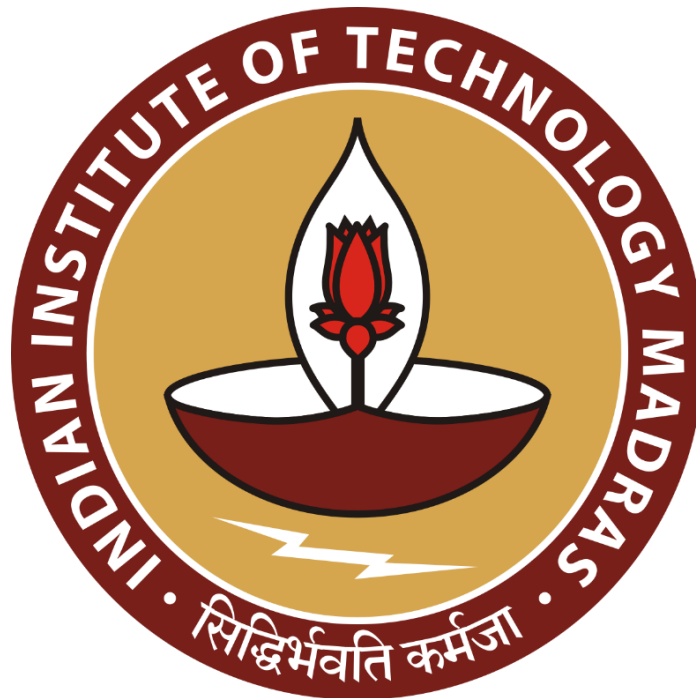# Optimizing Sales and Profitability for a Multi-Category Retail Superstore

A Proposal report for the BDM capstone Project

Submitted by

Name: Parth Kacha

Roll No: 23f1002650

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

Declaration Statement

I am working on a Project Title "Optimizing Sales and Profitability for a Multi-Category Retail Superstore". I extend my appreciation to Superstore Sales Team, for providing the necessary resources that enabled me to conduct my project.

I hereby assert that the data presented and assessed in this project report is genuine and precise to the utmost extent of my knowledge and capabilities. The data has been collected from Kaggle's public data repository, and all analysis has been performed independently using Tableau.

Additionally, I affirm that all procedures employed for the purpose of data collection and analysis have been duly explained in this report. The outcomes and inferences derived from the data are an accurate depiction of the findings acquired through thorough analytical procedures.
I am dedicated to adhering to the information of academic honesty and integrity, and I am receptive to any additional examination or validation of the data contained in this project report.

I understand that the execution of this project is intended for individual completion and is not to be undertaken collectively. I thus affirm that I am not engaged in any form of collaboration with other individuals, and that all the work undertaken has been solely conducted by me. In the event that plagiarism is detected in the report at any stage of the project's completion, I am fully aware and prepared to accept disciplinary measures imposed by the relevant authority.

I agree that all the recommendations are business-specific and limited to this project exclusively, and cannot be utilized for any other purpose with an IIT Madras tag. I understand that IIT Madras does not endorse this.

Parth

Signature of Candidate: (**Digital Signature**)
Name: Parth Kacha
Date: 03/12/2025

1. Executive Summary

This project titled **"Optimizing Sales and Profitability for a Multi-Category Retail Superstore"** focuses on analyzing sales performance, profit trends, customer behavior, and operational challenges using the publicly available Superstore US Sales dataset. The dataset consists of multi-regional retail transactions across various product categories such as Technology, Office Supplies, and Furniture, recorded over a period of four years.

The primary objective of this project was to understand how key business drivers such as product category, geography, discounting decisions, customer segments, and shipping modes influence overall revenue and profitability. Tableau was used to design interactive dashboards for real-time pattern identification, while Python-based statistical analysis supported deeper validation of insights.

The analysis revealed that Furniture, especially Tables, consistently generated losses despite contributing significantly to sales. Technology emerged as the most profitable category. The Central and South regions showed comparatively lower profitability and require strategic intervention. Discounts were shown to have a strong negative correlation with profit, where excessive price reductions turned high-value items into loss-making transactions. Customer segmentation results indicated that the Consumer segment contributed the highest share of revenue, while Corporate and Home Office segments remained stable but less profitable.

Overall, this project delivered actionable insights to improve discount policies, optimize product mix, strengthen region-specific sales strategies, and enhance profitability through data-driven decision-making. These findings provide a strong foundation for improving inventory planning, marketing focus, and logistics efficiency in a highly competitive retail market.

