Dr. Vishwanath Karad
MIT WORLD PEACE UNIVERSITY | PUNE
TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS

Third Year B. Tech (EL & CE)

Semester: VI **Subject:** Data Science for Engineering

Name: Shreerang Mhatre **Class: TY**

Roll No: 52 **Batch: A2**

Experiment No: 08

Name of the Experiment: Data Visualization Techniques in Python

Performed on: 25/04/2024

Submitted on: 25/04/2024

Problem Statement:

Aim: Implement various data visualization techniques in Python using Matplotlib and Seaborn libraries to analyze and represent data effectively.

(Attempt any 5).

i. Read Total profit of all months and show it using a line plot.

ii. Generate above plot with following style

- a. Line Style dotted Line-color should be red
- b. Show legend at the lower right location.
- c. X label name = Month Number
- d. Y label name = Sold units number
- e. Add a circle marker.
- f Line marker color as read
- g. Line width should be 3

iii. Read the total profit of each month and show it using the histogram to see the most common profit ranges.

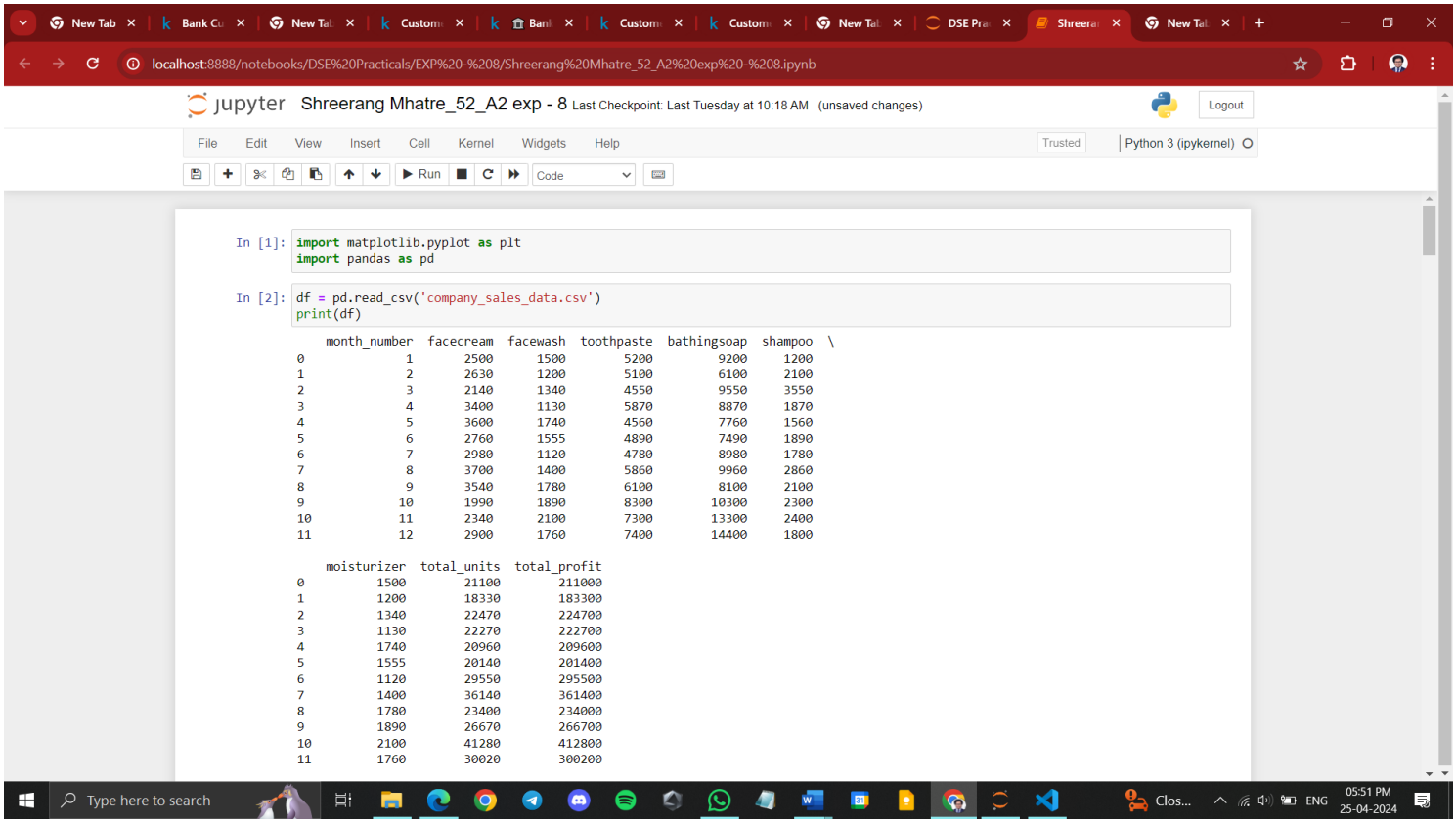
iv. Calculate total sale data for last Product and show it using a Pie

v. Read all product sales data and show it using the stack plot.

vi. Read all product sales data and show it using a multi line plot. Display the number of units sold per month for each product using multi line plots. (i.e., Separate Plotline for each product).

viii Read toothpaste sales data of each month and it using a scatter plot

ix Read face cream facewash sales data and show it using the bar chart



localhost:8888/notebooks/DSE%20Practicals/EXP%20-%208/Shreerang%20Mhatre_52_A2%20exp%20-%208.ipynb

jupyter Shreerang Mhatre_52_A2 exp - 8 Last Checkpoint: Last Tuesday at 10:18 AM (unsaved changes)

Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

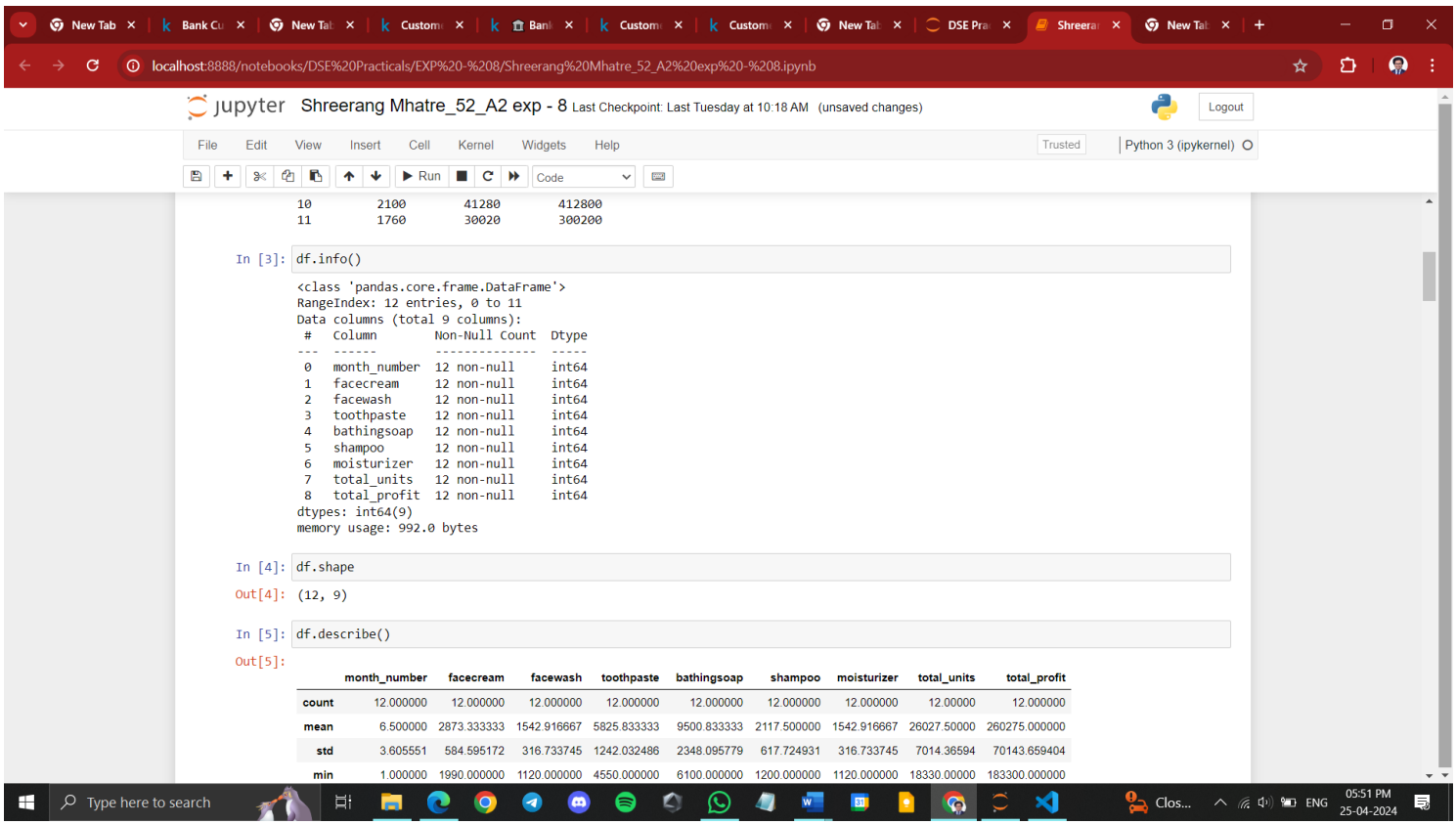
Run Code

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

In [2]:
```
df = pd.read_csv('company_sales_data.csv')
print(df)
```

```
    month_number  facecream  facewash  toothpaste  bathingsoap  shampoo  \
0              1       2500      1500        5200         9200     1200
1              2       2630      1200        5100         6100     2100
2              3       2140      1340        4550         9550     3550
3              4       3400      1130        5870         8870     1870
4              5       3600      1740        4560         7760     1560
5              6       2760      1555        4890         7490     1890
6              7       2980      1120        4780         8980     1780
7              8       3700      1400        5860         9960     2860
8              9       3540      1780        6100         8100     2100
9             10       1990      1890        8300        10300     2300
10            11       2340      2100        7300        13300     2400
11            12       2900      1760        7400        14400     1800

    moisturizer  total_units  total_profit
0          1500        21100        211000
1          1200        18330        183300
2          1340        22470        224700
3          1130        22270        222700
4          1740        20960        209600
5          1555        20140        201400
6          1120        29550        295500
7          1400        36140        361400
8          1780        23400        234000
9          1890        26670        266700
10         2100        41280        412800
11         1760        30020        300200
```



