Exploratory Data Analysis Report on XYZ Supermarket Sales Data

Summary

The provided dataset contains sales data from a fictional company, XYZ, which owns a supermarket chain across three major cities in Nigeria: Abuja, Lagos, and Port Harcourt. The data covers a period of three months and includes information on various aspects of sales transactions.

Business objective

This analysis aims to understand sales trends and provide insights to help the company make informed decisions for growth.

Data Exploration

- The dataset consists of sales data from three branches: A (Lagos), B (Abuja), and C (Port Harcourt).
- The data were merged into a single dataset for analysis.
- Null values were checked, and no missing values were found in any of the columns.
- The dataset includes both numerical and categorical columns.
- Date and time columns were converted to the appropriate datetime format.
- Additional features were extracted from the datetime columns, including hour, day, and month.

Key Findings

- Branches and Cities: The branches (A, B, C) correspond to the cities (Lagos, Abuja, Port Harcourt) respectively. The number of entries for each branch and city is the same, indicating a one-to-one correspondence.
- Customer Type and Gender: Both customer types ("Member" and "Normal") and genders ("Male" and "Female") are present in the dataset. Not all members are female, and not all regular customers are male.
- Product Line Performance: The product line "Fashion Accessories" has the highest number of sales, while "Health and Beauty" has the lowest. However, the difference in sales between the most and least purchased product lines is only 2.6%, suggesting that there are no significantly underperforming product lines.
- Branch and City Performance: Port Harcourt City (Branch C) has the highest average price per item and total gross income. Abuja (Branch B) has the lowest average rating.
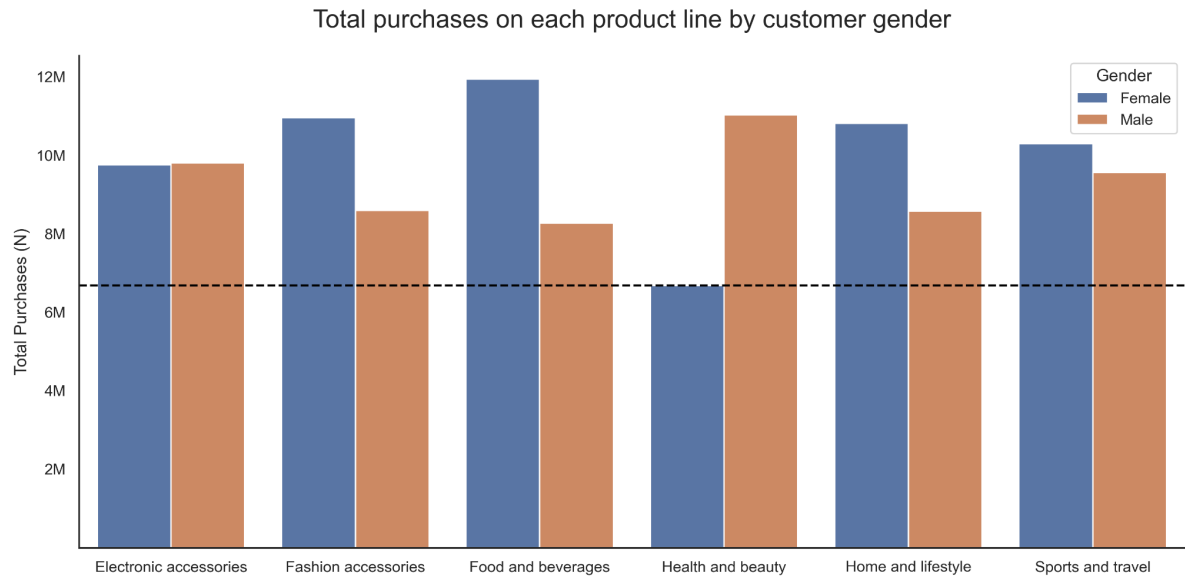
Data Visualization

Number of sales by Product line
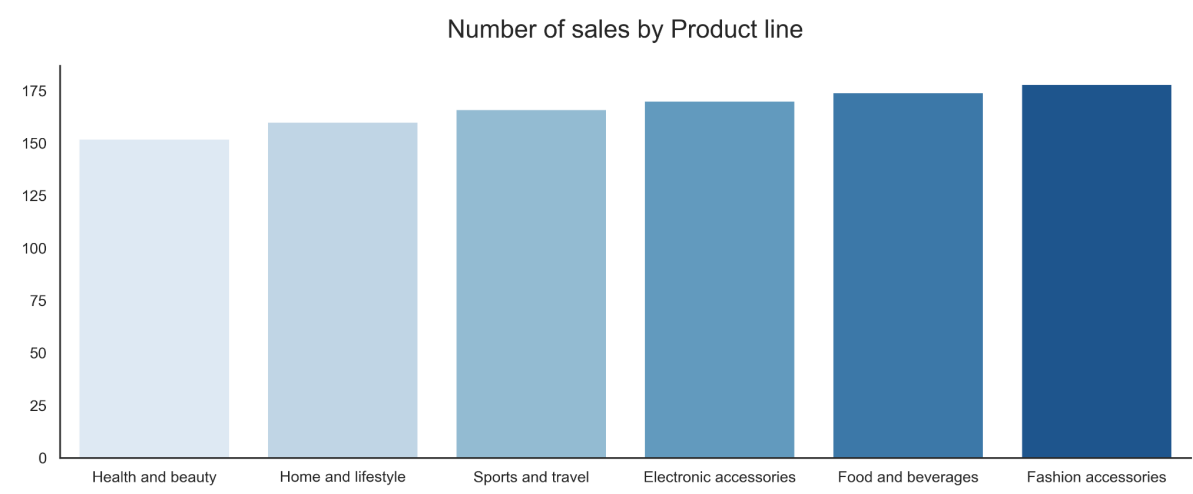
The above chart shows the number of sales for each product line, indicating that "Fashion Accessories" has the highest sales volume.
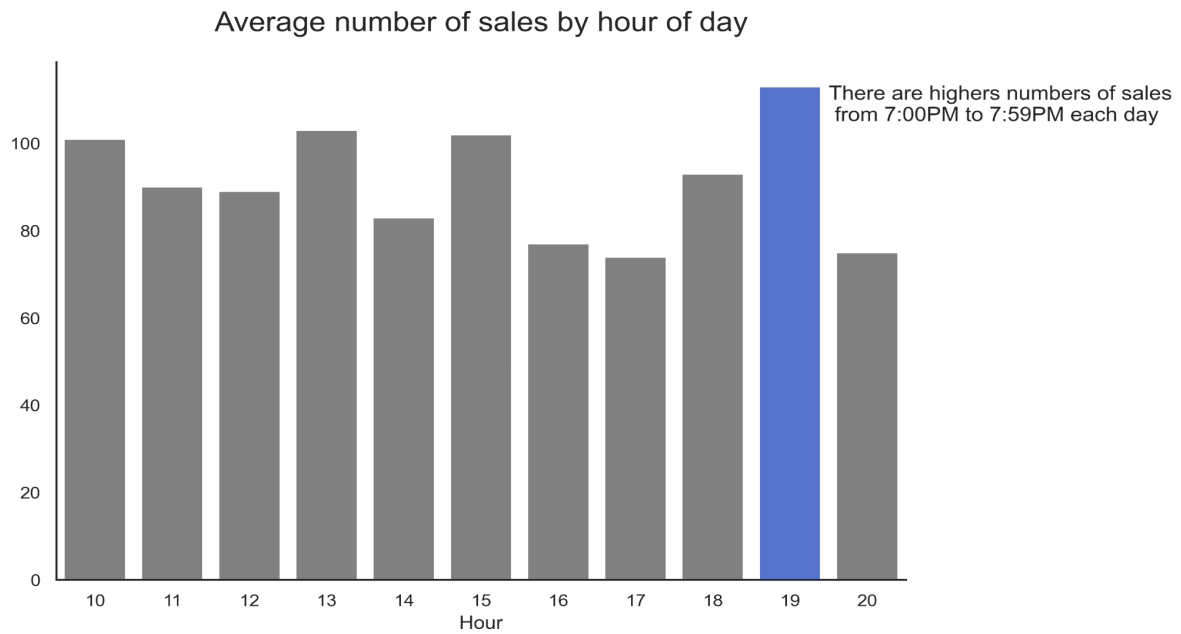
Total sales by product line
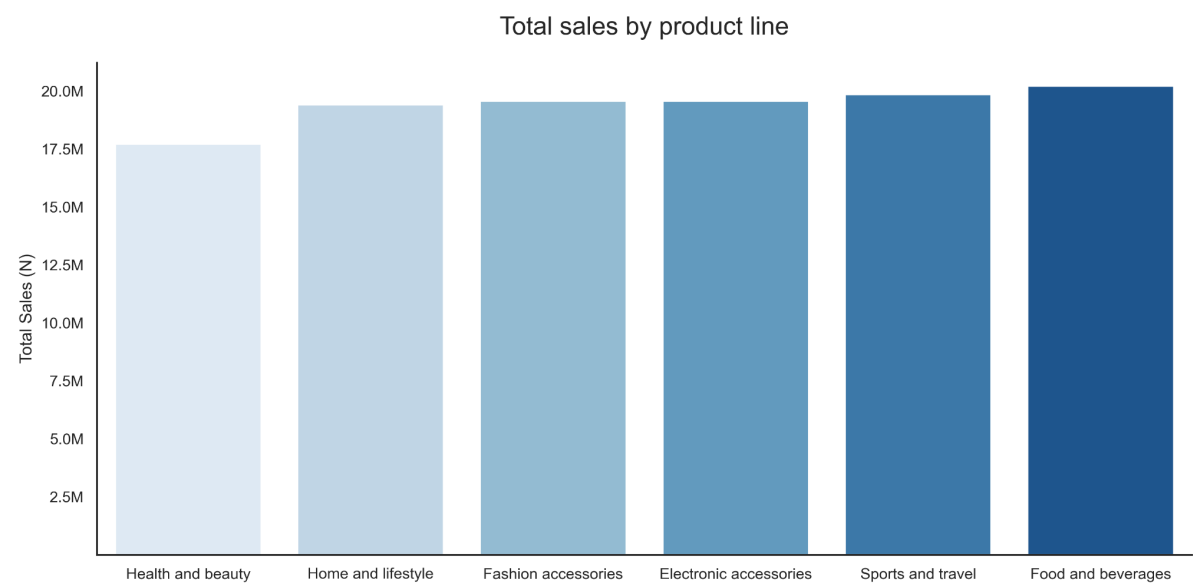
The Health & Beauty product line has the lowest sales compared to other product lines.

