Introduction to Matplotlib

Course Code: CPE 031	**Program**: Computer Engineering
Course Title: Visualization and Data Analysis	**Date Performed**: 10/22/24
Section: CPE 21S4	**Date Submitted**: 10/22/24
Name: Dominic Joseph P. Virtucio	**Instructor**: Ma’am Sayo

Intended Learning Outcomes (ILO):

By the end of this laboratory session, learners will be able to:

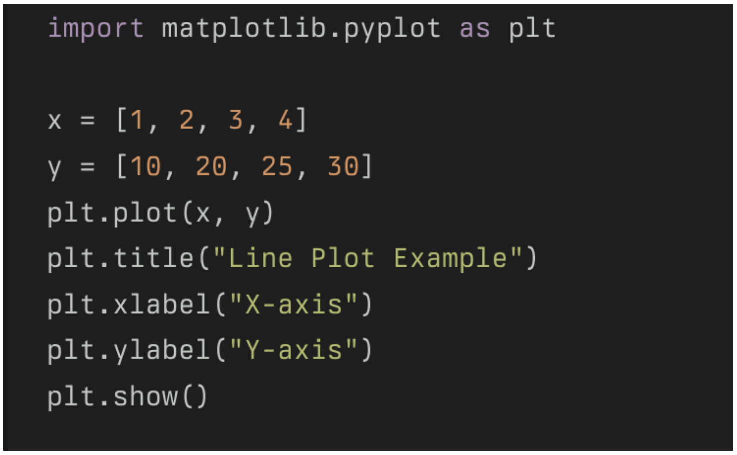
1. Utilize Matplotlib's pyplot interface to create a variety of visualizations, including line plots, scatter plots, histograms, and box plots, demonstrating an understanding of the library's syntax and functionality.

2. Customize visual elements such as titles, labels, and legends to enhance the clarity and aesthetics of their plots, applying best practices in data visualization.

3. Analyze and interpret visual data representations to extract meaningful insights, effectively communicating findings through well-structured graphical presentations.

Part 1: Perform the following codes, and understand the difference between line plot, scatter plot, histogram, bar chart, box plot, and pie chart using matplotlib’s pyplot sub-module. **(Provide a screenshot of your output.)**

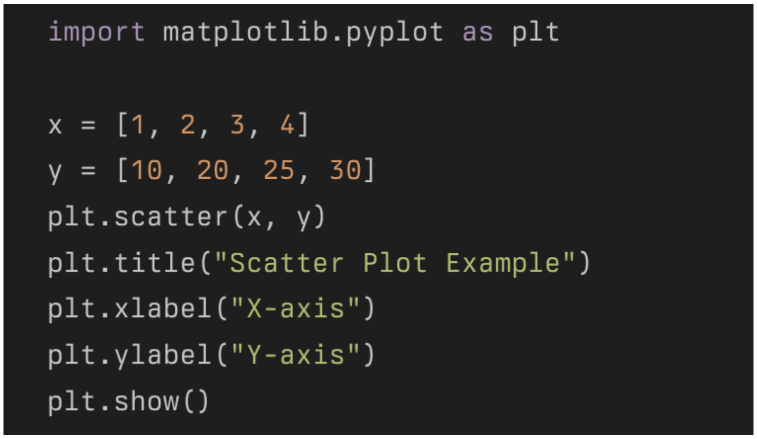
1. Line Plot



```
import matplotlib.pyplot as plt

x = [1, 2, 3, 4]
y = [10, 20, 25, 30]
plt.plot(x, y)
plt.title("Line Plot Example")
plt.xlabel("X-axis")
plt.ylabel("Y-axis")
plt.show()
```

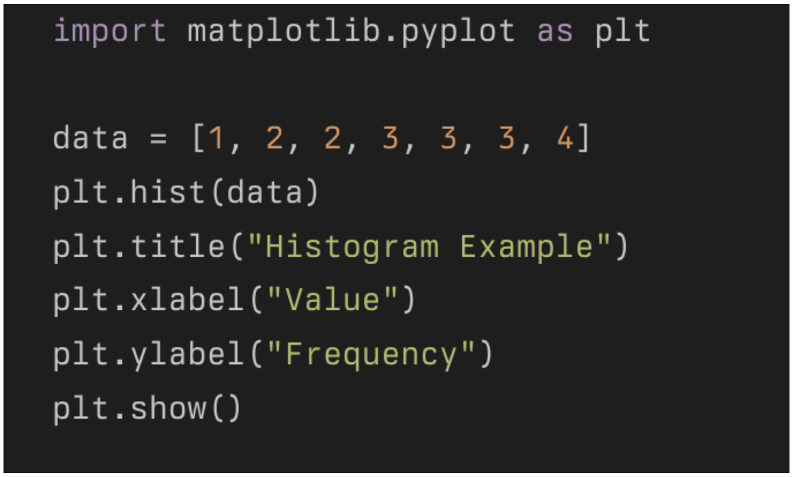
2. Scatter Plot



```python
import matplotlib.pyplot as plt

x = [1, 2, 3, 4]
y = [10, 20, 25, 30]
plt.scatter(x, y)
plt.title("Scatter Plot Example")
plt.xlabel("X-axis")
plt.ylabel("Y-axis")
plt.show()
```