localhost:8888/notebooks/DSE%20Practicals/EXP%20-%208/Shreerang%20Mhatre_52_A2%20exp%20-%208.ipynb

jupyter Shreerang Mhatre_52_A2 exp - 8 Last Checkpoint: Last Tuesday at 10:18 AM (unsaved changes)

Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Run Code

```
10         2100        41280        412800
11         1760        30020        300200
```

In [3]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12 entries, 0 to 11
Data columns (total 9 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   month_number  12 non-null     int64
 1   facecream     12 non-null     int64
 2   facewash      12 non-null     int64
 3   toothpaste    12 non-null     int64
 4   bathingsoap   12 non-null     int64
 5   shampoo       12 non-null     int64
 6   moisturizer   12 non-null     int64
 7   total_units   12 non-null     int64
 8   total_profit  12 non-null     int64
dtypes: int64(9)
memory usage: 992.0 bytes
```

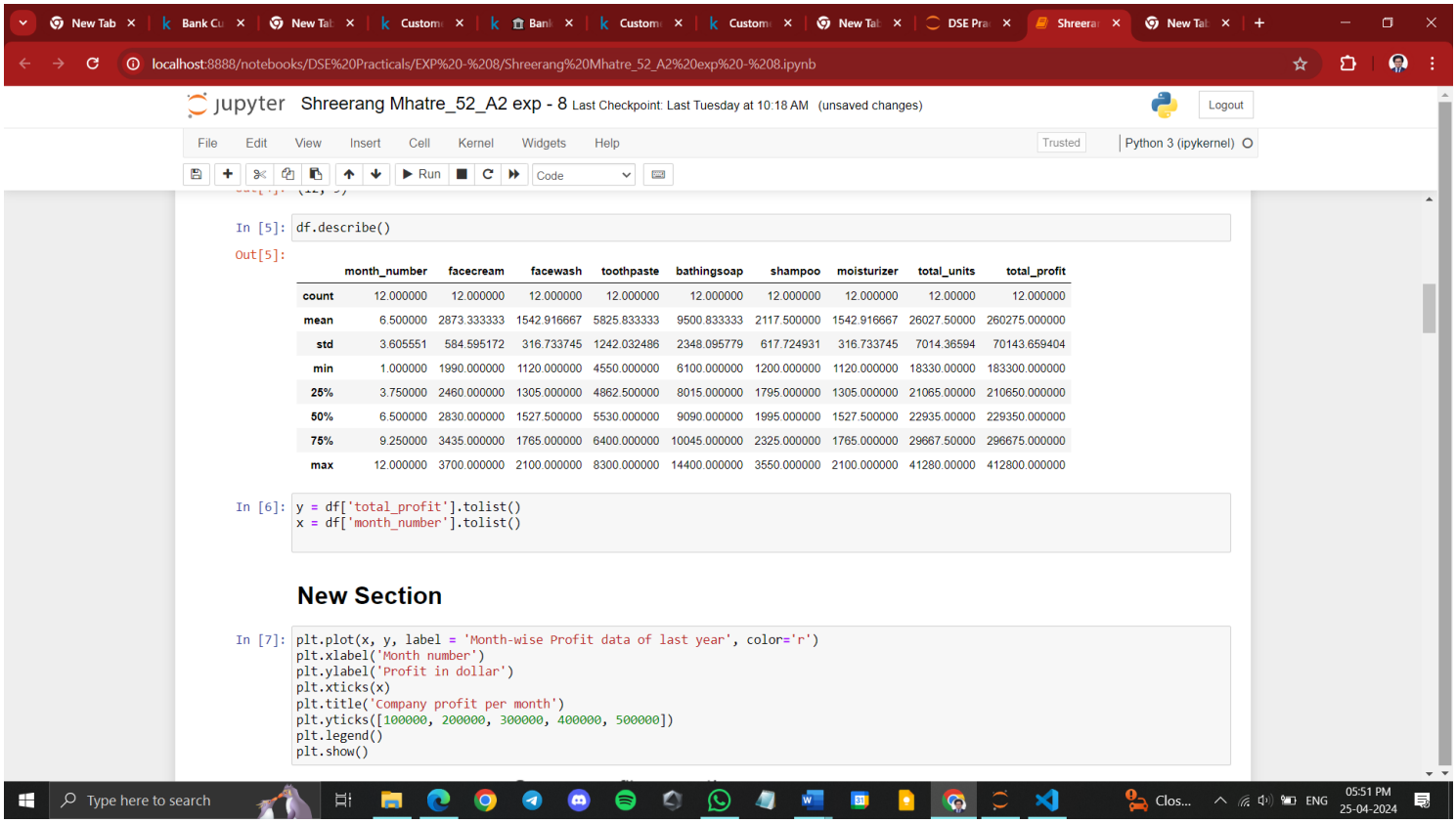
In [4]:
```
df.shape
```

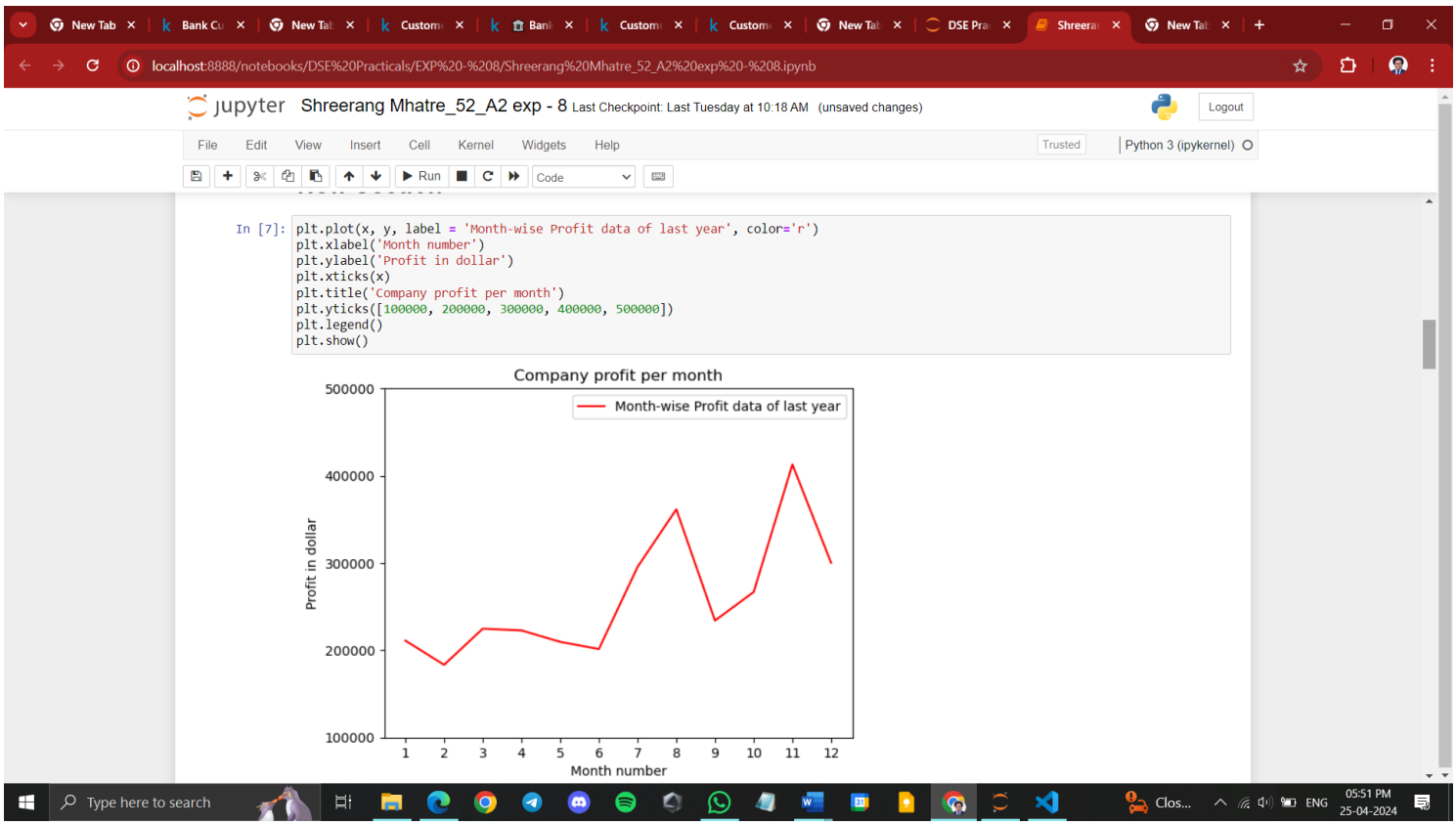
Out[4]: (12, 9)

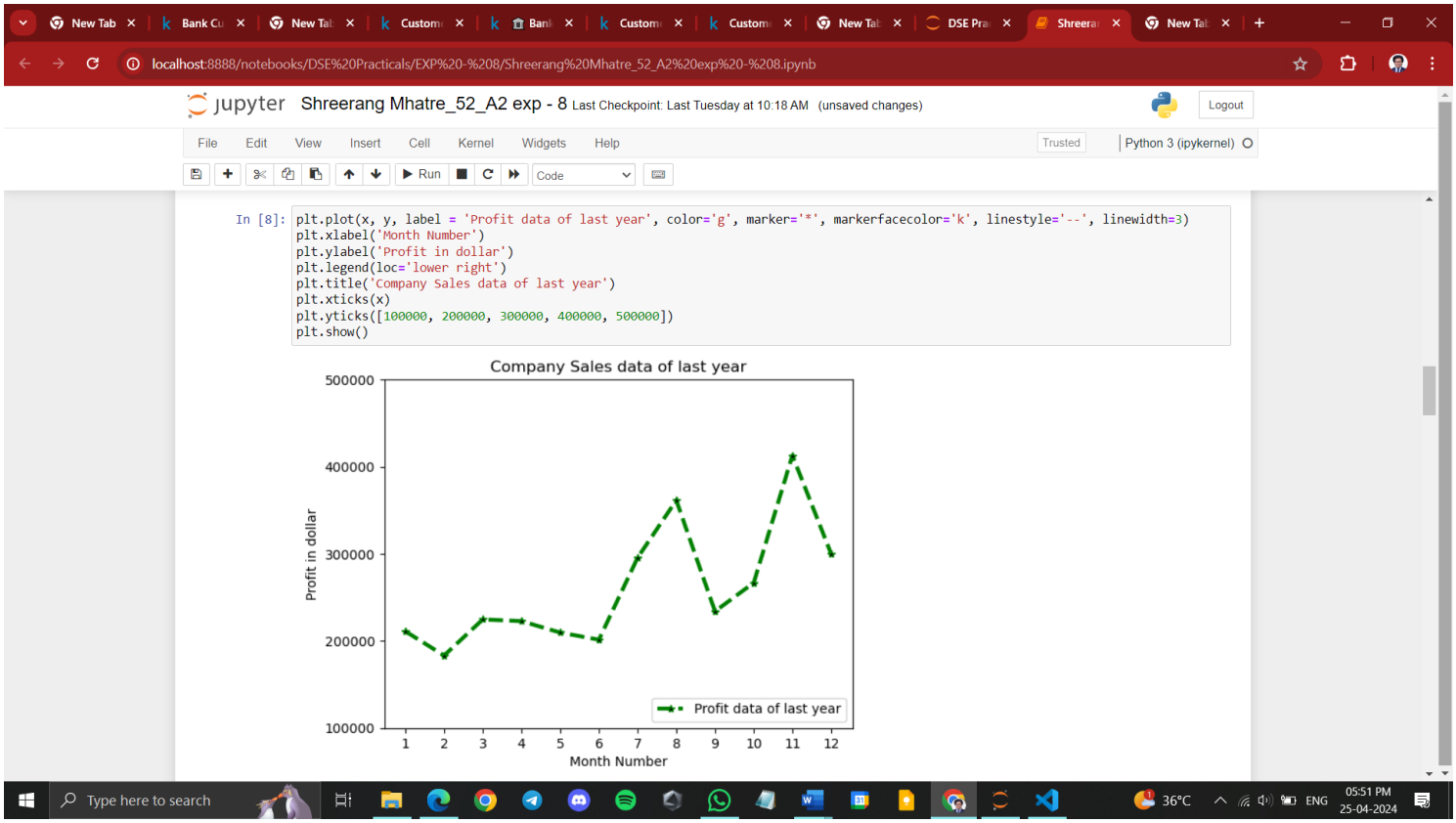
In [5]:
```
df.describe()
```

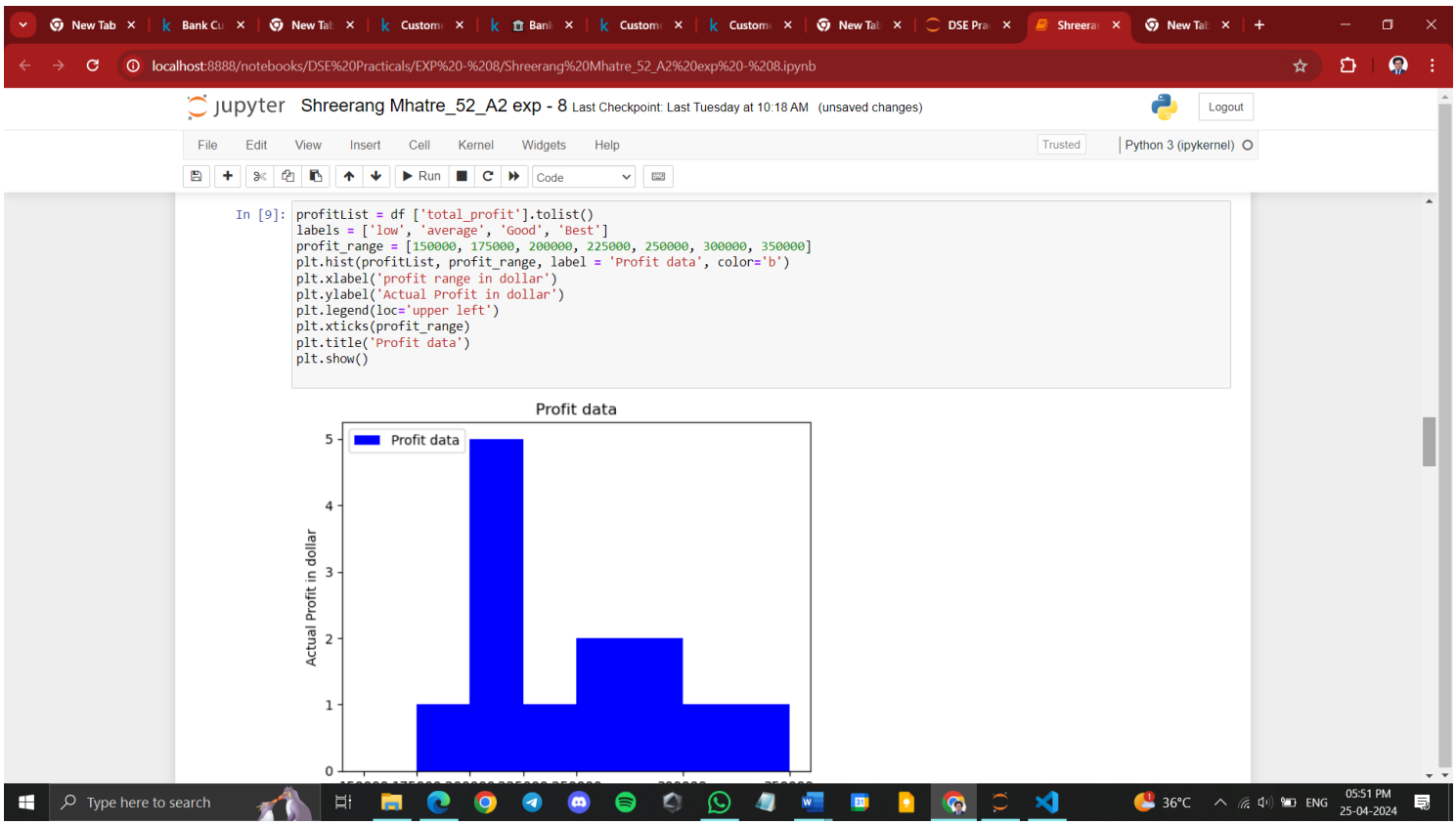
Out[5]:

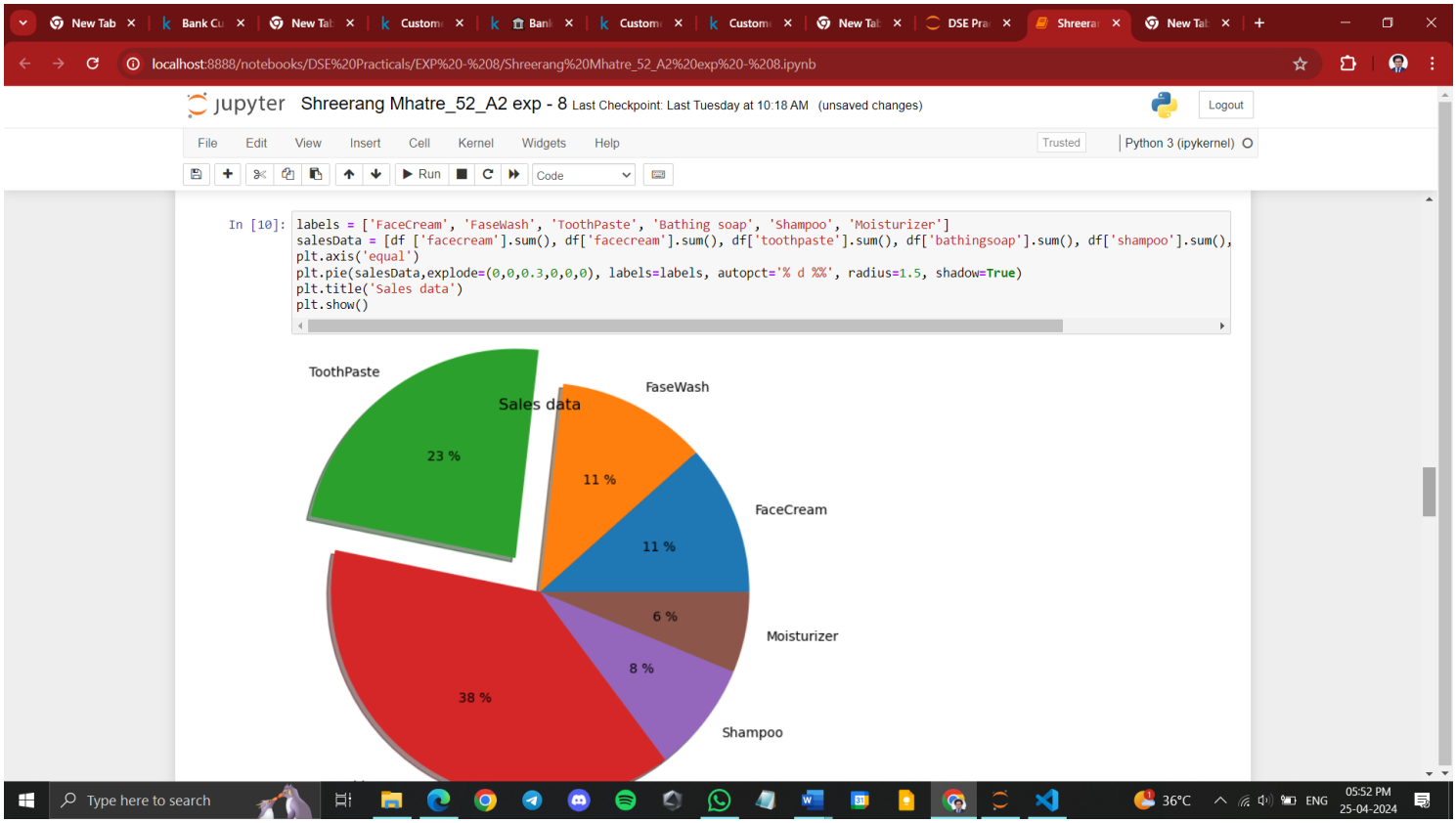
	month_number	facecream	facewash	toothpaste	bathingsoap	shampoo	moisturizer	total_units	total_profit
count	12.000000	12.000000	12.000000	12.000000	12.000000	12.000000	12.000000	12.00000	12.000000
mean	6.500000	2873.333333	1542.916667	5825.833333	9500.833333	2117.500000	1542.916667	26027.50000	260275.000000
std	3.605551	584.595172	316.733745	1242.032486	2348.095779	617.724931	316.733745	7014.36594	70143.659404
min	1.000000	1990.000000	1120.000000	4550.000000	6100.000000	1200.000000	1120.000000	18330.00000	183300.000000

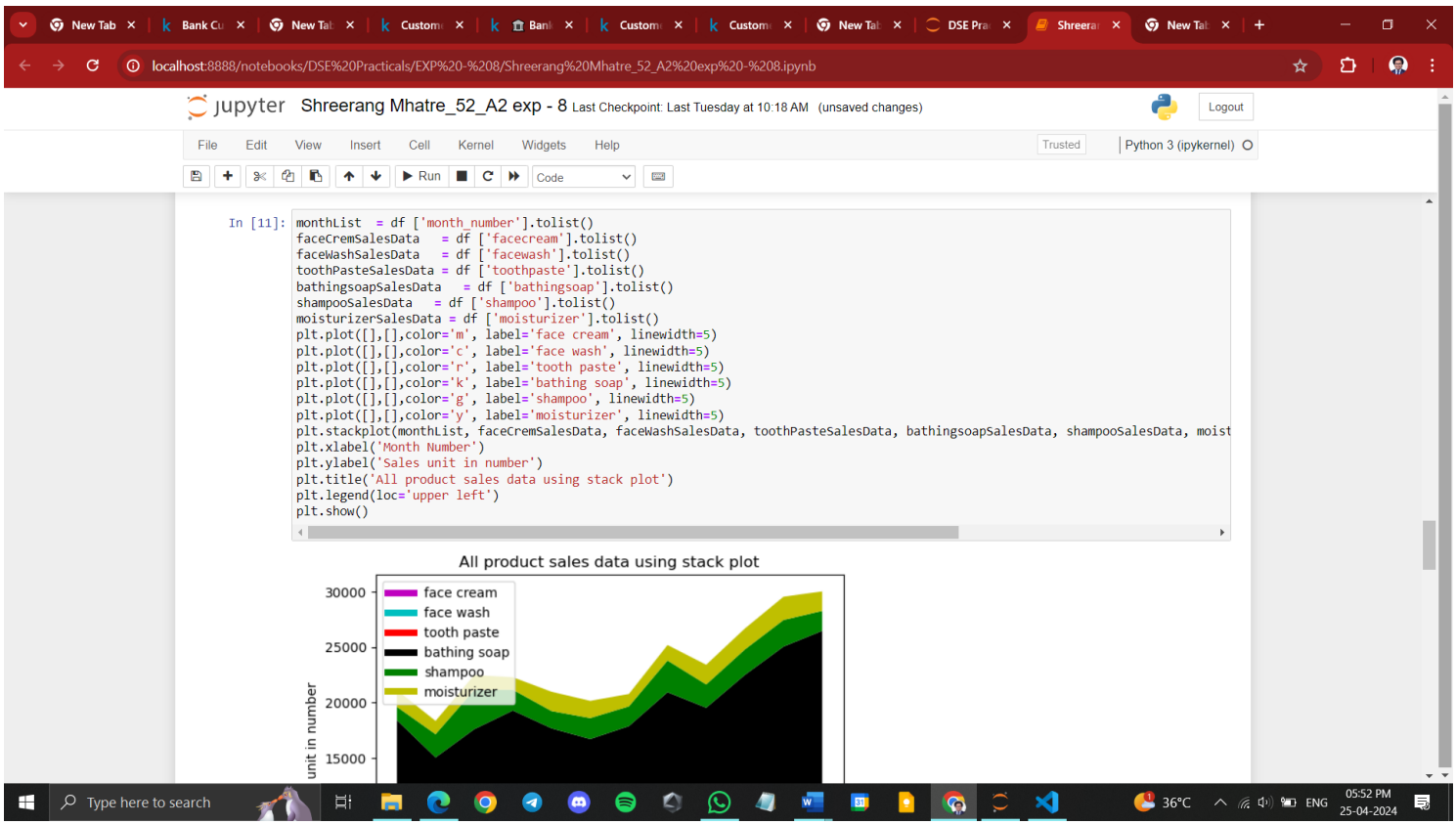

localhost:8888/notebooks/DSE%20Practicals/EXP%20-%208/Shreerang%20Mhatre_52_A2%20exp%20-%208.ipynb
jupyter Shreerang Mhatre_52_A2 exp - 8 Last Checkpoint: Last Tuesday at 10:18 AM (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [5]: df.describe()
Out[5]:
month_number facecream facewash toothpaste bathingsoap shampoo moisturizer total_units total_profit
count 12.000000 12.000000 12.000000 12.000000 12.000000 12.000000 12.000000 12.00000 12.000000
