**EXPLORATORY DATA ANALYSIS (EDA) REPORT**

This report serves as a roadmap for leveraging data analytics to enhance operational efficiency, drive revenue growth, and gain a competitive advantage. By embracing data-driven decision making, Walmart In. can strengthen its position in the competitive retail landscape and pave the way for a more prosperous future. This project focuses on using data analytics to optimize retail operations for Walmart Inc., aiming to increase sales and profitability. The goal is to provide actionable insights from exploratory data analysis (EDA) that Walmart Inc. can implement to enhance performance and gain a competitive edge by increasing sales. Through analysing variables such as total sales, customer segmentation (Customer type and Gender), product lines (for inventory management strategies) and pricing strategies from the Walmart Inc. dataset, the aim is to uncover opportunities for improvement.

The raw Walmart Inc. dataset was extracted from Kaggle and opened in excel. It is comprised of numerical and categorical data. There are 17 attributes and 101 observations uniquely identified by Invoice Id, one of the variables. According to (Sharda, 2017) , the second phase of data processing is data cleaning where values in the dataset are identified and dealt with. At first glance the dataset doesn’t seem to consist of any missing values, however it was still transformed and clean to ensure consistency of the data to improve data quality.

A screenshot of a computer

Description automatically generated

Figure 1

Figure 1 shows a successfully transformed and cleaned Walmart Inc. dataset.

***DESCRIPTIVE SUMMARY OF STATISTICS***

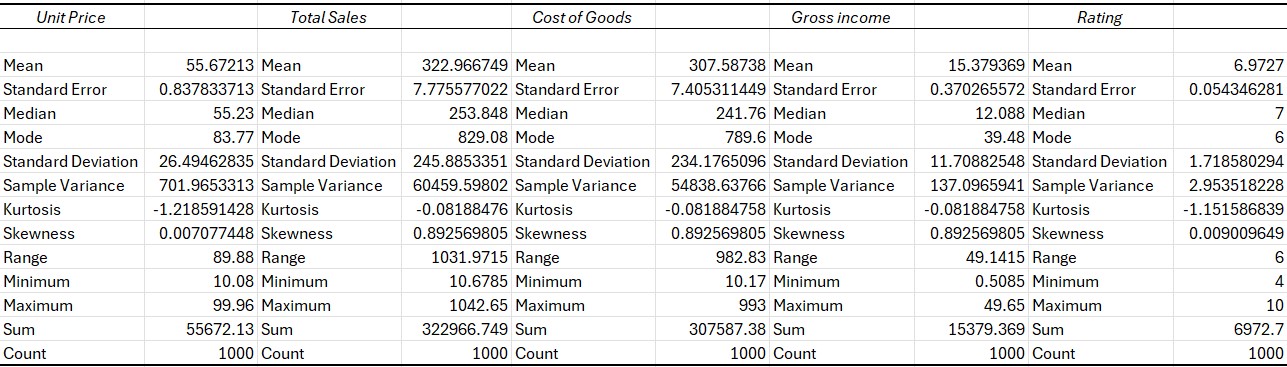


Figure 3

***INTERPRETATION OF THE DESCRIPTIVE SUMMARY OF STATISTICS***

**Unit Price**: With an average of approximately $55.67, the unit price of products sold exhibits moderate variability, indicated by a standard deviation of around $26.49. The dataset encompasses a diverse range of product pricing, with unit prices ranging from $10.08 to $99.96. Symmetrical distribution with a slight positive skewness suggests that most unit prices cluster around the mean.

**Total Sales**: **The average total sales amount per transaction** stands at approximately $322.97, reflecting significant variability with a standard deviation of around $245.89. A wide range of total sales, spanning from $10.68 to $1042.65, indicates varying transaction volumes. Positive skewness in the distribution suggests the presence of transactions with exceptionally high sales amounts.