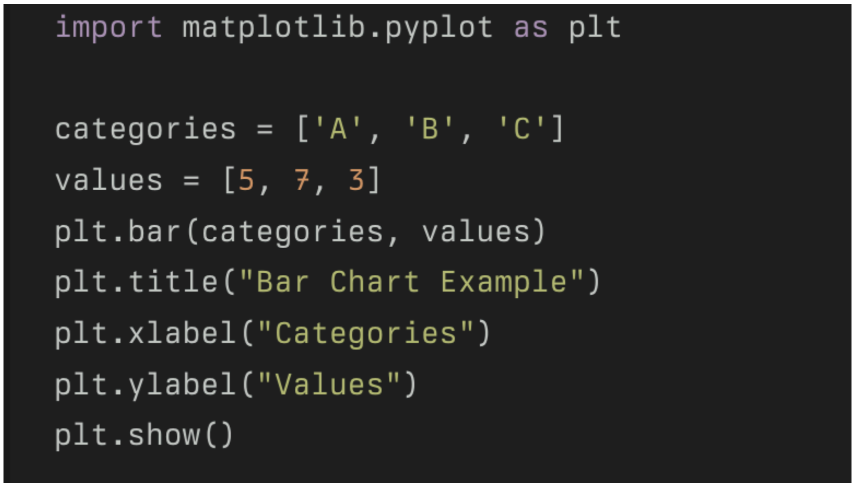
3. Histogram



```python
import matplotlib.pyplot as plt

data = [1, 2, 2, 3, 3, 3, 4]
plt.hist(data)
plt.title("Histogram Example")
plt.xlabel("Value")
plt.ylabel("Frequency")
plt.show()
```

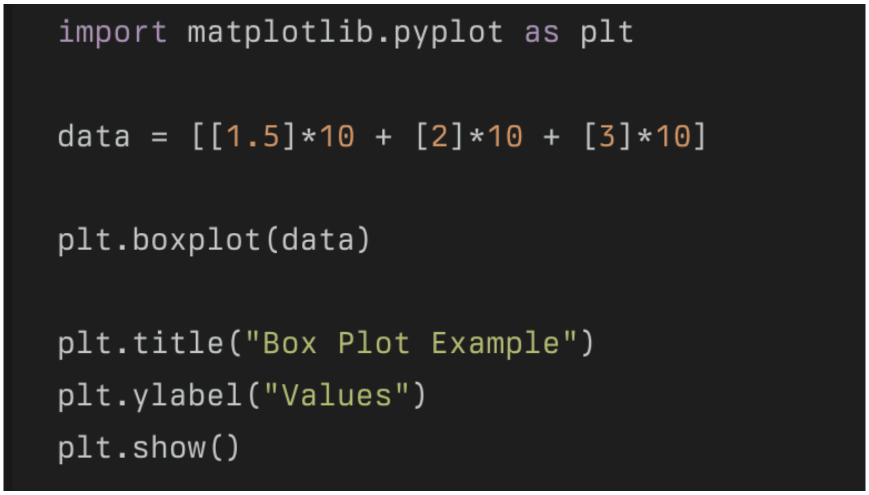
4. Bar Chart



```python
import matplotlib.pyplot as plt

categories = ['A', 'B', 'C']
values = [5, 7, 3]
plt.bar(categories, values)
plt.title("Bar Chart Example")
plt.xlabel("Categories")
plt.ylabel("Values")
plt.show()
```

