

DAY 3 LAB EXPERIMENTS

NAME: GUTTI CHAITANYA

Reg No : 192424185

Date : 18/12/2025

EXP_11 You have been tasked with creating a simple plot to show the sales of a product over time.

```
EXP_11 A Data Scientist working for a Company that sells products online.

1 | import pandas as pd
2 | import matplotlib.pyplot as plt

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

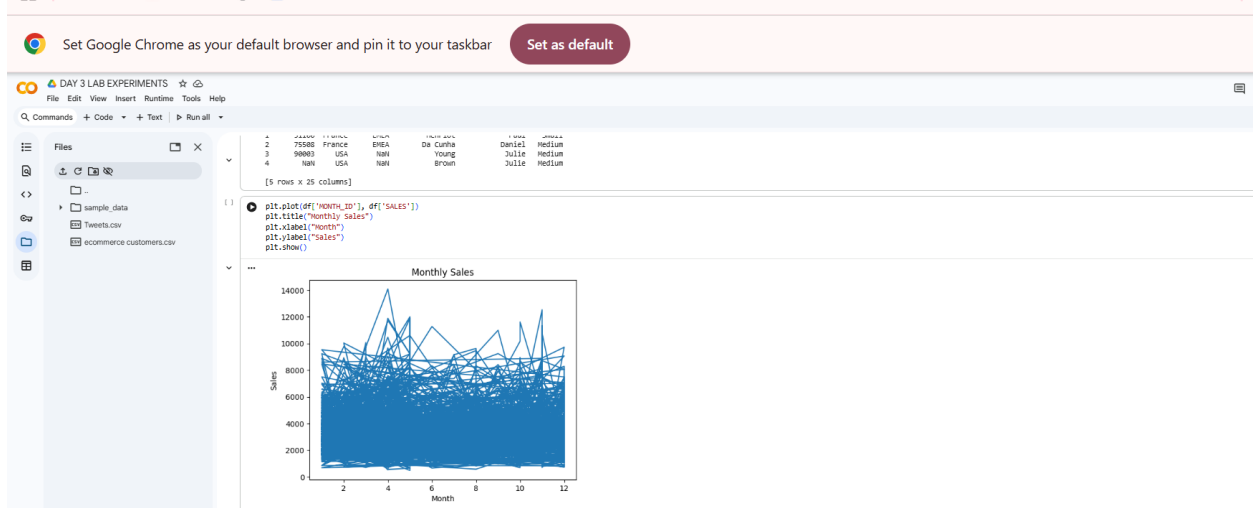
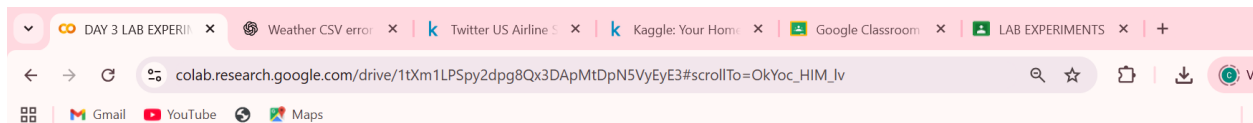
5 |
6 | ORDERNUMBER  QUANTITYORDERED  PRICEEACH  ORDERLINENUMBER  SALES
7 | 0 38587 30 95.70 2 3871.00
8 | 1 38121 34 81.35 5 2765.90
9 | 2 38134 41 84.74 2 3884.34
10 | 3 38145 45 83.26 6 3748.70
11 | 4 38159 49 380.00 14 1505.27

12 |
13 | ORDERDATE  STATUS  QTR_ID  MONTH_ID  YEAR_ID  ...
14 | 0 2/24/2003 0:00 Shipped 1 2 2003 ...
15 | 1 5/17/2003 0:00 Shipped 2 5 2003 ...
16 | 2 7/1/2003 0:00 Shipped 3 7 2003 ...
17 | 3 8/15/2003 0:00 Shipped 3 9 2003 ...
18 | 4 10/10/2003 0:00 Shipped 4 10 2003 ...

19 |
20 | ADDRESSLINE1  ADDRESSLINE2  CITY  STATE
21 | 0 807 Long Airport Avenue  NaN  NYC  NY
22 | 1 59 rue de l'Abbaye  NaN  Nantes  NaN
23 | 2 27 rue du Colonel Pierre Aude  NaN  Paris  NaN
24 | 3 78334 Hillside Dr.  NaN  Pasadena  CA
25 | 4 7734 Strong St.  NaN  San Francisco  CA

26 |
27 | POSTALCODE  COUNTRY  TERRITORY  CONTACTLASTNAME  CONTACTFIRSTNAME  DEALSIZE
28 | 0 98022  USA  NaN  Vu  Kari  Small
29 | 1 51188  France  EMEA  Henriot  Paul  Small
30 | 2 75585  France  EMEA  Du Cunha  Daniel  Medium
31 | 3 90003  USA  NaN  Young  Julie  Medium
32 | 4  NaN  USA  NaN  Brown  Julie  Medium

[5 rows x 25 columns]
```



