**Hypermart Data analysis Report**

Name: Aveek Bandyopadhyay

Registration number: 240957118

# Objective

# The main goal of this analysis is to derive meaningful business insights from the Hypermart dataset. By utilizing Python libraries such as pandas, matplotlib, and seaborn, the objective is to understand key performance metrics and behaviors related to customer segments, product categories, regional demand, and discount sensitivity. These insights will assist in strategic decision-making to enhance revenue and operational efficiency.

# Initial Data Exploration

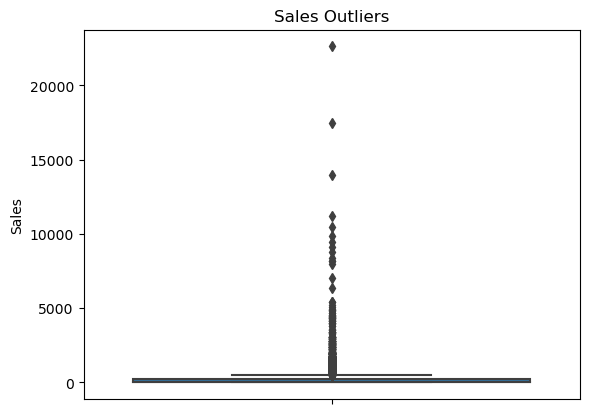
The dataset contained 9994 rows and 21 columns with various types of information, including order and shipping dates, customer details, product information, and financial metrics such as sales, discounts, and profit. No missing values were detected in any of the columns, which made it suitable for direct analysis. The sales values ranged from as low as 0.44 to as high as 22,638.48, while the profit varied from a loss of -6599.98 to a gain of 8399.97.

# Q1. Are there any duplicate or unnecessary attributes?

There were no duplicate rows in the dataset. The 'Row ID' column was deemed redundant since each entry already had a unique identifier through the Order ID. Additionally, the 'Country' column had only one value—United States—so it was excluded from further analysis. Columns such as 'Ship Mode', 'Segment', 'Region', 'Category', and 'Sub-Category' were converted to categorical data types to reduce memory usage and improve processing speed.

# Q2. Are there any missing values or inconsistencies?

The dataset was complete with no null values. Columns containing dates were converted into datetime format to support time-based analysis. Formatting inconsistencies in category names were corrected to maintain uniformity. Overall, the dataset was clean and well-prepared for analysis.



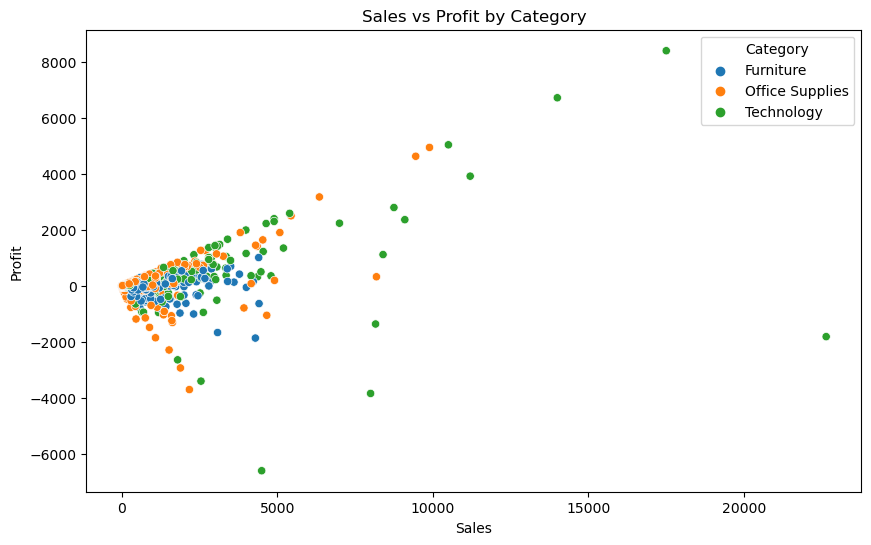
Boxplot of sales column.

According to the boxplot, sales column has many outliers, which could be bulk orders. it represents that some orders were very cheap while others were very expensive. These may also represent high priced products. The plot confirmed that while most sales transactions were of modest value, a significant number of large transactions pushed the upper range of data. Understanding these outliers is important for inventory and pricing strategies.