**Cost of Goods Sold (COGS)**: On average, the cost of goods sold per transaction is approximately $307.59, exhibiting moderate variability with a standard deviation of around $234.18. A wide range of COGS, from $10.17 to $993, suggests varying costs associated with producing goods. Like total sales, the distribution of COGS is positively skewed, indicating the presence of transactions with exceptionally high COGS values.

**Gross Income**: The average gross income per transaction is approximately $15.38, with moderate variability reflected by a standard deviation of around $11.71. A relatively wide range of gross income, from $0.51 to $49.65, suggests varying levels of profitability. Positive skewness in the distribution indicates the presence of transactions with exceptionally high gross income values.

**Rating**: Customers provide an average rating of approximately 6.97, with moderate variability indicated by a standard deviation of around 1.72. Ratings range from 4 to 10, reflecting a range of customer satisfaction levels. The distribution of ratings appears to be approximately symmetrical, with a slight negative kurtosis, suggesting a relatively normal distribution.

Overall, these interpretations offer insights into the characteristics and variability of the variables in the dataset.

Having gained valuable insights into the characteristics and variability of the variables in the dataset through the summary of descriptive statistics, the focus shifts to the **analysis of visualizations**. By delving deeper into the data through visual exploration, the aim is to uncover actionable patterns, trends, and relationships that will inform strategic decision-making and contribute to achieving this project's overarching goal.

***ANALYSIS OF VISUALIZATIONS***

|  |
| --- |
| [Average Total sales by Date](https://app.powerbi.com/MobileRedirect.html?action=OpenReport&reportObjectId=acc11a67-5467-4972-891d-de4e5e122c4b&ctid=ee292977-7b58-4df4-8138-50d0de7d28d8&reportPage=ReportSection&pbi_source=copyvisualimage)  Figure 4 f |
| [Open in Power BI](https://app.powerbi.com/MobileRedirect.html?action=OpenReport&reportObjectId=acc11a67-5467-4972-891d-de4e5e122c4b&ctid=ee292977-7b58-4df4-8138-50d0de7d28d8&reportPage=ReportSection&pbi_source=copyvisualimage) |

This visualization in figure 4 shows the trend of Average total sales over a period of 3 months.

1. **Recent Trend**: The recent uptrend in Average total sales, particularly between Tuesday, February 26, 2019, and Saturday, March 2, 2019, corresponds to an increase in sales volume by 96.89 units. This trend mirrors the overall average sales amount per transaction, suggesting a period of heightened sales activity during this timeframe.
2. **Steep Trend**: Similarly, the steep upward trend in Average total sales observed during the same period indicates a significant rise in sales, further corroborating the positive sales trend identified. This steep increase in sales aligns with the range of total sales, indicating a notable spike in sales volumes during this period.
3. **Long Trend**: The prolonged upward trend in Average total sales between Sunday, January 6, 2019, and Wednesday, January 16, 2019, with a rise of 27.33 units, reflects a sustained period of growth in sales. This long-term trend is consistent with the average sales amount per transaction, indicating a gradual but steady increase in sales over the specified period.
4. **Anomalies**: The anomalies detected in Average total sales highlight instances where sales deviated significantly from the expected range.

A graph on a screen

Description automatically generatedA graph with blue lines

Description automatically generated

Figure 5

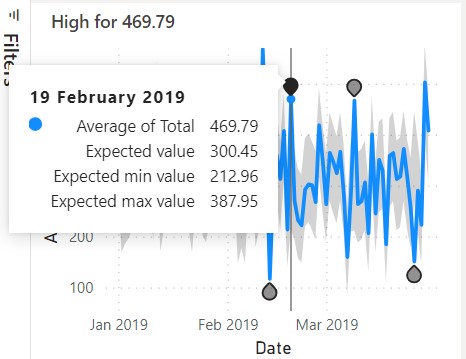


Figure 6

For example, as illustrated in figure 5 and 6, the unexpectedly low sales on Tuesday, March 26, 2019, and unusually high sales on Tuesday, February 19, 2019, suggest potential anomalies in sales performance. These anomalies may warrant further investigation to understand the underlying factors contributing to these deviations from the expected range.

\***possible factors driving changes in the observed trends from the line chart**…

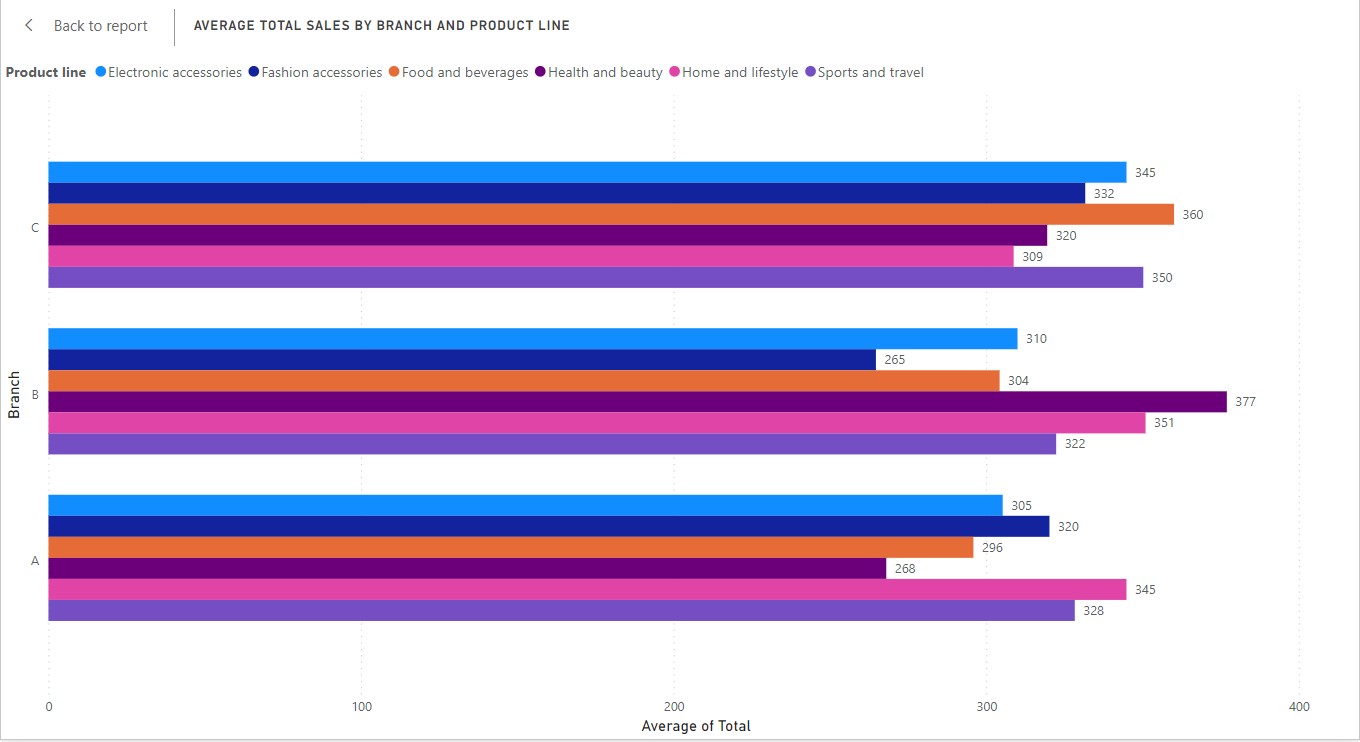


Figure 7

This chart in figure 7 compares the average total sales across different branches and product lines.

1. **Top-Performing Branches and Product Lines**: Branch B appears to have the highest average total sales across all product lines, particularly in the categories of Health and beauty at $377 and Home and Lifestyle at $351. Among the product lines, Food and Beverages consistently show high average total sales across all branches.