DAY 3 LAB EXPERIMENTS | Weather CSV error | Twitter US Airline | Kaggle: Your Home | Google Classroom | LAB EXPERIMENTS

colab.research.google.com/drive/1tXm1LPSPy2dpg8Qx3DApMtDpN5VyEyE3#scrollTo=OkYoc_HIM_lv

Set Google Chrome as your default browser and pin it to your taskbar

DAY 3 LAB EXPERIMENTS

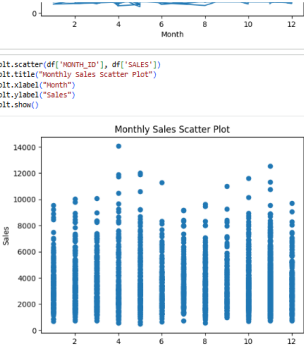
File Edit View Insert Runtime Tools Help

Commands + Code + Text + Run all

Files

- sample_data
- Tweets.csv
- e-commerce customers.csv

```
plt.scatter(df['MONTH_ID'], df['SALES'])
plt.title('Monthly Sales Scatter Plot')
plt.xlabel('Month')
plt.ylabel('Sales')
plt.show()
```



Monthly Sales Scatter Plot

The scatter plot shows the relationship between the month (x-axis, 1 to 12) and sales (y-axis, 0 to 14000). The data points are scattered, indicating a weak correlation between the month and sales.

DAY 3 LAB EXPERIMENTS | Weather CSV error | Twitter US Airline | Kaggle: Your Home | Google Classroom | LAB EXPERIMENTS

colab.research.google.com/drive/1tXm1LPSPy2dpg8Qx3DApMtDpN5VyEyE3#scrollTo=OkYoc_HIM_lv

Set Google Chrome as your default browser and pin it to your taskbar

DAY 3 LAB EXPERIMENTS

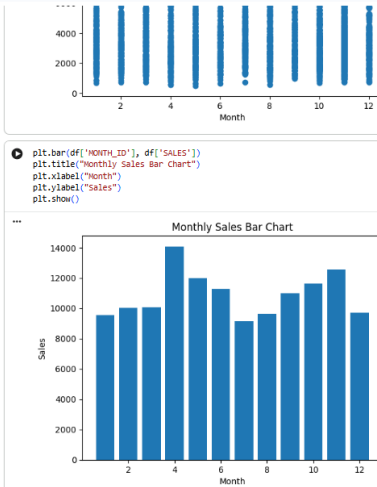
File Edit View Insert Runtime Tools Help

Commands + Code + Text + Run all

Files

- sample_data
- Tweets.csv
- e-commerce customers.csv

```
plt.bar(df['MONTH_ID'], df['SALES'])
plt.title('Monthly Sales Bar Chart')
plt.xlabel('Month')
plt.ylabel('Sales')
plt.show()
```



Monthly Sales Bar Chart

The bar chart shows the monthly sales (y-axis, 0 to 14000) for each month (x-axis, 1 to 12). The sales are relatively stable, with a slight increase in the middle of the year (months 4-6).

