VIT-AP UNIVERSITY, ANDHRA PRADESH

Lab Sheet :4

DW&DM

Academic year: 2022-2023 Branch/ Class: B.Tech

Semester: Fall Date: 18/9/2022

Student name: M.Jaswanth **Reg. no.: 20BCD7171**

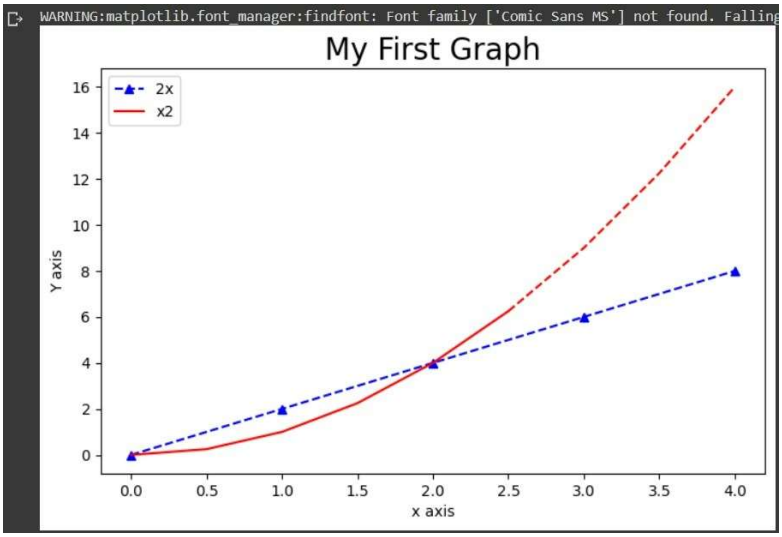
---

1. Draw Basic Graph using following data and instructions:

- • Data: x = [0,1,2,3,4], y = [0,2,4,6,8] ○ Plot the first line with a size (8,5), density per inch is 100, color is red, marker is ^, dashed line and the label is '2x'.
-
- • Data: x2 is a NumPy array of equal distributed from 0 to 4.5 with a space of 0.5 ○ Plot the part of second line (up to 5 data items) with a red color and label it x^2. Here you need to calculate the appropriate y values based on the label.
- ○ Plot the remaining part of second line with a red color and dashed line.
-
- • Name the entire graph with title "My First Graph" with a 'Comic Sans MS' font name and font size is 20.
- • Label the X axis and Y axis with appropriate ticks
- • Save the figure as "MyFirstGraph.png" and display in the same Jupyter notebook.