2. **Areas Requiring Attention**: Branch A shows relatively lower average total sales compared to Branches B and C, particularly in the categories of Health and beauty and Home and Lifestyle. Fashion accessories in Branch B and Food and Beverages in Branch C also exhibit lower average total sales compared to other product lines within the same branches.

3. **Potential Strategies for Improvement**: To replicate success in underperforming areas, such as Branch A and certain product lines, strategies like cross-selling or product bundling could be explored. Leveraging the success of high-performing product lines, such as Food and Beverages, across all branches could contribute to overall sales growth.

4. **Correlations Between Branches and Product Lines**: Branch B consistently performs well across various product lines, indicating a potential correlation between the branch's operational efficiency and sales performance. The relatively lower average total sales in Branch A and C may suggest areas for improvement in marketing strategies or product offerings specific to those branches.

5. **Actionable Insights**: Identifying top-performing branches and product lines enables Walmart Inc. to allocate resources effectively and focus on strategies that capitalize on existing strengths. Analysing underperforming areas provides opportunities for targeted interventions to enhance sales performance and drive overall profitability.

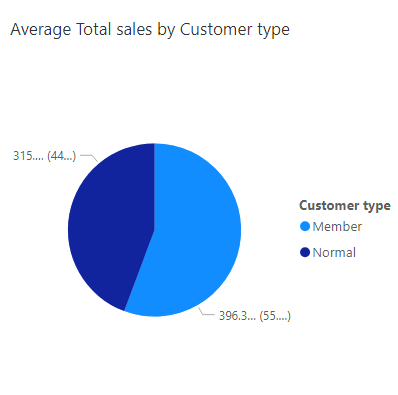
[](https://app.powerbi.com/MobileRedirect.html?action=OpenReport&reportObjectId=acc11a67-5467-4972-891d-de4e5e122c4b&ctid=ee292977-7b58-4df4-8138-50d0de7d28d8&reportPage=ReportSection&pbi_source=copyvisualimage)

Figure 8.Pie chart showing Average Toatl

|  |
| --- |
| [Open in Power BI](https://app.powerbi.com/MobileRedirect.html?action=OpenReport&reportObjectId=acc11a67-5467-4972-891d-de4e5e122c4b&ctid=ee292977-7b58-4df4-8138-50d0de7d28d8&reportPage=ReportSection&pbi_source=copyvisualimage) |

The pie chart in Figure 8 displays the distribution of average total sales between different customer types. The percentage of average total sales contributed by members and non-members allows for a comparison of their respective shares in the overall sales revenue.

1. **Member vs. Non-Member Spending**: Members account for approximately 44% of the total pie chart, with an average total sales value of $315.12.

Non-members represent the majority, constituting around 55% of the pie chart, with an average total sales value of $396.30.

2.**Purchasing Behaviour Differences**: The higher average total sales value for non-members suggests that they may make larger purchases compared to members, despite members accounting for a smaller portion of the total sales.

Members, on the other hand, contribute a significant portion of sales but may have more frequent, smaller transactions compared to non-members.

3.**Relationship to Summary Statistics**:

4.**Identifying Trends and Patterns**:

Overall, by analysing the dfiistribution of total sales by customer type and its relationship to summary statistics and customer segmentation, Walmart Inc. can gain valuable insights into purchasing behaviour, customer preferences, and opportunities for enhancing customer engagement and satisfaction.

