

Assignment 3

Problem Statement:-

Write and execute python code using the pandas ,seabourne and matplotlib to visualize the a dataset in 8 different ways.

S/W Packages and H/W apparatus used:

OS: Ubuntu/Windows, Tools:

Google Colab

Packages: Numpy, Pandas, Matplotlib and Seaborn

Theory:-

1. Methodology:

1. **Load the Dataset:** Begin by loading the dataset into a pandas DataFrame.
2. **Explore the Dataset:** Gain an understanding of the dataset's structure, including the number of rows and columns, data types, and summary statistics.
3. **Data Cleaning:** Handle missing values, duplicate entries, and outliers as necessary to ensure data quality.
4. **Visualization Techniques:**
 - Scatter Plot: Visualize the relationship between two numerical variables.
 - Line Plot: Display trends over time or sequential data.
 - Histogram: Illustrate the distribution of a single numerical variable.
 - Bar Plot: Compare categorical variables or aggregate numerical data into groups.

- **Box Plot:** Depict the distribution of numerical data, highlighting median, quartiles, and outliers.
- **Pair Plot:** Show pairwise relationships between multiple numerical variables.
- **Heatmap:** Display correlation between numerical variables using color intensity.
- **Violin Plot:** Combine the features of a box plot and kernel density estimation to show the distribution of data.

5. **Execute Visualization:**

- Utilize matplotlib and seaborn libraries to create and customize the visualizations according to the chosen techniques.

2. **Advantages and Disadvantages & Limitation/Example:**

1. **Advantages:**

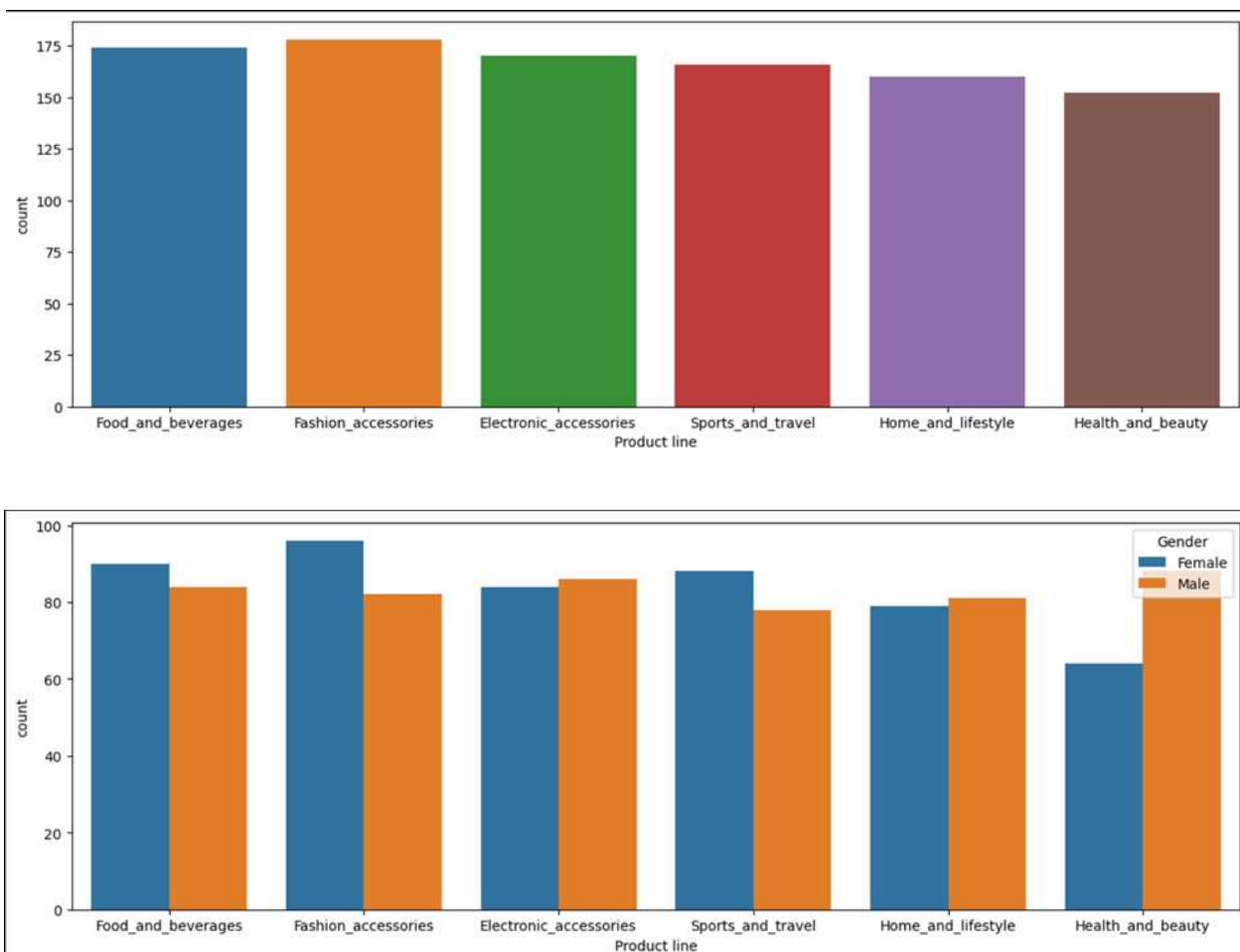
- **Enhanced Understanding:** Visualizations provide a clear and intuitive understanding of the dataset's characteristics and relationships.
- **Effective Communication:** Visual representations facilitate effective communication of insights to stakeholders and decision-makers.
- **Identification of Patterns:** Visualizations help in identifying patterns, trends, and outliers in the data, aiding in hypothesis generation and validation.
- **Insight Generation:** Through visual exploration, new insights and hypotheses can be generated, guiding further analysis and exploration.