5. Box plot



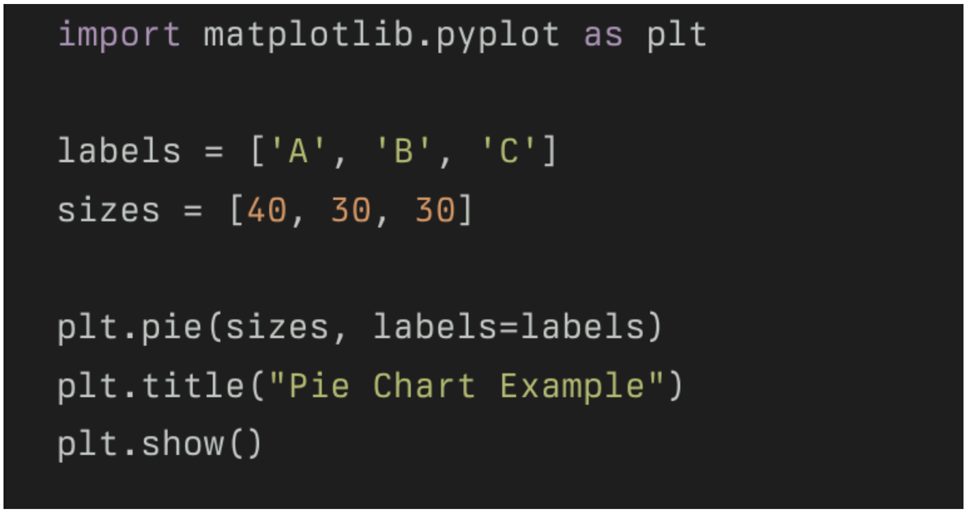
```python
import matplotlib.pyplot as plt

data = [[1.5]*10 + [2]*10 + [3]*10]

plt.boxplot(data)

plt.title("Box Plot Example")
plt.ylabel("Values")
plt.show()
```

6. Pie chart



```
import matplotlib.pyplot as plt

labels = ['A', 'B', 'C']
sizes = [40, 30, 30]

plt.pie(sizes, labels=labels)
plt.title("Pie Chart Example")
plt.show()
```

Part 2: Refer to the instructions below.

1. **Find a dataset for this activity**: Please visit Kaggle and look for a new dataset that would allow you to perform visualization and analysis using matplotlib.

2. **Creating a dataframe from your CSV file:** Once you have successfully loaded your dataset, you need to create a dataframe from your uploaded CSV file

3. **Import the matplotlib.pyplot**

4. **Based on your chosen dataset, you will develop three questions that you will answer using pyplot visualizations. This means that you will need to produce at least three pyplot visualizations. You are also required to make certain customizations on your data vizes.**

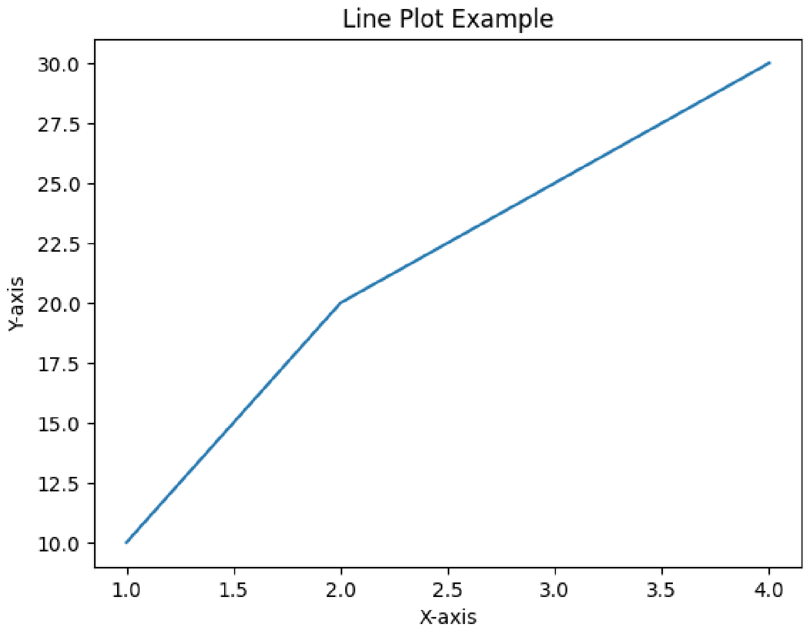
5. Provide observations for each of your data viz, then **produce one insight not longer than five sentences given your three observations**. Your output shall follow this outline:
    a. Introduction (Describe your dataset)
    b. Questions
    c. Visualization and Observation
    d. Insight

6. Your grade will depend on the quality of the question, difficulty/complexity of the visualization, and value-add of the insight that you will generate.

