DAY – 3

DSA0410 – Fundamentals of Data Science

Lab Experiments:

Name: B Harish Balaji

Reg no: 192424386

Slot: D

11. 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.

Code:

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt


# Create dataset (40 records)

days = np.arange(1, 41)

sales = np.random.randint(500, 2000, 40)


sales_df = pd.DataFrame({

    "Day": days,

    "Sales": sales

})
```

```python
# Save to Excel
sales_df.to_excel("monthly_sales_data.xlsx", index=False)


# ---------------- Line Plot ----------------
plt.figure()
plt.plot(sales_df["Day"], sales_df["Sales"])
plt.xlabel("Day")
plt.ylabel("Sales")
plt.title("Line Plot of Monthly Sales")
plt.show()


# ---------------- Scatter Plot ----------------
plt.figure()
plt.scatter(sales_df["Day"], sales_df["Sales"])
plt.xlabel("Day")
plt.ylabel("Sales")
plt.title("Scatter Plot of Monthly Sales")
plt.show()


# ---------------- Bar Plot ----------------
plt.figure()
plt.bar(sales_df["Day"], sales_df["Sales"])
plt.xlabel("Day")
plt.ylabel("Sales")
plt.title("Bar Plot of Monthly Sales")
plt.show()
```

sample output:

Line Plot of Monthly Sales

Scatter Plot of Monthly Sales

Bar Plot of Monthly Sales
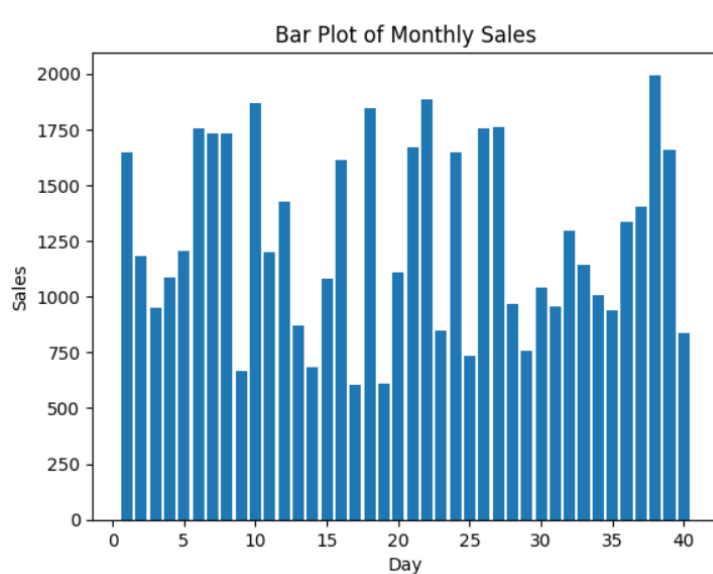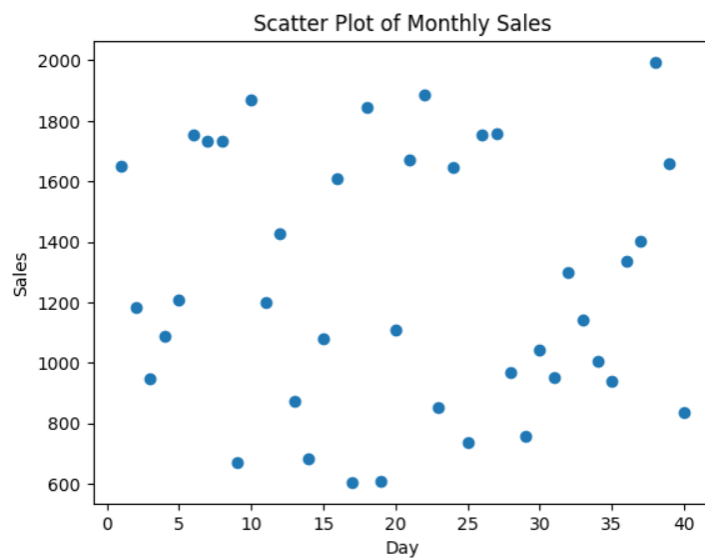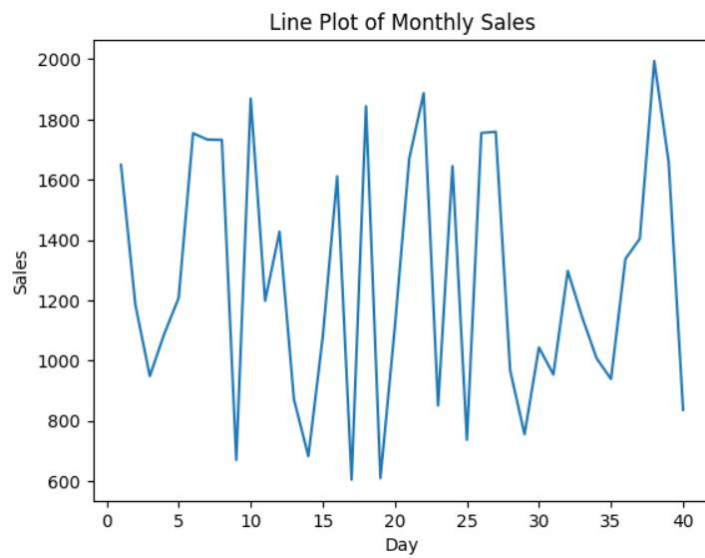
12. Scenario: You are working on a data analysis project that involves analyzing the monthly

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.

CODE:

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

Create dataset with 40 records

months = np.arange(1, 41)

temperature = np.random.randint(15, 40, 40)

rainfall = np.random.randint(20, 300, 40)

weather_df = pd.DataFrame({

   "Month": months,

   "Temperature (°C)": temperature,

