DAY 3 LAB EXPERIMENTS

Name: Rakshitha V B

Reg.no: 192324004

Dept: B.Tech AI & DS

Scenario : You are a data scientist working for a company that sells products online. You have been tasked with creating a simple plot to show the sales of a product over time.

Question:

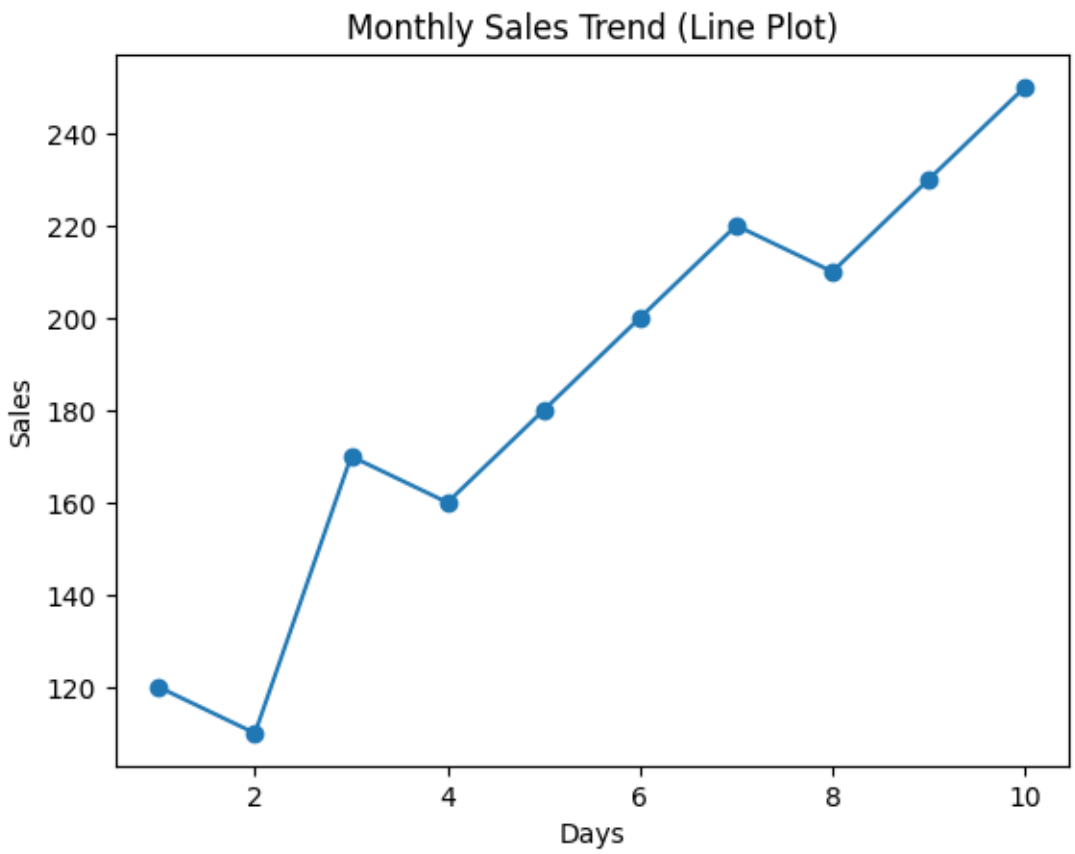
1. Write code to create a simple line plot in Python using Matplotlib to predict sales happened in a month?