2. Proof of Originality and Data Source Details

The dataset used in this project is publicly available and sourced from Kaggle, an open data-sharing platform widely used for analytics, academic research, and machine learning experimentation. The dataset contains transactional records from a fictional U.S.-based retail superstore, covering sales operations across multiple product categories, customer segments, shipping modes, and geographical regions.

Dataset Information

- Repository Name: Superstore Dataset – Final
- Dataset Contributor: Vivek468

- Source: Kaggle Public Data Repository

- Total Records: 9,994 rows

- Geographical Scope: United States (across four major regions)

- License: Open and permitted for educational and research use

- Dataset link: https://drive.google.com/file/d/1lcCsWYVJUGGlbUI27R2TZVqFOf-8LQzR/view?usp=sharing

- Colab link: https://colab.research.google.com/drive/10KbDV_lmzIUPdPz3ZUSTsauxLZt_Hulv?usp=sharing

No modifications were made to distort the original dataset. All data preprocessing, statistical analysis, visualizations, and business interpretations have been carried out independently by the student as part of this BDM capstone project.

Kaggle Reference:

Superstore Dataset — Vivek468

https://www.kaggle.com/datasets/vivek468/superstore-dataset-final

3. Meta Data and Descriptive Statistics

3.1 Metadata (Data Dictionary)

Dataset: Retail Sales for a Multi-Category U.S. Superstore (Kaggle)

Coverage: 4 calendar years of transactional records (2014–2017)

Total Records: 9,994 rows, 21 attributes

Missing Values: Minimal — no missing values in profit-critical features

The dataset consists of 9,994 transactional records capturing information related to orders, customers, products, geographical distribution, sales performance, and profitability attributes.

The dataset provides a comprehensive overview of retail sales operations in the United States, enabling multi-dimensional analysis such as product-level performance, regional trends, discount impact, and profit variations.

Field	Characteristics	Data Type
Row ID	Unique transaction identifier	int64
Order ID	Unique order identifier	string/object
Order Date	Date of Purchase	datetime64
Ship Date	Date product shipped	datetime64
Ship Mode	Shipping service type	string

Field	Characteristics	Data Type
Customer ID	Unique customer code	string/object
Customer Name	Customer's full name	string
Segment	Customer type	string
Country	Country name	string
State	State name	string
City	City name	string
Postal Code	Area postal code	int64
Region	Sales zone	string
Product ID	Unique product identifier	string/object
Category	Main product classification	string
Sub-Category	Detailed product classification	string
Product Name	Item name	string
Sales	Revenue from product classification	float64
Quantity	Number of units sold	int64
Discount	Price discount applied	float64
Profit	Profit or loss from sale	float64

Table-1 Metadata – Complete Attribute Summary

3.2 Descriptive Statistics

Statistical evaluation of key business variables is performed by using Python.

The dataset shows a wide variability in sales and margins, typical of retail commerce.

Sales (USD)

- Mean: $229.86
- Median: $54.49 → indicates right-skewed distribution driven by premium items
- Range: $0.44 – $22,638.48

Major variance due to high-value technology and furniture products.

Significant dependency on fewer large-ticket orders.

Profit (USD)

- Average Profit: $28.66
- Median Profit: $8.67 → most transactions yield very low earnings

- Minimum Profit: –$6,599.98 → indicates loss-making transactions caused by heavy discounting.

High volatility in profit suggests pricing and discount strategy issues.

Quantity

- Mean Quantity: 3.79 items per transaction

Office supplies drive higher volume but contribute less to profit.

Discount

- Average Discount: 15.6%
- Maximum Discount: 80%

Excessive discounting is strongly linked to negative profit occurrences.

Discounts beyond 20% frequently cause direct losses.

Variable	Data Type	Unit	Count	Min	Max	Mean	Median	Std. Dev.
Sales	float64	USD	9,994	0.44	22,638.48	229.86	54.49	623.25
Profit	float64	USD	9,994	-6,599.98	8,399.98	28.66	8.67	234.26
Discount	float64	Ratio(0-1)	9,994	0.00	0.80	0.156	0.20	0.206
Quantity	int64	Units	9,994	1	14	3.79	3	2.22

Table-2 Statistical Highlights

3.3 Category Level Summary

Category	Total Sales	Total Sales	Avg Sales	Avg Profit	Avg Profit
Furniture	7,41,999.80	349.83	18,451.27	8.70	2,121
Office Supplies	7,19,047.03	119.32	1,22,490.80	8.70	6,026
Technology	8,36,154.03	452.71	1,45,454.95	78.75	1,847

Table-3 Category Summary

It provides an overview of how each major product category performs in terms of sales and profit. From the category-wise summary, we can see that:

- **Technology** has the highest total sales and the highest average profit per order. This indicates strong market demand and better pricing strength for premium products like Phones and Accessories.