mean 6.500000 2873.333333 1542.916667 5825.833333 9500.833333 2117.500000 1542.916667 26027.50000 260275.000000
std 3.605551 584.595172 316.733745 1242.032486 2348.095779 617.724931 316.733745 7014.36594 70143.659404
min 1.000000 1990.000000 1120.000000 4550.000000 6100.000000 1200.000000 1120.000000 18330.00000 183300.000000
25% 3.750000 2460.000000 1305.000000 4862.500000 8015.000000 1795.000000 1305.000000 21065.00000 210650.000000
50% 6.500000 2830.000000 1527.500000 5530.000000 9090.000000 1995.000000 1527.500000 22935.00000 229350.000000
75% 9.250000 3435.000000 1765.000000 6400.000000 10045.000000 2325.000000 1765.000000 29667.50000 296675.000000
max 12.000000 3700.000000 2100.000000 8300.000000 14400.000000 3550.000000 2100.000000 41280.00000 412800.000000
In [6]: y = df['total_profit'].tolist()
x = df['month_number'].tolist()
New Section
In [7]: plt.plot(x, y, label = 'Month-wise Profit data of last year', color='r')
plt.xlabel('Month number')
plt.ylabel('Profit in dollar')
plt.xticks(x)
plt.title('Company profit per month')
plt.yticks([100000, 200000, 300000, 400000, 500000])
plt.legend()
plt.show()
Type here to search
ENG
05:51 PM
25-04-2024

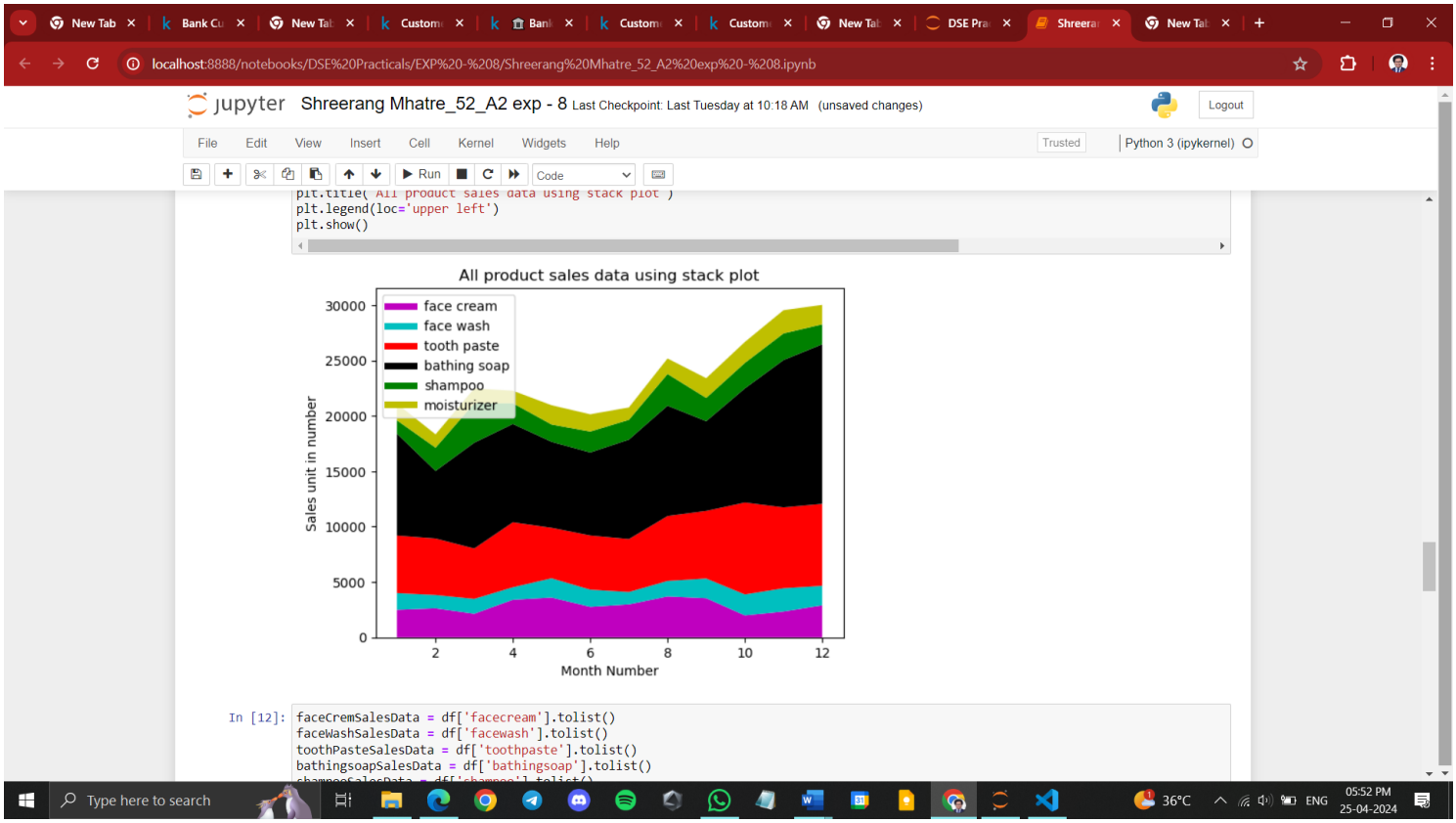

localhost:8888/notebooks/DSE%20Practicals/EXP%20-%208/Shreerang%20Mhatre_52_A2%20exp%20-%208.ipynb
jupyter Shreerang Mhatre_52_A2 exp - 8 Last Checkpoint: Last Tuesday at 10:18 AM (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [7]: plt.plot(x, y, label = 'Month-wise Profit data of last year', color='r')
plt.xlabel('Month number')
plt.ylabel('Profit in dollar')
plt.xticks(x)
plt.title('Company profit per month')
plt.yticks([100000, 200000, 300000, 400000, 500000])
plt.legend()
plt.show()
Company profit per month
Month-wise Profit data of last year
Profit in dollar
500000
400000
300000
200000
100000
1 2 3 4 5 6 7 8 9 10 11 12
Month number
Type here to search
ENG
05:51 PM
25-04-2024

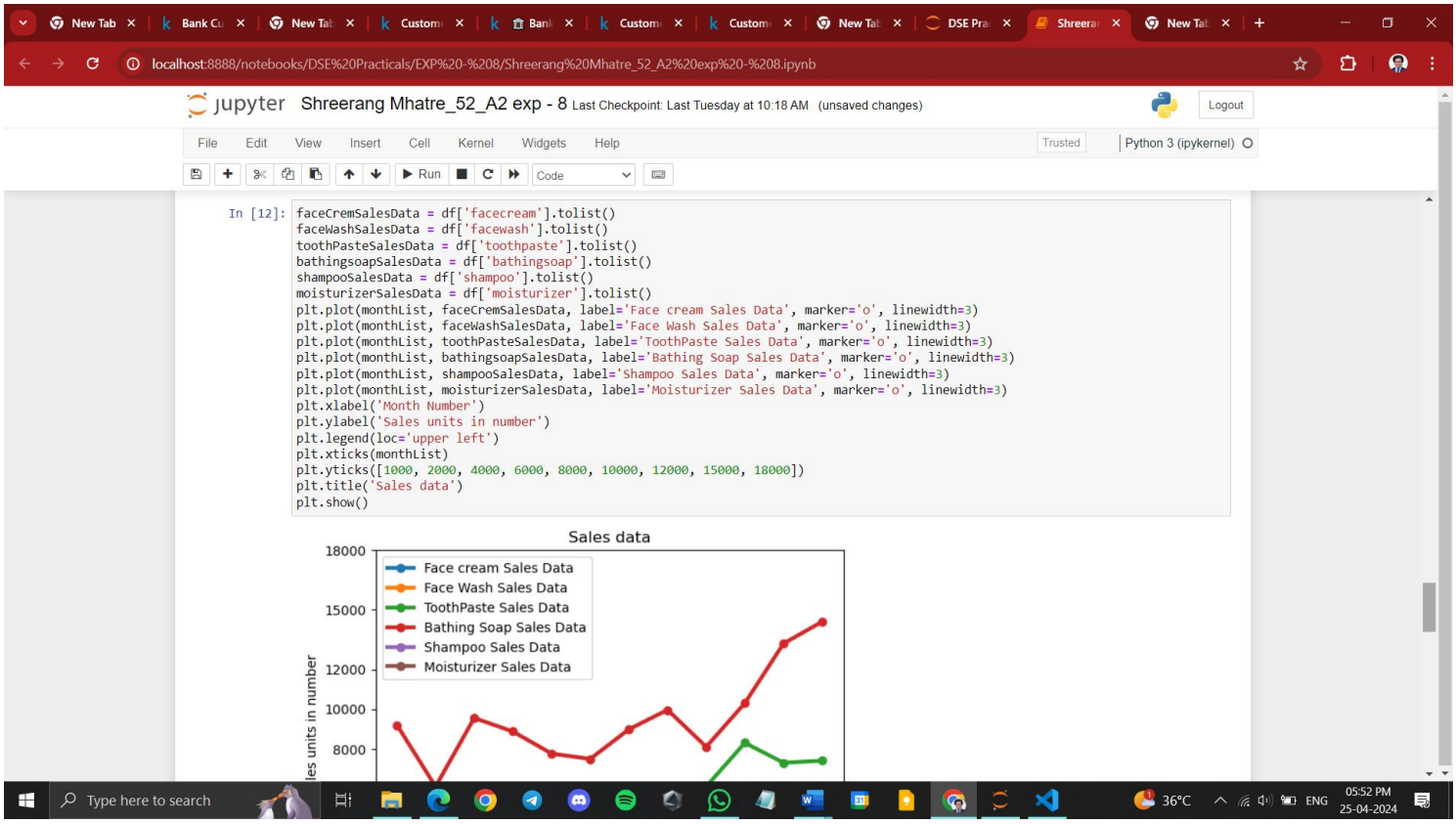

localhost:8888/notebooks/DSE%20Practicals/EXP%20-%208/Shreerang%20Mhatre_52_A2%20exp%20-%208.ipynb
jupyter Shreerang Mhatre_52_A2 exp - 8 Last Checkpoint: Last Tuesday at 10:18 AM (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [8]:
