

## DAY 3 LAB EXPERIMENTS

NAME: LARAVIND

Reg No : 192424080

Date : 18/12/2025

**EXP\_11 You have been tasked with creating a simple plot to show the sales of a product over time.**

```
File Edit View Insert Runtime Tools Help
Commands + Code + Text Run all

Files
sample_data
Tweets.csv
ecommerce_customers.csv

In [1]: import pandas as pd
import matplotlib.pyplot as plt

df=pd.read_csv("monthly_sales.csv",encoding="latin1")
print(df.head())

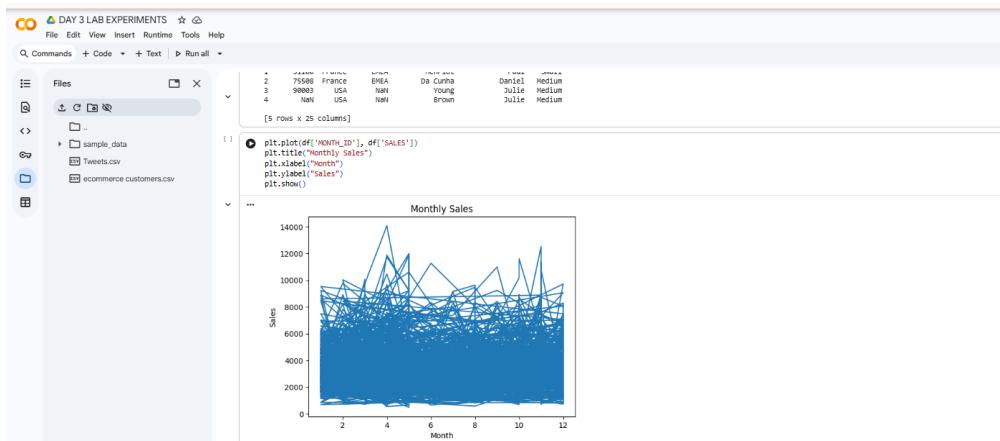
Out[1]:
ORDERNUMBER QUANTITYORDERED PRICEEACH ORDERLINENUMBER SALES \
0 180221 34 81.35 1 2765.90
1 180314 41 94.74 2 3884.34
2 180405 45 82.25 3 3672.75
3 180519 49 188.00 4 9215.27

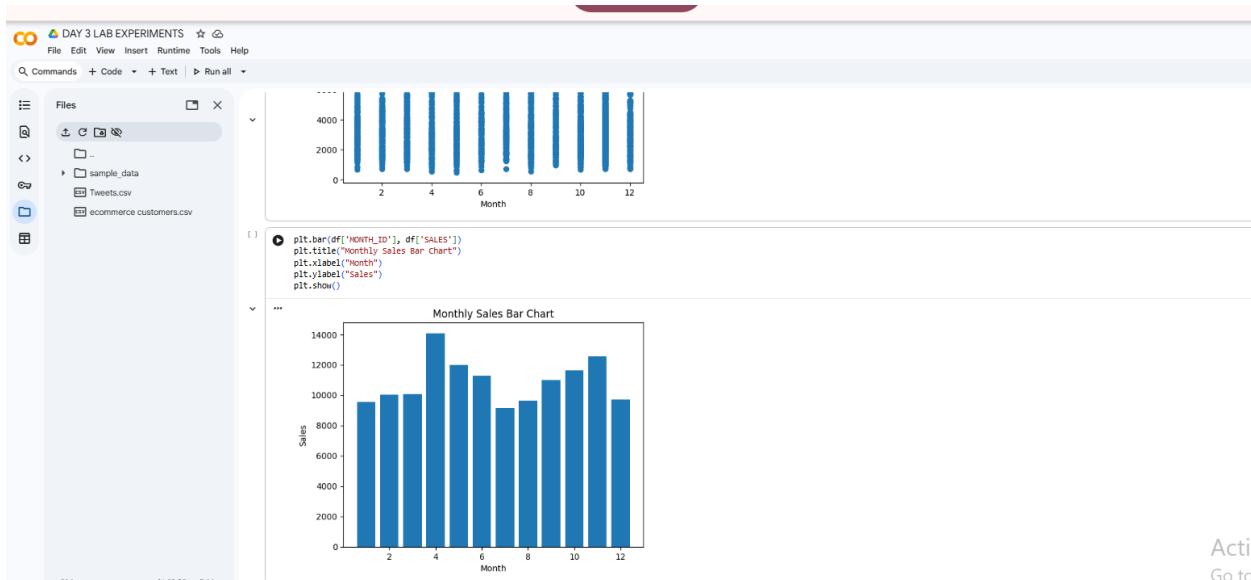
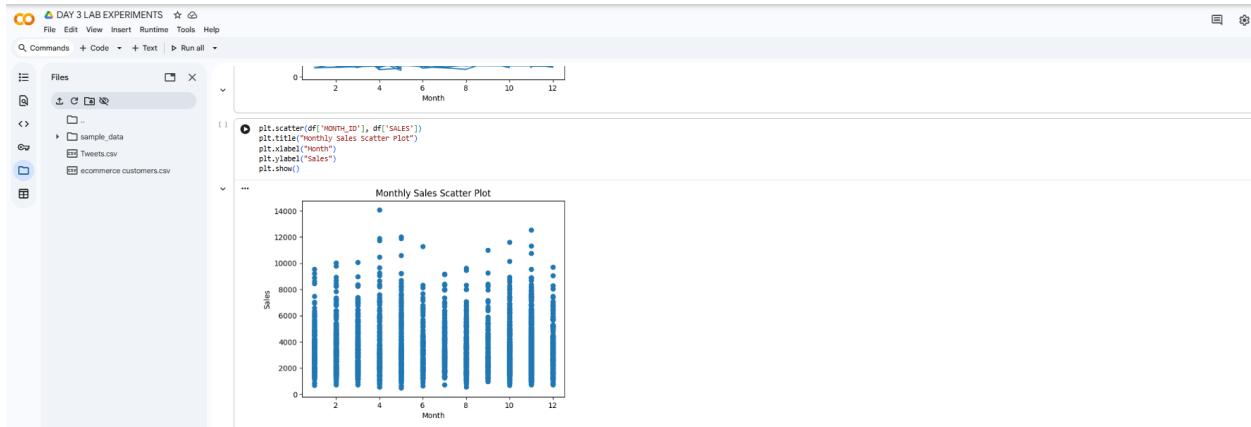
ORDERDATE STATUS QTR_ID MONTH_ID YEAR_ID ...
0 2/24/2003 8:00 Shipped 1 2 2003 ...
1 3/1/2003 8:00 Shipped 1 3 2003 ...
2 7/1/2003 8:00 Shipped 3 7 2003 ...
3 8/25/2003 8:00 Shipped 3 8 2003 ...
4 8/26/2003 8:00 Shipped 4 8 2003 ...

ADRESSLINE1 ADRESSLINE2 CITY STATE ...
0 887 Long Airport Avenue NaN NYC NY
1 123 Main Street NaN Boston MA
2 27 rue du Colonel Pierre Aria NaN Paris NaN
3 78934 Hillside Dr. NaN Pasadena CA
4 774 Strong St. NaN San Francisco CA

POSTALCODE COUNTRY TERRITORY CONTACTLASTNAME CONTACTFIRSTNAME SALES12
0 10002 USA NaN Yu Kuan Small
1 51100 France EMEA Herriot Paul Small
2 51100 France EMEA De Costa Daniel Medium
3 99000 USA NaN Young Julie Medium
4 NaN USA NaN Brown Julie Medium

[5 rows x 25 columns]
```