- **Furniture** has high total sales but very low average profit. This means that although customers purchase Furniture items frequently, the profit earned on each transaction is very small due to high delivery costs and discount usage.

- **Office Supplies** contribute steady and consistent sales, along with better average profit than Furniture. These products are usually low-risk items purchased in bulk by customers.

Overall, Technology performs the best in terms of profitability, Office Supplies support recurring revenue, and Furniture needs improvement in cost and pricing efficiency.

3.4 – Regional Performance Summary

Region	Total Sales	Avg Sales	Total Profit	Avg Profit	Orders
East	6,78,781.24	238.34	91,522.78	32.14	2,848
West	7,25,457.82	226.49	1,08,418.45	33.85	3,203
Central	5,01,239.89	215.77	39,706.36	17.09	2,323
South	3,91,721.91	241.80	46,749.43	28.86	1,620

Table 4 Regional Summary

The regional summary highlights how the Superstore performs in different parts of the United States.

- **West Region** shows the highest overall profitability, making it the strongest performing market.

- **East Region** also performs well with high revenue and stable profits.

- **Central Region** has moderate sales but very low profit compared to the other regions. Heavy discount usage and weaker customer demand may be the reason.

- **South Region** generates the lowest total sales and profit among all regions, showing a need for targeted marketing to increase growth.

In simple terms, East and West are strong profit zones, while Central and South regions require improved strategies for better revenue and margin results.

4. Detailed Explanation of Analysis Process / Method

4.1 Introduction

This section describes the complete analytical workflow followed in the project. The objective was to understand revenue and profitability behaviour in a U.S.-based retail superstore across products, regions, customer segments, and discounting strategies. The methodology included data quality checking, statistical summarisation using Google Colab (Python), and advanced business visualization using Tableau.

4.2 Tools and Technologies Used

Two primary tools were used for the analysis:

- **Google Colab (Python):** for dataset loading, cleaning, descriptive statistics, correlation, and identifying loss-making items.
- **Tableau:** for interactive dashboards focusing on key managerial insights such as category profitability, regional sales variation, and discount impact.

The two tools made sure the analysis was accurate and easy to understand for business decisions.

4.3 Data Preparation and Cleaning

The dataset was imported into Colab using Pandas. Structural checks confirmed correct datatypes for numerical and categorical variables. There were no missing values in critical metrics such as Sales, Profit, Discount, and Quantity, allowing direct analysis without imputation. Outliers, especially deep-loss transactions, were preserved as they reflect real business risks associated with excessive discounts and weak demand.

4.4 Exploratory Data Analysis (EDA) in Python

Initial EDA included shape analysis, distribution inspection, summary statistics, and grouping based on business attributes.

Key insights were derived through code such as:

- Category-wise sales and profit
- Region-wise performance
- Loss-leading products
- Correlation between discount, sales, and profit

A correlation matrix revealed that discount is negatively correlated with Profit, confirming that aggressive discounting often results in operational losses.

4.5 Business Performance Profiling

Using Python results, profit drivers and risk zones were identified:

- Technology had the highest avg. sales but also major losses due to discounted Machines.
- Office Supplies generated high volume and contributed strongly to recurring revenue.
- A list of top 10 negative-profit transactions confirmed Machines and Binders as high-risk products.

These findings formed the basis for deeper visualization in Tableau.

4.6 Dashboard Design Strategy in Tableau

Five dashboards were created to satisfy project objectives:

1. **Time Trend of Sales & Profit** – Yearly/Monthly performance stability
2. **Category and Sub-Category Performance** – Identifying profitable vs. loss-making product lines
3. **Customer Segment Contribution** – Understanding spending behavior of Consumer, Corporate & Home Office
4. **Shipping Mode Efficiency** – Analyzing logistics-driven margin effects
5. **Region & State Performance** – Regional hotspots for sales & profitability

Each visualization was selected to uncover a specific business problem stated in the proposal.