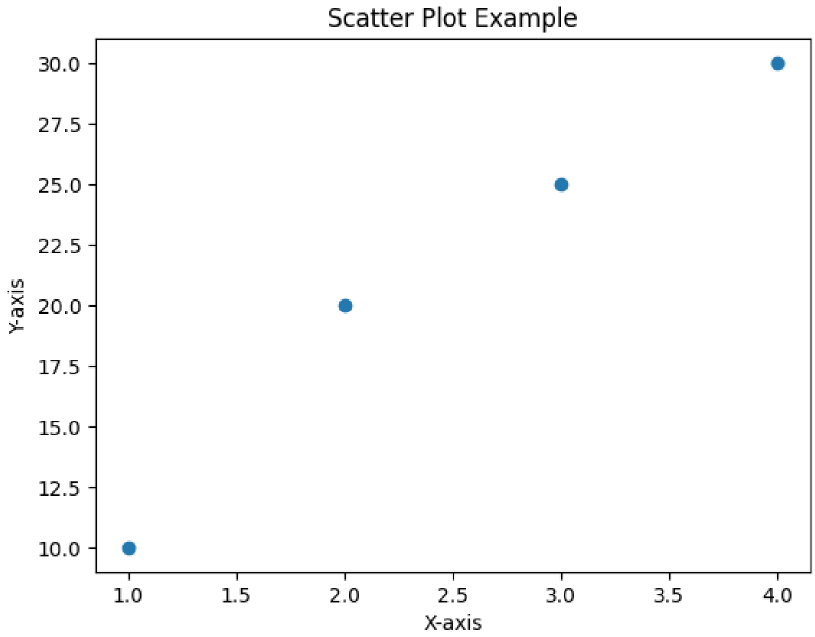
Output

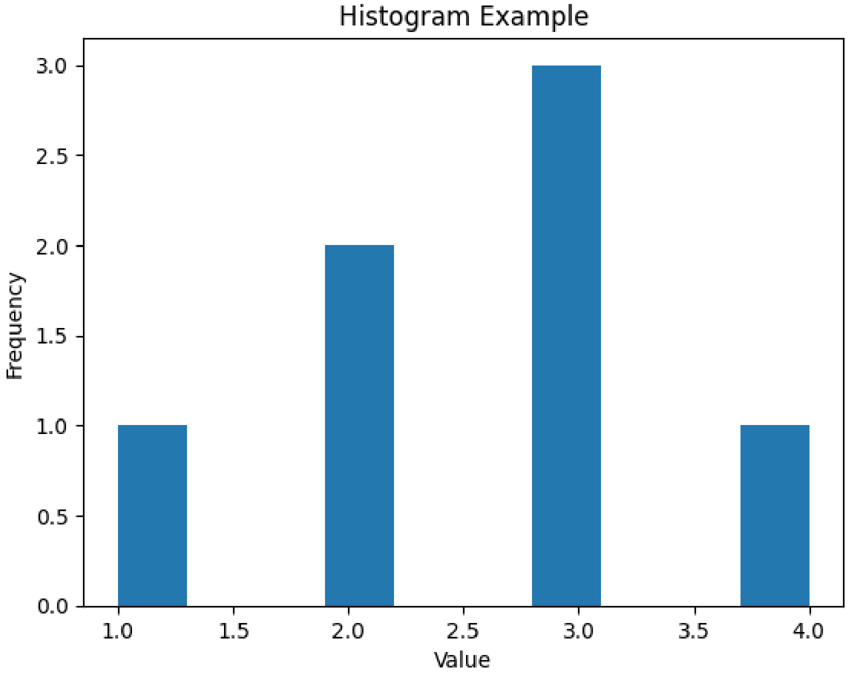
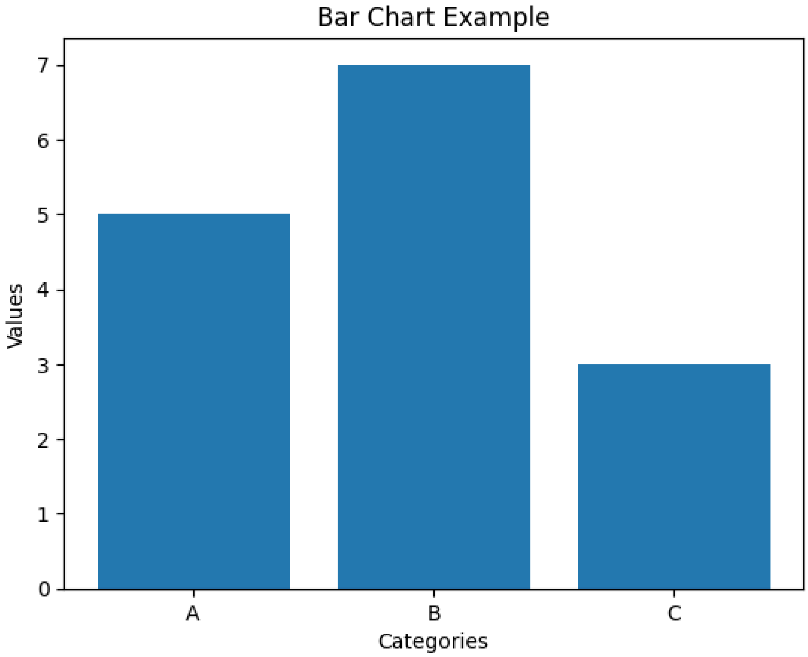
PART 1