Action  
Go to

## EXP\_12 To Develop a Python program that generates line plots and scatter plots to visualize the temperature and rainfall data.

DAY 3 LAB EXPERIMENTS

```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv(
    "weather_data.csv",
    encoding="latin1",
    on_bad_lines="skip"
)

print(df.head(10))

   Location          Date_Time Temperature_C Humidity_pct \
0  San Diego 2024-01-15 20:12:46  18.683908  41.19754 \
1  San Diego 2024-01-15 20:13:04  18.683908  39.319107 \
2  San Diego 2024-01-11 09:38:59  11.632436  38.826175 \
3 Philadelphia 2024-02-26 17:32:39 -8.628976  54.074474 \
4  San Antonio 2024-04-29 13:23:51  39.808213  72.899908

   Precipitation_mm Wind_Speed_kph \
0           4.020119      8.233549 \
1           9.111623     27.715161 \
2           4.607511     26.732951 \
3           3.183724     26.367303 \
4           5.362424     29.898622

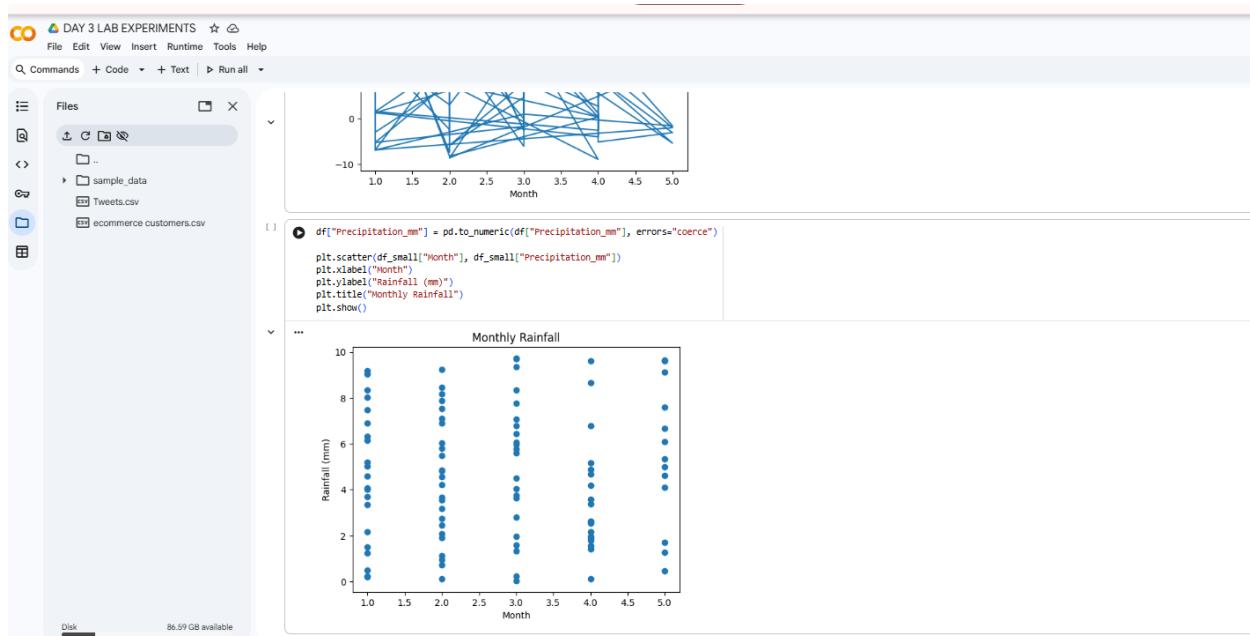
/tmipython-input-1235396151.py:3: DtypeWarning: Columns (2,3,4) have mixed types. Specify dtype option on import or set low_memory=False.
df = pd.read_csv(


df["Date_Time"] = pd.to_datetime(df["Date_Time"], errors="coerce")
df["Month"] = df["Date_Time"].dt.month
```

df["Date\_Time"] = pd.to\_datetime(df["Date\_Time"], errors="coerce")  
df["Month"] = df["Date\_Time"].dt.month

```
df["Temperature_C"] = pd.to_numeric(df["Temperature_C"], errors="coerce")
df_small = df.head(100)

plt.plot(df_small["Month"], df_small["Temperature_C"])
plt.xlabel("Month")
plt.ylabel("Temperature (°C)")
plt.title("Monthly Temperature")
plt.show()
```



## **EXP\_13 To Develop a Python program that reads the text document, processes the text, and generates a frequency distribution of the words.**

The screenshot shows a Jupyter Notebook interface with the following details:

- Title Bar:** DAY 3 LAB EXPERIMENTS
- File Tree:** Files, sample\_data (containing Tweets.csv and ecommerce customers.csv).
- Code Cell:** EXP\_13 You are working on a text analysis project and need to determine the frequency distribution of words in a given text document.

```
from collections import Counter

with open("sample_data.txt", "r") as file:
    text = file.read()
    words = text.lower().split()
    word_freq = Counter(words)
    for i in word_freq:
        print(f'{i}: {word_freq[i]}')
print(1)
```
- Output Cell:** Displays the word frequency distribution from the script.
- Bottom Status:** Disk 86.59 GB available.

