

## **Experiment No. - 2**

**Student Name:** ANIKET KUMAR

**Branch:** CSE

**Semester:** 5<sup>th</sup>

**Subject Name:** Machine Learning Lab

**UID:** 20BCS5306

**Section/Group:** 20BCS\_WM-703 / B

**Date of Performance:** 06<sup>th</sup> Sep, 2022

**Subject Code:** 20CSP-317

### **Aim:** Data Visualization

**Objective** – To analyze the data for certain trends, patterns may become difficult if the data is in its raw format. To overcome this data visualization comes into play. Data visualization provides a good, organized pictorial representation of the data which makes it easier to understand, observe, analyze.

### **Steps Involved-**

Python provides various libraries that come with different features for visualizing data.

All these libraries come with different features and can support various types of graphs. The four main libraries are: Matplotlib, Seaborn, Bokeh, Plotly

**S/W Requirement: -** Windows 7 or above, GoogleCollab or VS Code.

### **Input and Output –**

```
import pandas as pd
```

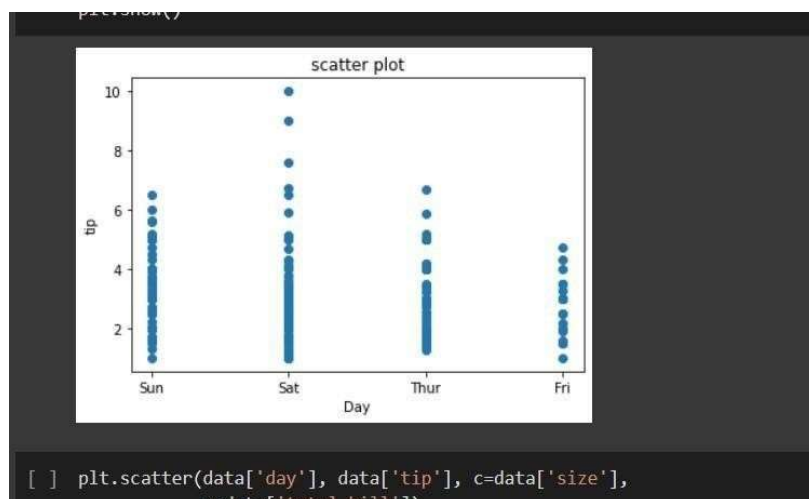
```
data=pd.read_csv("/content/tips.csv")  
data.head(10)
```

	total_bill	tip	sex	smoker	day	time	size	price_per_person	Payer Name	CC Number	Payment ID
0	16.99	1.01	Female	No	Sun	Dinner	2	8.49	Christy Cunningham	3560325168603410	Sun2959
1	10.34	1.66	Male	No	Sun	Dinner	3	3.45	Douglas Tucker	4478071379779230	Sun4608
2	21.01	3.50	Male	No	Sun	Dinner	3	7.00	Travis Walters	6011812112971322	Sun4458
3	23.68	3.31	Male	No	Sun	Dinner	2	11.84	Nathaniel Harris	4676137647685994	Sun5260
4	24.59	3.61	Female	No	Sun	Dinner	4	6.15	Tonya Carter	4832732618637221	Sun2251
5	25.29	4.71	Male	No	Sun	Dinner	4	6.32	Erik Smith	213140353657882	Sun9679
6	8.77	2.00	Male	No	Sun	Dinner	2	4.38	Kristopher Johnson	2223727524230344	Sun5985
7	26.88	3.12	Male	No	Sun	Dinner	4	6.72	Robert Buck	3514785077705092	Sun8157
8	15.04	1.96	Male	No	Sun	Dinner	2	7.52	Joseph Mcdonald	3522866365840377	Sun6820
9	14.78	3.23	Male	No	Sun	Dinner	2	7.39	Jerome Abbott	3532124519049786	Sun3776

## Scatter Plot

Scatter plots are used to observe relationships between variables and uses dots to represent the relationship between them. The scatter() method in the matplotlib library is used to draw a scatterplot.