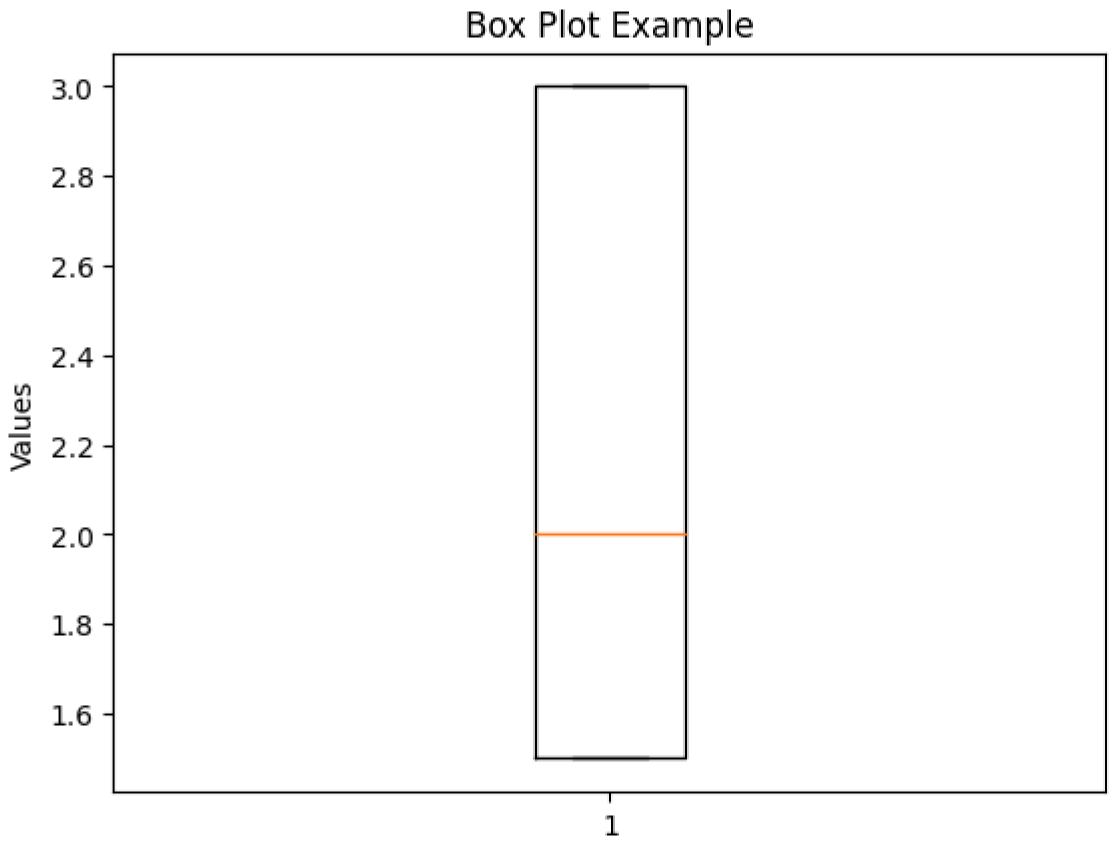
Line Plot



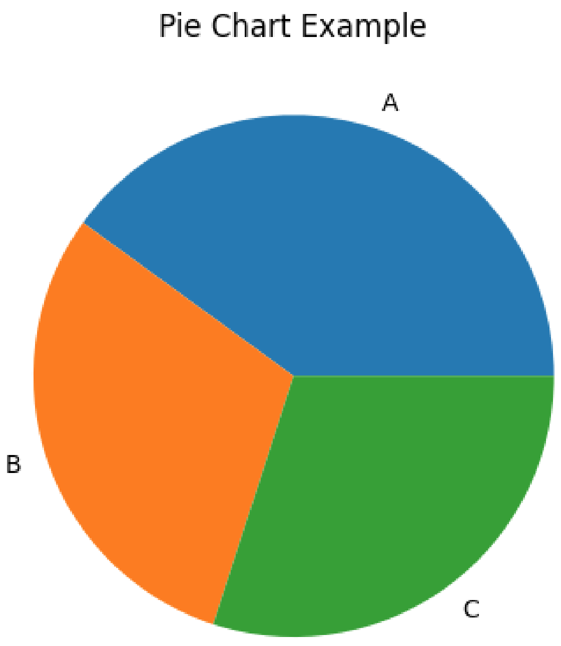
Scatter plot



HIstogram

Bar chart

Box plot

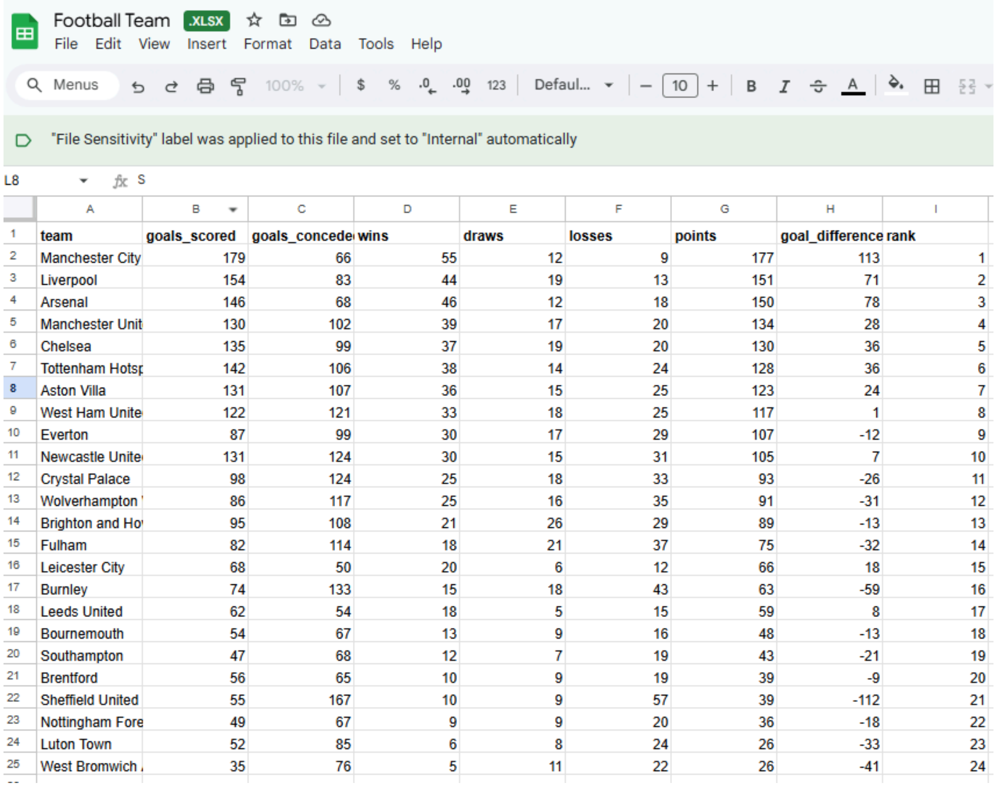

Box Plot Example
Values
3.0
2.8
2.6
2.4
2.2
2.0
1.8
1.6
1

Pie chart


Pie Chart Example
A
B
C

PART 2

Find a dataset for this activity

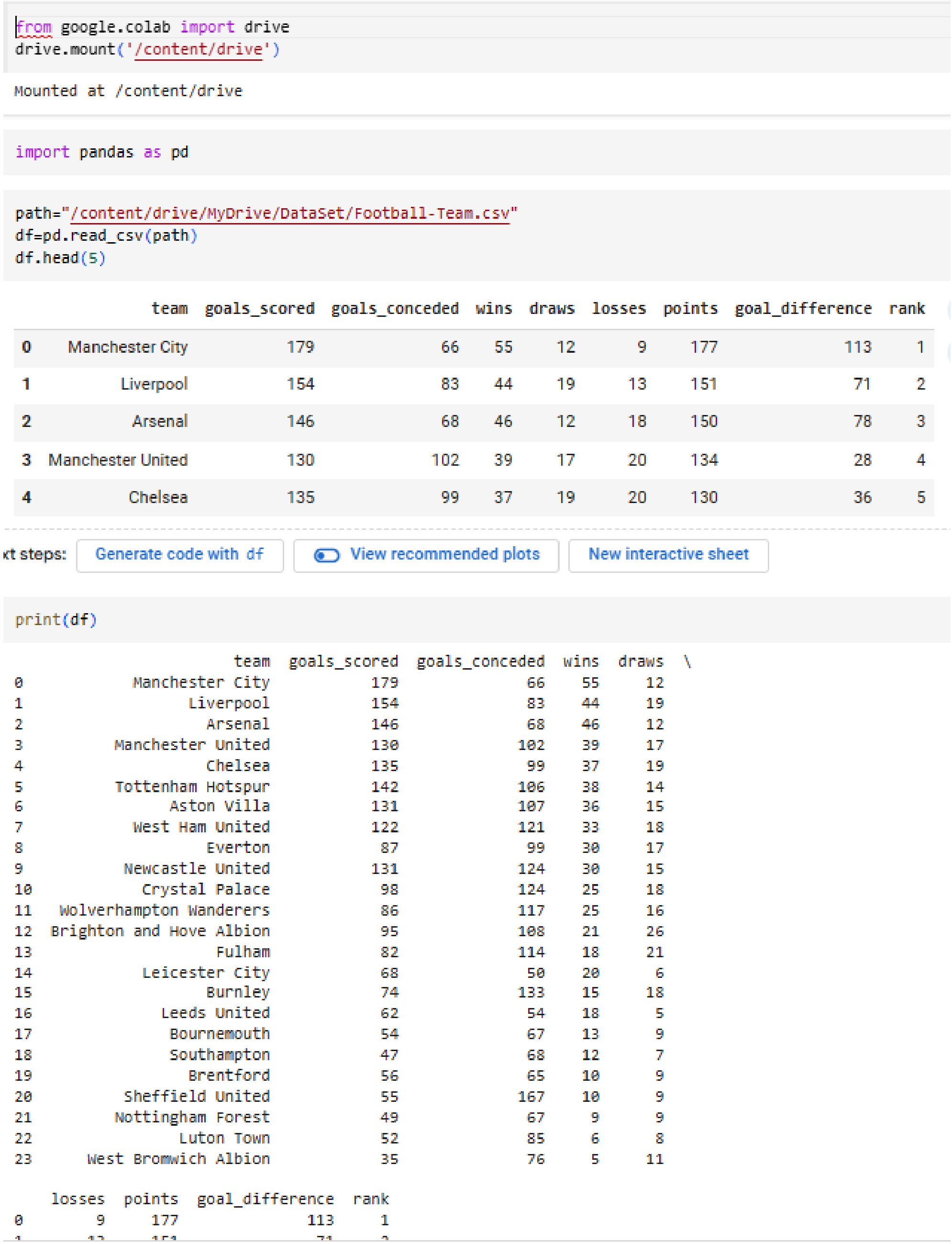


"File Sensitivity" label was applied to this file and set to "Internal" automatically

L8 fx S

	A	B	C	D	E	F	G	H	I
1	team	goals_scored	goals_concede	wins	draws	losses	points	goal_difference	rank
2	Manchester City	179	66	55	12	9	177	113	1
3	Liverpool	154	83	44	19	13	151	71	2
4	Arsenal	146	68	46	12	18	150	78	3
5	Manchester Unit	130	102	39	17	20	134	28	4
6	Chelsea	135	99	37	19	20	130	36	5
7	Tottenham Hotsp	142	106	38	14	24	128	36	6
8	Aston Villa	131	107	36	15	25	123	24	7
9	West Ham Unite	122	121	33	18	25	117	1	8
10	Everton	87	99	30	17	29	107	-12	9
11	Newcastle Unite	131	124	30	15	31	105	7	10
12	Crystal Palace	98	124	25	18	33	93	-26	11
13	Wolverhampton	86	117	25	16	35	91	-31	12
14	Brighton and Ho	95	108	21	26	29	89	-13	13
15	Fulham	82	114	18	21	37	75	-32	14
16	Leicester City	68	50	20	6	12	66	18	15
17	Burnley	74	133	15	18	43	63	-59	16
18	Leeds United	62	54	18	5	15	59	8	17
19	Bournemouth	54	67	13	9	16	48	-13	18
20	Southampton	47	68	12	7	19	43	-21	19
21	Brentford	56	65	10	9	19	39	-9	20
22	Sheffield United	55	167	10	9	57	39	-112	21
23	Nottingham Fore	49	67	9	9	20	36	-18	22
24	Luton Town	52	85	6	8	24	26	-33	23
25	West Bromwich	35	76	5	11	22	26	-41	24

Creating a dataframe from your CSV file:



```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
import pandas as pd
```

```
path="/content/drive/MyDrive/DataSet/Football-Team.csv"
df=pd.read_csv(path)
df.head(5)
```