## **EXP\_14 To 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.**

The screenshot shows a Jupyter Notebook interface with the following details:

- Title Bar:** DAY 3 LAB EXPERIMENTS
- File Tree:** Files, sample\_data (containing Tweets.csv and ecommerce customers.csv).
- Code Cell:** EXP\_14 To Find the frequency distribution of the ages of the customers who have made a purchase in the past month.

```
import pandas as pd

df=pd.read_csv("ecommerce customers.csv")
print(df.head())
print(df.columns)

CustomerID  Genre  Age  Annual Income (k$)  Spending Score (1-100)
0           1   Male   19             15            39
1           2   Male   21             15            81
2           3 Female   20             16              6
3           4 Female   23             16            77
4           5 Female   31             17            40
Index(['CustomerID', 'Genre', 'Age', 'Annual Income (k$)', 'Spending Score (1-100)', dtype='object')

age_frequency = df["Age"].value_counts()
print(age_frequency)
```
- Output Cell:** Displays the age frequency distribution from the script.
- Bottom Status:** Disk 86.59 GB available.

The screenshot shows a Jupyter Notebook interface with the following details:

- Header:** DAY 3 LAB EXPERIMENTS
- File Menu:** File, Edit, View, Insert, Runtime, Tools, Help
- Toolbar:** Commands, Code, Text, Run all
- File Browser:** Shows a tree view of files in the current directory:
  - ..
  - sample\_data
  - Tweets.csv
  - ecommerce customers.csv
- Code Cell Output:** Displays a list of integers from 49 down to 56, each followed by a value of 1 or 2.

Value	Count
49	7
23	6
46	6
36	6
27	6
47	6
38	6
58	5
28	5
21	5
29	5
48	5
34	5
67	4
54	4
24	4
59	4
18	4
28	4
22	3
25	3
46	3
43	3
39	3
68	3
45	3
37	3
33	3
68	3
58	2
41	2
52	2
57	2
65	2
53	2
51	2
42	2
78	2
44	2
26	2
63	2
66	2
64	1
69	1
55	1
56	1
- Disk Status:** 86.59 GB available

**EXP\_15 To Develop a Python program that calculates the frequency distribution of likes among the posts.**

CO DAY 3 LAB EXPERIMENTS ⚡

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all

Files

.. sample\_data Tweets.csv ecommerce customers.csv

EXP\_15 To Develop a Python program that calculates the frequency distribution of likes among the posts.

```
[1] import pandas as pd
```

```
[2] df=pd.read_csv("Tweets.csv",encoding="latin1")
print(df.head())
```

```
[3] ... tweet_id airline_sentiment airline_sentiment_confidence \
0 579306133977769513 neutral 1.0000
1 579301138088122368 positive 0.3486
2 579301063672813571 neutral 0.6837
3 579301031487624196 negative 1.0000
4 579300817074462722 negative 1.0000
```

```
[4] negativereson negativereson_confidence \
0 NAN NAN Virgin America
1 NAN 0.0000 Virgin America
2 NAN NAN Virgin America
3 Bad Flight 0.7033 Virgin America
4 Can't Tell 1.0000 Virgin America
```

```
[5] airline_sentiment_gold name negativereson_gold retweet_count \
0 NAN cairdin NAN 0
1 NAN jnardino NAN 0
2 NAN yvonnalynn NAN 0
3 NAN jnardino NAN 0
4 NAN jnardino NAN 0
```

```
[6] text tweet_coord \
0 @VirginAmerica What @dhepburn said. NAN
1 @VirginAmerica plz you've added commercial t... NAN
2 @VirginAmerica i didn't today... Must me I n... NAN
3 @VirginAmerica it's really aggressive to blast... NAN
4 @VirginAmerica and it's a really big bad thing... NAN
```

```
[7] tweet_created tweet_location user_timezone
0 2015-02-24 11:35:52 -0800 NAN Eastern Time (US & Canada)
1 2015-02-24 11:15:59 -0800 NAN Pacific Time (US & Canada)
2 2015-02-24 11:15:48 -0800 Lets Play Central Time (US & Canada)
3 2015-02-24 11:15:36 -0800 NAN Pacific Time (US & Canada)
4 2015-02-24 11:14:45 -0800 NAN Pacific Time (US & Canada)
```

```
[8] likes_frequency = df["retweet_count"].value_counts()
```

Disk 86.59 GB available

CO DAY 3 LAB EXPERIMENTS ⚡

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all

Files

.. sample\_data Tweets.csv ecommerce customers.csv

```
[1] print("The Frequency Distribution of Likes are:")
print(likes_frequency)
```

```
[2] ... The Frequency Distribution of Likes are:
retweet_count
0 13873
1 640
2 66
3 22
4 17
5 5
7 3
6 3
22 2
18 1
15 1
31 1
11 1
8 1
9 1
28 1
32 1
44 1
Name: count, dtype: int64
```