4.7 Analysis Logic and Interpretation Approach

Insights were derived by evaluating relationships among variables:

- Profit margin gaps → analyzed using bar charts & scatter plots
- High-discount & low-profit clusters → identified visually in scatter distribution
- Regional disparities → measured using heat maps & profit overlays
- Customer segmentation → enabled marketing & retention strategy insights

This approach ensured that observations made in Python were validated through meaningful visual patterns.

4.8 Scope and Limitations

The dataset is fictional and limited to 4 years of U.S. operations. It does not include inventory cost structure or promotional campaign data, which could provide deeper margin attribution. However, the

analysis methods used are scalable to real-world retail datasets and support strong, data-driven business recommendations.

5. Results and Findings

The findings from the Superstore retail sales data were generated by using Tableau dashboards to reveal business insights, and then using Python statistical outputs to validate those findings.Results are organized around the four major business objectives defined earlier: (1) profitability improvement, (2) product mix performance, (3) customer and shipping efficiency, and (4) geographical opportunities and risk zones. Each visualization directly supports decision-making for managerial actions.

5.1 Time Trend of Sales & Profit

This dashboard visualizes the year-wise trend of Sales and Profit for the Superstore across the complete data timeline. Sales exhibit a gradual upward movement year after year, indicating strong and growing market demand. Profit, however, shows fluctuations, with visible sharp drops during specific discount periods. The overall profit contribution does not scale proportionately with the rising revenue trend, highlighting inefficiencies in pricing and operational planning.
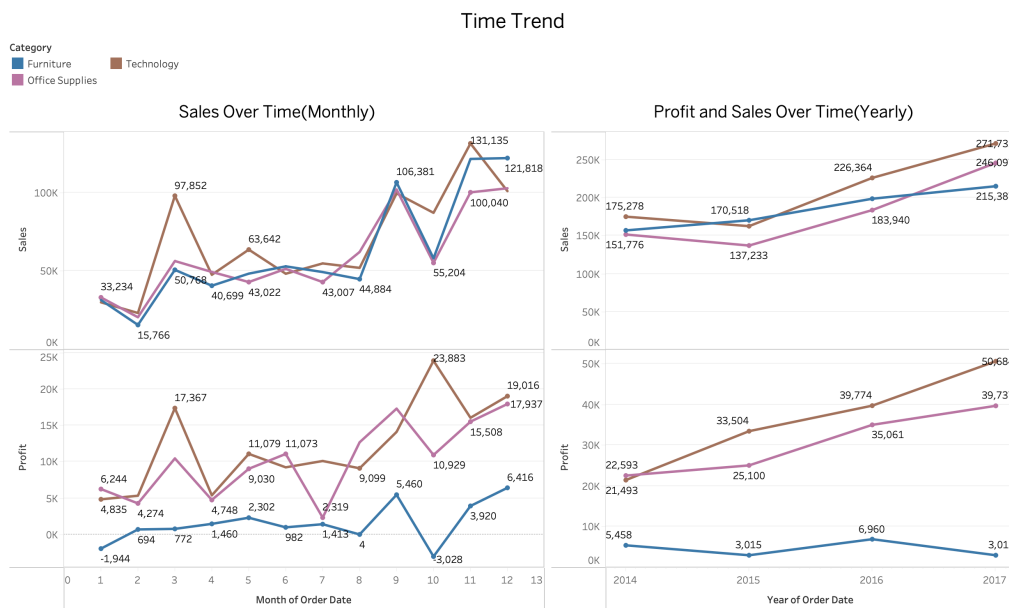

Figure 1. Time Trend Analysis

Interpretation:

Even though the business is expanding and generating more sales, the organization's profitability remains inconsistent. This suggests that increased revenue is being achieved at the cost of margin deterioration. Profit-centric decision making such as discount control and cost efficiency must become a priority.