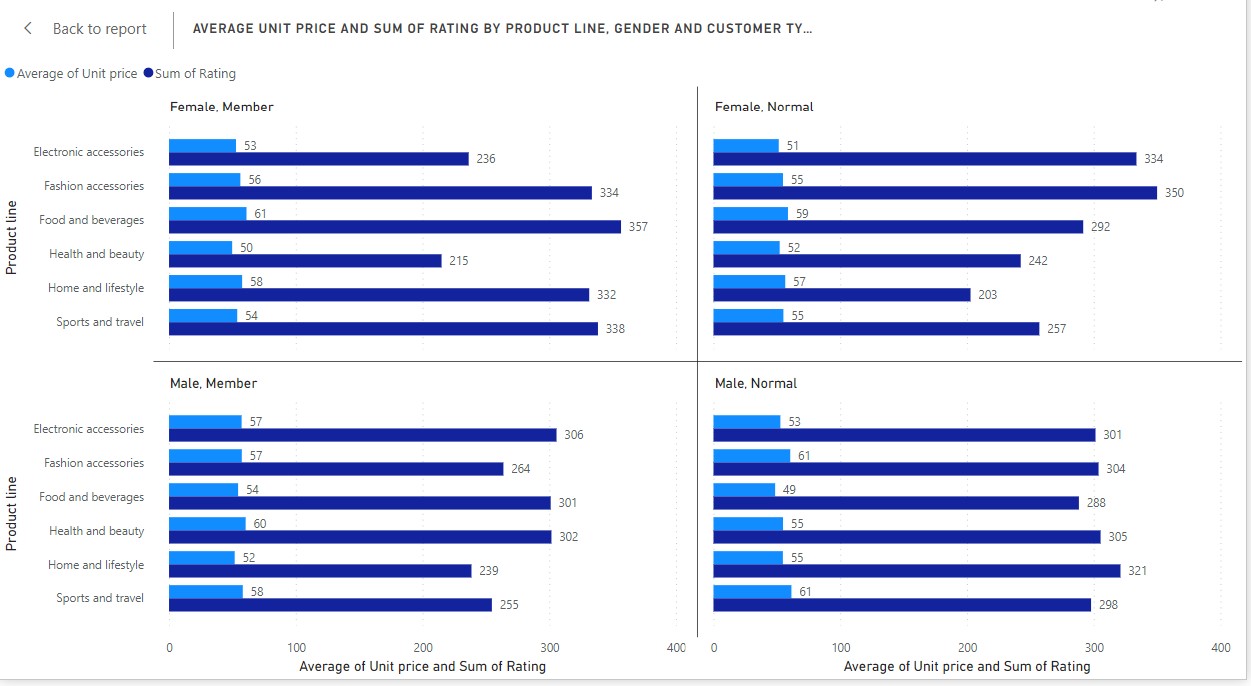


Figure 9 Clustered column chart

**Pricing Strategies**: Across different product lines, there are variations in the average unit prices. For instance, "Food and Beverages" and "Home and Lifestyle" have relatively higher average unit prices compared to "Electronic accessories" and "Health and beauty. "Analysing pricing strategies involves understanding how these variations align with customer preferences and market demands. For example, higher-priced products may target a niche market segment or offer premium quality, while lower-priced items may aim for broader accessibility.

**Pricing Strategies Analysis**:

Examining the average unit prices across various product lines reveals intriguing insights:

"Electronic accessories" boast average unit prices ranging from $51 to $57. "Fashion accessories" exhibit prices ranging from $55 to $61. "Food and Beverages" showcase prices ranging from $59 to $61. "Health and beauty" products vary from $50 to $60. "Home and Lifestyle" items range from $52 to $58. "Sports and Travel" products span prices from $54 to $61.

**Customer Satisfaction Levels:** The sum of ratings provides insights into customer satisfaction levels across different product lines, genders, and customer types.Comparing the sum of ratings across product lines allows us to identify which categories consistently receive higher ratings, indicating greater customer satisfaction.For instance, "Fashion accessories" and "Sports and Travel" demonstrate relatively higher sum of ratings across all customer segments, suggesting higher satisfaction levels in these product categories.

**Customer Satisfaction Levels Analysis**:

Analysing the sum of ratings sheds light on customer satisfaction across various product lines: "Fashion accessories" and "Food and Beverages" generally boast higher customer satisfaction ratings. "Health and beauty" products also show relatively high satisfaction levels.

