Final Report:

Profitability Analysis and Customer Segmentation

By

Adwoa Darkoa Nicol

Spirit ID: 1054990

Table of Contents

[Project Problem Statement 3](#_Toc198068412)

[Introduction 3](#_Toc198068413)

[Background 3](#_Toc198068414)

[Problem 3](#_Toc198068415)

[Key Points 4](#_Toc198068416)

[Data Set 4](#_Toc198068417)

[Methodology 4](#_Toc198068418)

[Results/Findings 4](#_Toc198068419)

[Recommendations 5](#_Toc198068420)

[Conclusion 5](#_Toc198068421)

[References 5](#_Toc198068422)

[Appendix 5](#_Toc198068423)

Project Problem Statement  
The company needs to understand why some areas of the business are unprofitable despite strong sales and identify customer segments or product categories that most influence profitability.

Introduction  
This project analyzes sales, discounts, and profit margins to determine what drives profitability for a retail business using the Superstore dataset. The goal is to reveal actionable insights that help optimize pricing, marketing, and customer targeting strategies.

Key subjects discussed include:

* The impact of discounting on profit
* Customer segmentation based on purchasing behavior
* Regional and product category performance

Background  
Profitability in retail is not solely driven by high sales volume. Prior research (Kumar & George, 2007; Pandey & Chawla, 2019) indicates that discounts often reduce profits and that regional performance and product mix have significant impacts. Data analytics, particularly regression and clustering, have been used to uncover these patterns.

This study builds upon those findings using the Superstore dataset to investigate how customer behavior, regional sales, and discount strategies affect overall profitability.

Problem  
While Superstore reports strong sales numbers, profitability remains inconsistent. Some products and regions lose money, and most customers deliver low profit. There is a need to:

* Identify valuable customer segments
* Understand why certain categories or regions underperform
* Reduce reliance on discounting that harms profit

## Key Points

**Key Point 1:** Discounting Reduces Profit  
Correlation (-0.22) and regression analysis show that high discounts significantly lower profit, confirming that discounting is a major issue.

**Key Point 2**: Not All Customers Are Equal  
Clustering analysis revealed that only a small portion of customers (Cluster 2) are high-value. Most customers buy frequently but with low or no profit, making segmentation essential.

**Key Point 3**: Regional and Product Variation  
Regions like West and East are consistently profitable, especially in Q4. Technology products drive the most profit, while Furniture often results in losses.

## Data Set

* Source: Superstore Dataset from Kaggle (Vivek468)
* Content: 9,000+ orders including variables like Sales, Profit, Discount, Category, Region, Segment, and Quantity
* Treatment: Data was cleaned and grouped for analysis. Sales, discounts, and profit were the primary variables used for correlation, regression, and clustering.

## Methodology

* Imported and cleaned the dataset using Excel.
* Calculated correlations between Sales, Profit, and Discount.
* Performed linear regression to quantify impact of Sales, Quantity, and Discount on Profit.
* Applied KMeans clustering (k=4) to segment customers.
* Created bar, pie, and line charts to visualize regional, category, and discount-related trends.

## Results/Findings

* Discounts have a strong negative effect on profit.
* High sales don’t guarantee profit; bulk orders may hurt margins.
* Cluster 2 customers (41 users) are the most profitable.
* The West and East regions perform best; South and Central need improvement.
* Technology is the most profitable category. Furniture often loses money.

## Recommendations

* Reduce or eliminate heavy discounting.
* Focus on high-value customer segments.
* Reallocate resources to profitable regions (East, West).
* Review pricing and cost structures for unprofitable categories.

Conclusion  
This study highlights that retail profitability is driven more by smart segmentation and operational efficiency than just sales volume. By using data analysis, Superstore can shift from a volume-based approach to a value-focused strategy. Future work could involve real-time tracking and customer lifecycle value modeling.

## References

* Kumar, V., & George, M. (2007). Measuring and maximizing customer lifetime value: a review and future directions. Journal of Interactive Marketing.
* Pandey, R., & Chawla, D. (2019). Customer segmentation and strategy development based on customer lifetime value. Journal of Business Research.
* Superstore Dataset. (n.d.). Retrieved from <https://www.kaggle.com/datasets/vivek468/superstore-dataset-final>

## Appendix

#### Correlation

A screenshot of a computer

AI-generated content may be incorrect.

* **Discount and Profit correlation**: -0.22, confirming a weak but consistent negative relationship.
* **Sales and Profit correlation:** 0.48, a moderate positive correlation, meaning that while higher sales generally lead to more profit, it’s not a guarantee.

#### Regression Analysis

A screenshot of a data

AI-generated content may be incorrect.