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
x = [0,1,2,3,4]
y = [0,2,4,6,8]
x2 = np.arange(0,4.5,0.5)
y2 = x2**2
plt.figure(figsize = (8,5), dpi=100)
plt.plot(x,y,"b^--",label = "2x")
plt.plot(x2[0:6],y2[0:6], "r-", label = "x2")
plt.plot(x2[5:],y2[5:],"r--")
plt.xlabel("x axis")
plt.ylabel("Y axis")
plt.legend()
plt.title("My First Graph", fontname="Comic Sans MS", fontsize = 20)
plt.savefig("MyFirstGraph.png")
```

Output



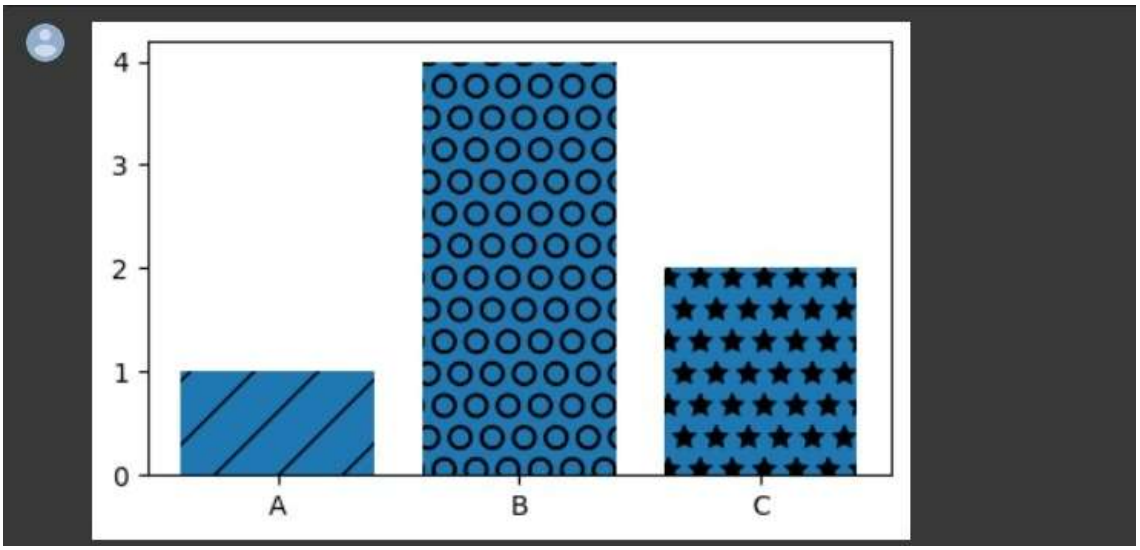
2.

Draw Bar Chart using following data and instructions:

- o Data: labels = ['A', 'B', 'C'], values = [1,4,2]
- o Size of the figure is (5,3) and dpi is 100. Show each bar with different patterns.
- o Save this bar chart as 'barchart.png' and display it

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

labels = ['A', 'B', 'C']
values = [1,4,2]
plt.figure(figsize=(5,3), dpi=100)
bars = plt.bar(labels, values)
patterns = ['/', 'O', '*']
for bar in bars:
    bar.set_hatch(patterns.pop(0))
plt.savefig('barchart.png', dpi=300)
plt.show()
```

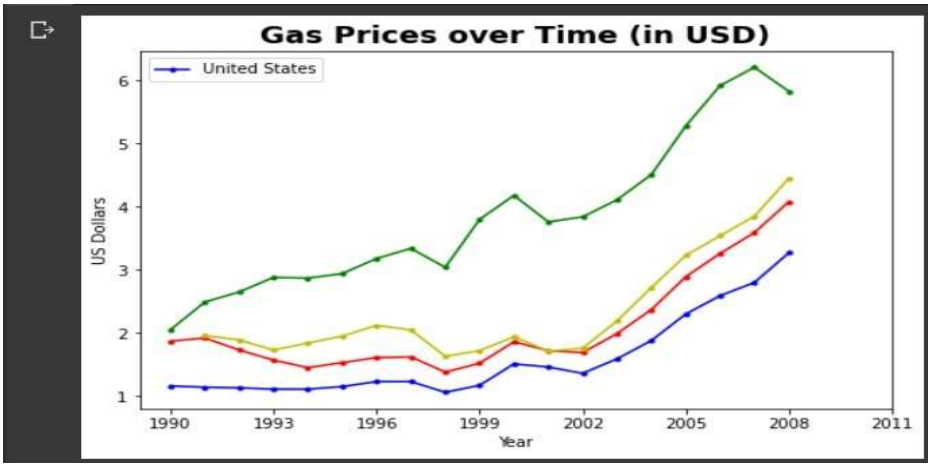
Output



3. Plot the “Gas prices over time” using the data set “gas_prices.csv”

```
import pandas as pd
import matplotlib.pyplot as plt
gas=pd.read_csv('gas_prices.csv')

plt.figure(figsize=(8,5))

plt.title('Gas Prices over Time (in USD)',fontdict={'fontweight':'bold','fontsize': 18})

plt.plot(gas.Year, gas.USA, 'b.-', label='United States')
plt.plot(gas.Year, gas.Canada, 'r.-')
plt.plot(gas.Year, gas['South Korea'], 'g.-')
plt.plot(gas.Year, gas.Australia, 'y.-')


plt.xticks(gas.Year[::3].tolist()+[2011])

plt.xlabel('Year')
plt.ylabel('US Dollars')

plt.legend()

plt.savefig('Gas_price_figure.png', dpi=300)

plt.show()
```