5.2 Category & Sub-Category Profitability

This dashboard compares profitability across three major categories. Technology records the highest sales (~$836K) and profit (~$145K), mainly driven by high-value items like Phones and Accessories, but certain Machine products incur losses due to heavy discounting. Office Supplies perform steadily with moderate sales (~$719K) and a strong profit contribution (~$122K), making them a low-risk revenue driver. In contrast, Furniture shows high sales (~$742K) but very low profit (~$18K), with sub-categories like Tables and Bookcases causing significant losses.
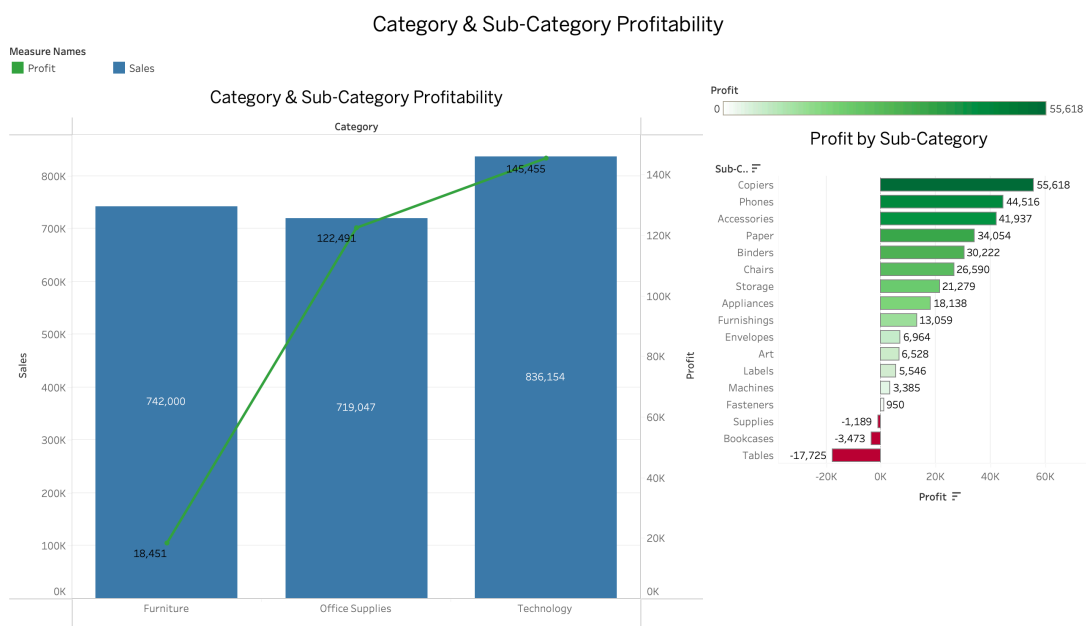

Figure 2 - Category wise Profitability

Interpretation:

Profit and revenue do not always move together the focus should shift from selling more products to selling the right products. Loss-making sub-categories require urgent price and cost restructuring, while profitable product lines like phones and accessories should be promoted more aggressively.

5.3 Customer / Segment Analysis

This dashboard illustrates sales and profit contribution by customer segments and profitability concentration in key buyers.The Consumer segment generates the highest sales (~$1.16M) but produces

the lowest average profit (~$25.8K), indicating thin margins due to price sensitivity and discount dependency.

In contrast, Corporate and Home Office customers, though fewer in orders, contribute higher profit per transaction(~$91K and ~$60K respectively).

The Top 10 Customers by Profit chart highlights a strong dependency on a few high-value customers, with Tamara Chand alone contributing ~$9K profit. This indicates Customer Lifetime Value (CLV) concentration risk.
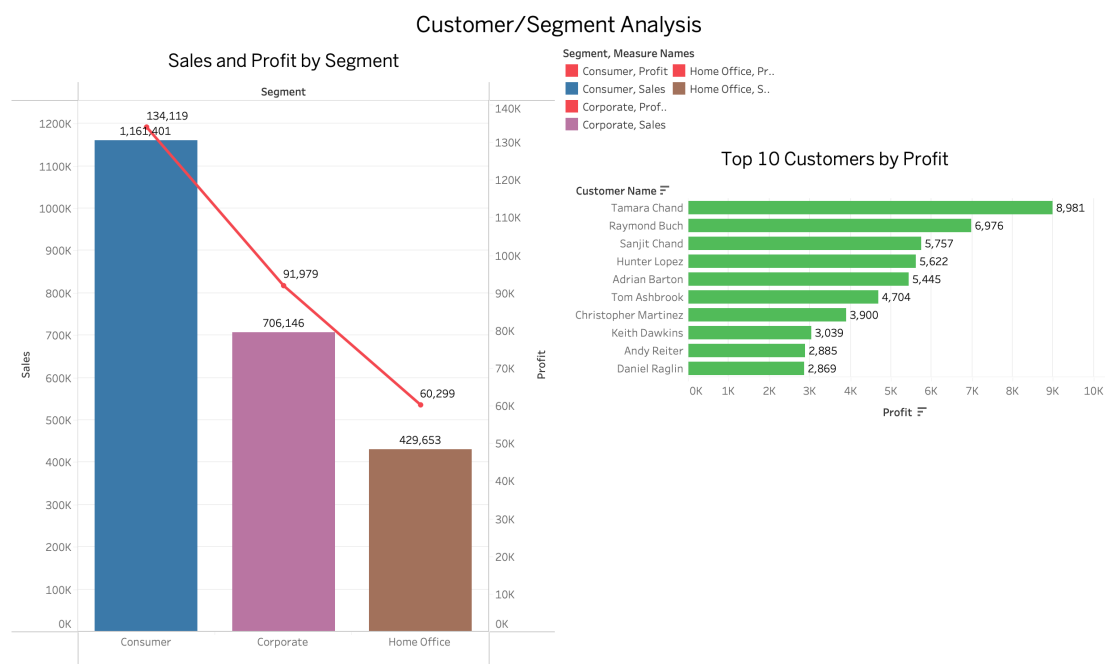