The regression analysis produced an R-squared value of 0.27, which means that approximately 27% of the variation in profit can be explained by the variables: sales, quantity, and discount. This suggests that while these variables do impact profit, 73% of the variation in profit is influenced by other factors, such as product type, shipping cost, or maybe regional operations.

The regression equation: Profit = (0.18 × Sales) - (2.96 × Quantity) - (233.46 × Discount) + 34.97

This shows:

* A positive relationship between Sales and Profit: more sales generally lead to more profit.
* A negative relationship between Quantity and Profit: large order volumes may lead to lower margins or higher costs.
* A strong negative impact of Discount on Profit: as discounts increase, profits drop significantly.

#### Clustering

Using the KMeans algorithm with k=4, customers were grouped based on **total sales, profit, and quantity purchased**. Here are the four key clusters:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Sales | Profit | Quantity | Number of Customers |
| Cluster |  |  |  |  |
| 0 | 5273.805487 | 503.259380 | 88.699029 | 103 |
| 1 | 1189.027789 | 49.571825 | 26.440729 | 329 |
| 2 | 10329.847032 | 2586.376173 | 67.951220 | 41 |
| 3 | 2935.265701 | 350.658601 | 53.912500 | 320 |

* **Cluster 0 (103 customers)**:
  + High quantity (89), moderate sales ($5,274), low profit ($503).
  + Interpretation: Frequent buyers, but low profitability. Possibly price-sensitive or high-discount customers.
* **Cluster 1 (329 customers)**:
  + Low sales ($1,189), very low profit ($50), and low quantity (~26).
  + Interpretation: Least valuable segment.
* **Cluster 2 (41 customers)**:
  + Very high sales ($10,330), high profit ($2,586), and medium quantity (~68).
  + Interpretation: Most valuable customers. They purchase a lot and generate strong returns.
* **Cluster 3 (320 customers)**:
  + Moderate sales ($2,935), moderate profit ($351), medium quantity (~54).
  + Interpretation: Average customers with steady contribution.

#### Bar chart showing total sales and profit by category

Technology had the highest profit and sales. Office Supplies and Furniture had lower profits. However, there is a significant difference between profit and sales meaning costs are high.

* Insight: Focus marketing or cross-sell on high-performing categories, Technology and find ways to reduce costs to bridge the profit – sales gap.

#### Pie chart of sales distribution by category

Sales are mostly from Technology and Office Supplies. Furniture contributes less.

* Insight: Diversify or reconsider pricing in weaker categories.

#### Sales distribution by Region

A chart with different colored circles

AI-generated content may be incorrect.

The West and East regions had the highest share of total sales.

* Insight: Prioritize support, inventory, or marketing in those regions and do more research to find out the cause of low sales in the other regions.

Discount Vs Profit  
  
A graph with orange lines

AI-generated content may be incorrect.  
Profit drops sharply as discounts increase.

* Insight: Avoid over-discounting. High discounts don’t guarantee profit and may even erode margins.

Profit by Sub-CategoryA graph with blue lines

AI-generated content may be incorrect.Sub-categories like Chairs and Tables have negative profit, while Copiers and Phones are highly profitable.

* Insight: Reevaluate sourcing, pricing, or bundling strategies for low-margin sub-categories.

Profit per Category per Region  
A graph with colorful bars

AI-generated content may be incorrect.

Technology consistently performs well across all regions, especially in the East and Central. Furniture shows low or even negative profit, particularly in the Central region. The West region’s Office Supplies category is highly profitable — a standout compared to the same category in other regions.

* Insight: The store should consider reducing or improving its Furniture operations in Central region. Meanwhile, Office Supplies in the West may offer an opportunity to replicate success elsewhere.

#### Profit per Region

A graph with different colored bars

AI-generated content may be incorrect.

West is the most profitable region. East follows closely behind. Central and South trail significantly in total profit.

* Insight: Investment in customer acquisition, marketing, and inventory should prioritize West and East. Central and South may need process optimization, cost control, or new product strategies.

#### Profit per Quarter year per region

A graph of a graph of a company

AI-generated content may be incorrect.

The East and West regions demonstrate strong and consistent growth over time, with noticeable profit spikes in Q4 each year, suggesting seasonal demand—likely due to year-end or holiday-related purchasing. The West region especially shows a steep rise in profit in late 2016 and continues a generally upward trend. In contrast, the South region remains consistently low in profitability with minimal fluctuation, indicating either market saturation or underperformance. The Central region appears volatile, with profits peaking in some quarters but sharply declining or even turning negative in others, pointing to operational inconsistencies or external disruptions.

* Insight: The East and West regions are the most dependable in terms of profit and should be prioritized for future investment and capacity planning, especially during Q4. Central’s volatility calls for a detailed operational review, while South’s stagnant performance may require a refreshed strategy—either by refining product offerings or improving marketing and customer engagement in that area.