```python
    "Rainfall (mm)": rainfall
})


# Save dataset to Excel
weather_df.to_excel("monthly_weather_data.xlsx", index=False)


# -------- Line Plot: Monthly Temperature --------
plt.figure()
plt.plot(weather_df["Month"], weather_df["Temperature (°C)"])
plt.xlabel("Month")
plt.ylabel("Temperature (°C)")
plt.title("Line Plot of Monthly Temperature")
plt.show()


# -------- Scatter Plot: Monthly Rainfall --------
plt.figure()
plt.scatter(weather_df["Month"], weather_df["Rainfall (mm)"])
plt.xlabel("Month")
plt.ylabel("Rainfall (mm)")
plt.title("Scatter Plot of Monthly Rainfall")
plt.show()
```
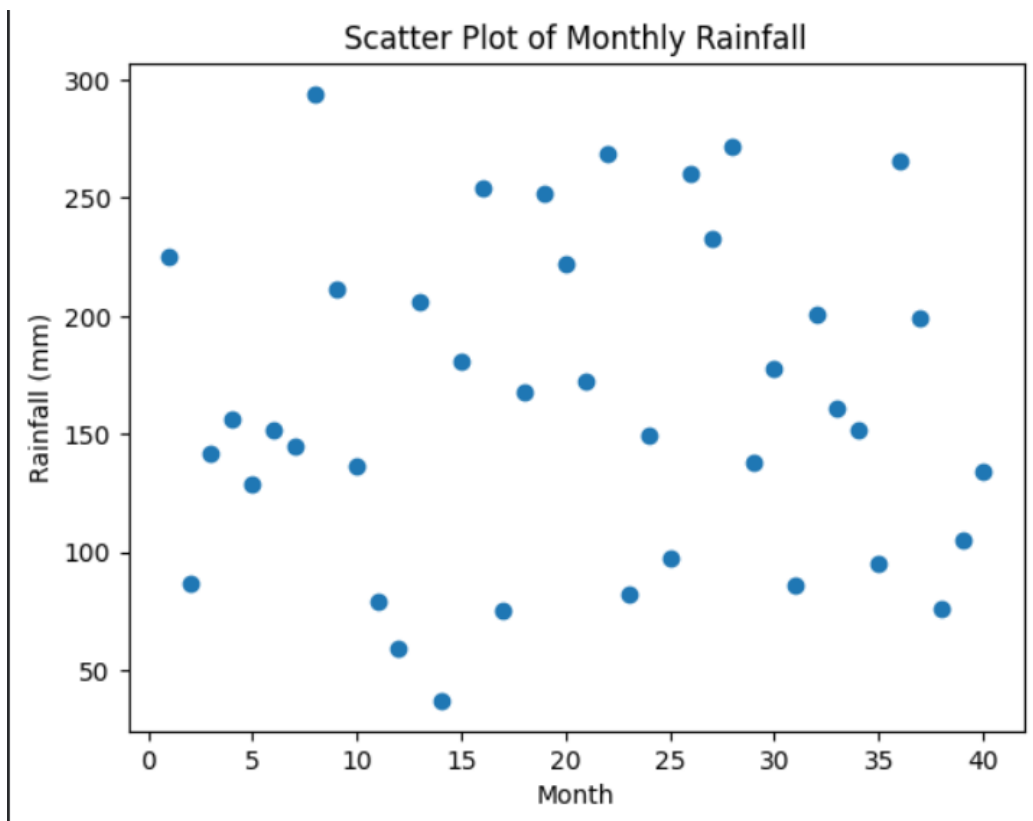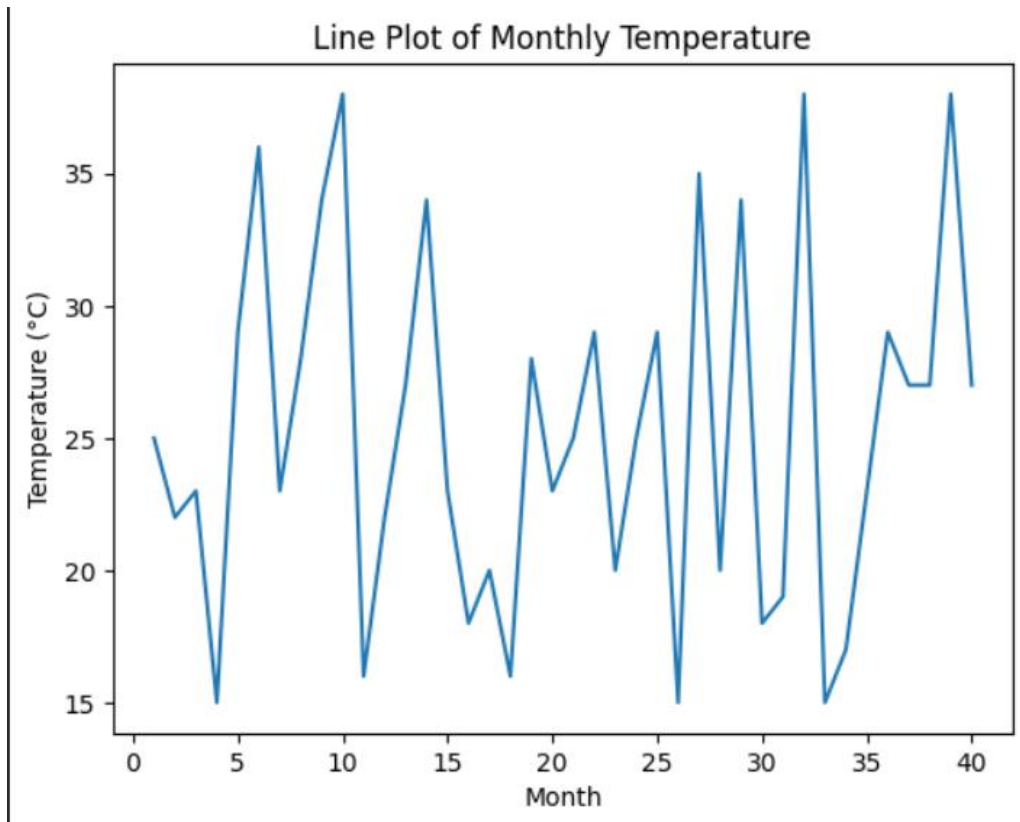
sample output:

Line Plot of Monthly Temperature

Scatter Plot of Monthly Rainfall

13. 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?

CODE:

```
import pandas as pd

import re

from collections import Counter

# Step 1: Create sample text document

sample_text = """

Data science is an interdisciplinary field that uses scientific methods,

processes, algorithms and systems to extract knowledge and insights
from data.