plt.plot(x, y, label = 'Profit data of last year', color='g', marker='*', markerfacecolor='k', linestyle='--', linewidth=3)
plt.xlabel('Month Number')
plt.ylabel('Profit in dollar')
plt.legend(loc='lower right')
plt.title('Company Sales data of last year')
plt.xticks(x)
plt.yticks([100000, 200000, 300000, 400000, 500000])
plt.show()
Company Sales data of last year
Profit in dollar
Month Number
Profit data of last year
Type here to search
36°C
ENG
05:51 PM
25-04-2024

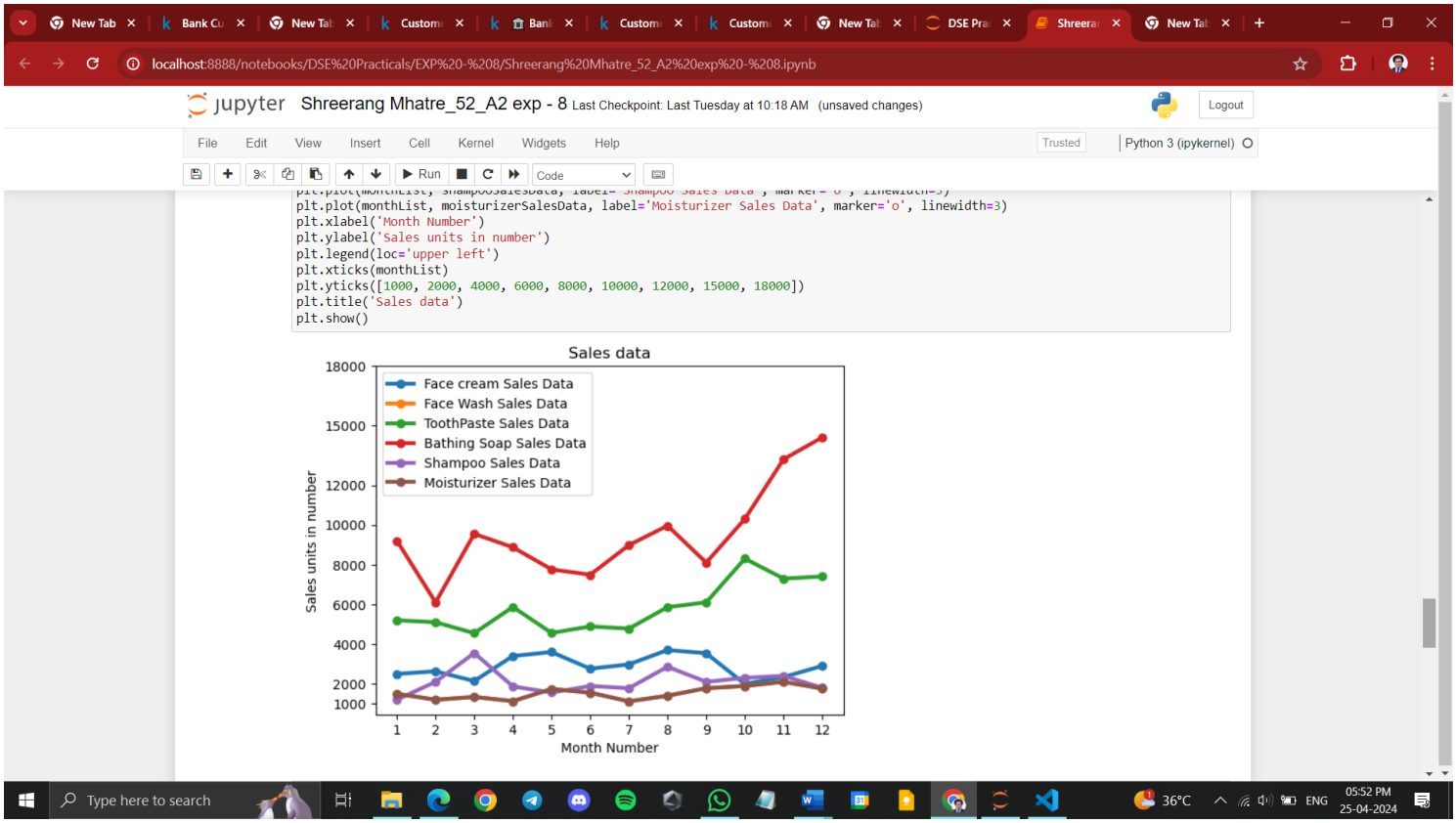

localhost:8888/notebooks/DSE%20Practicals/EXP%20-%208/Shreerang%20Mhatre_52_A2%20exp%20-%208.ipynb
jupyter Shreerang Mhatre_52_A2 exp - 8 Last Checkpoint: Last Tuesday at 10:18 AM (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [9]:
profitList = df ['total_profit'].tolist()
labels = ['low', 'average', 'Good', 'Best']
profit_range = [150000, 175000, 200000, 225000, 250000, 300000, 350000]
plt.hist(profitList, profit_range, label = 'Profit data', color='b')
plt.xlabel('profit range in dollar')
plt.ylabel('Actual Profit in dollar')
plt.legend(loc='upper left')
plt.xticks(profit_range)
plt.title('Profit data')
plt.show()
Profit data
Profit data
Actual Profit in dollar
Type here to search
36°C
ENG
05:51 PM
25-04-2024

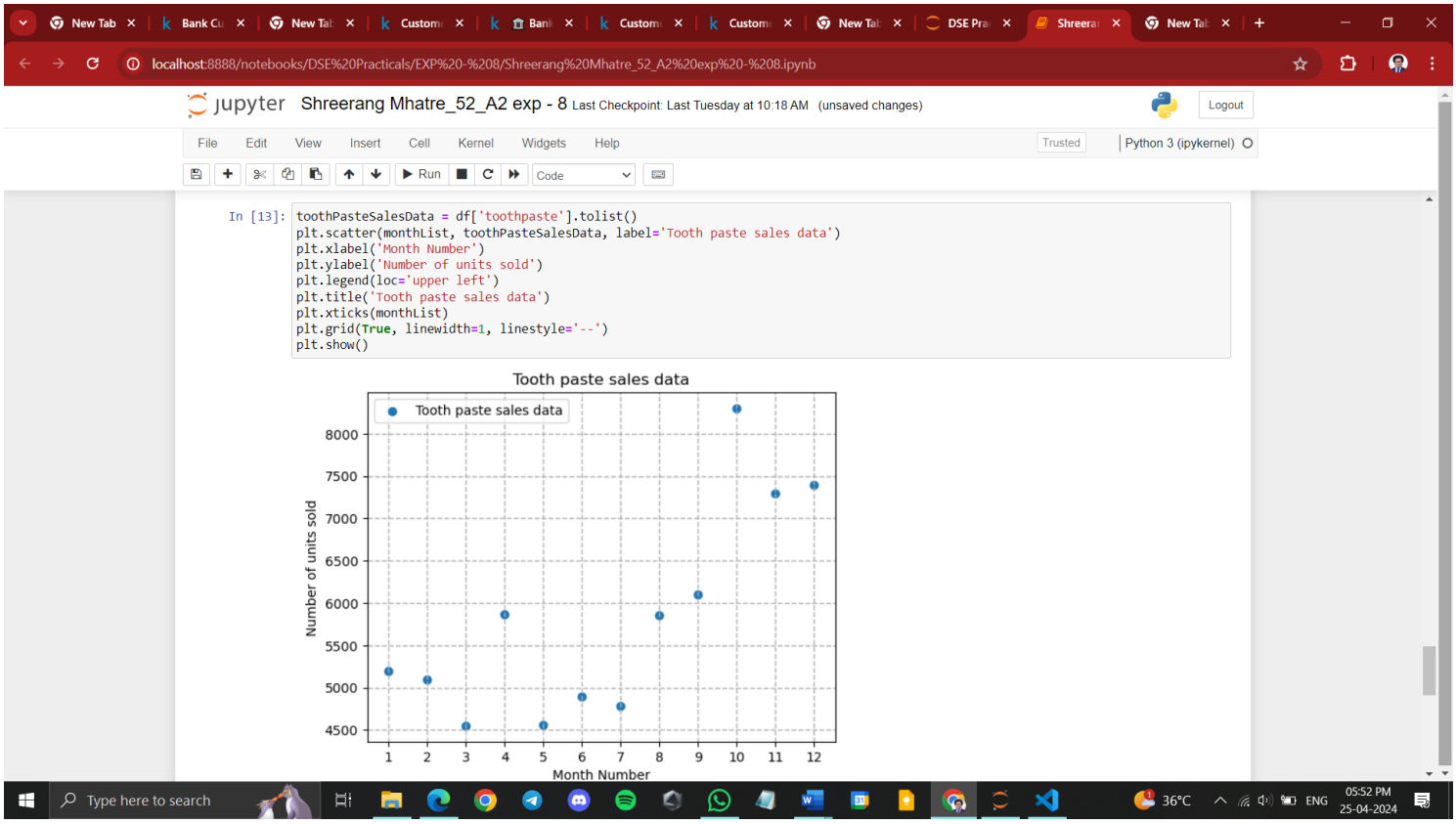

localhost:8888/notebooks/DSE%20Practicals/EXP%20-%208/Shreerang%20Mhatre_52_A2%20exp%20-%208.ipynb
jupyter Shreerang Mhatre_52_A2 exp - 8 Last Checkpoint: Last Tuesday at 10:18 AM (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [10]:
labels = ['FaceCream', 'FaseWash', 'ToothPaste', 'Bathing soap', 'Shampoo', 'Moisturizer']
salesData = [df ['facecream'].sum(), df['facecream'].sum(), df['toothpaste'].sum(), df['bathingsoap'].sum(), df['shampoo'].sum(),
plt.axis('equal')
plt.pie(salesData,explode=(0,0,0.3,0,0,0), labels=labels, autopct='% d %%', radius=1.5, shadow=True)
plt.title('Sales data')
plt.show()
Sales data
ToothPaste
FaseWash
FaceCream
Moisturizer
Shampoo
23 %
11 %
11 %
6 %
8 %
38 %
Type here to search
36°C
ENG
05:52 PM
25-04-2024

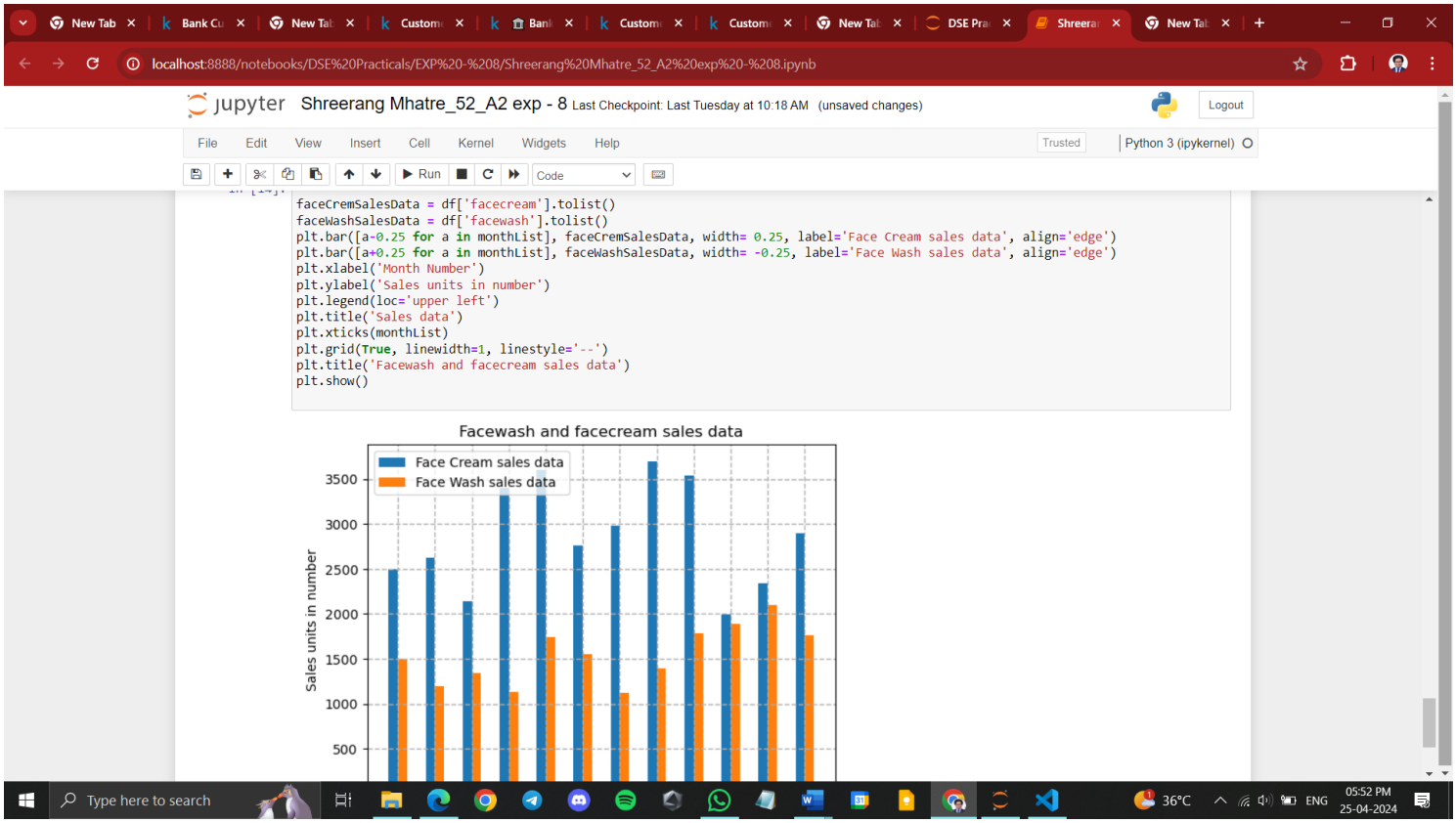

localhost:8888/notebooks/DSE%20Practicals/EXP%20-%208/Shreerang%20Mhatre_52_A2%20exp%20-%208.ipynb
