

School of Computer Science Engineering and Technology

Course-BTech
Course Code - CSET211
Year - Second
Date - 12/08/2024

Type - AI Core-1
Course Name - Statistical Machine Learning
Semester - ODD
Batch - CSE 3rd Semester

Lab Assignment - 2: Performing visualization using Matplotlib and Seaborn packages

CO- Mapping

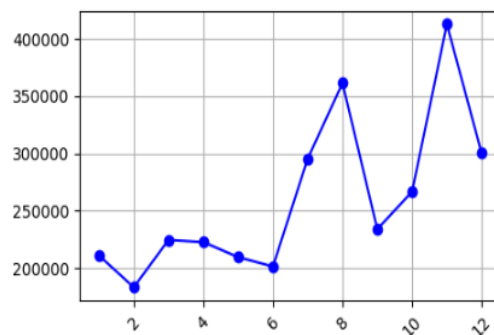
Section	CO1	CO2	CO3	CO4
Section 1: Q1-Q5	√			
Section 2: Q1-Q5	√			

Useful Links for Matplotlib & Seaborn

1. <http://www.matplotlib.org> - The project web page for matplotlib.
2. <https://github.com/matplotlib/matplotlib> - The source code for matplotlib.
3. <http://matplotlib.org/gallery.html> - A large gallery showcasing various types of plots matplotlib can create. Highly recommended!
4. <http://www.loria.fr/~rougier/teaching/matplotlib> - A good matplotlib tutorial.
5. Seaborn documentation: <https://seaborn.pydata.org/index.html>

Section 1: Matplotlib

1. Import matplotlib and pandas libraries.
2. Load the given dataset - 'ecommerce_data.csv'.
3. Create a simple line plot in Matplotlib to show total profit of the ecommerce website month wise. Use the Month column of the data on the x-axis and total profit on the y-axis.



4. Using the same plot as in Q3. Add the following style properties in the graph:

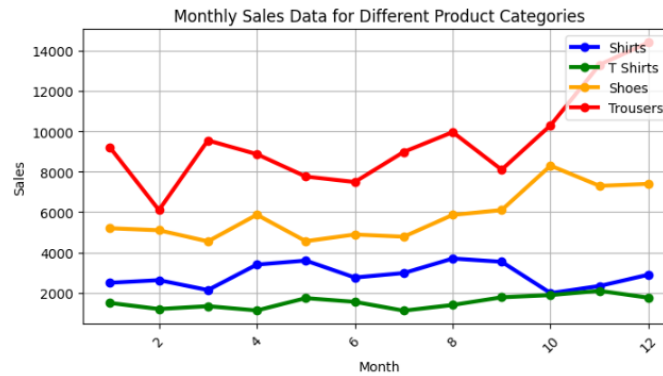
-
- The graph displays the daily new cases in the United States. The y-axis is labeled 'New cases' and ranges from 0 to 400,000 in increments of 50,000. The x-axis is labeled 'Year' and shows the years 2020, 2021, and 2022. The data points are connected by a red line, showing a general upward trend with significant fluctuations, including a major peak in early 2022.
- | Year | New cases |
|------|-----------|
| 2020 | 210,000 |
| 2021 | 180,000 |
| 2022 | 225,000 |
| 2023 | 225,000 |
| 2024 | 210,000 |
| 2025 | 200,000 |
| 2026 | 295,000 |
| 2027 | 360,000 |
| 2028 | 235,000 |
| 2029 | 265,000 |
| 2030 | 410,000 |
| 2031 | 300,000 |

-
- | Month | Total Profit |
|-------|--------------|
| 1 | 210000 |
| 2 | 180000 |
| 3 | 225000 |
| 4 | 220000 |
| 5 | 210000 |
| 6 | 300000 |
| 7 | 365000 |
| 8 | 235000 |
| 9 | 265000 |
| 10 | 415000 |
| 11 | 305000 |

-
- | Month | Total Profit |
|-------|--------------|
| 1 | 210000 |
| 2 | 185000 |
| 3 | 225000 |
| 4 | 225000 |
| 5 | 210000 |
| 6 | 200000 |
| 7 | 300000 |
| 8 | 365000 |
| 9 | 235000 |
| 10 | 265000 |
| 11 | 415000 |
| 12 | 300000 |

