("\nMarks column:")
print(df['Marks'])

# Adding a new column
df['PID'] = [221068, 221062, 221075, 221074]
print("\nDataFrame after adding PID column:")
print(df)

# Filtering data
filtered_df = df[df['Marks'] > 70]
print("\nFiltered DataFrame where Marks > 70:")
print(filtered_df)

# Sorting data
```

```
sorted_df = df.sort_values(by='PID', ascending=False)
print("\nSorted DataFrame by PID:")
print(sorted_df)
```

Output:

```
Original DataFrame:
   Name   Address  Rollno  Marks
0  Vishal   Jalgoan     63    90
1  Joseph  Ram Mandir     57    65
2   Kevin     Malad       5    95
3   Ajay   Borivali       4    75

Accessing and displaying data:
First row:
Name      Vishal
Address   Jalgoan
Rollno      63
Marks      90
Name: 0, dtype: object

Last 2 rows:
   Name   Address  Rollno  Marks
2  Kevin     Malad       5    95
3   Ajay  Borivali       4    75

Marks column:
0    90
1    65
2    95
3    75
Name: Marks, dtype: int64

DataFrame after adding PID column:
   Name   Address  Rollno  Marks  PID
0  Vishal   Jalgoan     63    90  221068
1  Joseph  Ram Mandir     57    65  221062
2   Kevin     Malad       5    95  221075
3   Ajay   Borivali       4    75  221074

Filtered DataFrame where Marks > 70:
   Name   Address  Rollno  Marks  PID
0  Vishal   Jalgoan     63    90  221068
2   Kevin     Malad       5    95  221075
3   Ajay  Borivali       4    75  221074

Sorted DataFrame by PID:
   Name   Address  Rollno  Marks  PID
2   Kevin     Malad       5    95  221075
3   Ajay  Borivali       4    75  221074
0  Vishal   Jalgoan     63    90  221068
1  Joseph  Ram Mandir     57    65  221062
```

F:\College Stuff\Vishal Mahajan SE IT SEM 4\Python Lab>