Data science is widely used in business analytics, machine learning,

and artificial intelligence.

"""
```

```python
with open("sample_text.txt", "w") as file:
    file.write(sample_text)

# Step 2: Read and preprocess text
with open("sample_text.txt", "r") as file:
    text = file.read().lower()

# Remove punctuation and tokenize
words = re.findall(r'\b\w+\b', text)

# Step 3: Frequency distribution
word_frequency = Counter(words)

# Select top 30 words
top_30_words = word_frequency.most_common(30)

# Step 4: Save to Excel
df = pd.DataFrame(top_30_words, columns=["Word", "Frequency"])
df.to_excel("word_frequency_30_records.xlsx", index=False)

# Display result
print("Word Frequency Distribution (Top 30):")
print(df)
```

sample output:

```
word_frequency_30_records (1).xlsx(application/vnd.openxmlformats-officedocument.spreadsheetml.sheet) - 5850 bytes, last modified: 12/18/2025 - 100% done
Saving word_frequency_30_records (1).xlsx to word_frequency_30_records (1).xlsx
Word Frequency Distribution (Top 30):
                Word  Frequency
0               data          3
1                and          3
2            science          2
3                 is          2
4                 an          1
5    interdisciplinary        1
6              field          1
7               that          1
8               uses          1
9         scientific          1
10           methods          1
11         processes          1
12        algorithms          1
13           systems          1
14                to          1
15           extract          1
16         knowledge          1
17          insights          1
18              from          1
19            widely          1
20              used          1
21                in          1
22          business          1
23         analytics          1
24           machine          1
25          learning          1
26        artificial          1
27      intelligence          1
```

14. 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.

CODE:

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

```python
# Create dataset with 30 different ages
ages = np.array([
    18, 19, 20, 21, 22, 23, 24, 25, 26, 27,
    28, 29, 30, 31, 32, 33, 34, 35, 36, 37,
    38, 39, 40, 41, 42, 45, 48, 50, 55, 60
])

sales_df = pd.DataFrame({
    "Customer_Age": ages
})

# Save to Excel
sales_df.to_excel("customer_age_sales.xlsx", index=False)

# Frequency distribution of ages
age_frequency =
sales_df["Customer_Age"].value_counts().sort_index()

print("Age Frequency Distribution:")
print(age_frequency)