Total purchases on each product line by customer gender


Female customers and members generate a higher sales total than male customers and members. The Food & Beverages product line has the highest combined sales total of #20.2M.

Average number of sales by day of week
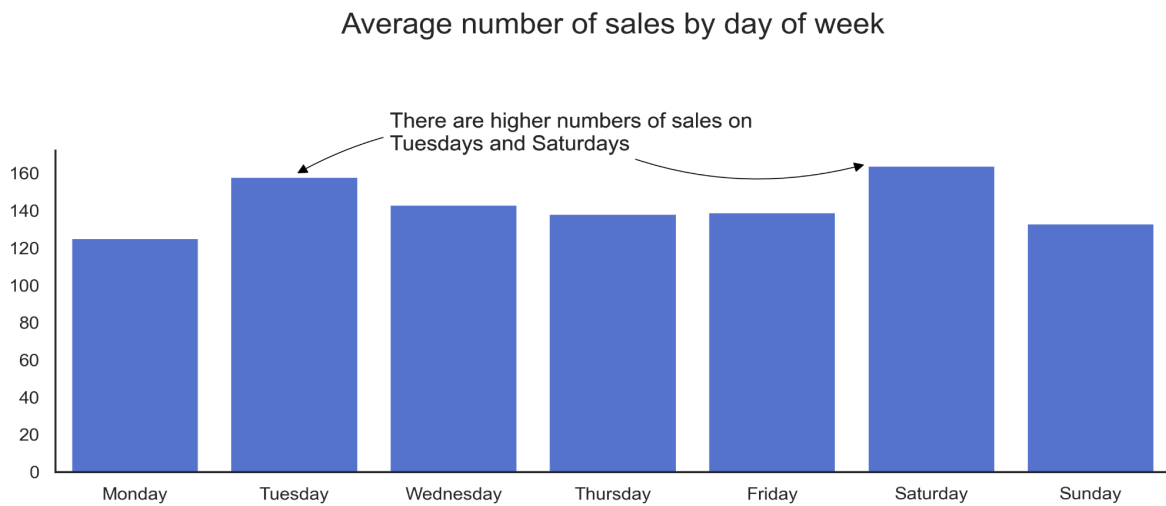
There are higher sales volumes on Tuesdays and Saturdays

