**DEPARTMENT OF INFORMATION TECHNOLOGY**

**Course: Data Mining & Business Intelligence Lab (ITL601)**

**B.Tech. (Information Technology) – Semester VI**

**Academic Year: 2023-24 (Even Semester)**

**PRACTICAL 4 -A**

**Aim:** Implementation of Data Visualization and Statistical Data Analysis using

Python / Scilab:

1. Plotting Bar Charts, Histograms, Scatter Plots, etc.

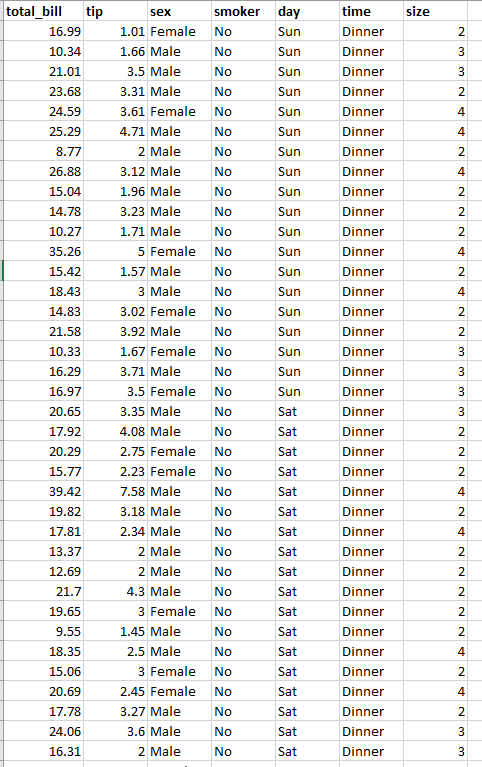
**Lab Objective:** Proficiently utilize Python, specifically Matplotlib and Seaborn, for creating diverse data visualizations and Scikit-learn for statistical analysis, fostering a comprehensive understanding of data exploration techniques.

**Theory:**

Data visualization and statistical data analysis are crucial aspects of gaining insights from datasets. In Python, libraries like Matplotlib and Seaborn provide powerful tools for creating various plots such as bar charts, histograms, and scatter plots. Bar charts are useful for comparing categorical data, histograms display the distribution of numerical data, and scatter plots reveal relationships between two variables. These visualizations aid in understanding patterns, trends, and outliers within the data. Additionally, Python's Scikit-learn library facilitates statistical data analysis, offering tools for tasks like regression and clustering. The integration of Python and Scilab, a scientific computing software, further enhances the capabilities for comprehensive data exploration and analysis, making it a versatile combination for researchers and data scientists.