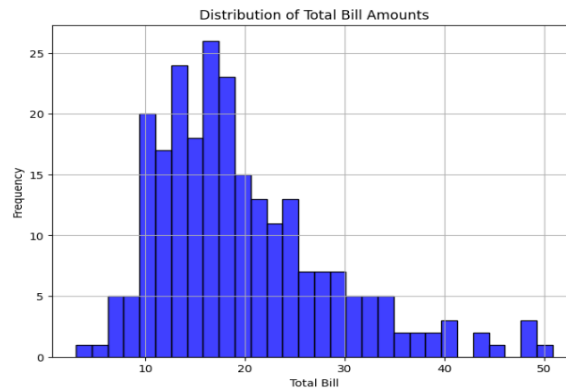
-
- | Month | Total Profit |
|-------|--------------|
| 1 | 210000 |
| 2 | 180000 |
| 3 | 225000 |
| 4 | 220000 |
| 5 | 205000 |
| 6 | 200000 |
| 7 | 295000 |
| 8 | 360000 |
| 9 | 235000 |
| 10 | 265000 |
| 11 | 420000 |
| 12 | 300000 |

5. For each month, plot the 'shirts', 't shirts', 'trousers', and 'shoes' data in one graph. Add the following style properties in the graph:
 - a. Each line should be of a different color.
 - b. Set the linewidth to be 3.
 - c. Set marker as 'o'.
 - d. Set the titles for x and y label axes.
 - e. Set the title of the graph.
 - f. Set the legend in the upper right corner inside the plot.

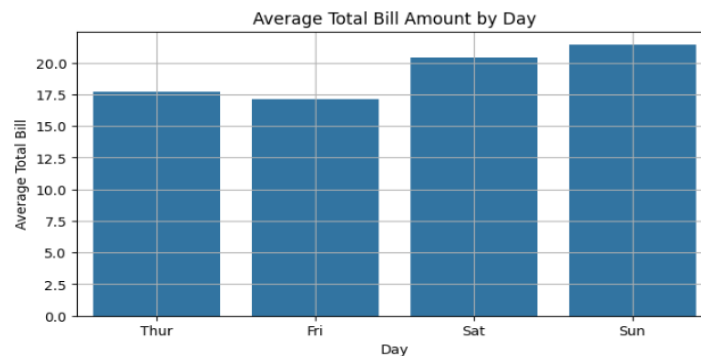


Section 2: Seaborn

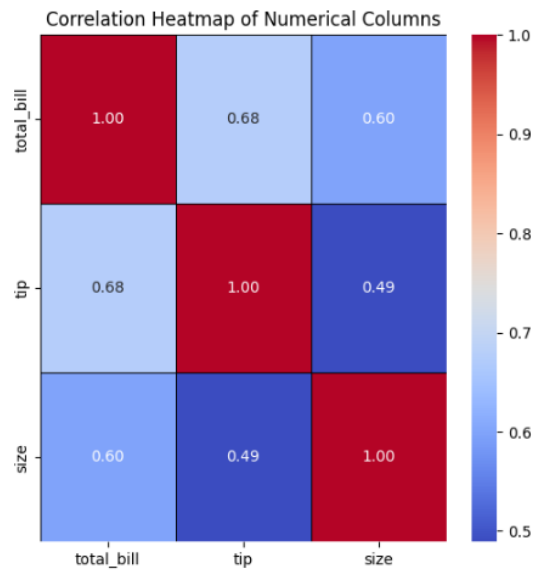
1. Import seaborn library.
2. Load the built-in 'tips' dataset in seaborn and check the columns.
3. Plot the data distribution of the column 'total_bill'.



4. Use the bar plot for 'total bill' and 'day' columns.



5. Draw Heatmap by finding the correlation of numerical columns in the data.



Platform Required: Anaconda, Editor: Jupyter/Spyder/Pycharm/Google Colab

Submission Instructions:

- Submission required .ipynb file only
- Submission is through LMS only.