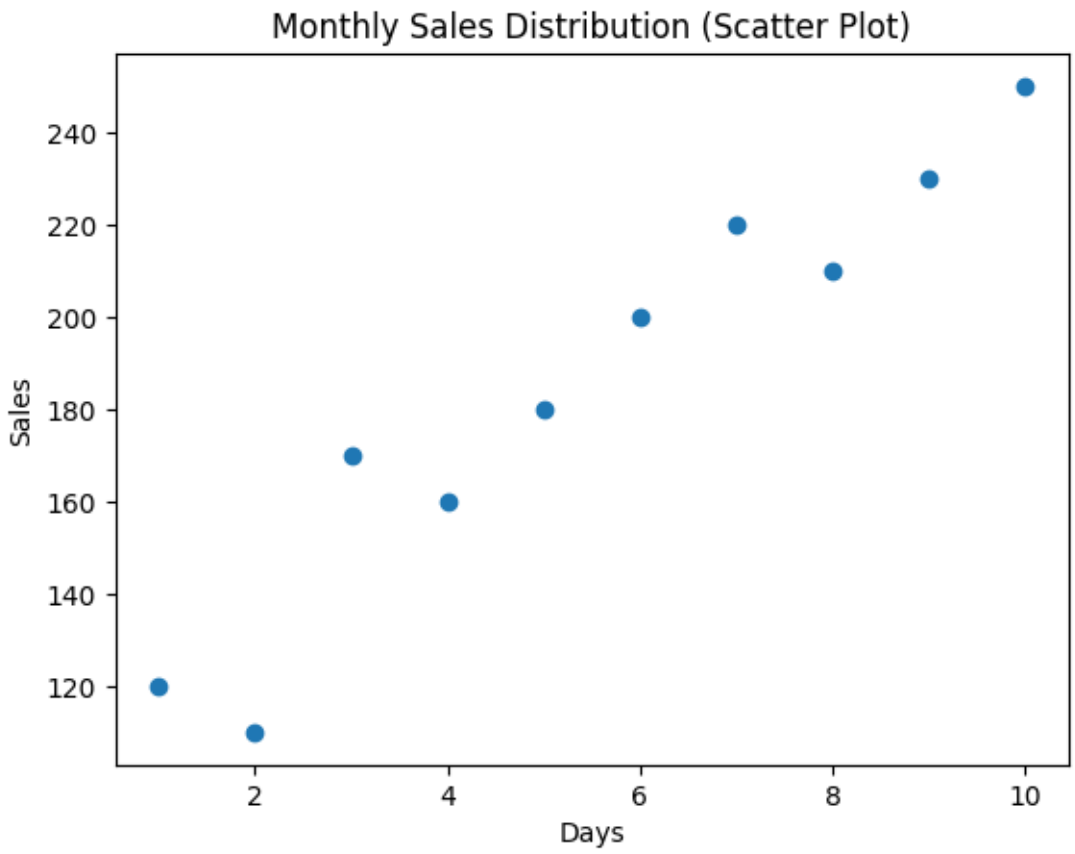
2. Write code to create a scatter plot in Python using Matplotlib to predict sales happened in a month?

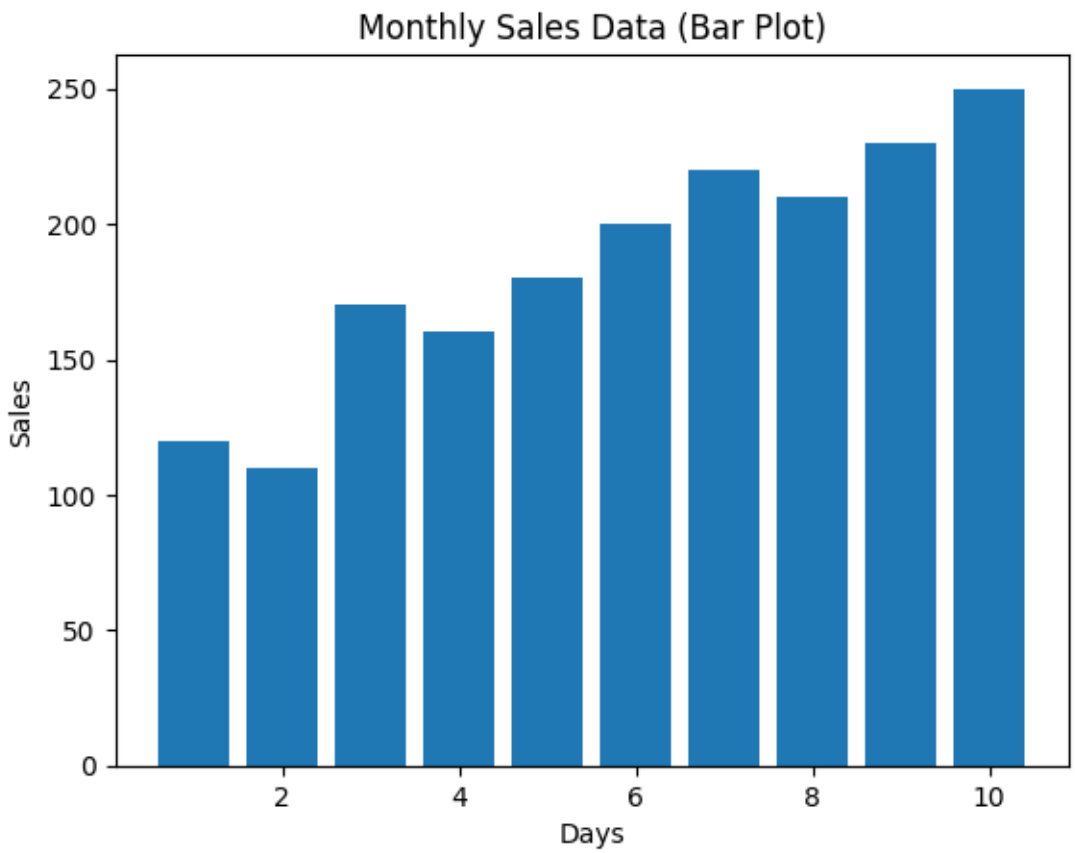
3. Develop a Python program to create a bar plot of the monthly sales data.

