Customer Data Analysis Using Python

Student Name: Subasri Nandakumar
Student Number: 20079001
Module: Programming for Analytics (B9BA200)
Lecturer: Dr. Muhammad Alli

---

1. Introduction

The main aim of this project is to analyse customer-related data using Python programming. The dataset used for this analysis was obtained from a *Data in Brief* article, which provides peer-reviewed datasets for academic research. Customer data analysis is important for organisations because it helps them understand customer behaviour and usage patterns. By analysing such data, businesses can make better decisions and improve their services.

In this project, Python libraries are used to clean the dataset, visualise the data, and identify meaningful patterns. The analysis focuses on understanding how customer values and categories are distributed within the dataset. This study also helps to demonstrate the practical use of Python for data analysis tasks.

---

2. Data Description

The dataset used in this project is a survey-based customer dataset obtained from a *Data in Brief* publication. It contains multiple records and several variables related to customer or service usage behaviour. The dataset includes both numerical and categorical variables, which makes it suitable for exploratory data analysis using Python.

The numerical variables represent customer-related values, while the categorical variables describe different customer groups or classifications. This dataset provides enough information to perform basic data analysis and visualisation, helping to understand trends and distributions within the data.

---

3. Method of Data Analysis

Python programming was used to carry out the data analysis for this project. The dataset was first loaded into the Python environment using the Pandas library. Initial checks were performed to understand the structure of the dataset and to identify any missing values.

For data visualisation, Matplotlib and Seaborn libraries were used. Different types of graphs were created to explore the distribution of values, compare categories, and examine relationships between variables. These visual tools make it easier to interpret the data and identify patterns in a clear and understandable way.

---

4. Results

The results of the analysis are presented using graphical visualisations.

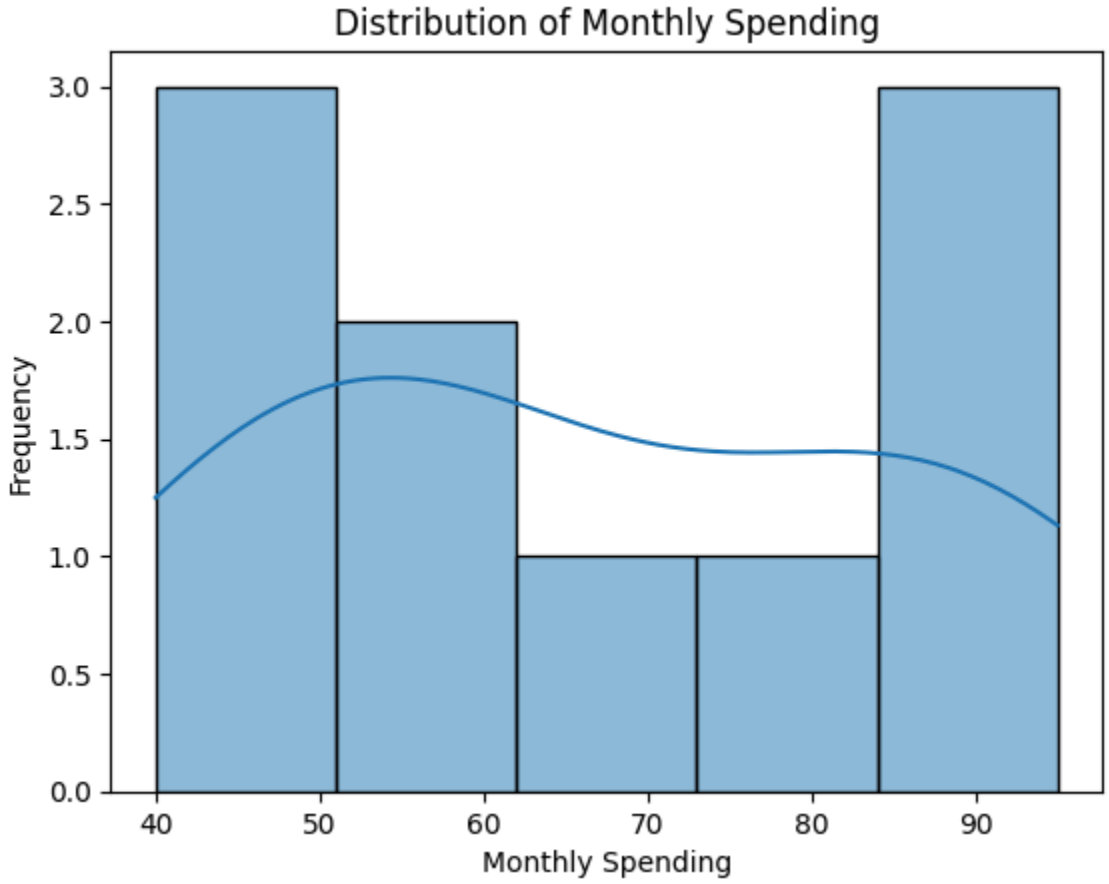
The first graph shows the distribution of a numerical customer-related variable. This graph helps to understand how the values are spread across different ranges and whether the data is evenly distributed or skewed.