Average number of sales by hour of day

There are highers numbers of sales from 7:00PM to 7:59PM each day

The above chart illustrates the average sales volume by hour of day, highlighting higher volumes between 7:00 PM to 7:59 PM each day.
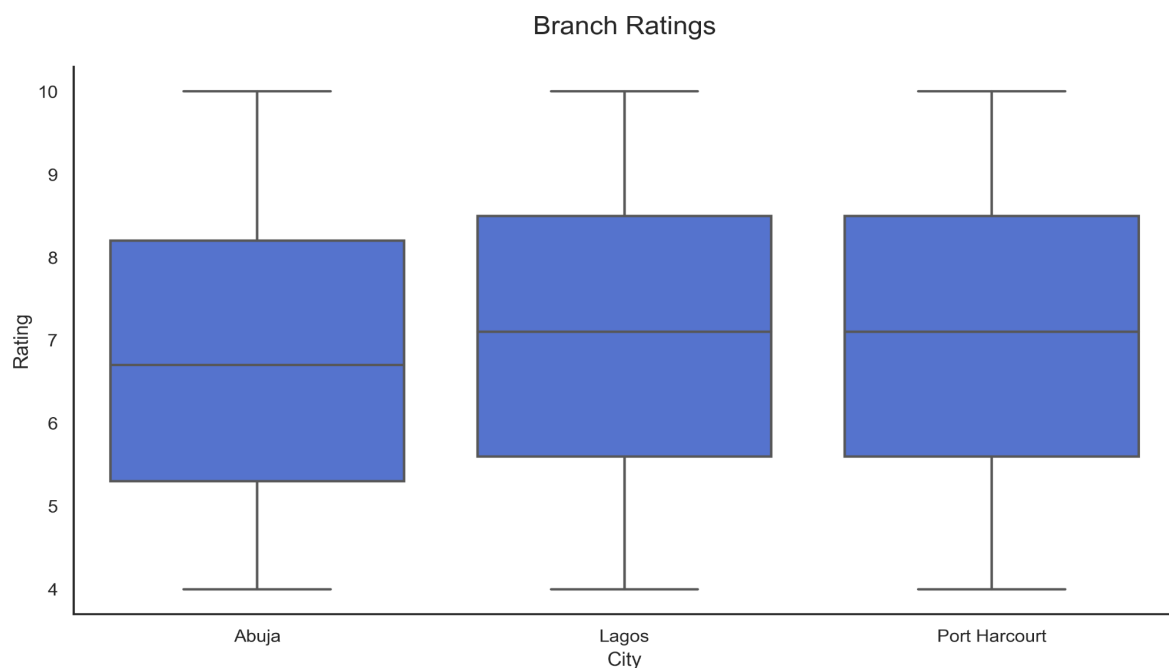
Branch Ratings

The boxplot compares the ratings of each branch, with Abuja (Branch B) having the lowest average rating.