Figure 3 - Customer Analysis

Interpretation:

Relying heavily on the Consumer segment is risky for profit growth. Targeted relationship-building campaigns for Corporate and Home Office clients can significantly enhance overall profitability without increasing operational load.

5.4 Shipping Mode & Discount Analysis

This dashboard evaluates how different shipping modes and discount levels impact profitability. Standard Class shipping, although most frequently used, results in the lowest profit per order due to higher operational and delivery costs. Discounts above 20% strongly correlate with negative profit, especially for Technology products.
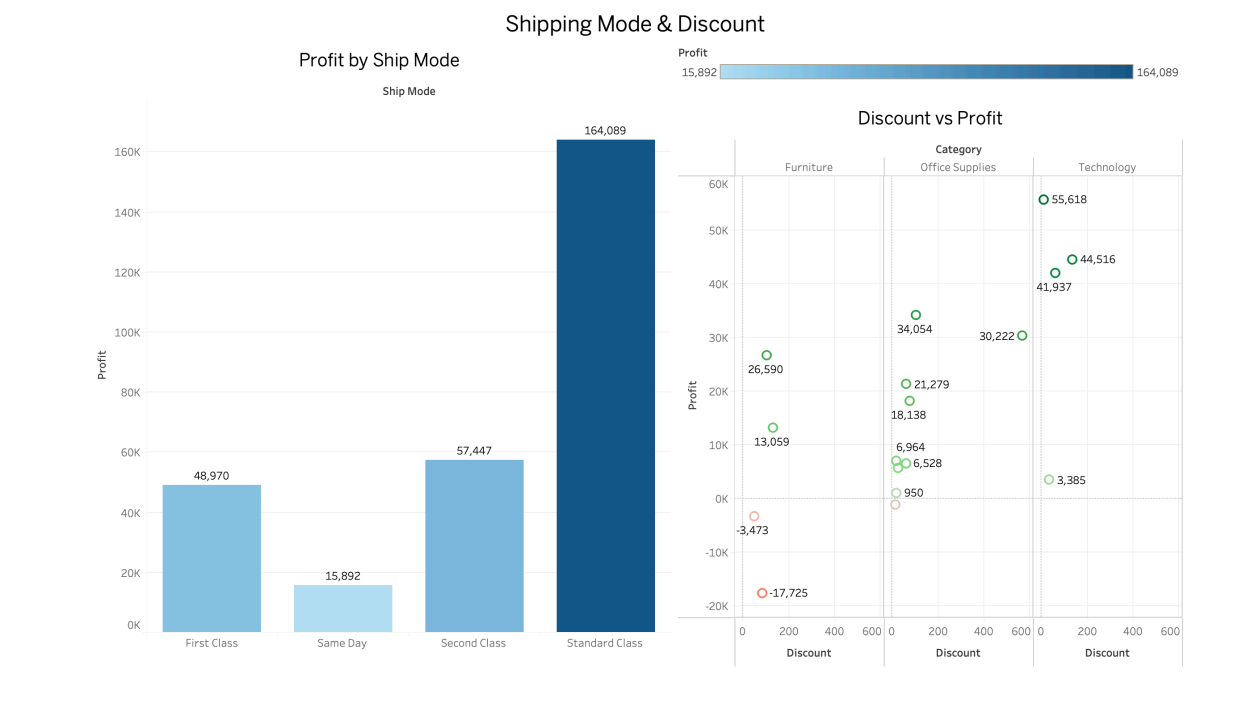
Figure 4 - Logistics Efficiency

Interpretation:

High shipping expenditure and uncontrolled discounting together are heavily eroding margins. By enforcing discount caps and encouraging more economical shipping options at checkout, significant profit recovery can be achieved.