Output



4. Draw histogram for the 'Overall' attribute of the data set 'fifa_data.csv'.

```
[ ] import pandas as pd
    import matplotlib.pyplot as plt
    fifa = pd.read_csv('fifa_data.csv')
    fifa.head(5)
```

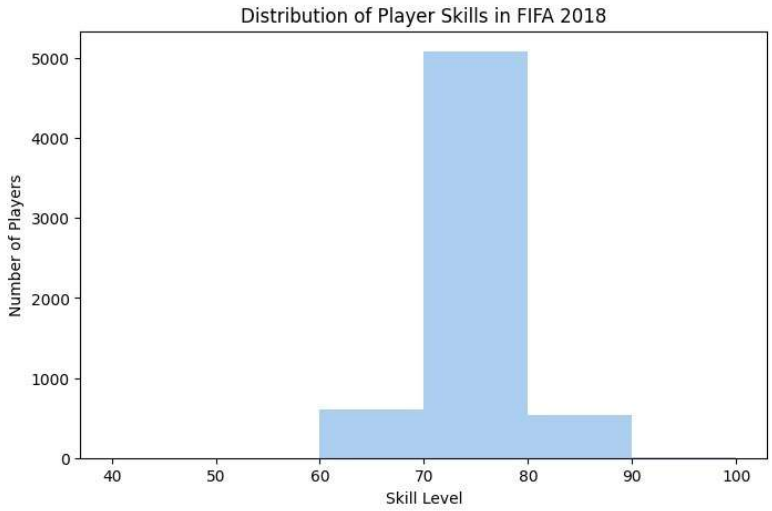
Output

	Unnamed: 0	ID	Name	Age	Photo	Nationality	Flag	Overall	Potential	Club	...	Composure
0	0	158023	L. Messi	31	https://cdn.sofifa.org/players/4/19/158023.png	Argentina	https://cdn.sofifa.org/flags/52.png	94	94	FC Barcelona	...	96.0
1	1	20801	Cristiano Ronaldo	33	https://cdn.sofifa.org/players/4/19/20801.png	Portugal	https://cdn.sofifa.org/flags/38.png	94	94	Juventus	...	95.0
2	2	190871	Neymar Jr	26	https://cdn.sofifa.org/players/4/19/190871.png	Brazil	https://cdn.sofifa.org/flags/54.png	92	93	Paris Saint-Germain	...	94.0
3	3	193080	De Gea	27	https://cdn.sofifa.org/players/4/19/193080.png	Spain	https://cdn.sofifa.org/flags/45.png	91	93	Manchester United	...	68.0
4	4	192985	K. De Bruyne	27	https://cdn.sofifa.org/players/4/19/192985.png	Belgium	https://cdn.sofifa.org/flags/7.png	91	92	Manchester City	...	88.0

5 rows × 89 columns

Marking	StandingTackle	SlidingTackle	GKDiving	GKHandling	GKKicking	GKPositioning	GKReflexes	Release Clause
33.0	28.0	26.0	6.0	11.0	15.0	14.0	8.0	€226.5M
28.0	31.0	23.0	7.0	11.0	15.0	14.0	11.0	€127.1M
27.0	24.0	33.0	9.0	9.0	15.0	15.0	11.0	€228.1M
15.0	21.0	13.0	90.0	85.0	87.0	88.0	94.0	€138.6M
68.0	58.0	51.0	15.0	13.0	5.0	10.0	13.0	€196.4M

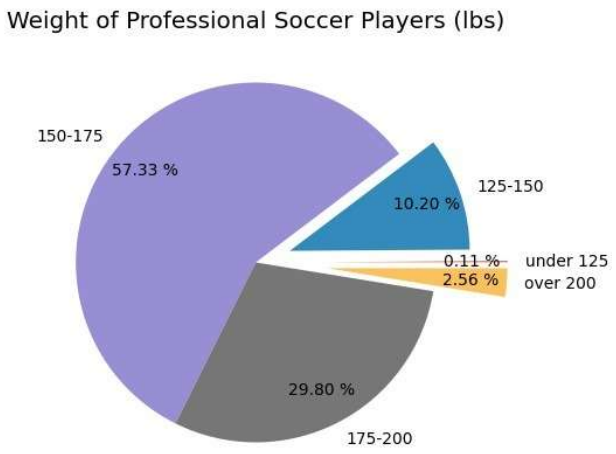
```
bins = [40,50,60,70,80,90,100]

plt.figure(figsize=(8,5))

plt.hist(fifa.Overall,bins=bins,color='#abcdef')

plt.xticks(bins)

plt.ylabel('Number of Players')
plt.xlabel('Skill Level')
plt.title('Distribution of Player Skills in FIFA 2018')
plt.savefig('histogram.png', dpi=300)

plt.show()
```



5. Draw the following pie chart which refers the Foot Preference of FIFA players.

```
import pandas as pd
import matplotlib.pyplot as plt
plt.figure(figsize=(8,5), dpi=100)

plt.style.use('ggplot')

fifa.Weight=[int(x.strip('lbs')) if type(x)==str else x for x in fifa.Weight]

light = fifa.loc[fifa.Weight < 125].count()[0]
light_medium = fifa[(fifa.Weight >= 125) & (fifa.Weight < 150)].count()[0]
medium = fifa[(fifa.Weight >= 150) & (fifa.Weight < 175)].count()[0]
medium_heavy = fifa[(fifa.Weight >= 175) & (fifa.Weight < 200)].count()[0]
heavy = fifa[fifa.Weight >= 200].count()[0]

weights = [light,light_medium, medium, medium_heavy, heavy]
label = ['under 125', '125-150', '150-175', '175-200', 'over 200']
explode = (.4,.2,0,0,.4)

plt.title('Weight of Professional Soccer Players (lbs)')

plt.pie(weights, labels=label, explode=explode, pctdistance=0.8,autopct='%.2f %%')
plt.show()
```

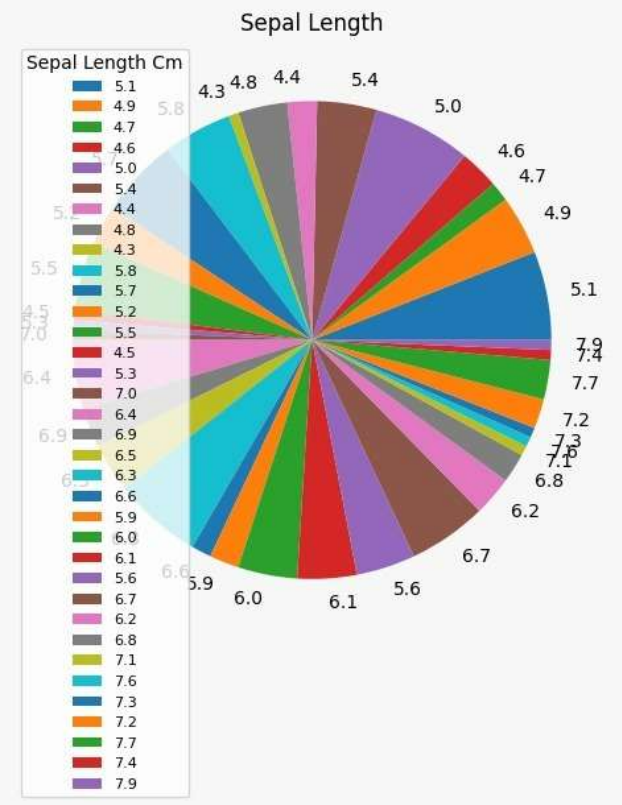
Output



6. Draw Sepal Length pie chart to refer frequency of each unique value in the Sepal Length attribute of the data set “iris_data.csv”

```
import pandas as pd
import matplotlib.pyplot as plt
df =pd.read_csv("iris_data.csv")
a=df["SepalLengthCm"].unique()

c=[]
for i in a:
  r = df.apply(lambda x : True
               if x['SepalLengthCm'] == i else False, axis=1)
  count = len(r[r== True].index)
  c.append(count)
plt.figure(figsize=(10,6),dpi=100)
plt.pie(c,labels=a)
plt.title("Sepal Length")
plt.legend(title = "Sepal Length Cm",prop={'size':8})
plt.show()
```

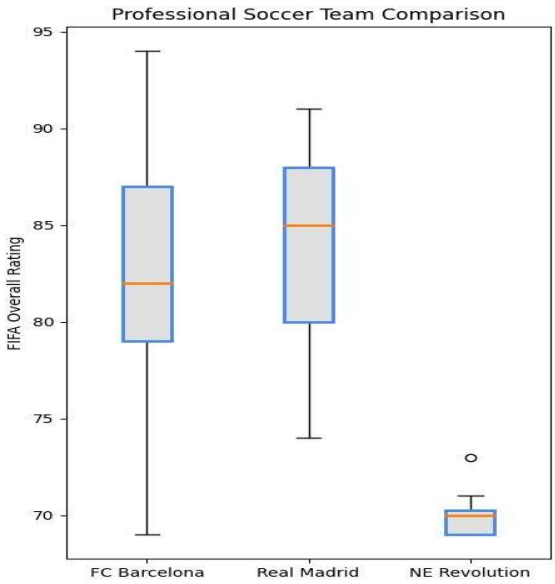
Output



7. Draw the Box and Whisker plot for the ‘Overall’ attribute of various clubs “FC Barcelona”, “Real Madrid”, and "New England Revolution". Set the Box color and face color to '#4286f4' and '#e0e0e0' respectively.

```
import pandas as pd
import matplotlib.pyplot as plt
plt.figure(figsize=(5,8), dpi=100)

plt.style.use('default')

barcelona=fifa.loc[fifa.Club=="FC Barcelona"]['Overall']
madrid = fifa.loc[fifa.Club == "Real Madrid"]['Overall']
revs = fifa.loc[fifa.Club == "New England Revolution"]['Overall']

bp = plt.boxplot([barcelona, madrid, revs], labels=['FC Barcelona','Real Madrid','NE Revolution'], patch_artist=True, medianprops={'linewidth': 2})

plt.title('Professional Soccer Team Comparison')
plt.ylabel('FIFA Overall Rating')

for box in bp['boxes']:
    box.set(color='#4286f4', linewidth=2)
    box.set(facecolor = '#e0e0e0' )
plt.show()
```

Output

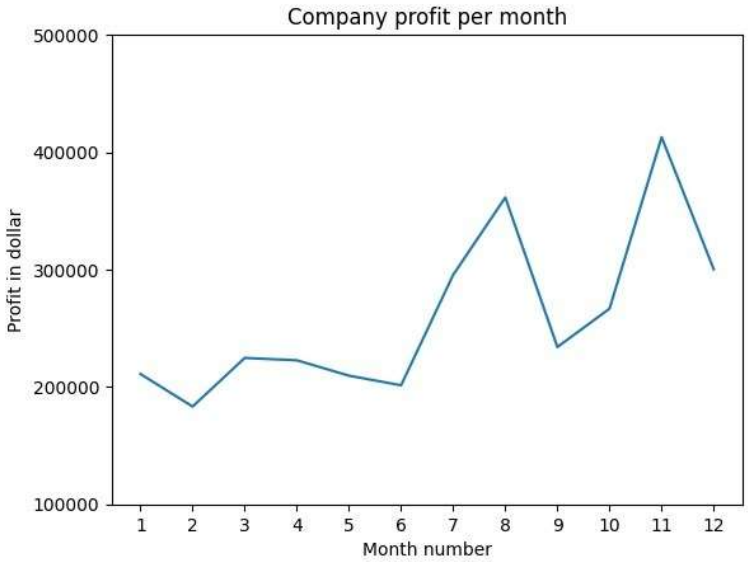


8. Use Company Sales Data(company_sales_data.csv) to draw the plots for the following:

a. Display total profit of all months using a line plot

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

df = pd.read_csv("company_sales_data.csv")
profitList = df ['total_profit'].tolist()
monthList  = df ['month_number'].tolist()

plt.plot(monthList, profitList, label = 'Month-wise Profit data of last year')
plt.xlabel('Month number')
plt.ylabel('Profit in dollar')
plt.xticks(monthList)
plt.title('Company profit per month')
plt.yticks([100000, 200000, 300000, 400000, 500000])
plt.show()
```

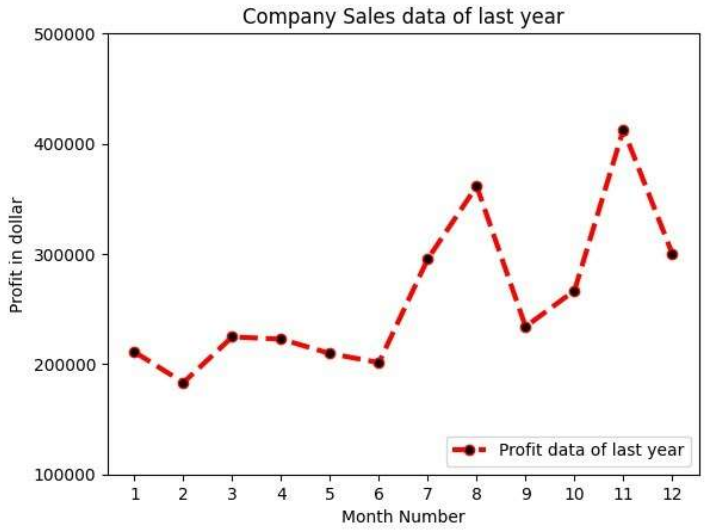
Output



b. Change the above plot like the following

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

     df = pd.read_csv("company_sales_data.csv")
     profitList = df ['total_profit'].tolist()
     monthList  = df ['month_number'].tolist()

     plt.plot(monthList, profitList, label = 'Profit data of last year',
           color='r', marker='o', 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(monthList)
     plt.yticks([100000, 200000, 300000, 400000, 500000])
     plt.show()
```

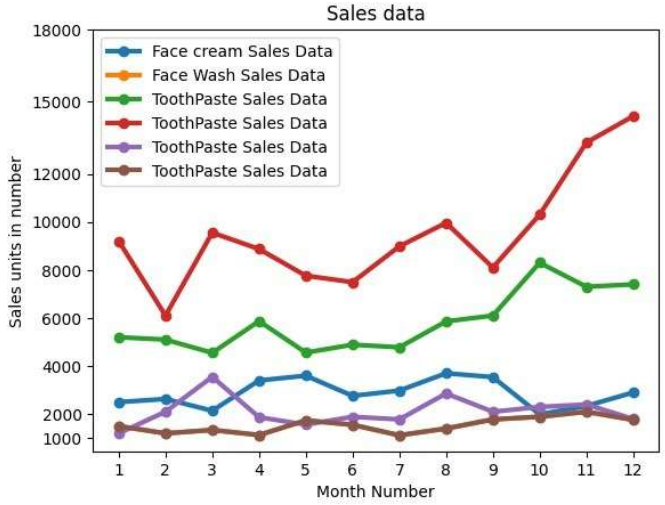
Output



c. Draw Multiline plot to display all product sales data.

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

df = pd.read_csv("company_sales_data.csv")
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(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 = 'ToothPaste Sales Data', marker='o', linewidth=3)
plt.plot(monthList, shampooSalesData, label = 'ToothPaste Sales Data', marker='o', linewidth=3)
plt.plot(monthList, moisturizerSalesData, label = 'ToothPaste 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()
```

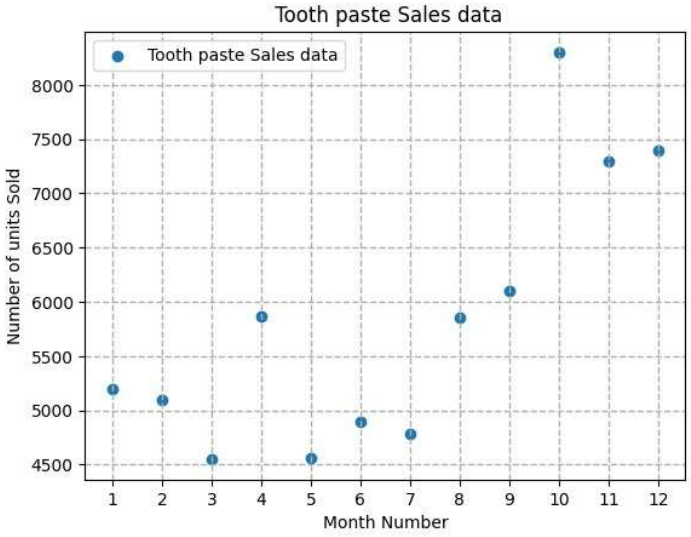
Output



d) Draw Scatter plot to represent toothpaste sales data of each month (Add grid lines for clear understanding)

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

df = pd.read_csv("company_sales_data.csv")
monthList  = df ['month_number'].tolist()
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()
```

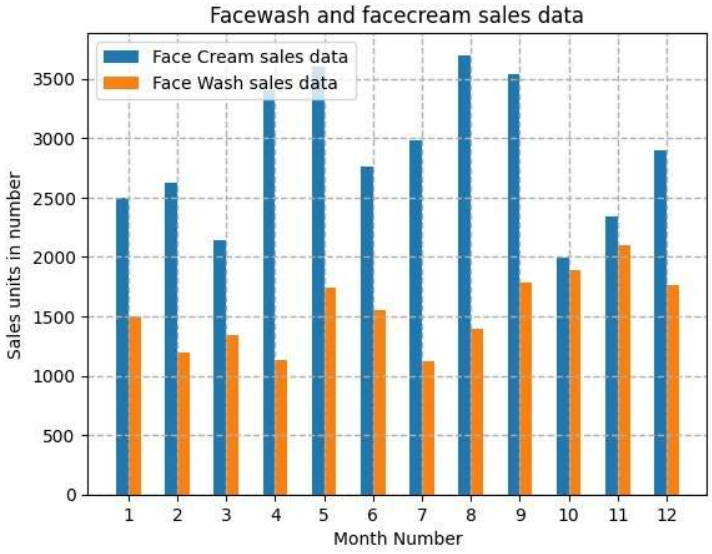
Output



e. Draw Bar Chart for face cream and facewash product sales data

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

df = pd.read_csv("company_sales_data.csv")
monthList  = df ['month_number'].tolist()
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()
```

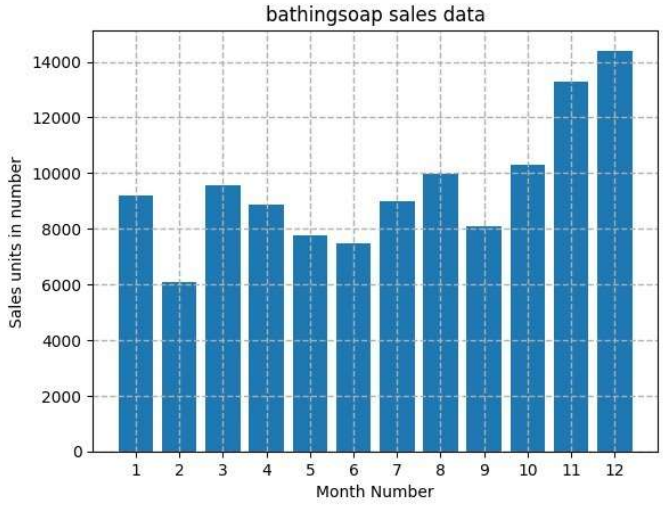
Output



f) Draw bar chart for the Bathing soap sales of all months.

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

df = pd.read_csv("company_sales_data.csv")
monthList  = df ['month_number'].tolist()
bathingsoapSalesData   = df ['bathingsoap'].tolist()
plt.bar(monthList, bathingsoapSalesData)

plt.xlabel('Month Number')
plt.ylabel('Sales units in number')
plt.title(' Sales data')
plt.xticks(monthList)
plt.grid(True, linewidth= 1, linestyle="--")
plt.title('bathingsoap sales data')

plt.savefig('sales_data_plot.png', dpi=150)
plt.show()
```