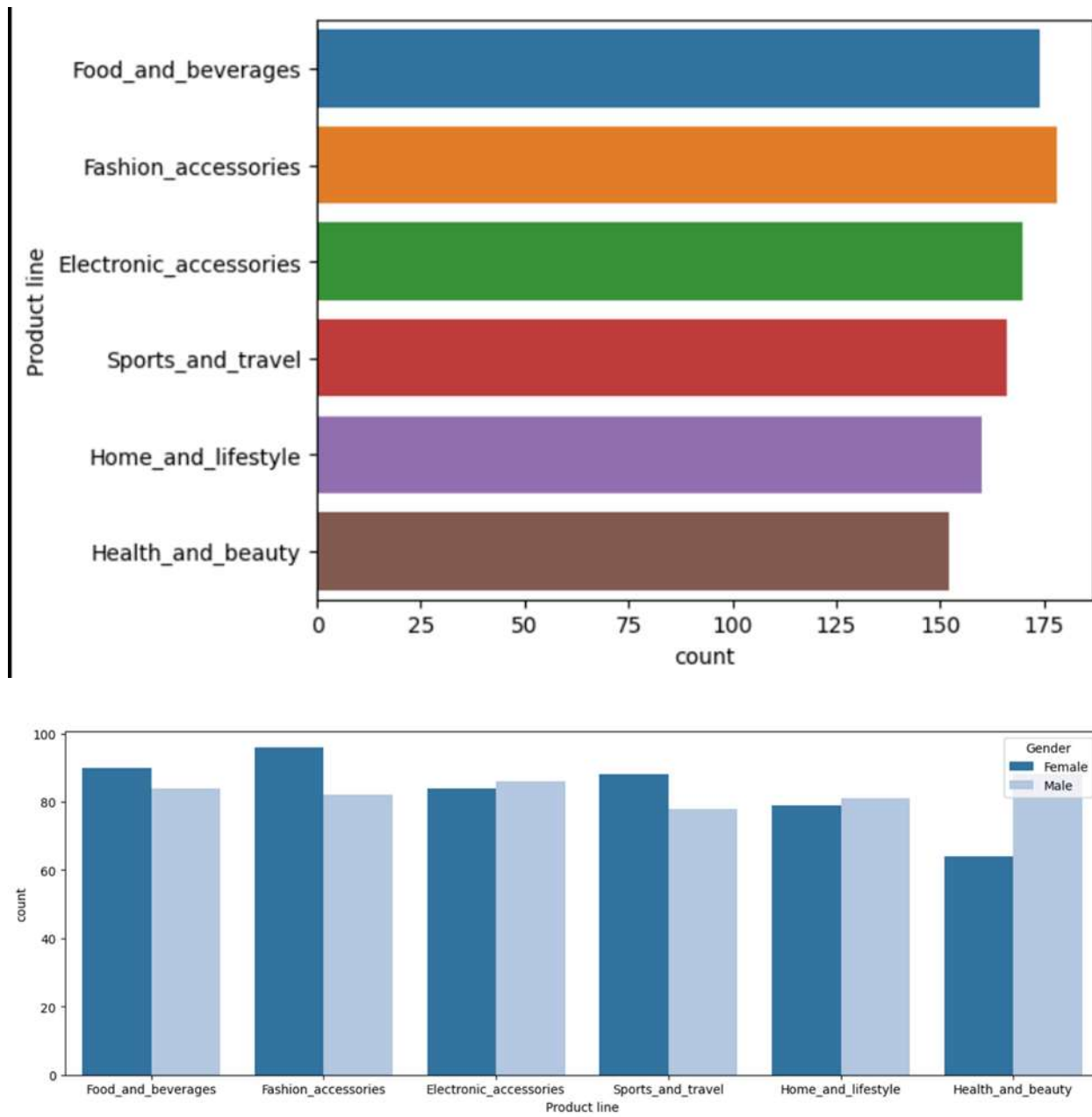
2. **Disadvantages & Limitations/Example:**