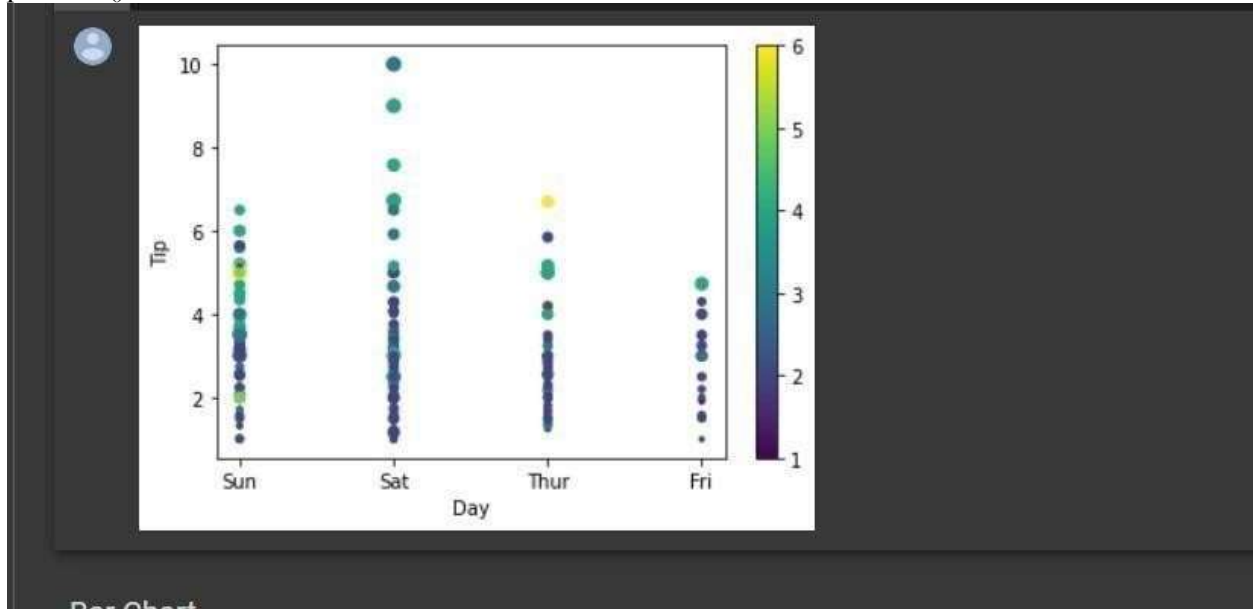
```
pip install matplotlib
import matplotlib.pyplot as plt
plt.scatter(data['day'], data['tip'])
plt.title('scatter plot')
plt.xlabel('Day')
plt.ylabel('tip')
plt.show()
```



```
plt.scatter(data['day'], data['tip'], c=data['size'],
            s=data['total_bill'])
plt.xlabel('Day')
plt.ylabel('Tip')

plt.colorbar()

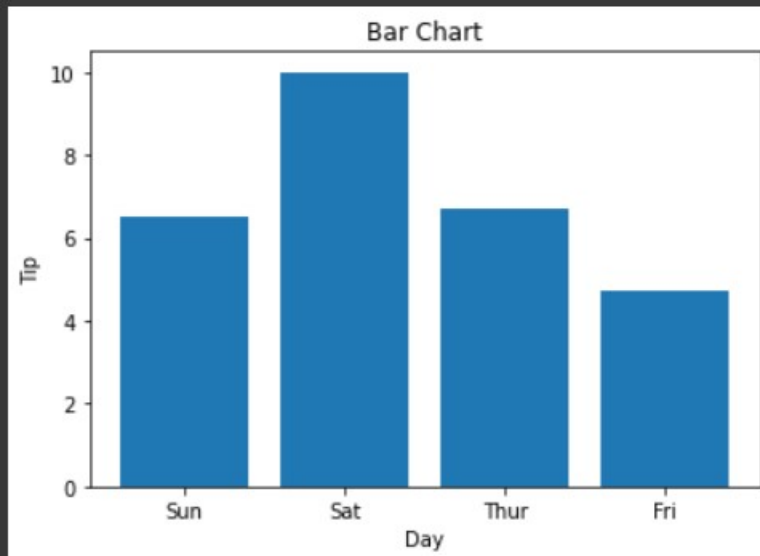
plt.show()
```



## Bar Chart

A bar plot or bar chart is a graph that represents the category of data with rectangular bars with lengths and heights that is proportional to the values which they represent. It can be created using the `bar()` method.

