Supermarket Sales Dataset

Data analysis and visualization, Alessandro Sciorilli

The following report presents a comprehensive analysis and visualization of the Supermarket Sales Dataset. The dataset contains information on sales, gross income, payment methods, customer demographics, and ratings for three branches of a supermarket located in different cities: Yangon, Mandalay, and Naypyitaw. The objective of this analysis is to gain valuable insights into the performance of the supermarkets and identify areas for improvement and growth.

- **Plot 1 - Gross income by City.**
  The first aspect I want to visualize is the gross income performance of each supermarket. This information is valuable for various reasons, including analyzing the metrics of the top-performing branch and applying the same strategies to other branches, rewarding the managers of the leading branch with a cash bonus, and determining whether it is beneficial to keep certain branches open. Although all three branches reported similar gross income results, the branch in Naypyitaw emerged as the top performer.
- **Plot 2 - Sales by product line.**
  I have created a pie chart to visualize the breakdown of sales across different product lines. Understanding which product line sells the most is crucial for making decisions regarding future investments and balancing the range of items offered to clients in order to maximize returns. Despite the similar performance of each product line, the top-selling category is “Food and beverage”.
- **Plot 3 - Total sales by city and product line.**
  This horizontal bar chart is useful for identifying the differences between the three branches in terms of the product lines they sell. With the help of this chart, I can determine which product lines perform better at each branch and develop an investment strategy based on that information.
- **Plot 4 - Gross Income by product line and city.**
  This chart provides information on how each product line contributes to the generation of gross income for each branch. I can see that there is an issue with the health and beauty sector at the Yangon branch, whereas the food and beverage product line at the Naypyitaw branch significantly boosts gross income.
- **Plot 5 - Gross Income over time.**
  A graphical visualization that shows the variation in gross income over a three-month period. I can observe that there is significant volatility, indicating high risks associated with the operations.
- **Plot 6 - Ratings by product line**
  A boxplot is used to understand customer satisfaction for each product line.
- **Plot 7 - Ratings by city.**
  A swarm plot is used to understand customer’s satisfaction for each branch.
- **Plot 8 - Payment methods in each city**
  The objective is to understand which payment method is more prevalent in each location, in order to improve customers' payment experience. This can be achieved by either providing more cash exchange options at specific locations or allocating a larger budget for the maintenance of electronic infrastructures and service fees in other locations.

- **Plot 9 - Payment method by payment amount.**
  My goal is s to gain insights into whether the payment method is linked to the amount spent. Specifically, the hypothesis is that Cash is more likely to be used for small payments, while E-wallets and Credit cards are more commonly used for larger payments. If this hypothesis holds true, it would be beneficial to increase the availability of electronic payment methods to incentivize higher spending.
- **Plot 10 - Gender by City, Customer Type by City.**
  Two bar charts are used to better understand the demographics and composition of clients, as well as to develop corresponding selling strategies**.**
- **Plot 11 - Total amount spent by customer type.**
  I have created two violin plots to visualize whether member customers spend more than normal customers. However, the data does not show any significant difference. Therefore, further efforts are needed to enhance customer loyalty among the member segment.
- **Plot 12 - Pairwise Scatter Plot of Numerical Features.**
  To obtain a more comprehensive and precise understanding of the variations in numerical performance between the customer types "Member" and "Normal," a pairwise scatter plot of the numerical features can be created. This plot will provide a clearer visual representation of the differences.
- **Plot 13 - Missing Data**
  Heatmap is used to show whether there is missing data in the dataset.