**<Code with Output>:**

**Dataset .csv**

****

1. **Histogram:**

**Code:**

**import pandas as pd**

**import matplotlib.pyplot as plt**

**data = pd.read\_csv("tips.csv")**

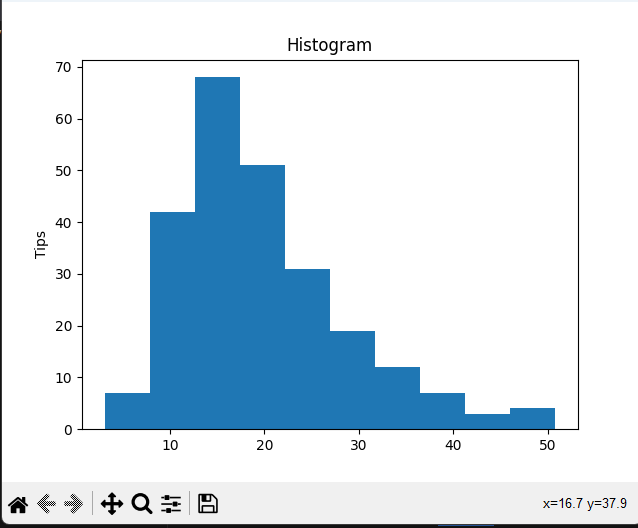
**plt.hist(data['total\_bill'])**

**plt.title("Histogram")**

**plt.ylabel('Tips')**

**plt.show()**

**Output:**

****

1. **Bar charts**

**import pandas as pd**

**import matplotlib.pyplot as plt**

**data = pd.read\_csv("tips.csv")**

**plt.bar(data['day'], data['tip'])**

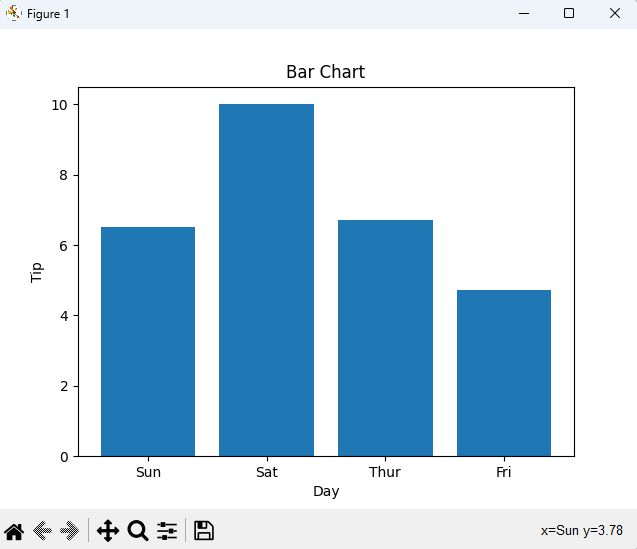
**plt.title("Bar Chart")**

**plt.xlabel('Day')**

**plt.ylabel('Tip')**

**plt.show()**

**Output:**

****

1. **Scatter plot:**

**from bokeh.plotting import figure, output\_file, show**

**from bokeh.palettes import magma**

**import pandas as pd**

**graph = figure(title = "Bokeh Scatter Graph")**

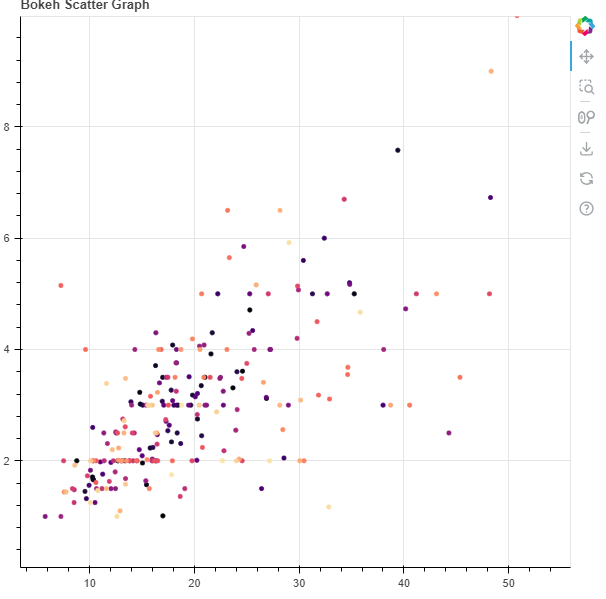
**data = pd.read\_csv("tips.csv")**

**color = magma(256)**

**graph.scatter(data['total\_bill'], data['tip'], color=color)**

**show(graph)**

**Output:**

****

1. **Box plots:**

**import seaborn as sns**

**import matplotlib.pyplot as plt**

**import pandas as pd**

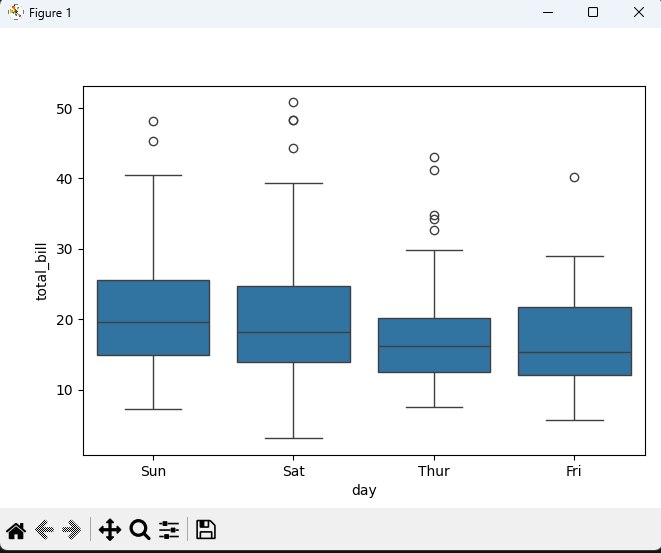
**data = pd.read\_csv("tips.csv")**

**df = pd.DataFrame(data)**

**sns.boxplot(x="day", y="total\_bill", data =df)**

**plt.show()**

**Output:**

****

**Conclusion:**

In conclusion, leveraging Python with libraries like Matplotlib and Seaborn enables the creation of insightful visualizations, while Scikit-learn enhances statistical data analysis capabilities. The seamless integration of Python and Scilab provides a robust platform for researchers and data scientists to explore, interpret, and derive meaningful insights from diverse datasets.

**Lab Outcome**: Implementation of data visualization and statistical analysis techniques in Python, showcasing skills in creating meaningful Bar Charts, Histograms, Scatter Plots, and more for comprehensive data exploration.

**Submitted Details -**

**Name of Student: Parth Malviya**

**Roll No.: 22**

**Date of Performance: 30 Jan 2024**

**Date of Submission: 27 feb 2024**