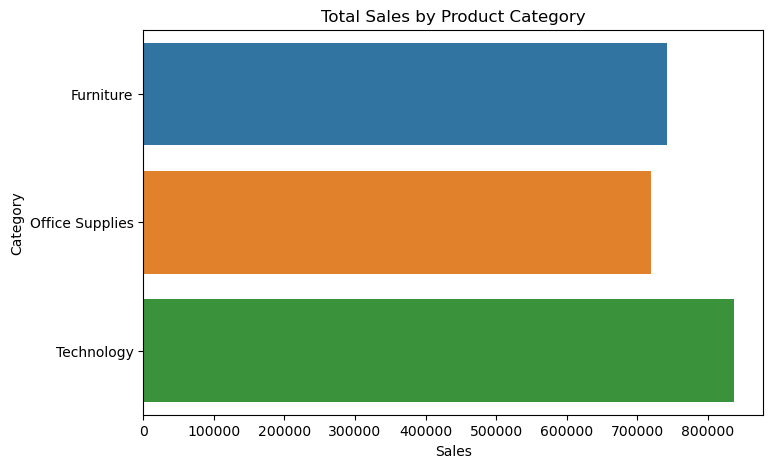
# Q3. Exploratory Data Analysis

## (a) Data Distribution and Trends

Analysis of segment-wise sales and profit revealed that the Consumer segment brought in the highest sales and profit overall. However, not all segments performed equally. Scatter plots indicated that many Office Supplies and some Technology products were sold at low or negative profits. This suggested over-discounting or pricing inefficiencies. Technology was the leading category in terms of overall sales volume. Shipping analysis showed clear differences in delivery times among modes. Same-day delivery had the shortest shipping time, while standard class took the longest—typically 4–6 days.



From the scatterplot, inference is that most office supplies and some technology items are bought at low or negative profits. The issue could be a high discount.



The bar plot tells us that total sales from all three categories was highest in the technology category.

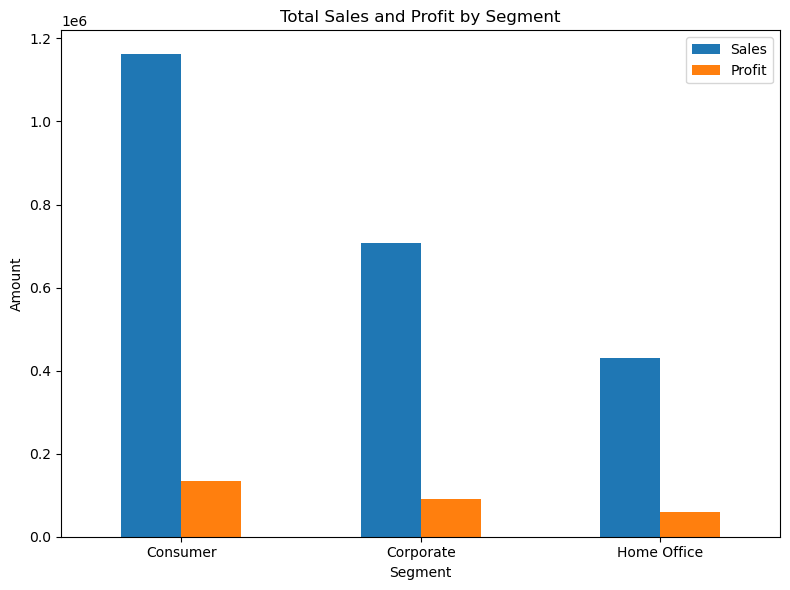


The boxplot here shows us that shipping mode is clearly impacted by time, with shipping on same day having a median of 0-1.

First class takes 2-3 days, second class takes 2-4 days, standard class takes 4-6 days in general.

# b. Customer Segmentation Insights

Consumer customers generated the highest overall sales and profits. However, when comparing profit-to-sales ratios, the Home Office segment performed most efficiently. Corporate and Home Office segments showed increased losses when discount levels were high—especially at a 50% discount level. This suggests that these segments are more sensitive to aggressive discounting. The Consumer segment demonstrated more resilience to discounts, indicating a better response to price promotions.