5.5 Region & State Performance

This dashboard maps the geographic distribution of sales and profit across the four U.S. regions — East, West, Central, and South — along with detailed state-level contribution. The East and West regions show strong profitability, while Central and Southern regions reflect lower demand and discount-dependent sales.

East & West - Strong revenue & profit leads to continue investment

Central - Discount dependency, poor profit leads to efficiency intervention

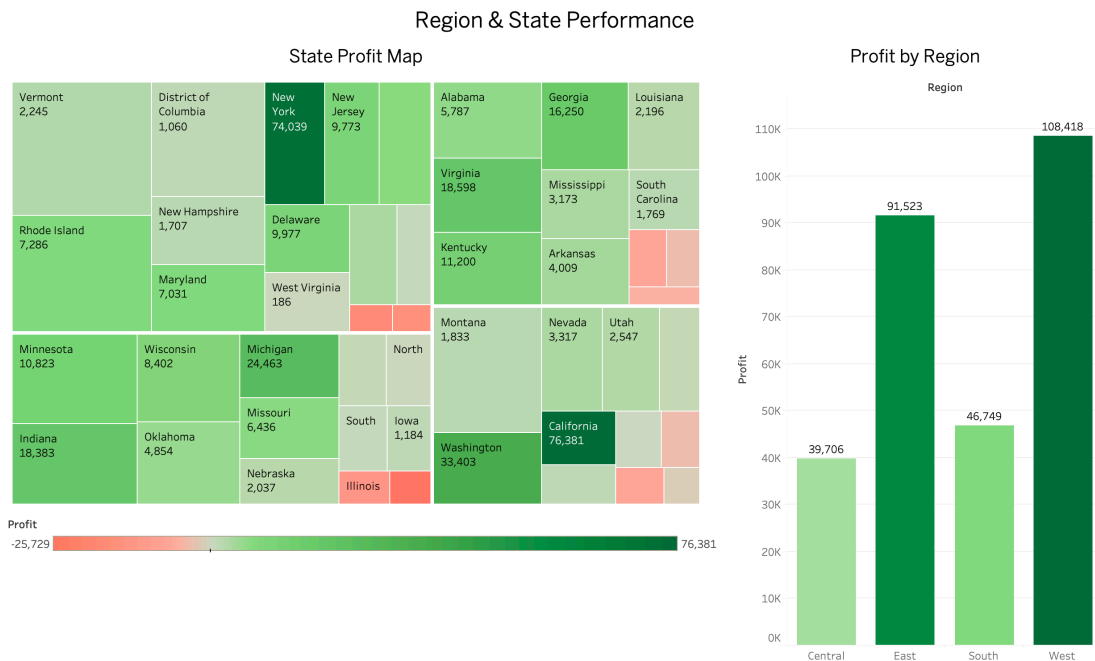South - Soft demand leads to targeted promotions required

Figure 5 - Geographic Profitability

Interpretation:

A uniform market strategy is inefficient. Region-specific decisions are required:

1. Expansion & product push in East & West.

2. Efficiency improvements and pricing review in Central.

3. Targeted promotions to boost weak demand in South.

6. Interpretation of Results and Recommendations

This section explains what the results mean from a business point of view and provides recommendations that the Superstore can follow to improve its profits, customer focus, and operations.

6.1 Interpretation of Results

From the analysis and dashboards, the following important observations were made:

Product and Profitability Insights

- Technology is the most profitable category with high demand for Phones and Accessories.
- Office Supplies generate steady sales and also provide good profits due to large order volumes.
- Furniture shows high sales, but profits are almost negligible. This is mainly because sub-categories like Tables and Bookcases create heavy losses due to high shipping cost and large discounts.

This means that selling more products does not always give more profit — the type of product matters.

Effect of Discounts and Shipping

- Discounts higher than 20% mostly result in negative profit, especially for expensive technology items and furniture.
- Standard Class is the most used shipping mode, but it brings the least profit due to high transportation and delivery expenses.
- Same Day shipments have the lowest profit contribution, showing that faster services add extra cost without enough return.

Therefore, unnecessary discounts and costly delivery options are reducing the overall margin.

Regional and Customer Performance

- The West and East regions are strong profit makers and should be supported more.
- The Central region needs attention because its profit is low even when sales are decent.
- The Consumer segment generates the highest number of orders but brings very low profit per order.
- Corporate and Home Office customers give fewer orders but much larger profit per sale.