Solution:

```
import matplotlib.pyplot as plt
days = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
sales = [120, 110, 170, 160, 180, 200, 220, 210, 230, 250]
plt.plot(days, sales,marker='o')
plt.xlabel("Days")
plt.ylabel("Sales")
plt.title("Monthly Sales Trend (Line Plot)")
plt.show()
plt.scatter(days, sales)
plt.xlabel("Days")
plt.ylabel("Sales")
plt.title("Monthly Sales Distribution (Scatter Plot)")
plt.show()
plt.bar(days, sales)
plt.xlabel("Days")
plt.ylabel("Sales")
plt.title("Monthly Sales Data (Bar Plot)")
plt.show()
```


Monthly Sales Trend (Line Plot)
Sales
Days
120
140
160
180
200
220
240
2
4
6
8
10


Monthly Sales Distribution (Scatter Plot)
Sales
Days
120
140
160
180
200
220
240
2
4
6
8
10


Monthly Sales Data (Bar Plot)
Sales
Days
0
50
100
150
200
250
2
4
6
8
10

Scenario: You are working on a data analysis project that involves analyzing the monthly temperature and rainfall data for a city. You have a dataset containing the monthly temperature and rainfall values for each month of a year. Your task is to develop a Python program that generates line plots and scatter plots to visualize the temperature and rainfall data.

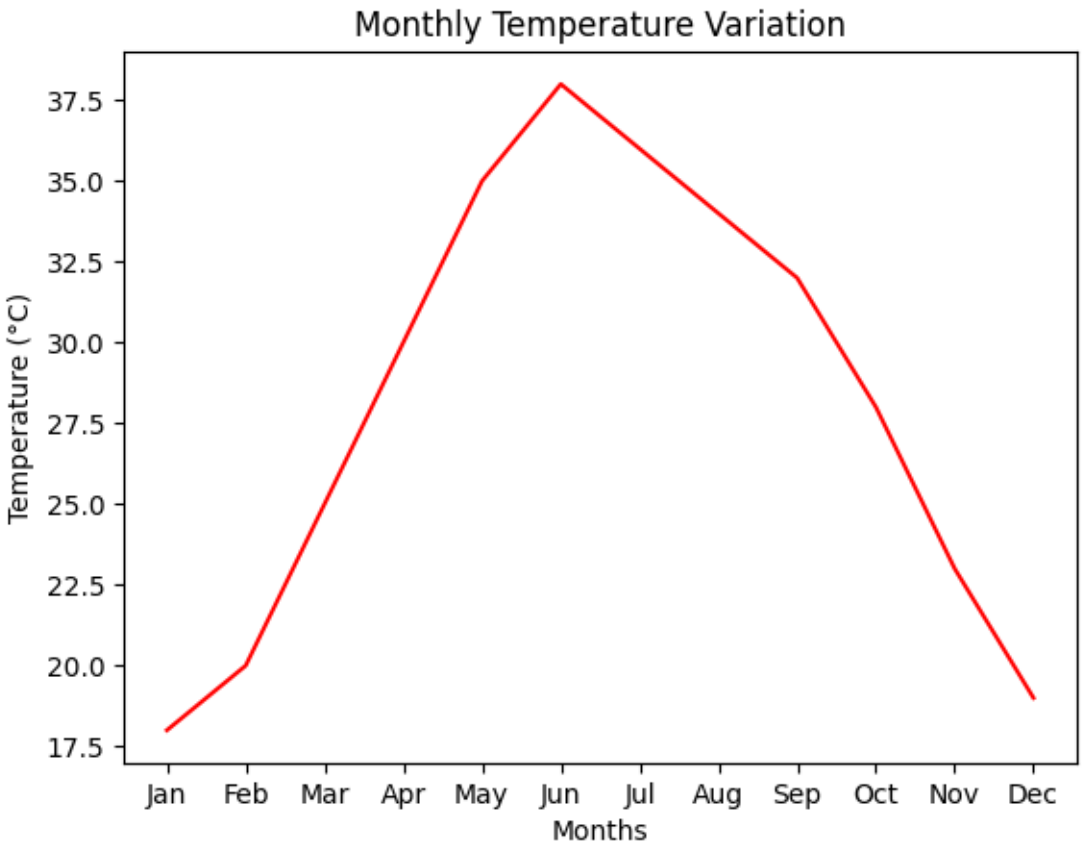
Question:

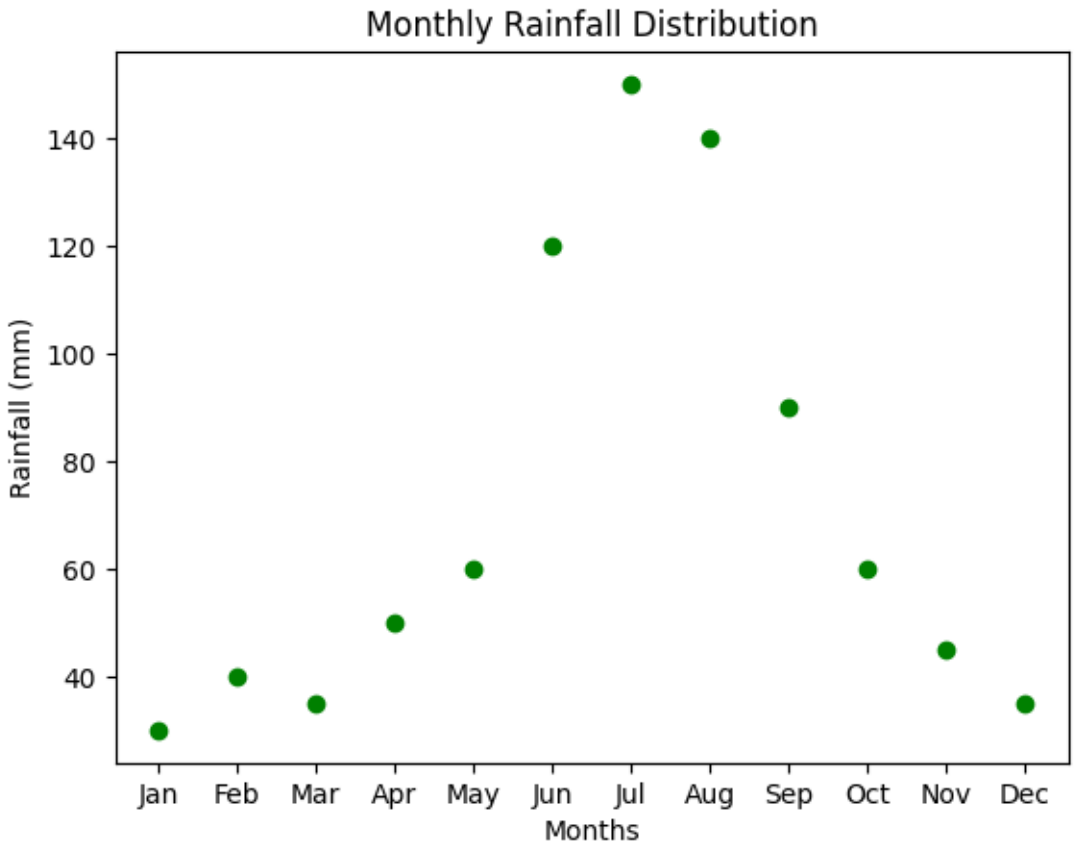
1. Develop a Python program to create a line plot of the monthly temperature data.