jupyter Shreerang Mhatre_52_A2 exp - 8 Last Checkpoint: Last Tuesday at 10:18 AM (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [11]:
monthList = df ['month_number'].tolist()
faceCremSalesData = df ['facecream'].tolist()
faceWashSalesData = df ['facewash'].tolist()
toothPasteSalesData = df ['toothpaste'].tolist()
bathingsoapSalesData = df ['bathingsoap'].tolist()
shampooSalesData = df ['shampoo'].tolist()
moisturizerSalesData = df ['moisturizer'].tolist()
plt.plot([],[],color='m', label='face cream', linewidth=5)
plt.plot([],[],color='c', label='face wash', linewidth=5)
plt.plot([],[],color='r', label='tooth paste', linewidth=5)
plt.plot([],[],color='k', label='bathing soap', linewidth=5)
plt.plot([],[],color='g', label='shampoo', linewidth=5)
plt.plot([],[],color='y', label='moisturizer', linewidth=5)
plt.stackplot(monthList, faceCremSalesData, faceWashSalesData, toothPasteSalesData, bathingsoapSalesData, shampooSalesData, moist
plt.xlabel('Month Number')
plt.ylabel('Sales unit in number')
plt.title('All product sales data using stack plot')
plt.legend(loc='upper left')
plt.show()
All product sales data using stack plot
face cream
face wash
tooth paste
bathing soap
shampoo
moisturizer
30000
25000
20000
15000
unit in number
Type here to search
36°C
ENG
05:52 PM
25-04-2024

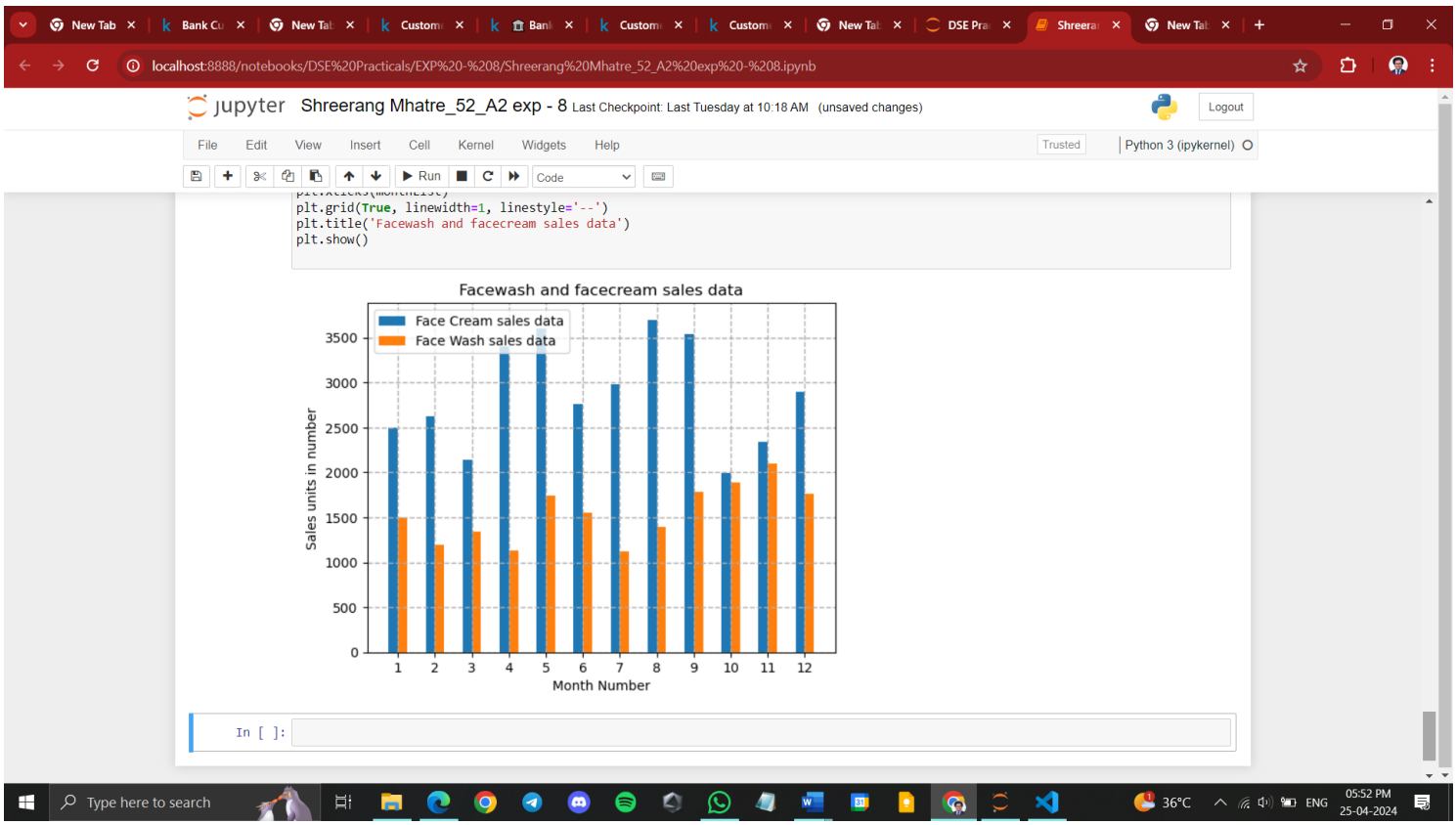

localhost:8888/notebooks/DSE%20Practicals/EXP%20-%208/Shreerang%20Mhatre_52_A2%20exp%20-%208.ipynb
jupyter Shreerang Mhatre_52_A2 exp - 8 Last Checkpoint: Last Tuesday at 10:18 AM (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
plt.title(All product sales data using stack plot)
plt.legend(loc='upper left')
plt.show()
All product sales data using stack plot
face cream
face wash
tooth paste
bathing soap
shampoo
moisturizer
Sales unit in number
Month Number
In [12]:
faceCremSalesData = df['facecream'].tolist()
faceWashSalesData = df['facewash'].tolist()
toothPasteSalesData = df['toothpaste'].tolist()
bathingsoapSalesData = df['bathingsoap'].tolist()
Type here to search
36°C
ENG
05:52 PM
25-04-2024


localhost:8888/notebooks/DSE%20Practicals/EXP%20-%208/Shreerang%20Mhatre_52_A2%20exp%20-%208.ipynb