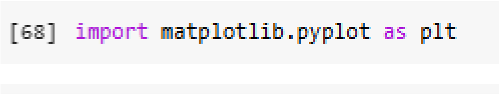
	team	goals_scored	goals_conceded	wins	draws	losses	points	goal_difference	rank
0	Manchester City	179	66	55	12	9	177	113	1
1	Liverpool	154	83	44	19	13	151	71	2
2	Arsenal	146	68	46	12	18	150	78	3
3	Manchester United	130	102	39	17	20	134	28	4
4	Chelsea	135	99	37	19	20	130	36	5

```
print(df)
```

```
                       team  goals_scored  goals_conceded  wins  draws  \
0           Manchester City           179              66    55     12   
1                 Liverpool           154              83    44     19   
2                   Arsenal           146              68    46     12   
3         Manchester United           130             102    39     17   
4                   Chelsea           135              99    37     19   
5         Tottenham Hotspur           142             106    38     14   
6               Aston Villa           131             107    36     15   
7           West Ham United           122             121    33     18   
8                   Everton            87              99    30     17   
9          Newcastle United           131             124    30     15   
10           Crystal Palace            98             124    25     18   
11  Wolverhampton Wanderers            86             117    25     16   
12   Brighton and Hove Albion          95             108    21     26   
13                   Fulham            82             114    18     21   
14           Leicester City            68              50    20      6   
15                  Burnley            74             133    15     18   
16             Leeds United            62              54    18      5   
17              Bournemouth            54              67    13      9   
18              Southampton            47              68    12      7   
19                Brentford            56              65    10      9   
20         Sheffield United            55             167    10      9   
21        Nottingham Forest            49              67     9      9   
22               Luton Town            52              85     6      8   
23     West Bromwich Albion            35              76     5     11   

    losses  points  goal_difference  rank  
0        9     177              113     1  
1       13     151               71     2  
```

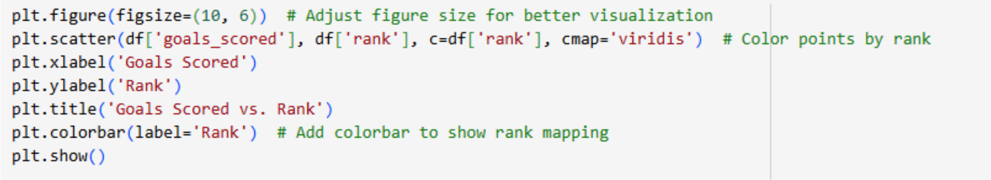
Import the matplotlib.pyplot



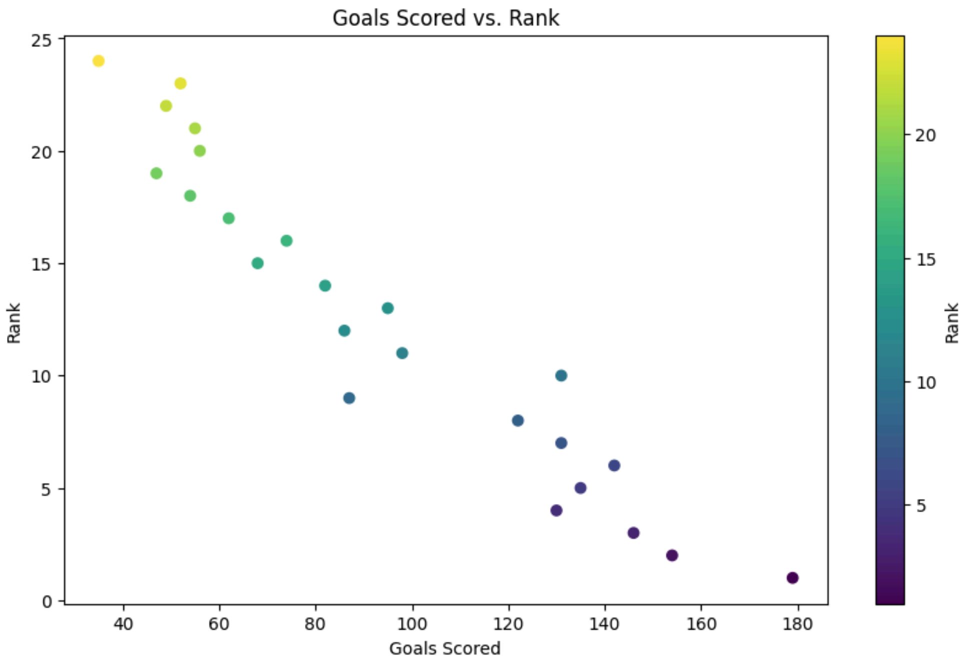
```
[68] import matplotlib.pyplot as plt
```

Developing three questions

1. Question 1: Does a team's ability to score goals consistently correlate with their final league position?



```
plt.figure(figsize=(10, 6))  # Adjust figure size for better visualization
plt.scatter(df['goals_scored'], df['rank'], c=df['rank'], cmap='viridis')  # Color points by rank
plt.xlabel('Goals Scored')
plt.ylabel('Rank')
plt.title('Goals Scored vs. Rank')
plt.colorbar(label='Rank')  # Add colorbar to show rank mapping
plt.show()
```



2. Question 2: Is there a relationship between a team's goal difference and their number of wins?

3. Question 3: How does the distribution of points earned by teams in the top half of the league compare to the distribution of points earned by teams in the bottom half?

Introduction:
The "Football-Team.csv" dataset includes statistics on the play of 24 football teams in a league. It includes data on goals scored, goals conceded, wins, draws, losses, points, goal difference, and final league rank.