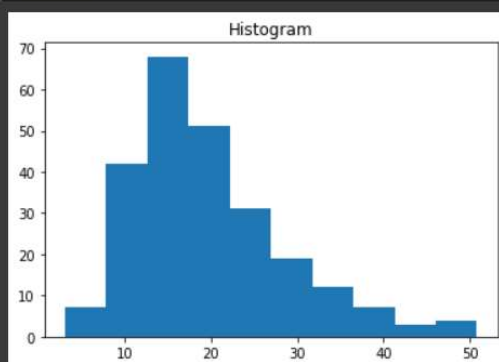
```
plt.bar(data['day'], data['tip'])
plt.title("Bar Chart")
plt.xlabel('Day')
plt.ylabel('Tip')
plt.show()
```



## Histogram

A histogram is basically used to represent data in the form of some groups. It is a type of bar plot where the X-axis represents the bin ranges while the Y-axis gives information about frequency. The `hist()` function is used to compute and create a histogram. In histogram, if we pass categorical data then it will automatically compute the frequency of that data i.e. how often each value occurred.

```
plt.hist(data['total_bill'])
plt.title("Histogram")
plt.show()
```



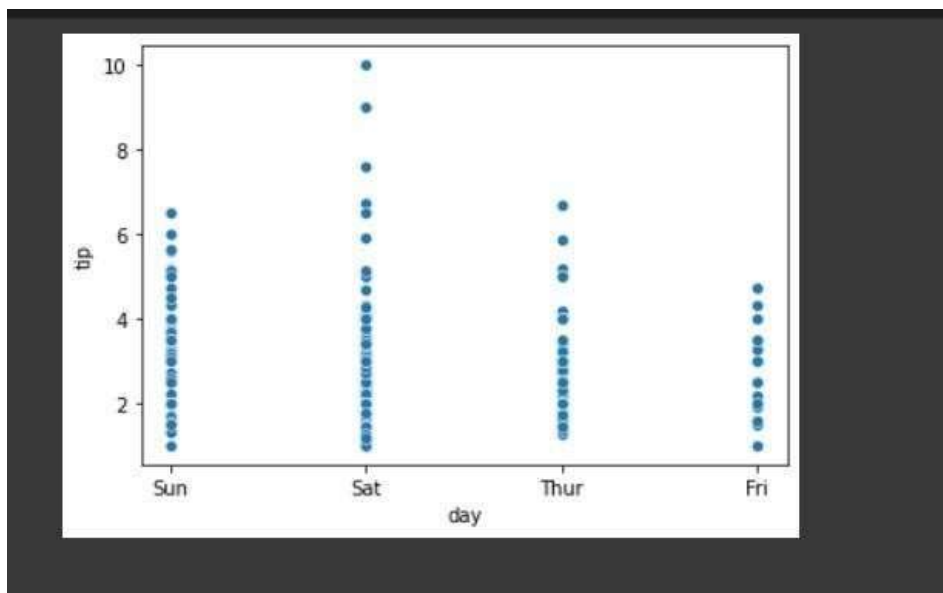
## Seaborn

Seaborn is a high-level interface built on top of the Matplotlib. It provides beautiful design styles and color palettes to make more attractive graphs.

