Introduction to Matplotlib

Course Code: CPE 031	**Program**: Computer Engineering
Course Title: Visualization and Data Analysis	**Date Performed**: 10/22/2024
Section: CPE21S4	**Date Submitted**: 10/22/2024
Name: Edson Ray E. San Juan	**Instructor**: Engr. Edcel B. Artificio

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.)
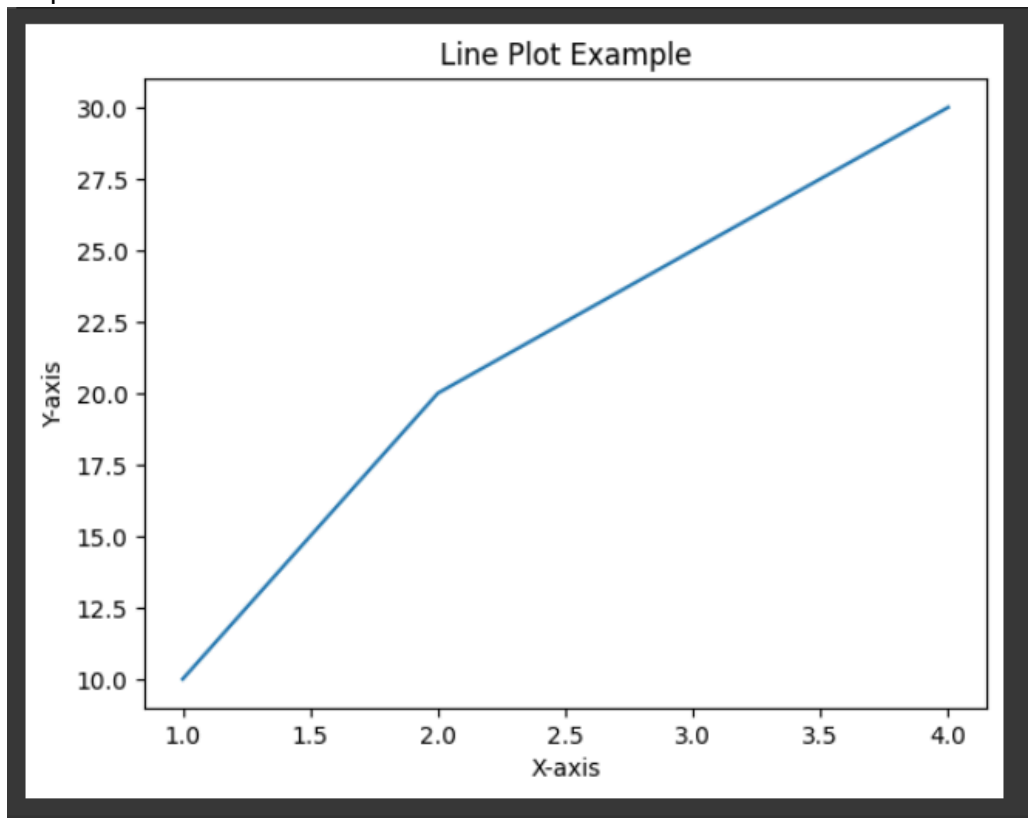
1. Line Plot

```python
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()
```
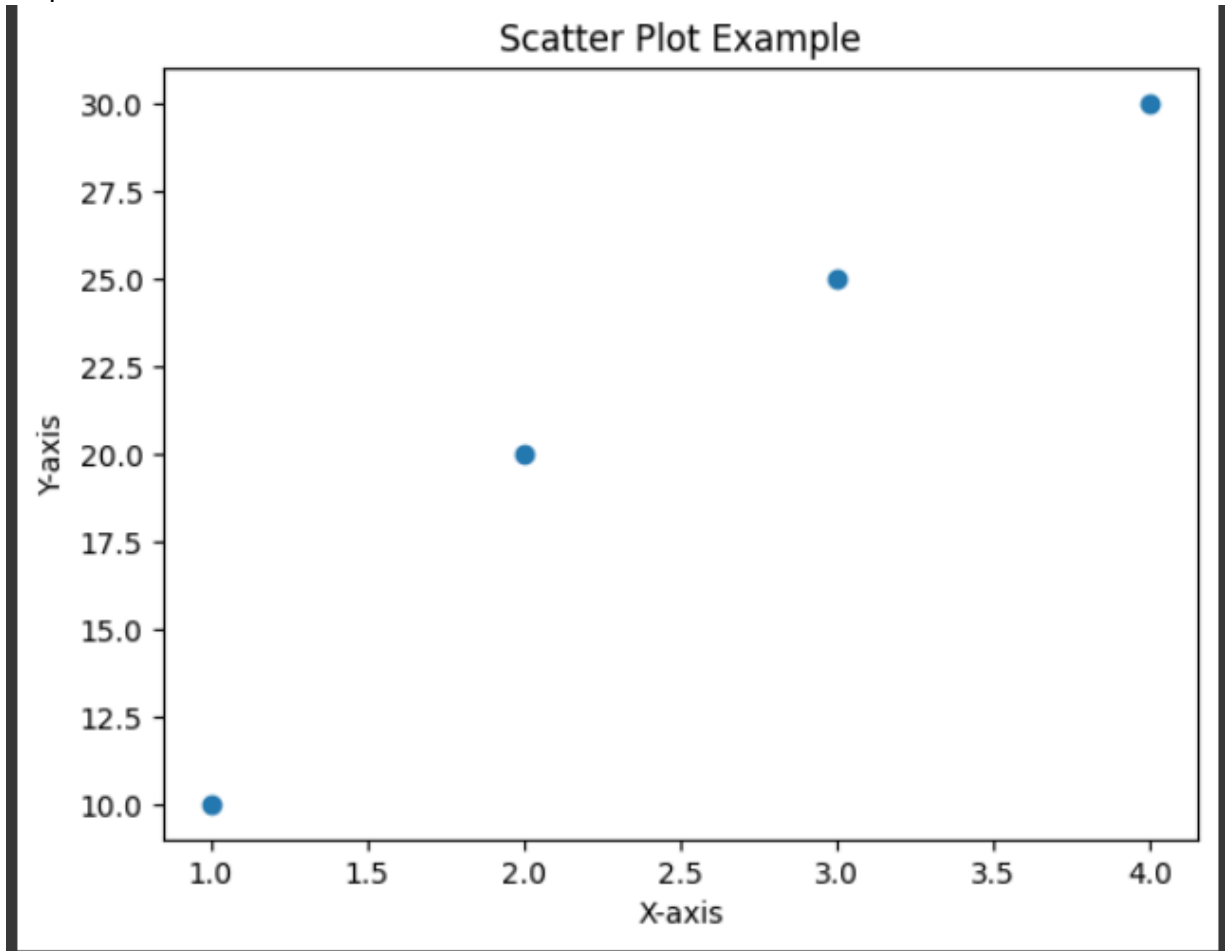
Output:

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()
```
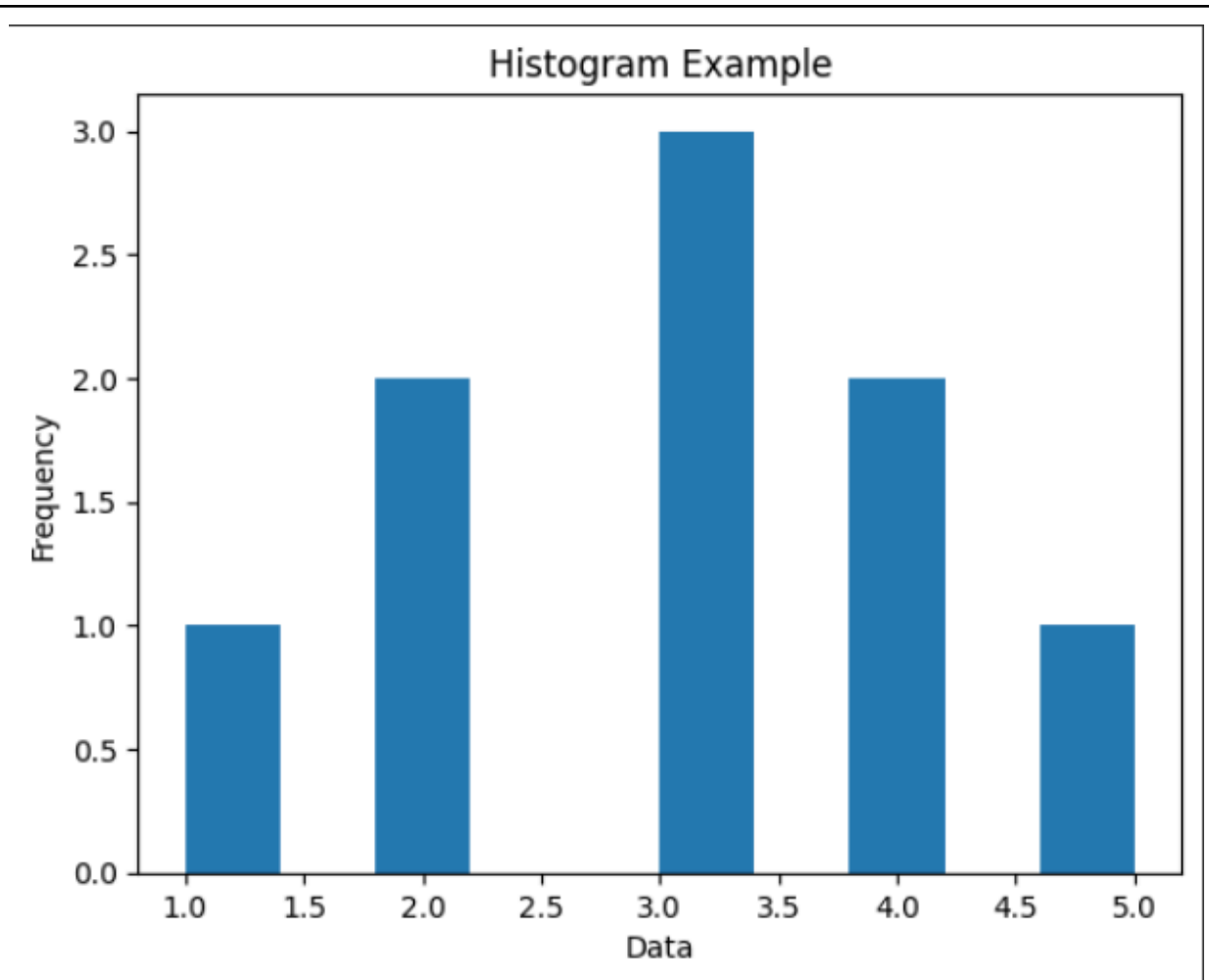
Output:

Scatter Plot Example

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()
```
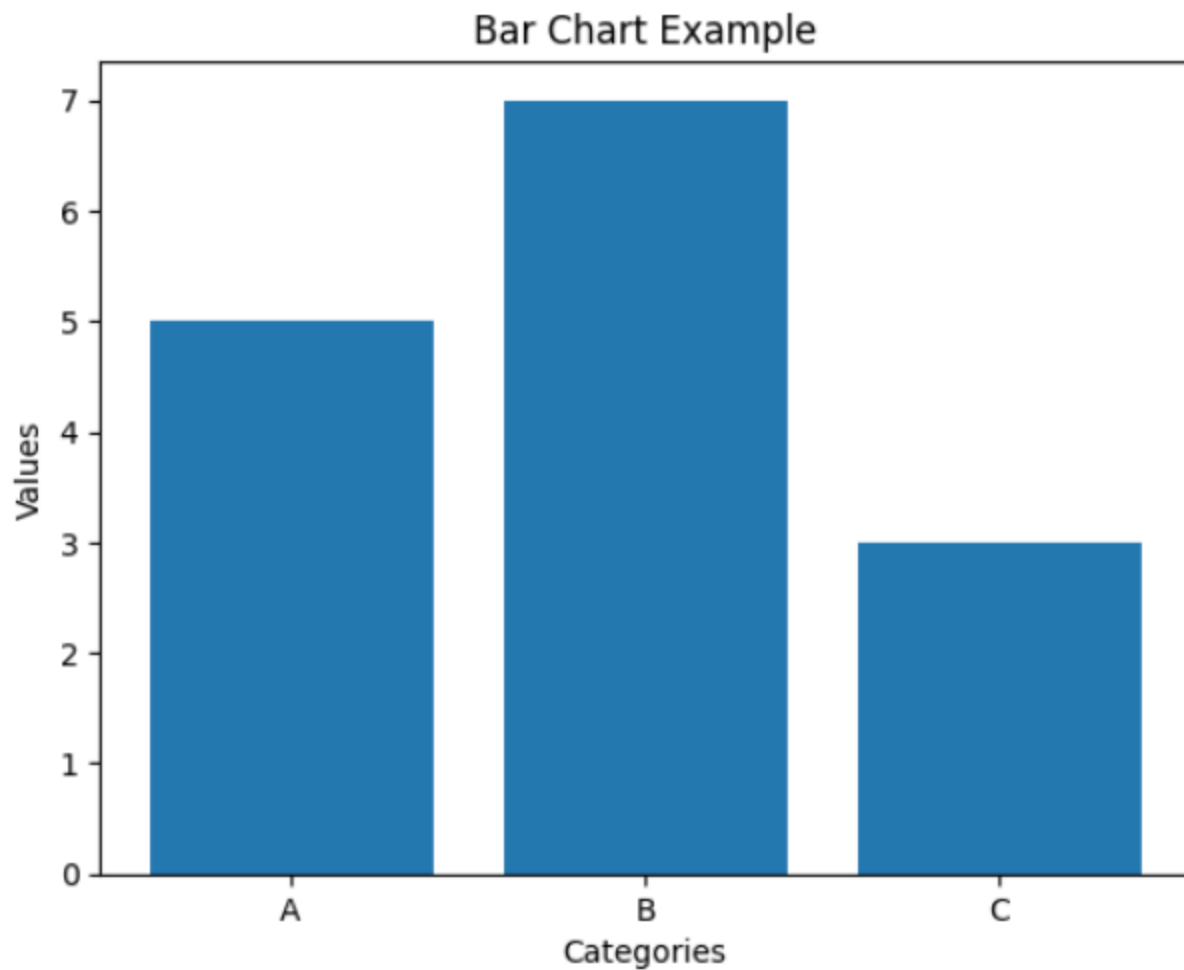
Output:

Histogram Example

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()
```
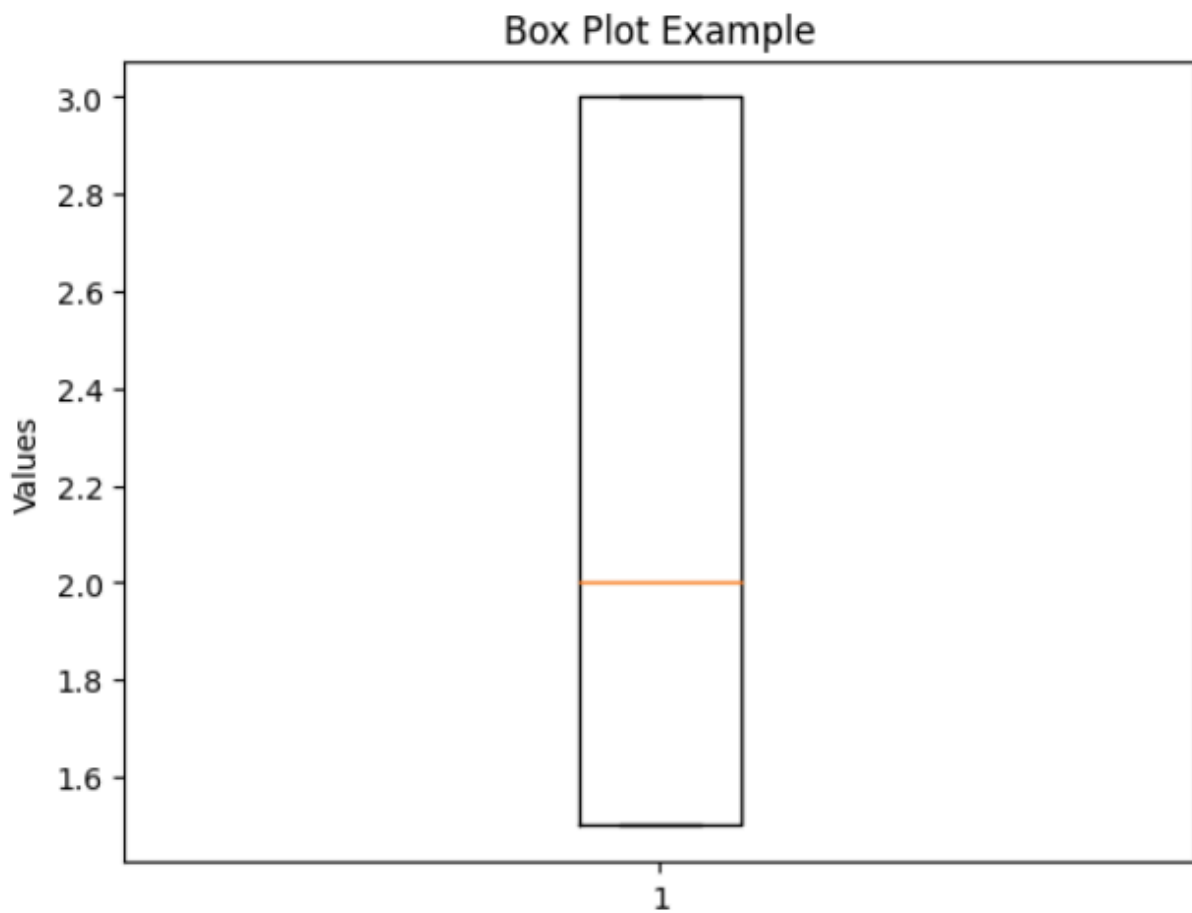
Output:

Bar Chart Example

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()
```
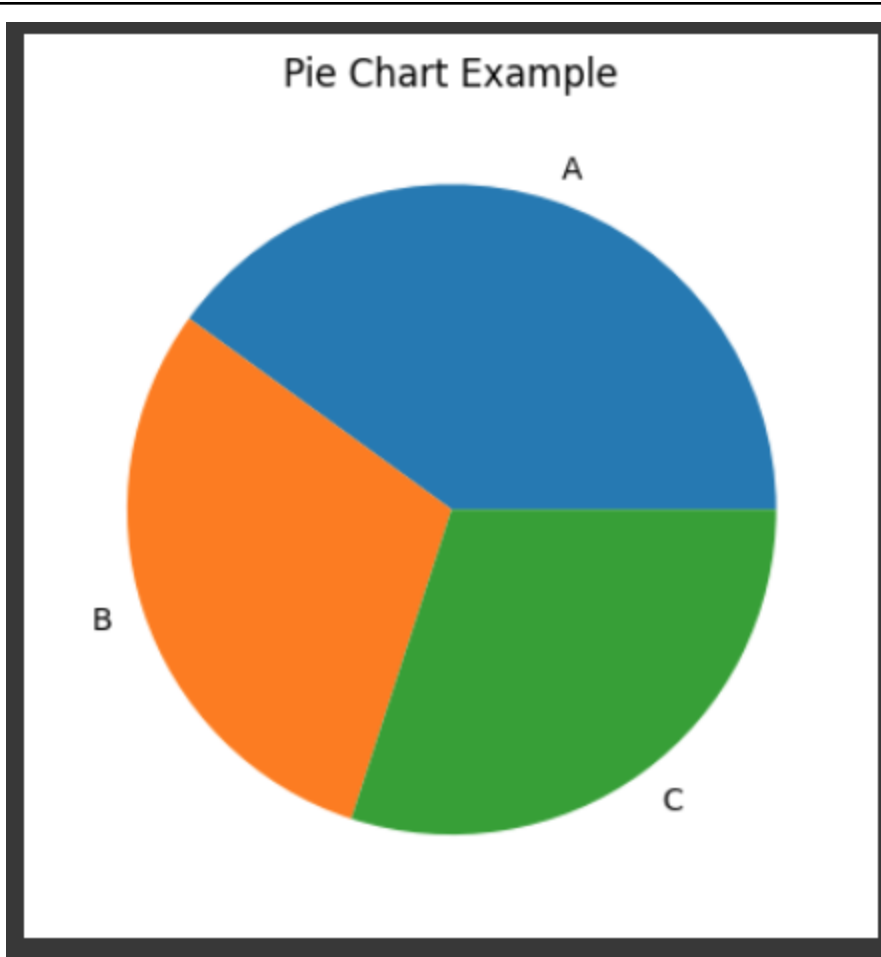
Output:

Box Plot Example

6. Pie chart

```python
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()
```

Output:

Pie Chart Example

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.

```
[101] from google.colab import files
      upload = files.upload()
```

Choose Files gym_memb...tracking.csv
• **gym_members_exercise_tracking**.csv(text/csv) - 65136 bytes, last modified: 10/22
done
Saving gym_members_exercise_tracking.csv to gym_members_exercise_track