jupyter Shreerang Mhatre_52_A2 exp - 8 Last Checkpoint: Last Tuesday at 10:18 AM (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [12]:
faceCremSalesData = df['facecream'].tolist()
faceWashSalesData = df['facewash'].tolist()
toothPasteSalesData = df['toothpaste'].tolist()
bathingsoapSalesData = df['bathingsoap'].tolist()
shampooSalesData = df['shampoo'].tolist()
moisturizerSalesData = df['moisturizer'].tolist()
plt.plot(monthList, faceCremSalesData, label='Face cream Sales Data', marker='o', linewidth=3)
plt.plot(monthList, faceWashSalesData, label='Face Wash Sales Data', marker='o', linewidth=3)
plt.plot(monthList, toothPasteSalesData, label='ToothPaste Sales Data', marker='o', linewidth=3)
plt.plot(monthList, bathingsoapSalesData, label='Bathing Soap Sales Data', marker='o', linewidth=3)
plt.plot(monthList, shampooSalesData, label='Shampoo Sales Data', marker='o', linewidth=3)
plt.plot(monthList, moisturizerSalesData, label='Moisturizer Sales Data', marker='o', linewidth=3)
plt.xlabel('Month Number')
plt.ylabel('Sales units in number')
plt.legend(loc='upper left')
plt.xticks(monthList)
plt.yticks([1000, 2000, 4000, 6000, 8000, 10000, 12000, 15000, 18000])
plt.title('Sales data')
plt.show()
Sales data
Face cream Sales Data
Face Wash Sales Data
ToothPaste Sales Data
Bathing Soap Sales Data
Shampoo Sales Data
Moisturizer Sales Data
Type here to search
36°C
ENG
05:52 PM
25-04-2024


localhost:8888/notebooks/DSE%20Practicals/EXP%20-%208/Shreerang%20Mhatre_52_A2%20exp%20-%208.ipynb
jupyter Shreerang Mhatre_52_A2 exp - 8 Last Checkpoint: Last Tuesday at 10:18 AM (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
plt.plot(monthList, moisturizerSalesData, label='Moisturizer Sales Data', marker='o', linewidth=3)
plt.xlabel('Month Number')
plt.ylabel('Sales units in number')
plt.legend(loc='upper left')
plt.xticks(monthList)
plt.yticks([1000, 2000, 4000, 6000, 8000, 10000, 12000, 15000, 18000])
plt.title('Sales data')
plt.show()
Sales data
Face cream Sales Data
Face Wash Sales Data
ToothPaste Sales Data
Bathing Soap Sales Data
Shampoo Sales Data
Moisturizer Sales Data
Sales units in number
Month Number
Type here to search
36°C
ENG
05:52 PM
25-04-2024