**Relationship between Pricing, Product Quality, and Customer Satisfaction**: While this analysis suggests a potential correlation between higher-priced products and customer satisfaction levels, it's important to note that this observation is based on the data available and may not fully capture the complexities of customer preferences and perceptions. Further research and validation are necessary to confirm the relationship and understand the underlying factors influencing customer satisfaction.

**Recommendations**:

Adjusting pricing strategies based on customer demographics could further enhance this relationship. For example: Offering competitive pricing on products with high satisfaction ratings could attract more customers and foster loyalty.

Tailoring pricing strategies to specific customer segments, such as offering discounts or promotions to incentivize purchases among certain demographics, could drive sales and strengthen customer relationships.

**Values in the Analysis**: Average unit prices and sum of ratings for each product line, segmented by gender and customer type, provide the foundation for this analysis. For example:

"Fashion accessories" have average unit prices ranging from $55 to $61, with corresponding sum of ratings ranging from 264 to 350 across different customer segments.

By leveraging these insights, Walmart Inc. can refine its pricing strategies to align with customer preferences, enhance product value perception, and ultimately drive sales and loyalty in the competitive retail market.

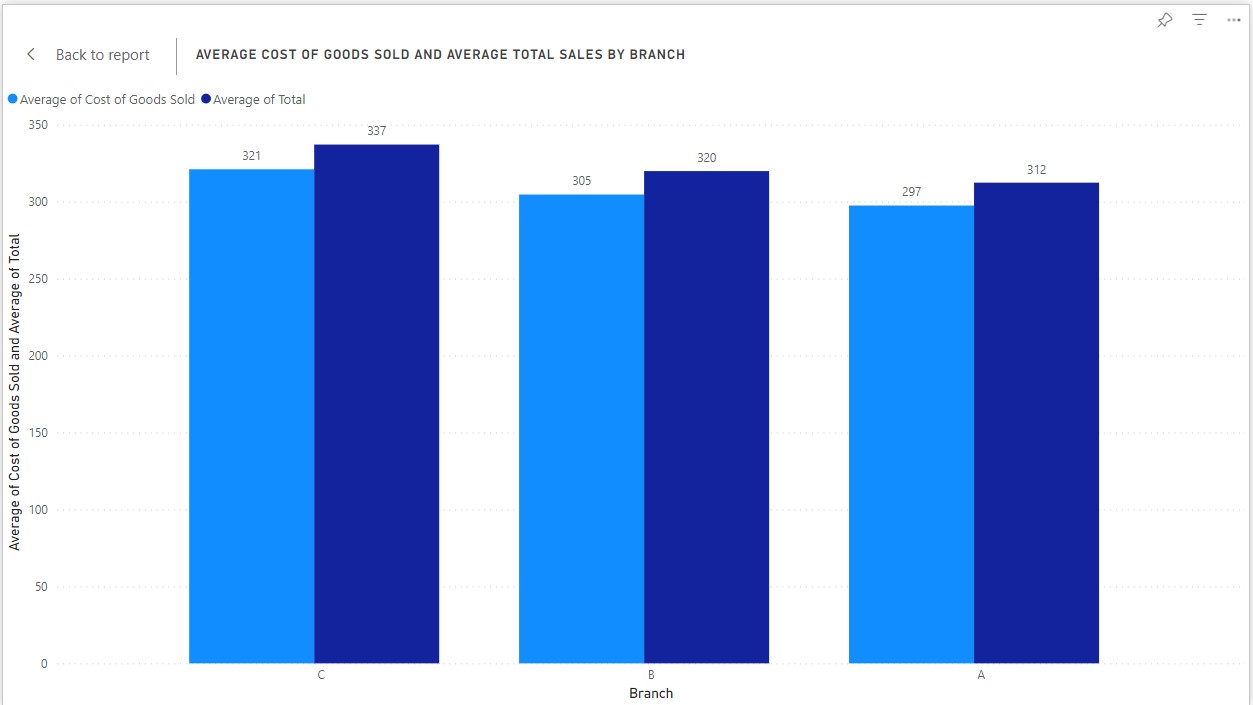


Figure 10

The chart in Figure 10 compares the average cost of goods sold and average total sales across different branches. This visualization offers insights into the cost-effectiveness of sales in each branch, shedding light on potential areas for improvement and optimization.