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

```
import pandas as pd
df = pd.read_csv('gym_members_exercise_tracking.csv')
sabcar = pd.DataFrame(df)
print(sabcar)
```
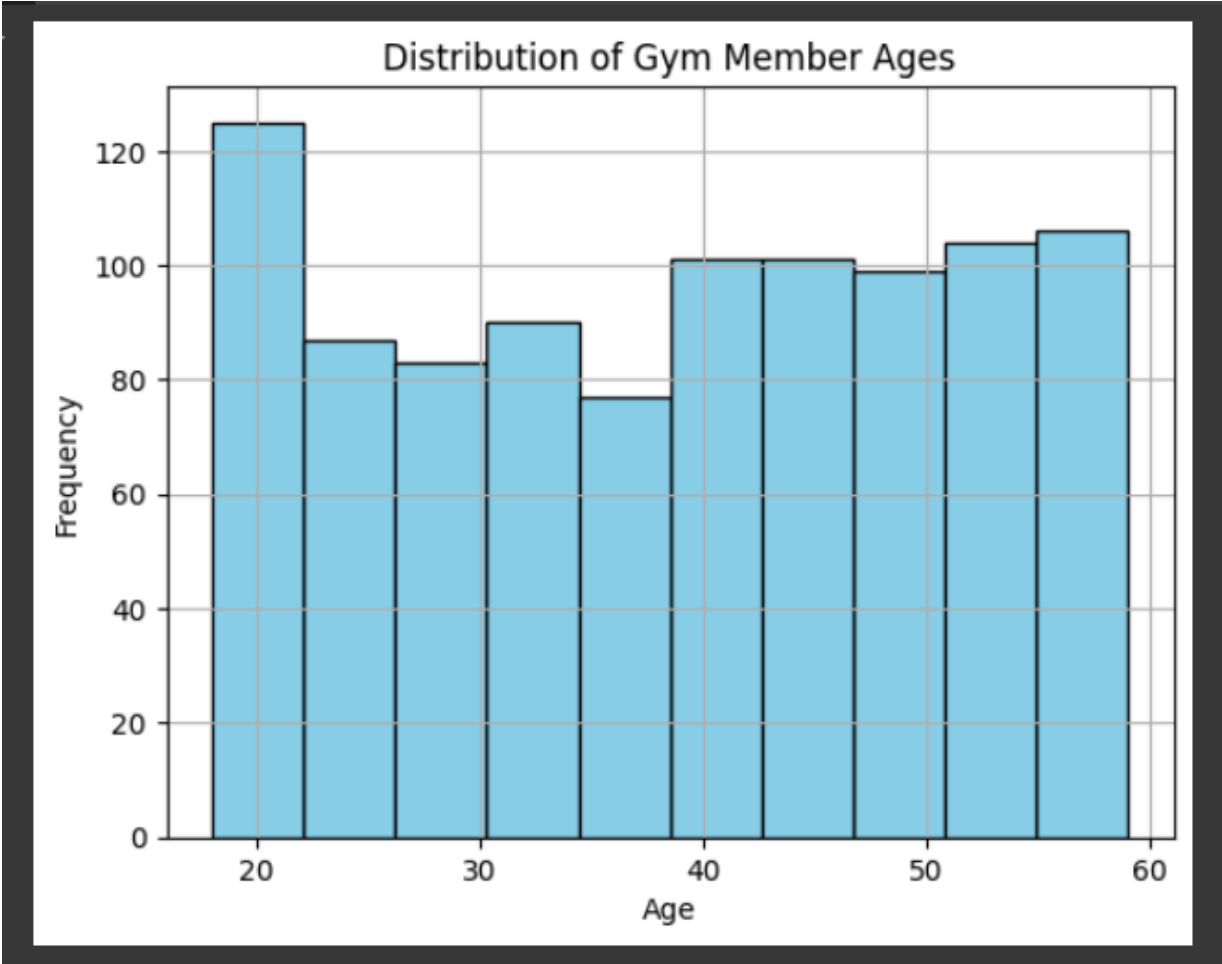
3. **Import the matplotlib.pyplot**

```
import matplotlib.pyplot as plt
```