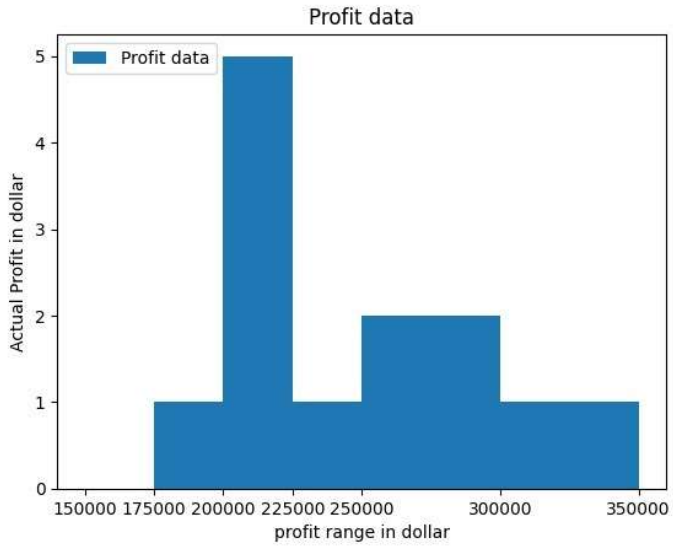
Output



g. Draw histogram to display total profit of each month to observe most common profit ranges.

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

df = pd.read_csv("company_sales_data.csv")
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')
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()
```

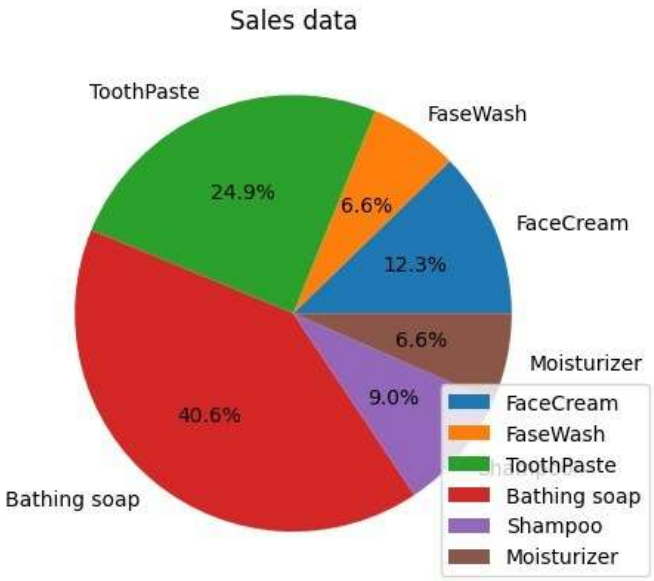
Output



h. Calculate total sale data for last year for each product and show it using a Pie chart. In Pie chart display Number of units sold per year for each product in percentage.

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

     df = pd.read_csv("company_sales_data.csv")
     monthList  = df ['month_number'].tolist()
     labels = ['FaceCream', 'FaseWash', 'ToothPaste', 'Bathing soap', 'Shampoo', 'Moisturizer']
     salesData   = [df ['facecream'].sum(), df ['facewash'].sum(), df ['toothpaste'].sum(),
             df ['bathingsoap'].sum(), df ['shampoo'].sum(), df ['moisturizer'].sum()]

     plt.axis("equal")
     plt.pie(salesData, labels=labels, autopct='%1.1f%%')
     plt.legend(loc='lower right')
     plt.title('Sales data')
     plt.show()
```

Output



i. Draw subplots for both Bathing soap and facewash sales of all months.

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

df = pd.read_csv("company_sales_data.csv")
monthList  = df ['month_number'].tolist()
bathingsoap   = df ['bathingsoap'].tolist()
faceWashSalesData   = df ['facewash'].tolist()

f, axarr = plt.subplots(2, sharex=True)
axarr[0].plot(monthList, bathingsoap, label = 'Bathingsoap Sales Data', color='k', marker='o', linewidth=3)
axarr[0].set_title('Sales data of  a Bathingsoap')
axarr[1].plot(monthList, faceWashSalesData, label = 'Face Wash Sales Data', color='r', marker='o', linewidth=3)
axarr[1].set_title('Sales data of  a facewash')
plt.xticks(monthList)
plt.xlabel('Month Number')
plt.ylabel('Sales units in number')
plt.show()
```

Output