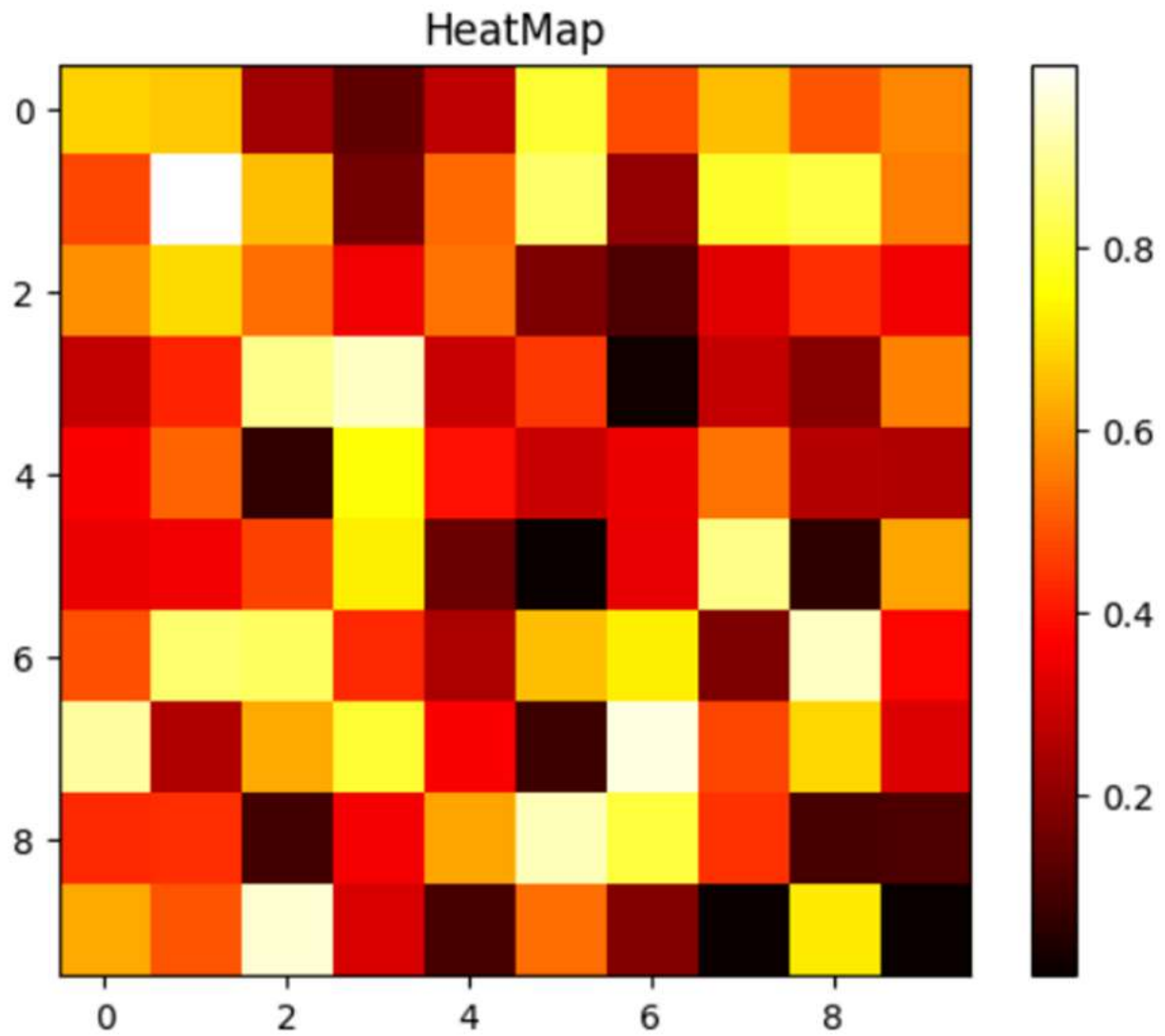
- **Subjectivity:** Interpretation of visualizations can be subjective and influenced by individual biases.

- **Limited Depth:** Visualizations may not capture all nuances and complexities present in the data, leading to potential oversimplification.
- **Data Quality:** Visualizations are only as good as the quality of the underlying data, and misleading visualizations can result from poor data quality.
- **Overwhelming Complexity:** In some cases, complex datasets may require sophisticated visualization techniques, which can be challenging to interpret.

Output :







Conclusion:

In conclusion, leveraging various visualization techniques such as scatter plots, line plots, histograms, bar plots, box plots, pair plots, heatmaps, and violin plots offers a comprehensive understanding of dataset characteristics and relationships. Through effective visualization, insights are communicated clearly to stakeholders, aiding in hypothesis generation, trend identification, and decision-making processes. While visualizations enhance data exploration and interpretation, their effectiveness depends on appropriate selection and interpretation, considering potential biases and limitations. Overall, employing a diverse range of visualization techniques facilitates insightful exploration and communication of dataset insights in a meaningful and accessible manner.