# Bar plot of age frequency
plt.figure()
plt.bar(age_frequency.index, age_frequency.values)
```
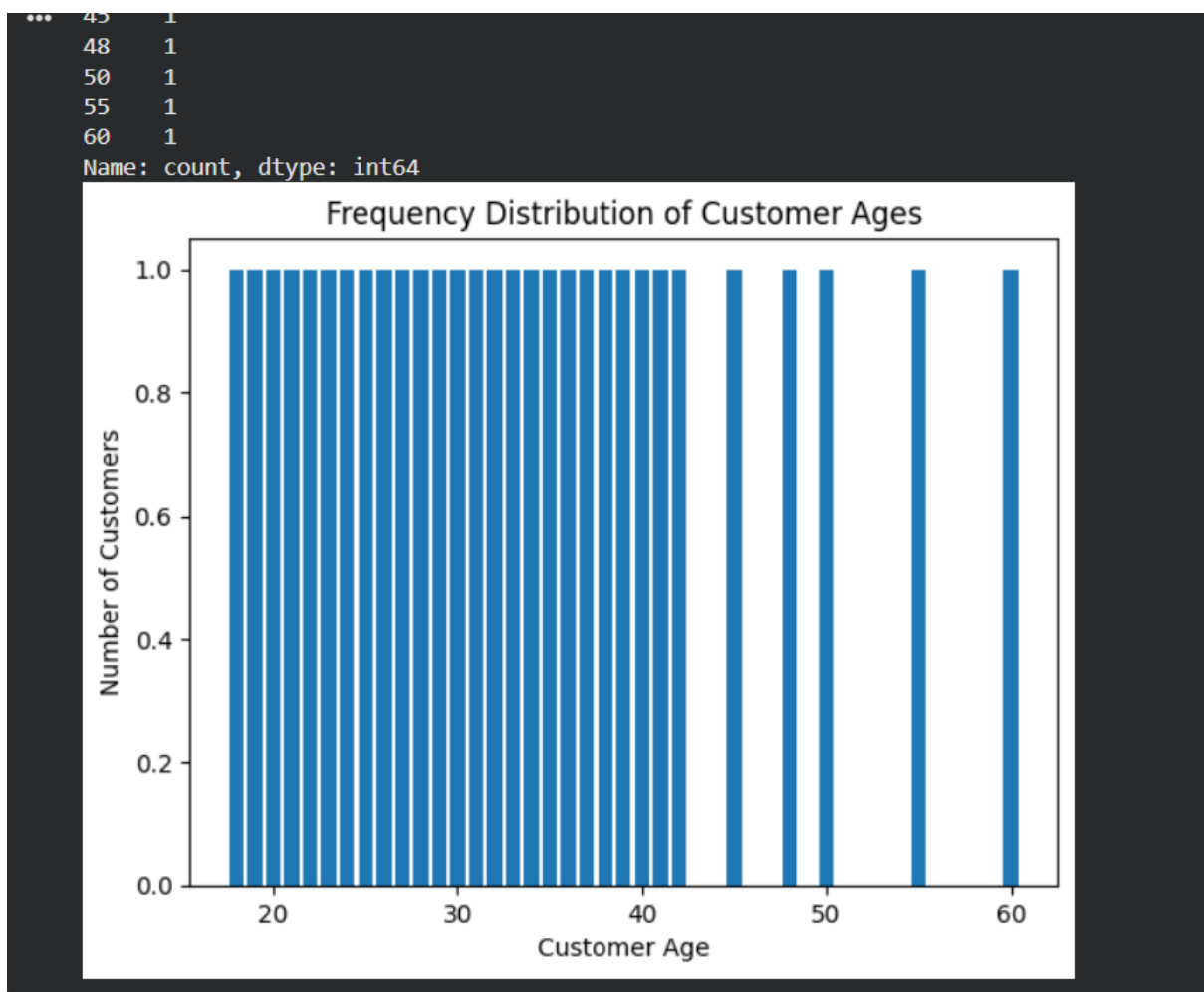
```
plt.xlabel("Customer Age")

plt.ylabel("Number of Customers")

plt.title("Frequency Distribution of Customer Ages")

plt.show()
```

sample output:

15. 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?

CODE:

```python
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt


# Create dataset: likes per post (30 records)

likes = np.array([

    5, 10, 10, 15, 20, 20, 20, 25, 30, 30,

    35, 40, 40, 45, 50, 50, 50, 60, 70, 75,

    80, 90, 100, 120, 150, 180, 200, 250, 300, 350

])

likes_df = pd.DataFrame({

    "Likes": likes

})

likes_df.to_excel("post_likes_data.xlsx", index=False)
```

```python
# Calculate frequency distribution

likes_frequency = likes_df["Likes"].value_counts().sort_index()


print("Frequency Distribution of Likes:")

print(likes_frequency)

# Plot frequency distribution

plt.figure()

plt.bar(likes_frequency.index, likes_frequency.values)

plt.xlabel("Number of Likes")

plt.ylabel("Number of Posts")

plt.title("Frequency Distribution of Likes per Post")

plt.show()
```

sample output: