במעבדה זו נבצע תרגילים להצגת תרשימים מסוגים שונים ובכלים שונים.

מצורפים קבצי נתונים בפורמט CSV ויש לעשות עיבוד מסוים כדי להגיע להצגה גרפית.

לצורך העיבוד (וכל פעולה אחרת) ניתן להשתמש בכל חבילה וספריה העולה על דעתכם - Pandas תתאים במיוחד למשימה.

תרגילים 1-10,12,17,18 יש לעשות עם matplotlib

את ההמשך יש לחלק בין seaborn ל bokeh (כחצי חצי לבחירתכם)

יש להגיש את המעבדה על גבי מחברת ג'ופיטר פורמט HTML או PDF.

1. Read Total profit of all months and show it using a line plot.
   a. X label name = Month Number
   b. Y label name = Total profit
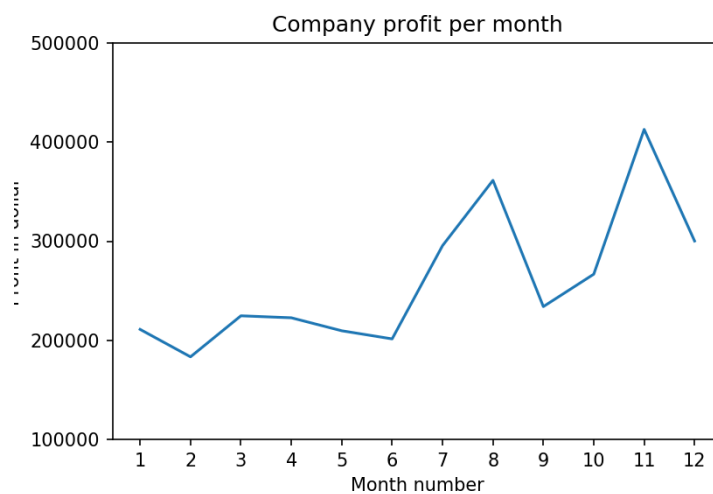
2. Get Total profit of all months and show line plot with the following Style properties

   a. Line Style dotted and 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
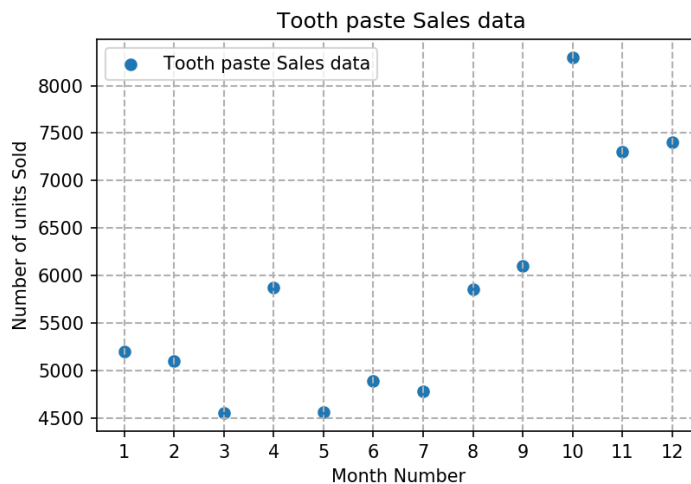
g. Line width should be 3

Company Sales data of last year
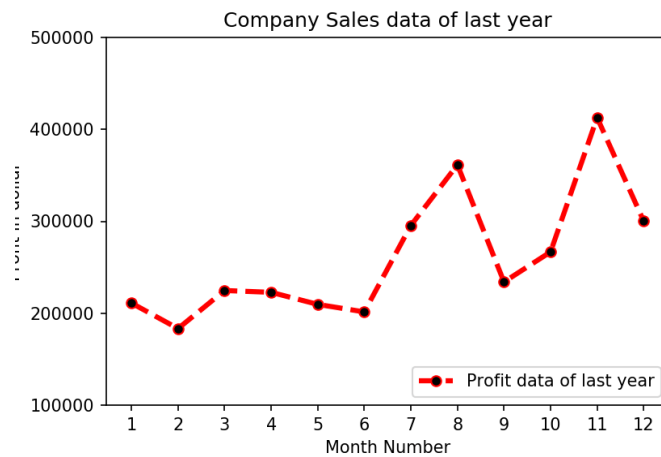
3. Read all product sales data and show it using a multiline plot
   Display the number of units sold per month for each product using multiline plots. (i.e.,
   separate Plotline for each product).
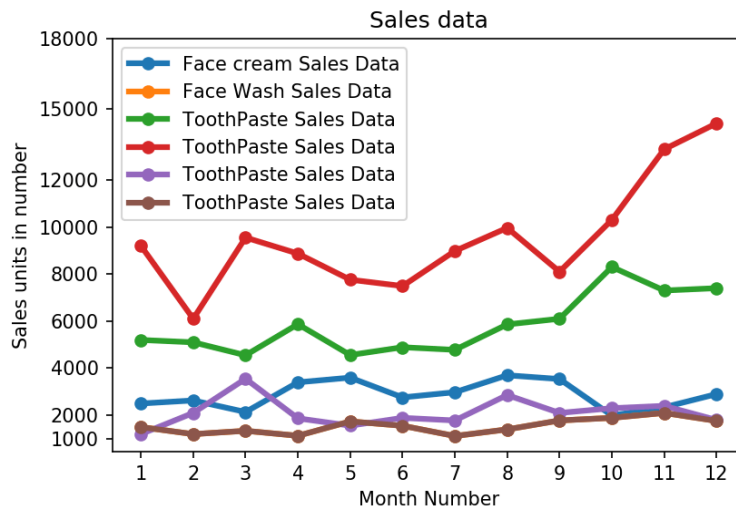
4. Read toothpaste sales data of each month and show it using a scatter plot
   add a grid in the plot. gridline style should "–".

5. Read face cream and facewash product sales data and show it using the bar chart
   The bar chart should display the number of units sold per month for each product.
   Add a separate bar for each product in the same chart.

Facewash and facecream sales data
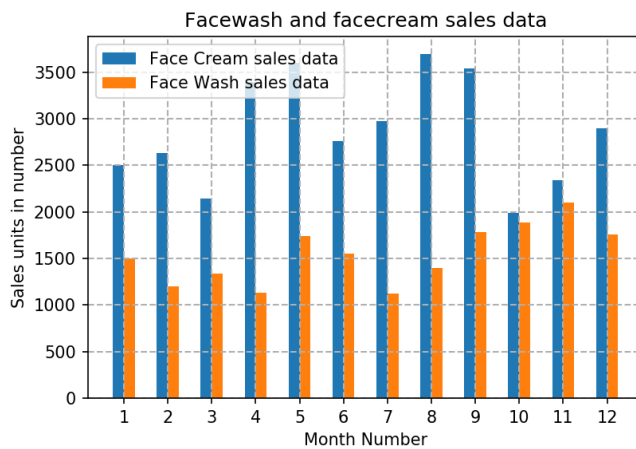
6. Read sales data of bathing soap of all months and show it using a bar chart.
   Save this plot to your hard disk.

bathingsoap sales data
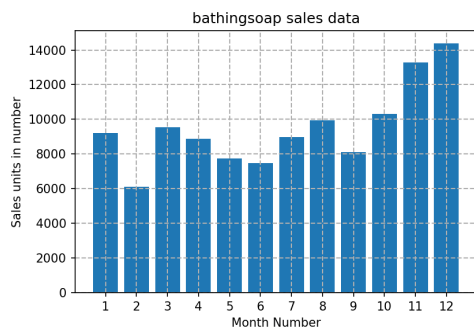
7. Read the total profit of each month and show it using the histogram to see most common profit ranges.

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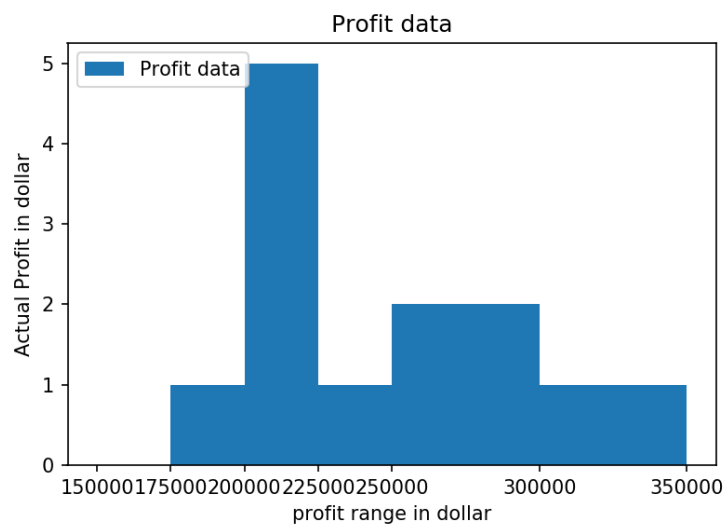
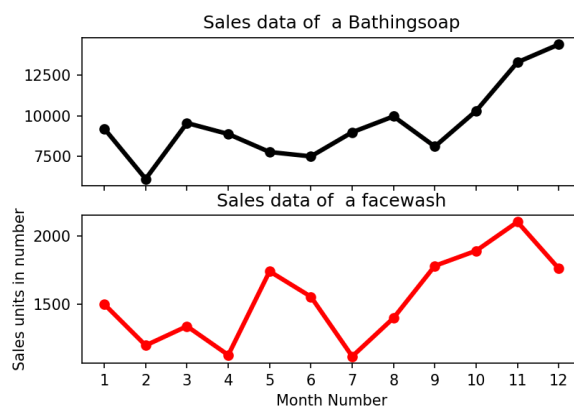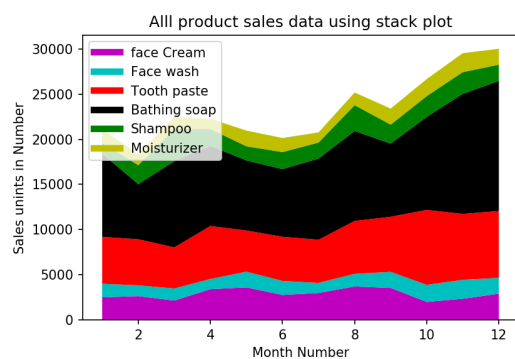8. Calculate total sale data for last year for each product and show it using a Pie chart.
    Note: In Pie chart display Number of units sold per year for each product in percentage.

9. Read Bathing soap facewash of all months and display it using the Subplot

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

11. Create a line plot showing the number of marriages and divorces per capita in the U.S. between 1867 and 2014.
    Label both of the lines and show the legend.
    Don't forget to label your axes!
12. Use the ggplot style on ex 11.
13. Create a vertical bar chart comparing the number of marriages and divorces per capita in the U.S. between 1900, 1950, and 2000.
14. Create a horizontal bar chart that compares the deadliest actors in Hollywood. Sort the actors by their kill count and label each bar with the corresponding actor's name.

15. Create a pie chart showing the fraction of all Roman Emperors that were assassinated. Make sure that the pie chart is an even circle, labels the categories, and shows the percentage breakdown of the categories.
16. Create a scatter plot showing the relationship between the total revenue earned by arcades and the number of Computer Science PhDs awarded in the U.S. between 2000 and 2009.
17. Use the FiveThirtyEight style to visualize the number of marriages and divorces per capita in the U.S. between 1867 and 2014.
18. Create a custom style that does the following:
Uses the color cycle: 1f77b4, ff7f0e
Sets the line width to 2.5
Sets the x- and y-tick labels to size 12
Use this example style file as a guide.

19. Create the same chart from Exercise 8 using your custom style.
20. Create a histogram showing the distribution of the Roman Emperor's reign lengths.What does this distribution tell you?
21. Create two box plots to compare the earnings of recent college graduates in majors that are primarily composed of women to the earnings of recent college graduates in majors that are primarily composed of men.
    a. Does there appear to be a significant difference? (Hint: Use notches.)
    b. Remember to label the categories!
22. Use subplots to create a figure with two separate line plots showing the number of marriages and divorces per capita in the U.S. between 1867 and 2014. Place the two line plots on top of each other so the years line up.
23. Use small multiples to compare the trends in gender ratio of 17 U.S. college majors between 1970 and 2011. Make sure all of the subplots have the same axes and ranges.
    a. Only label the y-axes of the subplots on the left
    b. only label the x-axes of the subplots on the bottom.

מקורות מידע

Matplotlib

https://towardsdatascience.com/matplotlib-tutorial-learn-basics-of-pythons-powerful-plotting-library-b5d1b8f67596
https://www.datacamp.com/community/tutorials/matplotlib-tutorial-python
https://www.tutorialspoint.com/matplotlib/index.htm
https://github.com/rougier/matplotlib-tutorial

Seaborn
https://elitedatascience.com/python-seaborn-tutorial
https://www.tutorialspoint.com/seaborn/index.htm
https://www.datacamp.com/community/tutorials/seaborn-python-tutorial
https://www.kaggle.com/kanncaa1/seaborn-tutorial-for-beginners
https://towardsdatascience.com/data-visualization-using-seaborn-fc24db95a850

Bokeh

https://towardsdatascience.com/data-visualization-with-bokeh-in-python-part-one-getting-started-a11655a467d4

https://programminghistorian.org/en/lessons/visualizing-with-bokeh

https://docs.bokeh.org/en/latest/docs/user_guide/jupyter.html

https://trenton3983.github.io/files/projects/2020-03-15_interactive_data_visualization_with_bokeh/2020-03-15_interactive_data_visualization_with_bokeh.html