Activ
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

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all

Files

sample_data

Tweets.csv

ecommerce customers.csv

EXP_12 A Data Analysis project that involves analyzing the monthly temperature and rainfall data for a city.

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

import pandas as pd

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

print(df.head())
```

	Location	Date_Time	Temperature_C	Humidity_pct
0	San Diego	2024-01-14 21:12:46	10.683001	41.195754
1	San Diego	2024-05-17 15:12:10	8.73414	58.319107
2	San Diego	2024-05-11 09:30:59	11.632436	38.820175
3	Philadelphia	2024-02-26 17:32:39	-8.628976	54.074474
4	San Antonio	2024-04-29 13:13:51	39.888213	72.899908

```

Precipitation_mm Wind_Speed_kmh
0      4.020119      8.233540
1     9.111623     27.715161
2     4.607511     28.732951
3     3.18372      26.367303
4     9.598282     29.898622
/tmp/ipython-input-1235356151.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()
```

Monthly Temperature

DAY 3 LAB EXPERIMENTS

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all

Files

sample_data

Tweets.csv

ecommerce customers.csv

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()

Monthly Temperature

Set Google Chrome as your default browser and pin it to your taskbar [Set as default](#)

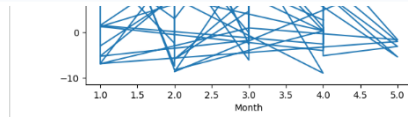
DAY 3 LAB EXPERIMENTS

File Edit View Insert Runtime Tools Help

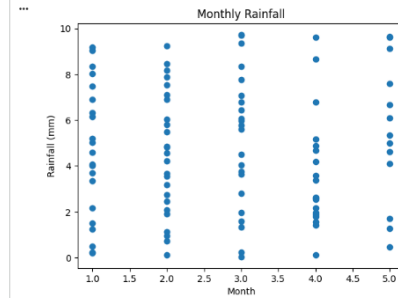
Commands + Code + Text Run all

Files

- sample_data
- Tweets.csv
- ecommerce customers.csv



```
df["Precipitation_mm"] = pd.to_numeric(df["Precipitation_mm"], errors="coerce")  
  
plt.scatter(df_small["Month"], df_small["Precipitation_mm"])  
plt.xlabel("Month")  
plt.ylabel("Rainfall (mm)")  
plt.title("Monthly Rainfall")  
plt.show()
```



Disk 86.59 GB available

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 Google Chrome browser window with several tabs open, including "DAY 3 LAB EXPERIMENT", "Weather CSV error", "Twitter US Airline S", "Kaggle: Your Home", "Google Classroom", and "LAB EXPERIMENT". The address bar shows the URL: `colab.research.google.com/drive/1tXm1LPSpy2dpg8Qx3DApMtDpN5VyEyE3#scrollTo=OkYoc_HIM_lv`. Below the browser window, there is a notification bar that says "Set Google Chrome as your default browser and pin it to your taskbar" with a "Set as default" button.

The main content is a Google Colab notebook titled "DAY 3 LAB EXPERIMENT". The notebook has a menu bar with "File", "Edit", "View", "Insert", "Runtime", "Tools", and "Help". Below the menu bar, there is a toolbar with "Commands", "+ Code", "+ Text", and "Run all". The notebook is divided into two main sections: "Files" on the left and "Code" on the right.

The "Files" section shows a file explorer with the following files:

- sample_data
- Tweets.csv
- ecommerce customers.csv

The "Code" section contains a Python program that reads a text file, processes the text, and generates a frequency distribution of the words. The code is as follows:

```
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()
```

The output of the program is a frequency distribution of the words in the text file:

```
welcome: 1
to: 2
natural: 1
language: 1
processing: 1
it: 1
is: 1
one: 1
of: 2
the: 1
most: 1
exciting: 1
research: 1
areas: 1
ss: 1
today: 1
we: 1
will: 1
see: 1
how: 1
python: 1
can: 1
be: 1
used: 1
work: 1
with: 1
text: 1
files: 1
files.
```

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.

DAY 3 LAB EXPERIMENTS x Weather CSV error x Twitter US Airline S x Kaggle: Your Home x

colab.research.google.com/drive/1tXm1LPsPy2dpg8Qx3DApMtDpN5VyEyE3#scrollTo=OkYoc_HIM

Gmail YouTube Maps

Set Google Chrome as your default browser and pin it to your taskbar [Set as default](#)

DAY 3 LAB EXPERIMENTS

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all

Files

- ..
- sample_data
- Tweets.csv
- e-commerce customers.csv

EXP_14 To Find the frequency distribution of the ages of the customers who have made a purchase in the past month.

```
[5] ✓ On
import pandas as pd

[6] ✓ On
df=pd.read_csv("e-commerce 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

```
[7] ✓ On
age_frequency = df["Age"].value_counts()
print(age_frequency)
```

```
Age
32    11
35     9
31     8
19     8
30     7
49     7
23     6
40     6
36     6
27     6
47     6
38     6
50     5
20     5
21     5
29     5
48     5
34     5
67     4
54     4
24     4
59     4
```

DAY 3 LAB EXPERIMENTS

Weather CSV error

Twitter US Airline S

Kaggle: Your Home

Google

colab.research.google.com/drive/1tXm1LP5py2dpg8Qx3DApMtDpN5VyEyE3#scrollTo=OkYoc_HIM_I

Gmail YouTube Maps

Set Google Chrome as your default browser and pin it to your taskbar

Set as default

DAY 3 LAB EXPERIMENTS

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all

Files

sample_data Tweets.csv ecommerce customers.csv

Disk 86.59 GB available

49 7

23 6

40 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

60 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

70 2

44 2

26 2

63 2

66 2

64 1

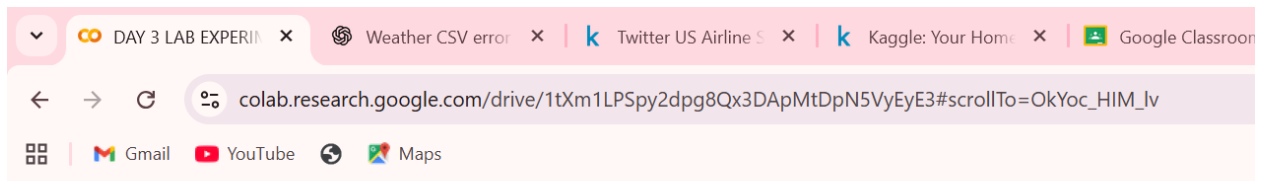
69 1

55 1

56 1

Name: count, dtype: int64

EXP_15 To Develop a Python program that calculates the frequency distribution of likes among the posts.



```
EXP_15 To Develop a Python program that calculates the frequency distribution of likes among the posts.

import pandas as pd

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

...
tweet_id airline_sentiment airline_sentiment_confidence \
0 570306133677760513 neutral 1.0000
1 570301130808122368 positive 0.3406
2 570301083672813571 neutral 0.6637
3 5703010831407624196 negative 1.0000
4 570300817074462722 negative 1.0000

negative_reason negative_reason_confidence airline \
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

airline_sentiment_gold name negative_reason_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

text tweet_coord \
0 @virginAmerica What @dhepburn said. NaN
1 @virginAmerica plus you've added commercials t... NaN
2 @virginAmerica I didn't today... Must mean I n... NaN
3 @virginAmerica it's really aggressive to blast... NaN
4 @virginAmerica and it's a really big bad thing... NaN

tweet_created tweet_location user_timezone
0 2015-02-24 11:15:52 -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)

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

```
DAY 3 LAB EXPERIMENTS

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all

Files
-
sample_data
Tweets.csv
ecommerce customers.csv

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

... 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
```