The second graph displays the count of different categories within the dataset. This allows a clear comparison of how many records fall into each category and highlights which groups are more common.

The third graph illustrates the relationship between a numerical variable and a categorical variable using a box plot. This graph helps to compare variations in values across different categories and identify any noticeable differences between groups.

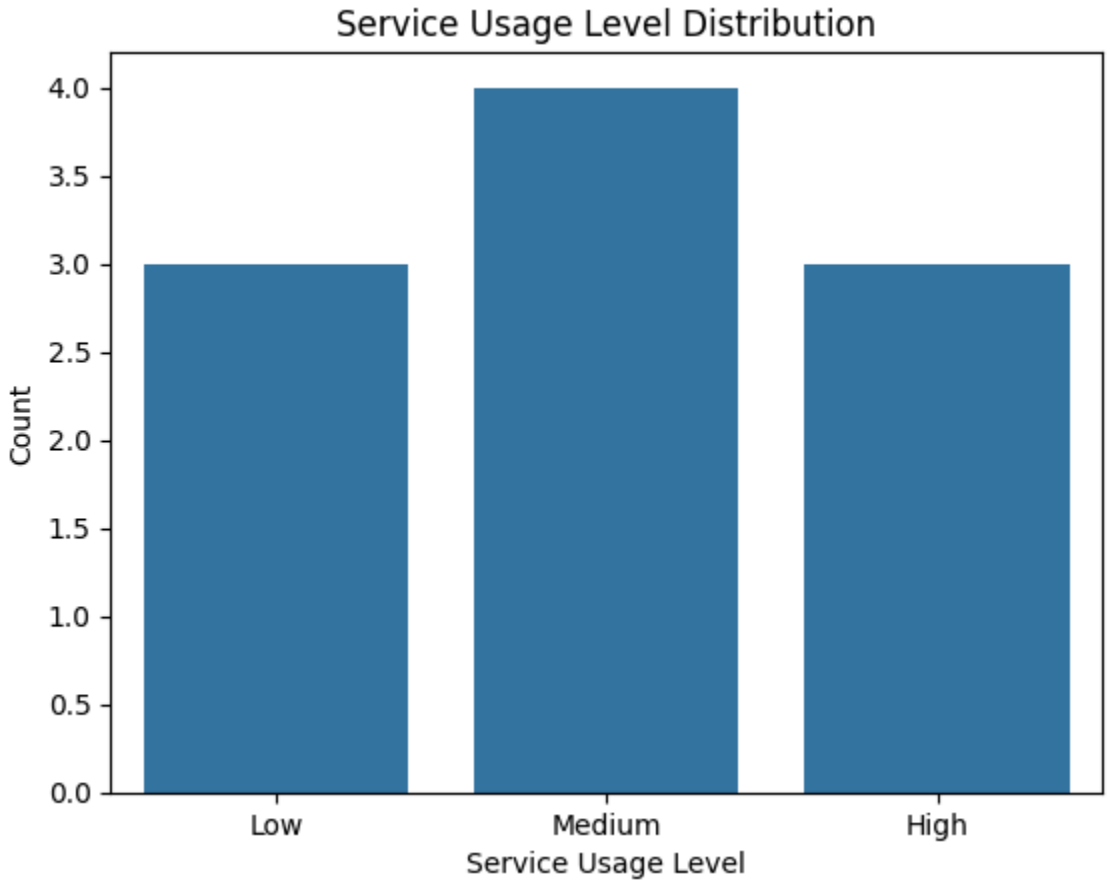
Overall, the graphs provide useful insights into the structure and patterns of the customer data.

Figure 1: Distribution of Customer Data Values.