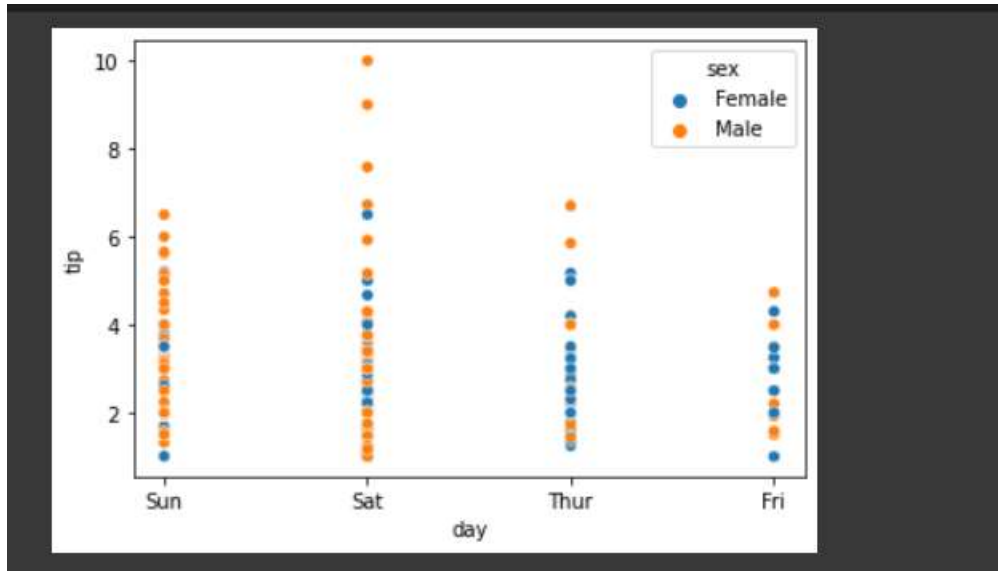
## Scatter Plot

Scatter plot is plotted using the scatterplot() method. This is similar to Matplotlib, but additional argument data is required.

```
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
sns.scatterplot(x='day', y='tip', data=data,)
plt.show()
```



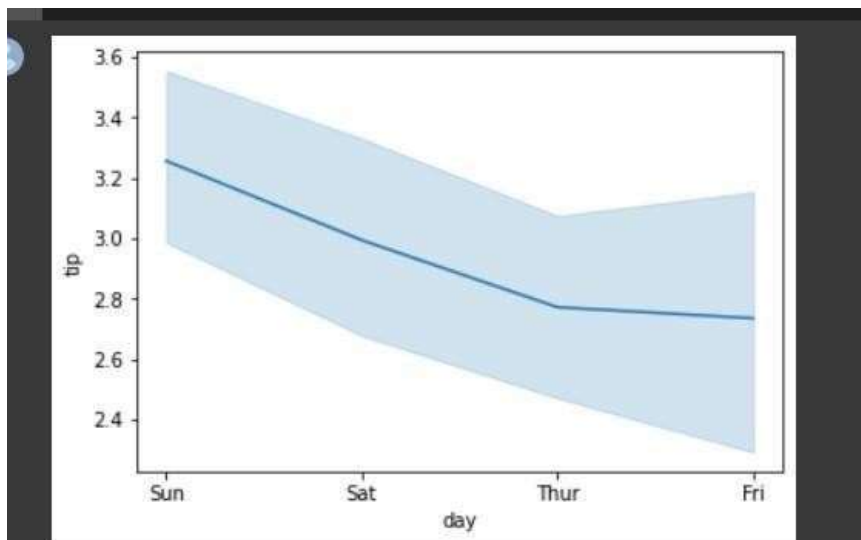
```
sns.scatterplot(x='day', y='tip', data=data, hue='sex')
plt.show()
```



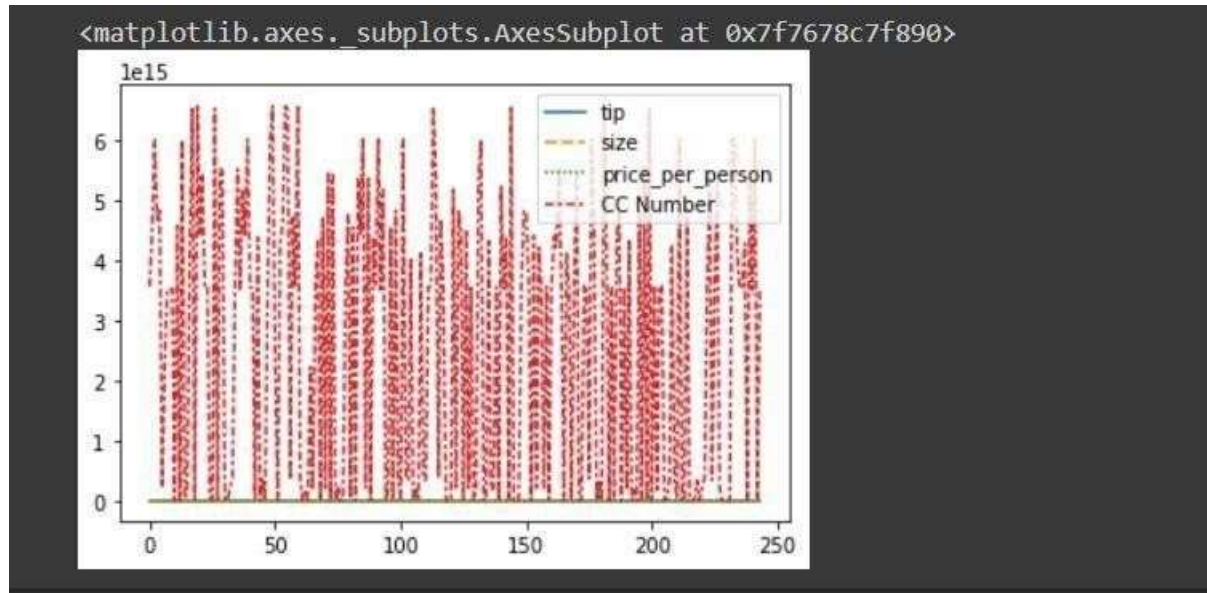
### Line Plot

Line Plot in Seaborn plotted using the `lineplot()` method. In this, we can pass only the `data` argument also.

```
sns.lineplot(x='day', y='tip', data=data)
plt.show()
```



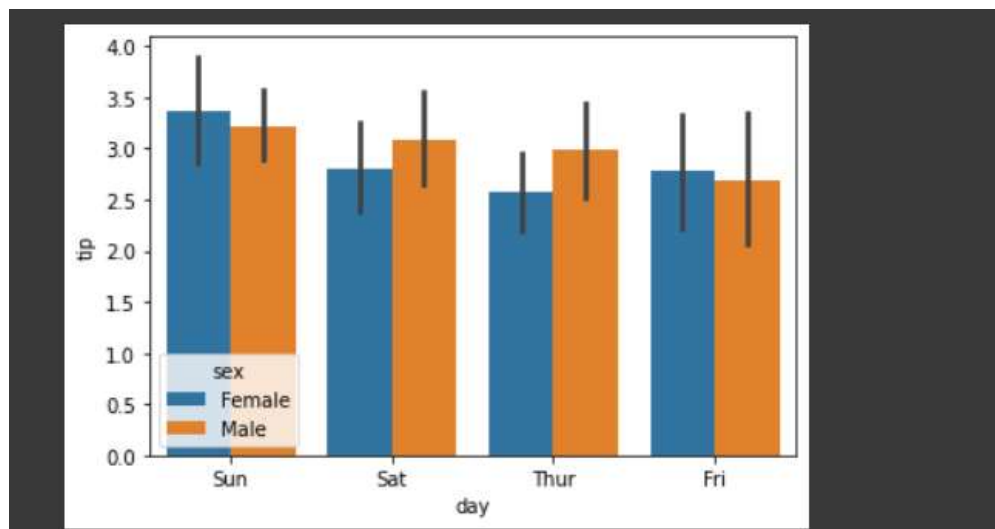
```
sns.lineplot(data=data.drop(['total_bill'], axis=1))
```



## Bar Plot

Bar Plot in Seaborn can be created using the `barplot()` method.

```
sns.barplot(x='day', y='tip', data=data, hue='sex')  
plt.show()
```