Total number of sales per month

352 January
303 February — February has the least number of sales
345 March
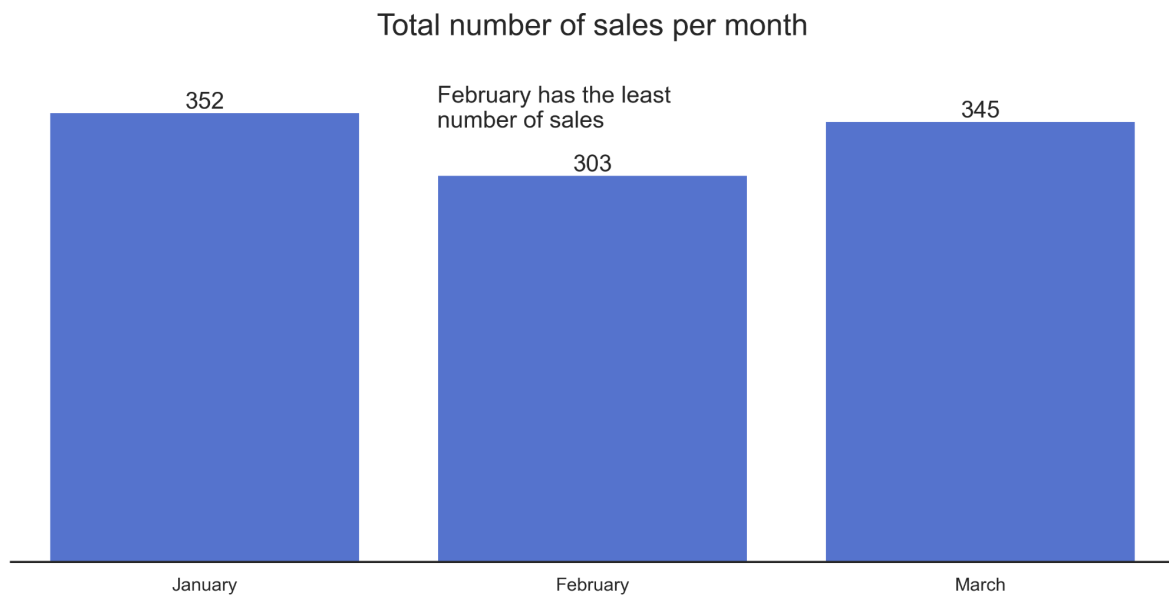
February has the lowest sales volume.

Gross Income per month by city

Though there's high competition across all Cities in terms of gross income generated, Port Harcourt city has the highest cumulative gross income across all cities.

Recommendation

Based on the analysis of the sales data, the following recommendations can be made to XYZ company:

1. **Branch Improvement**: Branch B (located in Abuja) has the lowest average rating among the three branches. The company should focus on identifying the reasons behind the lower ratings and take measures to improve customer satisfaction and experience in that branch.
2. **Product Line Optimization**: While all product lines are performing reasonably well, the company can consider strategies to further promote the "Health and Beauty" product line, which has the lowest sales compared to other product lines. This could include targeted marketing campaigns, offering discounts or promotions, and ensuring a diverse and attractive range of products within that category.
3. **Customer Segmentation**: The analysis reveals that females tend to generate a higher average total purchase than males, especially in certain product lines. The company can leverage this insight to tailor marketing and promotional activities to attract more female customers and encourage repeat purchases.
4. **Peak Sales Hours**: Sales are consistently higher from 7:00 PM to 7:59 PM each day. The company can optimize staffing and inventory during these peak hours to ensure a seamless shopping experience for customers and maximize sales opportunities.
5. **City Expansion Strategy**: Port Harcourt City has the highest gross income among other cities. This indicates potential for further expansion in that city.

Conclusion

Based on the analysis performed on the sales data, the following conclusions can be drawn:

- The three branches of XYZ's supermarket chain show consistent sales numbers, indicating balanced performance across the different cities.
- Product lines have relatively similar sales distribution, with slight variations in popularity.
- Female customers generate a higher average total purchase than male customers, with varying preferences across different product lines.
- Branch B (Abuja) has the lowest average rating, indicating a potential area for improvement in customer satisfaction.
- Sales are higher on Tuesdays and Saturdays, with a peak during the evening hours.
- February has the lowest number of sales compared to other months.
- Port Harcourt City consistently generates the highest gross income.