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.**
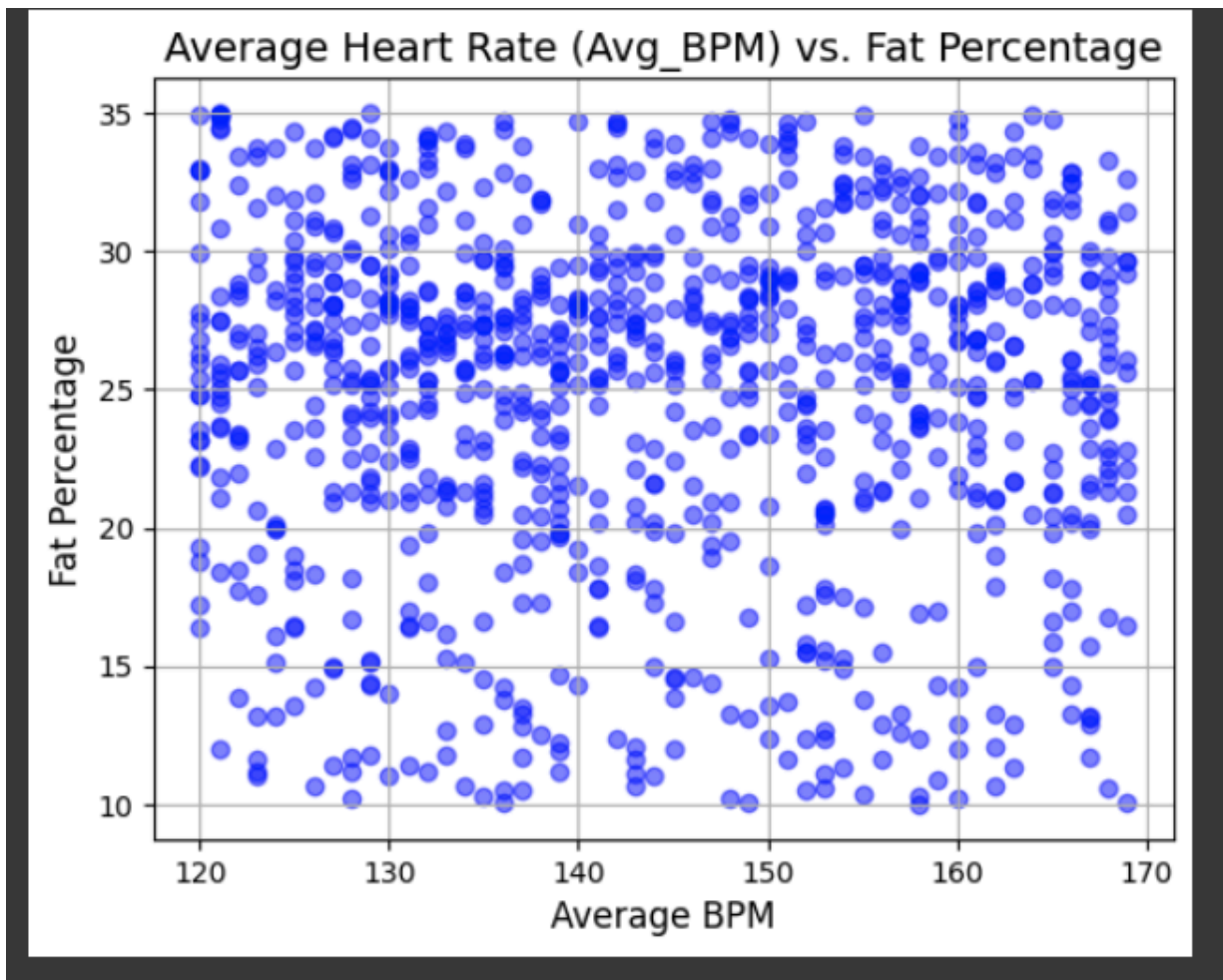
Questions:

- What is the distribution of sales amounts?
- Is there a relationship between Average Heart Rate and Fat Percentage?
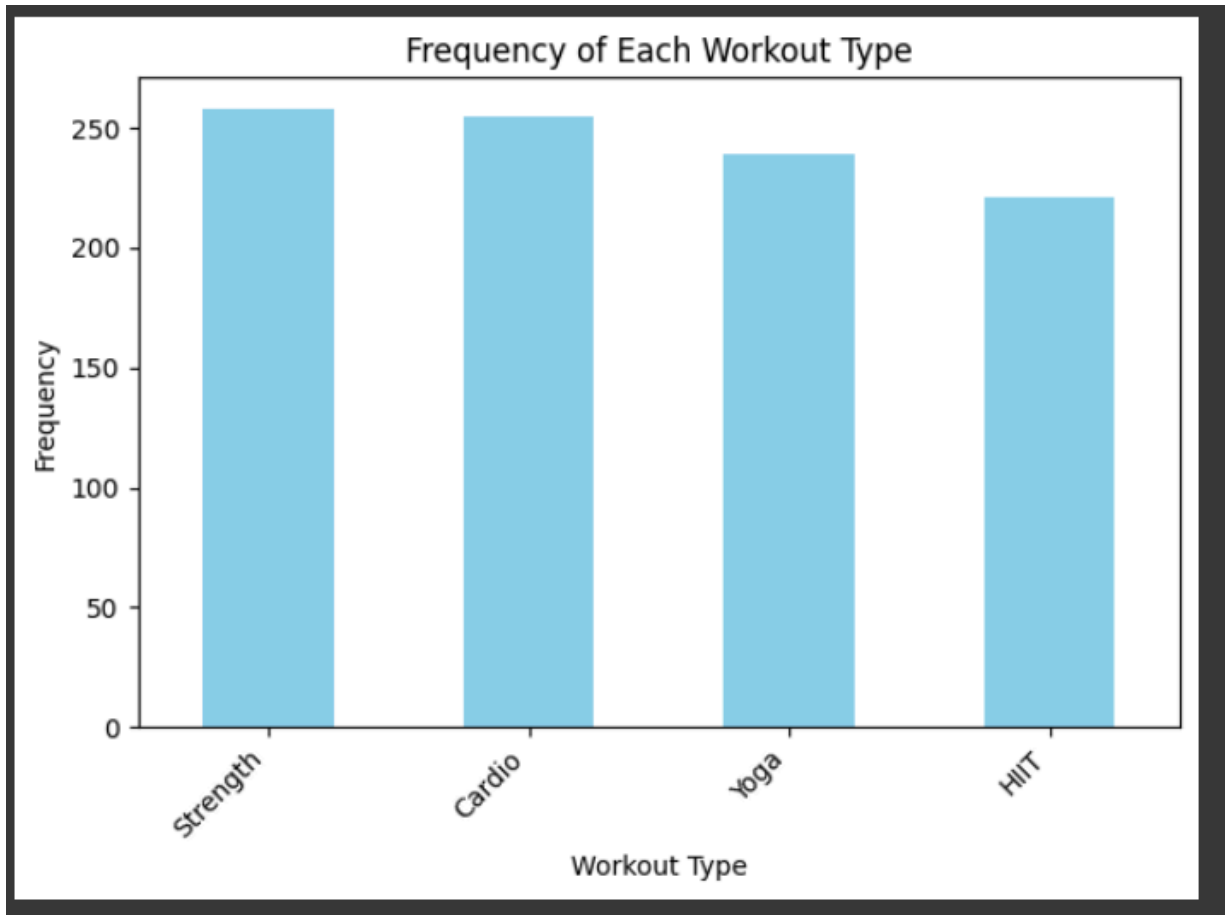- What is the most common work out type?

Histogram:

Distribution of Gym Member Ages

Scatter Plot:

Average Heart Rate (Avg_BPM) vs. Fat Percentage

Bar Chart:

Frequency of Each Workout Type

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)
      i. This dataset represents gym members' exercise tracking data, including various attributes such as age, gender, weight, height, heart rate metrics, exercise session details, body composition, and workout habits. Each record reflects an individual's fitness profile and exercise performance during a gym session.
   b. Questions
      i. How do session duration and workout type correlate with calories burned?
      ii. How does heart rate (both maximum and average BPM) relate to workout type?
      iii. Is there a notable difference in fat percentage between different workout types?

   c. Visualization and Observation
      i. **Session Duration vs. Calories Burned**

Observation: Individuals who spend more time exercising, particularly in high-intensity activities like HIIT and strength training, tend to burn significantly more calories.

ii. **Max BPM vs. Workout Type**
Observation: Strength and HIIT workouts tend to produce the highest maximum BPM, indicating a higher cardiovascular load during these workouts compared to others like yoga or cardio.

iii. **Fat Percentage vs. Workout Type**
Observation: Members who engage in strength training or HIIT tend to have lower fat percentages compared to those who primarily do yoga or cardio, indicating these workouts may be more effective in fat reduction.

d. Insight
i. High-intensity workouts like strength training and HIIT not only elevate heart rate more significantly but also result in greater caloric expenditure and lower fat percentages among participants, suggesting their efficiency in both cardiovascular conditioning and fat reduction.

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.