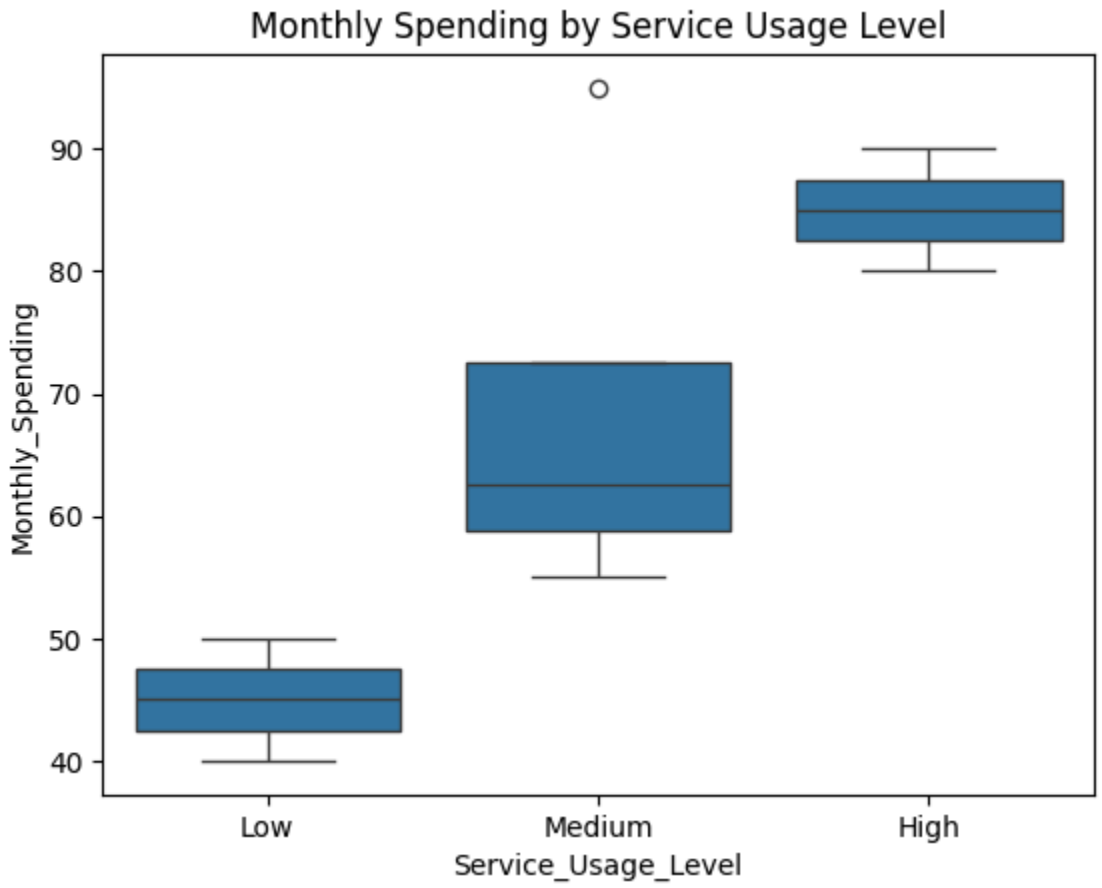
This figure shows the distribution of customer-related values in the dataset. It helps to understand how the values are spread across different ranges.

Figure 2: Category-wise Distribution of Customers.



This figure displays the number of records in each customer category, allowing a clear comparison between different groups.

Figure 3: Relationship Between Customer Categories and Values.



This figure illustrates how customer-related values vary across different categories in the dataset.

---

5. Discussion

From the analysis, it can be observed that the dataset contains noticeable variations across different customer categories. Some categories show higher or lower values compared to others, indicating differences in customer behaviour or usage patterns. The distribution graph also suggests that the data is not perfectly uniform, which is common in real-world datasets.

One limitation of this study is that the analysis is based on a limited number of variables. Including additional customer-related factors could provide deeper insights. However, for the purpose of demonstrating Python-based data analysis, the dataset is appropriate and effective.

---

6. Conclusion

This project successfully demonstrated the use of Python for analysing customer-related data obtained from a *Data in Brief* dataset. The analysis included data loading, basic checking, and visualisation using appropriate Python libraries. The graphs created during the analysis helped to clearly understand data distributions and relationships between variables.

This study highlights how Python can be used as a powerful tool for data analysis in an academic context. In the future, more advanced analytical techniques could be applied to similar datasets to gain deeper insights.

---

7. References

Data in Brief (2024). *Dataset on customer and service usage behaviour*. Elsevier.

Python Software Foundation (2024). *Python Documentation*. Available at: https://docs.python.org/3/

Seaborn Developers (2024). *Seaborn Documentation*. Available at: https://seaborn.pydata.org/