localhost:8888/notebooks/DSE%20Practicals/EXP%20-%208/Shreerang%20Mhatre_52_A2%20exp%20-%208.ipynb
jupyter Shreerang Mhatre_52_A2 exp - 8 Last Checkpoint: Last Tuesday at 10:18 AM (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [13]:
toothPasteSalesData = df['toothpaste'].tolist()
plt.scatter(monthList, toothPasteSalesData, label='Tooth paste sales data')
plt.xlabel('Month Number')
plt.ylabel('Number of units sold')
plt.legend(loc='upper left')
plt.title('Tooth paste sales data')
plt.xticks(monthList)
plt.grid(True, linewidth=1, linestyle='--')
plt.show()
Tooth paste sales data
Tooth paste sales data
Number of units sold
Month Number
Type here to search
36°C
ENG
05:52 PM
25-04-2024


localhost:8888/notebooks/DSE%20Practicals/EXP%20-%208/Shreerang%20Mhatre_52_A2%20exp%20-%208.ipynb
jupyter Shreerang Mhatre_52_A2 exp - 8 Last Checkpoint: Last Tuesday at 10:18 AM (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
faceCremSalesData = df['facecream'].tolist()
faceWashSalesData = df['facewash'].tolist()
plt.bar([a-0.25 for a in monthList], faceCremSalesData, width= 0.25, label='Face Cream sales data', align='edge')
plt.bar([a+0.25 for a in monthList], faceWashSalesData, width= -0.25, label='Face Wash sales data', align='edge')
plt.xlabel('Month Number')
plt.ylabel('Sales units in number')
plt.legend(loc='upper left')
plt.title('Sales data')
plt.xticks(monthList)
plt.grid(True, linewidth=1, linestyle='--')
plt.title('Facewash and facecream sales data')
plt.show()
Facewash and facecream sales data
Face Cream sales data
Face Wash sales data
Sales units in number
3500
3000
2500
2000
1500
1000
500
Type here to search
36°C
ENG
05:52 PM
25-04-2024


localhost:8888/notebooks/DSE%20Practicals/EXP%20-%208/Shreerang%20Mhatre_52_A2%20exp%20-%208.ipynb
jupyter Shreerang Mhatre_52_A2 exp - 8 Last Checkpoint: Last Tuesday at 10:18 AM (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
plt.grid(True, linewidth=1, linestyle='--')
plt.title('Facewash and facecream sales data')
plt.show()
Facewash and facecream sales data
Face Cream sales data
Face Wash sales data
Sales units in number
3500
3000
2500
2000
1500
1000
500
0
1
2
3
4
5
6
7
8
9
10
11
12
Month Number
In [ ]:
Type here to search
36°C
ENG
05:52 PM
25-04-2024

Exp 8 - DSE

Name - Shreerang Mhatre
Roll no - 52

* Post Lab Questions

Q1) Explain the importance of data visualization in data analysis.

→ Data visualization plays a crucial role in data analysis for several reasons -

① Insight Discovery -
Visualization allows analysts to identify patterns, trends, correlations, and outliers in the data that may not be apparent from raw numbers or tables.

② Communication of Findings
Visualisations are powerful tools for communicating findings and insights to stakeholders, decision-makers or non-technical audiences.

③ Data Quality Assesment
Visualization helps in assessing data quality by revealing inconsistencies, missing values, outliers & anomalies visually.

www.mitwpu.edu.in

Q2) Describe the differences between various types of plots used in this experiment.

→ ① Line plot -

A line plot is used to visualize data points connected by straight line segments ; typically used for showing trends or changes over time.

② Customized Line Plot -

Dotted line indicates a different style for the line segments. Red color makes the line stand out & draws attention to the data.

③ Histogram -

A histogram is used to visualize the distribution of data by grouping it into bins or intervals & showing the frequency of occurances in each bin.

④ Pie chart

A pie chart represents data in a circular graph, where each slice represents a proportion of the whole.

Q3) Discuss the advantages & limitations of using pie charts for data representation.

→ Pie charts are a common type of data visualization used to represent categorical data and show the proportional distribution of different categories within a whole. While pie charts have several advantages, they also come with limitations that need to be considered.

→ Advantages of Pie charts.
1. Easy to Understand
2. Effective for Comparing Proportions
3. Visually Appealing
4. Suitable for showing Parts of a whole.
5. Simplicity & clarity.

→ Limitations of Pie charts
1. Limited Categories
2. Difficulty in Comparing Small Slices
3. Lack of Precision
4. Misleading Aspect Ratios
5. Limited Context
6. Complex Interpretation with Many Categories