2: Develop a Python program to create a scatter plot of the monthly rainfall data.

Solution:

```
import matplotlib.pyplot as plt
months = ["Jan", "Feb", "Mar", "Apr", "May", "Jun",
        "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"]
temperature = [18, 20, 25, 30, 35, 38, 36, 34, 32, 28, 23, 19]
rainfall = [30, 40, 35, 50, 60, 120, 150, 140, 90, 60, 45, 35]
plt.plot(months, temperature,color='red')
plt.xlabel("Months")
plt.ylabel("Temperature (°C)")
plt.title("Monthly Temperature Variation")
plt.show()
plt.scatter(months, rainfall,color='green')
plt.xlabel("Months")
plt.ylabel("Rainfall (mm)")
plt.title("Monthly Rainfall Distribution")
plt.show()
```


Monthly Temperature Variation
Temperature (°C)
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Months


Monthly Rainfall Distribution
Rainfall (mm)
140
120
100
80
60
40
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Months

Scenario: You are working on a text analysis project and need to determine the frequency distribution of words in a given text document. You have a text document named "sample_text.txt" containing a paragraph of text. Your task is to develop a Python program that reads the text document, processes the text, and generates a frequency distribution of the words.

Question: How would you develop a Python program to calculate the frequency distribution of words in a text document?

Solution:

```
import string
import matplotlib.pyplot as plt
with open("/content/sample(pgm13).txt", "r") as file:
    text = file.read()
text = text.lower()
words = text.split()
word_frequency = {}

for word in words:
    if word in word_frequency:
        word_frequency[word] += 1
    else:
        word_frequency[word] = 1

for word, count in word_frequency.items():
    print(word, ":", count)
```

```
import string
import matplotlib.pyplot as plt
with open("/content/sample(pgm13).txt", "r") as file:
    text = file.read()
```

```
text = text.lower()
words = text.split()
word_frequency = {}

for word in words:
    if word in word_frequency:
        word_frequency[word] += 1
    else:
        word_frequency[word] = 1
```

```
for word, count in word_frequency.items():
    print(word, ":", count)
```

```
i : 3
like : 3
ms : 1
dhoni : 1
suriya : 1
michael : 1
jordan : 1
```

Scenario: You are a data analyst working for a company that sells products online. You have been tasked with analyzing the sales data for the past month. The data is stored in a Pandas data frame.

Question: Develop a code in python to find the frequency distribution of the ages of the customers who have made a purchase in the past month.

Solution:

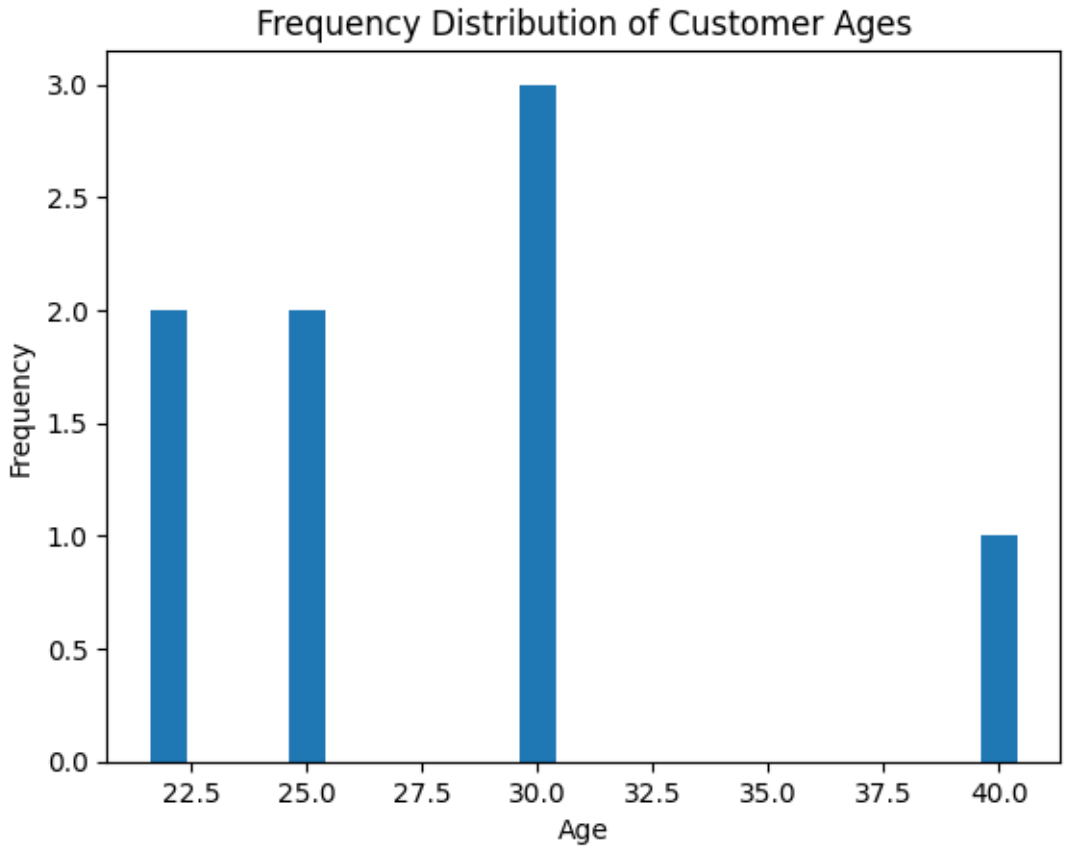
```
import pandas as pd
import matplotlib.pyplot as plt
data = {
    'Customer_ID': [101, 102, 103, 104, 105, 106, 107, 108],
    'Age': [25, 30, 22, 30, 25, 40, 22, 30],
    'Purchase_Amount': [500, 1200, 300, 800, 650, 1500, 400, 900]
}
df = pd.DataFrame(data)
age_frequency = df['Age'].value_counts().sort_index()
print("Frequency Distribution of Customer Ages:")
print(age_frequency)
plt.bar(age_frequency.index, age_frequency.values)
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.title('Frequency Distribution of Customer Ages')
plt.show()
```

```python
import pandas as pd
import matplotlib.pyplot as plt
data = {
    'Customer_ID': [101, 102, 103, 104, 105, 106, 107, 108],
    'Age': [25, 30, 22, 30, 25, 40, 22, 30],
    'Purchase_Amount': [500, 1200, 300, 800, 650, 1500, 400, 900]
}

df = pd.DataFrame(data)
age_frequency = df['Age'].value_counts().sort_index()

print("Frequency Distribution of Customer Ages:")
print(age_frequency)
```

```
Frequency Distribution of Customer Ages:
Age
22    2
25    2
30    3
40    1
```



Scenario: You are a data analyst working for a social media platform. As part of your analysis, you have a dataset containing user interaction data, including the number of likes received by each post. Your task is to develop a Python program that calculates the frequency distribution of likes among the posts.

Question: Develop a Python program to calculate the frequency distribution of likes among the posts?

Solution:

```
import pandas as pd
import matplotlib.pyplot as plt
data = {
    'Post_ID': [1, 2, 3, 4, 5, 6, 7, 8],
    'Likes': [120, 45, 120, 300, 45, 200, 120, 45]
}
df = pd.DataFrame(data)
likes_frequency = df['Likes'].value_counts().sort_index()
print("Frequency Distribution of Likes:")
print(likes_frequency)
plt.plot(likes_frequency.index, likes_frequency.values)
plt.xlabel("Number of Likes")
plt.ylabel("Number of Posts")
plt.title("Frequency Distribution of Likes Among Posts")
plt.show()
```

```python
import pandas as pd
import matplotlib.pyplot as plt
data = {
    'Post_ID': [1, 2, 3, 4, 5, 6, 7, 8],
    'Likes': [120, 45, 120, 300, 45, 200, 120, 45]
}
```

```python
df = pd.DataFrame(data)
likes_frequency = df['Likes'].value_counts().sort_index()
```

```python
print("Frequency Distribution of Likes:")
print(likes_frequency)
```

```
Frequency Distribution of Likes:
Likes
45     3
120    3
200    1
300    1
```