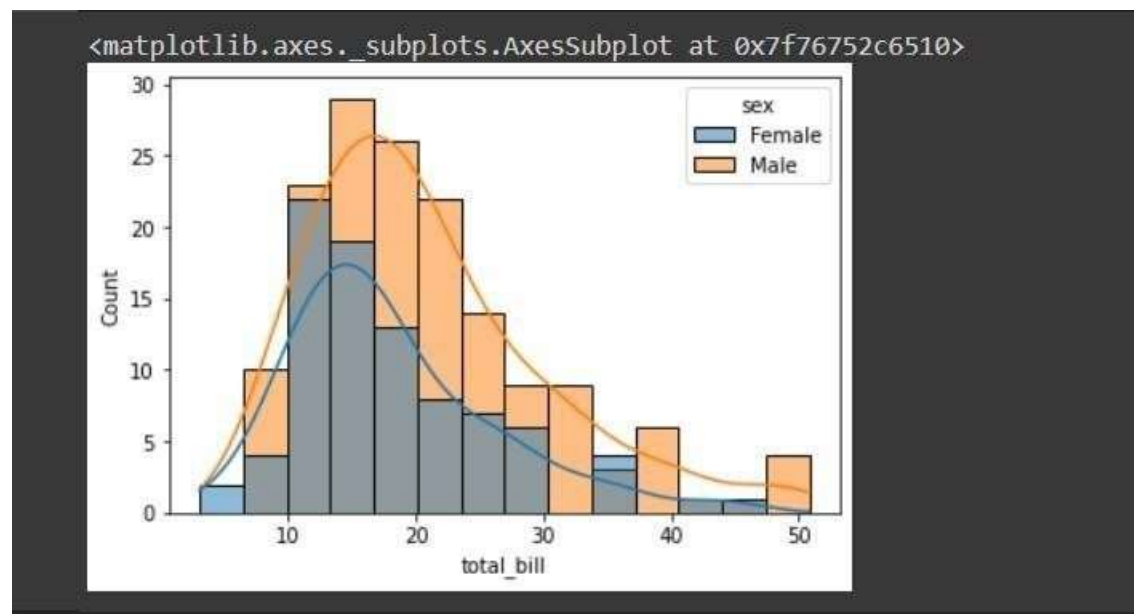




## Histogram

The histogram in Seaborn can be plotted using the `histplot()` function.

```
sns.histplot(x='total_bill', data=data, kde=True, hue='sex')
```



## Bokeh

Bokeh is mainly famous for its interactive charts visualization. Bokeh renders its plots using HTML and JavaScript that uses modern web browsers for presenting elegant, concise construction of novel graphics with high-level interactivity.

## Scatter Plot

Scatter Plot in Bokeh can be plotted using the `scatter()` method of the plotting module. Here pass the x and y coordinates respectively.

```
pip install bokeh
```

```
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(245)
graph.scatter(data['total_bill'], data['tip'], color=color)
show(graph)
```

```
show(graph)
```

```
BokehUserWarning: ColumnDataSource's columns must be of the same length. Current lengths: ('line_color', 245), ('x', 244), ('y', 244)
BokehUserWarning: ColumnDataSource's columns must be of the same length. Current lengths: ('fill_color', 245), ('line_color', 245), ('x', 244), ('y', 244)
```

## Line Chart

A line plot can be created using the `line()` method of the plotting module.

```
# importing the modules
from bokeh.plotting import figure, output_file, show
import pandas as pd

# instantiating the figure object
graph = figure(title = "Bokeh Bar Chart")

# reading the database
data = pd.read_csv("tips.csv")

# Count of each unique value of tip column
df = data['tip'].value_counts()
df = data['total_bill'].value_counts()

# plotting the graph
graph.line(df, data['tip'])
graph.line(df, data['total_bill'])

# displaying the model
show(graph)
```

```
BokehUserWarning: ColumnDataSource's columns must be of the same length. Current lengths: ('x', 229), ('y', 244)
BokehUserWarning: ColumnDataSource's columns must be of the same length. Current lengths: ('x', 229), ('y', 244)
```