**From the data presented**:

Branch A has an average cost of goods sold of $321 and average total sales of $337.

Branch B shows an average cost of goods sold of $305 and average total sales of $320.

Branch C indicates an average cost of goods sold of $297 and average total sales of $312.

**Analysis**:

Branch A demonstrates robust sales performance, with an average unit price of approximately $55.67 and an average total sales amount per transaction of approximately $322.97. However, the average cost of goods sold, approximately $307.59, suggests that a significant portion of the sales revenue is allocated towards covering the costs of goods sold.

In comparison, Branch B exhibits slightly lower average total sales compared to Branch A, but also lower average cost of goods sold. With similar average unit price and total sales figures, Branch B may be more cost-effective in managing its inventory and production costs, potentially resulting in improved profitability.

Branch C demonstrates the lowest average total sales among the three branches, accompanied by a relatively low average cost of goods sold. While the cost of goods sold is relatively low, indicating efficient cost management, the lower sales figures raise questions about sales performance and potential opportunities for revenue growth.

While Branches A and C appear to have a relatively favourable cost-effectiveness ratio, Branch B stands out with slightly lower average total sales compared to its average cost of goods sold. This suggests that while sales revenue is generated, it may not be sufficient to cover the associated costs effectively.

To enhance profitability and increase sales in Branch B, exploring cost-saving measures or operational improvements is crucial. Optimizing inventory management practices or negotiating better supplier deals could help reduce costs. Additionally, analysing sales strategies specific to Branch B may reveal opportunities for targeted marketing efforts or product diversification to stimulate sales growth and improve cost-effectiveness.

In conclusion, this Exploratory data analysis has unearthed valuable insights into Walmart Inc.'s retail operations, underscoring the pivotal role of data analytics in driving profitability and operational efficiency.

**Key Findings**:

The line chart depicting average total sales by date revealed notable trends in sales performance over time, highlighting periods of growth and potential anomalies.

Branch and product line comparisons from the clustered bar chart underscored variations in sales performance across different segments, indicating areas of strength and opportunities for improvement.

The pie chart illustrating the distribution of average total sales by customer type shed light on the contribution of member and non-member customers to overall sales, providing insights into customer segmentation strategies.

Analysis of pricing strategies revealed variations in average unit prices across product lines, prompting considerations of how pricing aligns with customer preferences and market dynamics.

Comparisons of average cost of goods sold and average total sales across branches highlighted differences in cost-effectiveness and profitability, signalling areas for optimization.

**Actionable Recommendations**:

Utilize insights from sales trends to inform inventory management and promotional strategies, capitalizing on periods of growth and addressing anomalies promptly.

Tailor marketing efforts and product offerings based on branch and product line performance, optimizing resources to maximize returns.

Implement targeted customer segmentation strategies to cater to the distinct preferences and behaviours of member and non-member customers, fostering loyalty and driving sales.

Refine pricing strategies to align with market demands and customer expectations, balancing premium offerings with accessible options to capture diverse consumer segments.

Enhance operational efficiency by identifying and addressing cost inefficiencies, leveraging data-driven insights to streamline processes and improve profitability.

By leveraging these insights and recommendations, Walmart Inc. can unlock new opportunities for growth, strengthen its competitive position, and ultimately achieve its goal of optimizing retail operations to drive profitability and success in the dynamic retail landscape.

# References

Sharda, R. D. D. a. T. E., 2017. *Business Intelligence, analytics and Dara Science: A Managerial Approach.* Boston: s.n.