The bar plot shows us that consumer category has highest sales and highest profit. But the ratio of sales to profit is highest for home office.

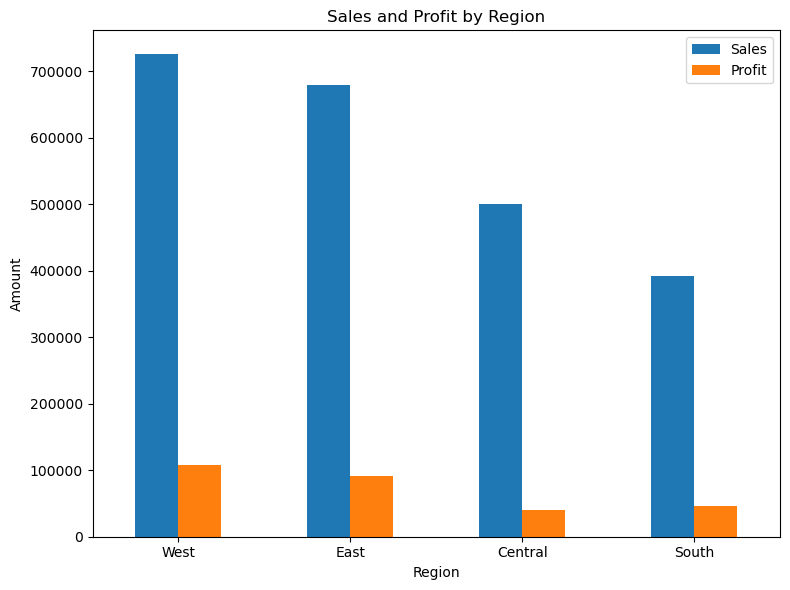


The plot between discount and average profit of different categories shows us that corporate and home office segments have the highest losses, especially the records where discount is 0.5

Consumer segment hence responds better to discounts.

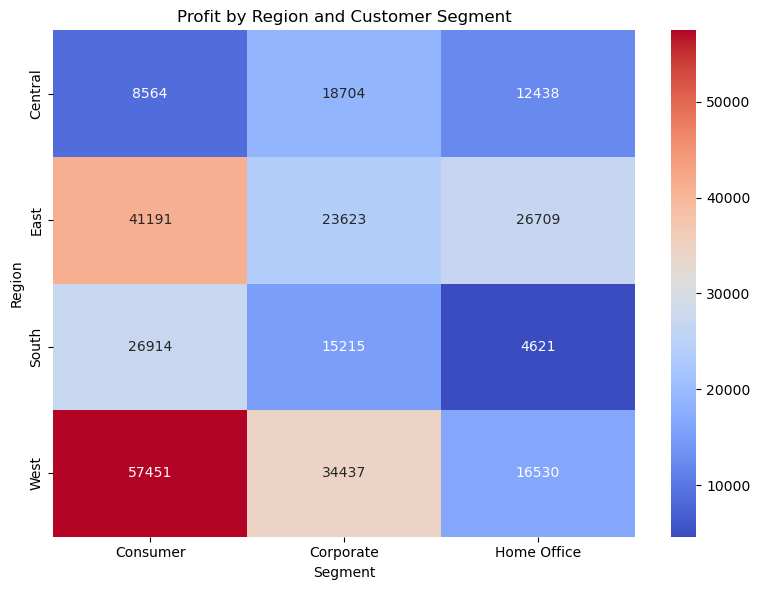
# C. Regional & Growth Opportunities

Regional analysis highlighted the West as the most profitable region across all customer segments. The East region followed in profitability. The Central and South regions underperformed, especially in the Home Office segment. This presents an opportunity for growth by improving strategies in these areas. Tactics could include tailored discounts, localized marketing, or service enhancements to increase engagement and profitability in those regions.



We evaluate regional performance to identify areas for potential business expansion and strategy optimization.

We see that profit and sales in the central and south region are less.



From the heatmap we see West has the highest overall profit in all three customer segments

East is the second most profitable region

# Conclusion

To maximize business performance, Hypermart should continue focusing on the Consumer segment and the West and East regions. Discounting strategies should be adjusted carefully, especially for Corporate and Home Office customers, where high discounts significantly reduce profit. Regions like the South and Central should be considered for business development through improved targeting, service upgrades, and cost management strategies.