## Bar Chart



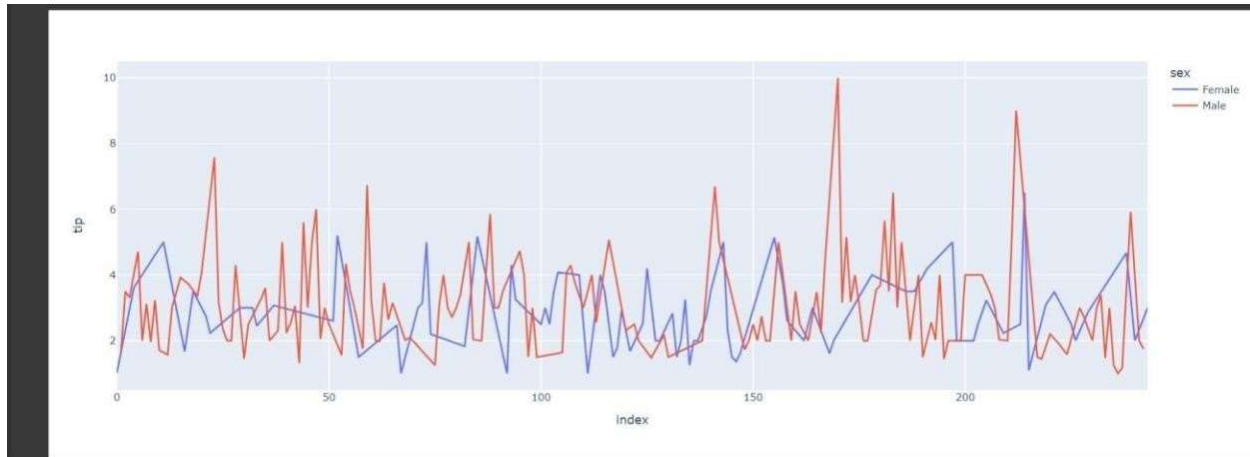
Line plot in Plotly is much accessible and illustrious annexation to plotly which manage a variety of types of data and assemble easy-to-style statistic. With `px.line` each data position is represented as a vertex

```
import plotly.express as px
import pandas as pd

# reading the database
data = pd.read_csv("tips.csv")

# plotting the scatter chart
fig = px.line(data, y='tip', color='sex')

# showing the plot
fig.show()
```



```
import plotly.express as px
import pandas as pd

# reading the database
data = pd.read_csv("tips.csv")

# plotting the scatter chart
fig = px.bar(data, x='day', y='tip', color='sex')

# showing the plot
fig.show()
```

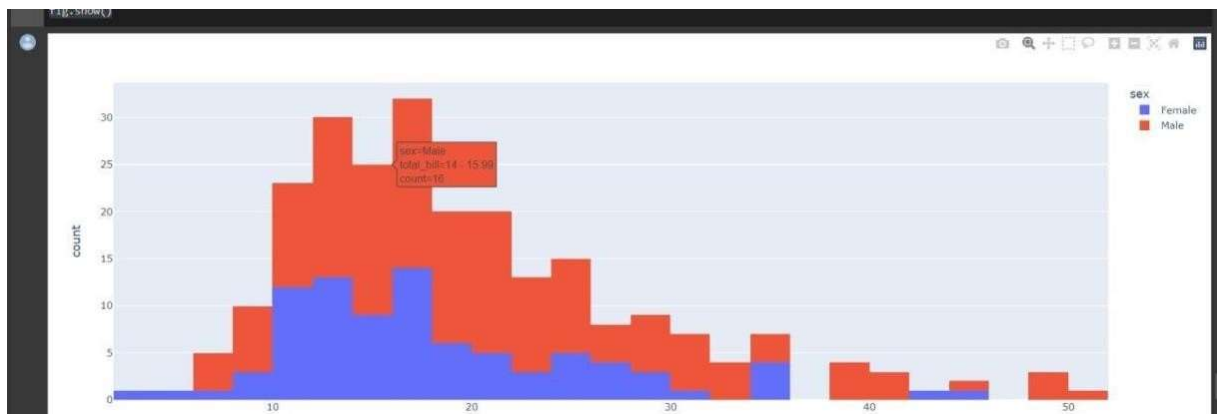


```
import plotly.express as px
import pandas as pd
```

```
# reading the database
data = pd.read_csv("tips.csv")
```

```
# plotting the scatter chart
fig = px.histogram(data, x='total_bill', color='sex')
```

```
# showing the plot
fig.show()
```



---

**Learning outcomes (What I have learnt) -**

1. How to establish relationship between different columns (variables) in a dataset using Graphs.
2. To draw different kind of graphs and charts by using different kind of libraries.
3. To find Relationship between different variables and map different type of Graphs.
4. Detailed introduction to visualization using matplotlib , seaborn , plotly and bokeh libraries.

**Evaluation Grid (To be created as per the SOP and Assessment guidelines by the faculty):**

Sr. No.	Parameters	Marks Obtained	Maximum Marks
1.			
2.			
3.			