This shows that not all customers provide equal value to the company.

Growth Over Time

- Sales have continuously increased with growing customer demand.
- But profit does not increase at the same speed because of poor pricing and discount decisions.

This indicates that the business is growing, but not in the most profitable way.

6.2 Recommendations for Business Improvement

Based on the results, the following actions are recommended:

Product Strategy

- Give more focus to high-profit products like Phones, Accessories, and Office Supplies.

- Review pricing and stock decisions for loss-making items such as Tables and Bookcases.

- Reduce costly items that frequently lead to losses.

Discount Strategy

- Set a maximum discount limit (eg.15–20%) to prevent profit leakage.

- Avoid big discounts on all items. Offer price reductions only where needed or for loyal customers.

Customer Relationship Strategy

- Provide special offers, loyalty programs, and dedicated services for Corporate and Home Office customers.

- Improve profit margins from Consumer segment by reducing discount dependency.

Logistics and Shipping Strategy

- Promote lower-cost shipping like Second Class with clear benefits to customers.

- Improve negotiation with delivery partners to reduce shipping cost for bulky items.

Regional Growth Strategy

- Continue investing in East and West regions because they show strong returns.

- Provide targeted promotions in the South region to boost demand.

- Improve operational efficiency in Central region where discount usage is high.

6.3 Overall Conclusion from Interpretation

The main learning from this project is that:

More sales do not guarantee more profit. Superstore needs to focus on profitable products, cost control, and right customers to grow successfully.

If the above recommendations are followed:

- Profit margins will improve

- Business expansion will be targeted and effective

- Discount and shipping losses will reduce

- Customer handling will become more value-based

With this strategy, the Superstore can create long-term financial growth and better decision-making using data insights.

6.4 Business Impact of Recommended Actions

If the suggested improvements are implemented properly, they can create several benefits for the Superstore:

- Better Profit Margins: By controlling discounts and focusing on profitable products, the company can earn more from the same number of sales.

- Improved Customer Strategy: Targeting customers who bring higher profits will increase long-term revenue.

- Cost Reduction: Better shipping mode selection will reduce logistics costs, especially for bulky items.

- Balanced Regional Growth: Investments in high-performing regions and corrective actions in weak regions will make business performance more stable.

- Smarter Inventory Planning: Less stock of loss-making items will reduce dead stock and wastage.

Overall, these actions will shift the business strategy from sales-focused to profit-focused, helping the company grow in a healthy and sustainable way.

6.5 Risks and Precautions

While applying the above recommendations, the company must be careful of the following risks:

- Customer Dissatisfaction from Reduced Discounts: Some customers may expect big discounts. So, price changes should be gradual and well-communicated.

- Stockouts of High-Demand Products: If the focus is too much on profitable products, shortages might occur. Inventory monitoring must be strong.

- Shipment Delays with Cheaper Shipping Modes: Promoting low-cost shipping should not affect customer satisfaction.

- Regional Strategy Misalignment: One plan may not fit all regions; decisions must consider local demand and competition.

- Too Much Dependence on Large Customers: Relying only on a few profitable customers can be risky if they reduce purchases suddenly.

Therefore, while improvements are necessary, they must be applied with proper balance and planning.

6.6 Future Scope of Analysis

This project mainly focused on sales, profit, customers, regions, and discounts. In the future, the analysis can be expanded to include:

- Forecasting future sales and demand using Machine Learning models
- Customer behaviour analysis such as buying frequency and churn prediction
- Supply chain analytics to reduce delays and improve stock availability
- Market comparison with competitors to build better pricing strategies
- Profitability at product level including cost of storage, manufacturing, and returns
- Seasonal and festival impact study for better promotion planning

7. Presentation and Legibility of the Report

This report is written in a simple and easy-to-read format. All the sections are arranged in a proper order so that the reader can understand the project step by step. Headings and subheadings are added to make the content clear. The pages are not crowded, and enough spacing is kept between paragraphs.

All the charts and dashboards used in the report are placed close to the explanation so the reader can easily understand what the graph shows. The visuals have clear labels, titles, and values, which makes the information easy to read. The font size and style are kept the same throughout the report to maintain a clean look.

The language used in the report is simple and understandable for everyone. Technical terms are explained wherever needed. Overall, the report looks neat, well - organised, and easy to follow without any difficulty.