****

**Student full name:**

Mehrshsd Salimnezhad

**Student’s programme:**

MSC in IT for BUSINESS DATA ANALYTICS

**Module:**

Business Intelligence for Data-Driven Management

**Academic Year / Semester:**

Autumn, Spring 2024

**Word Count:**

2470

**Due Date:**

10 December 2024

Contents

[1. Introduction 3](#_Toc184682980)

[2. Methodology 3](#_Toc184682981)

[2.1 Data Source and Description 3](#_Toc184682982)

[2.2 Data Cleaning and Preparation 4](#_Toc184682983)

[2.3 Analytical Approach 5](#_Toc184682984)

[2.4 Tools and Techniques 5](#_Toc184682985)

[2.5 Insights and Interpretation 6](#_Toc184682986)

[2.6 Challenges and Limitations 6](#_Toc184682987)

[3. Data Analysis and Insights 6](#_Toc184682988)

[3.1 Analysis of State-Level Profitability 7](#_Toc184682989)

[3.2 Analysis of Product Categories and Subcategories 7](#_Toc184682990)

[3.3 Insights on Customer Segmentation 7](#_Toc184682991)

[3.4 Time-Based Trends 8](#_Toc184682992)

[3.5 Observations on Discounts and Profitability 8](#_Toc184682993)

[3.6 Limitations in the Analysis 8](#_Toc184682994)

[3.7 Summary of Insights 8](#_Toc184682995)

[4. Recommendations and Conclusion 9](#_Toc184682996)

[4.1 Store Locations 9](#_Toc184682997)

[4.2 Product Focus 9](#_Toc184682998)

[4.3 Operational Strategies 9](#_Toc184682999)

[4.4 Conclusion 10](#_Toc184683000)

[References 10](#_Toc184683001)

# 1. Introduction

In today’s fast-paced retail industry, businesses must constantly adapt to shifting market conditions, evolving customer preferences, and increasing competition. Making smart decisions about where to open new stores and which products to prioritize can mean the difference between success and failure. This report takes on those challenges, using data-driven insights to recommend the best strategies for expanding into new markets.

As a data analytics team, our mission was to identify the most profitable states and product categories for opening five new retail stores in the United States. To achieve this, we analyzed the Sample Superstore dataset—a rich source of transactional data that sheds light on sales, profits, product trends, and geographical performance. We dug into this data to uncover patterns and trends that could guide the company’s expansion strategy.

The stakes are high. Expanding into new markets is a costly, resource-heavy effort. Picking the wrong locations or products could lead to significant losses. On the flip side, a smartly executed expansion can boost revenue, strengthen brand presence, and create a solid foundation for future success. Recognizing this, we used a systematic, data-driven approach to identify the best opportunities.

Our analysis highlighted five states—California, Washington, Virginia, New York, and Michigan—as top choices for new stores. These states have consistently demonstrated strong profitability and steady sales. Additionally, within the Office Supplies category, items like Paper, Art, Labels, and Envelopes stood out as highly profitable, making them excellent options for the new store inventory.

This report goes beyond just presenting data insights. It also connects these findings to academic research, providing a broader perspective and deeper understanding of retail strategies. This report offers a comprehensive approach to retail expansion by blending practical analysis with theoretical insights. In the sections that follow, we’ll outline the steps we took, share our findings, review relevant literature, and provide actionable recommendations to help the company achieve sustainable growth.

# 2. Methodology

A robust and structured methodology was essential to ensure that the insights derived from the Sample Superstore dataset were accurate, actionable, and aligned with the company’s goals. The following steps were taken to analyze the data and derive insights that would support the decision to open five new retail stores in the United States, focusing on profitability and customer demand.

## 2.1 Data Source and Description

The Sample Superstore dataset was selected as the basis of this analysis. This dataset provides a detailed snapshot of business operations, capturing transactional data related to sales, profit, discounts, and product categories. It also includes geographical information at the state level, which is critical for identifying optimal store locations (Zwilling, M. 2022).

Key fields in the dataset include:

* Sales: Total revenue generated from transactions.
* Profit: Net profit after subtracting costs from revenue.
* Category and Subcategory: Broad and specific classifications of products, such as Office Supplies, Furniture, and Technology, with subcategories like Paper, Art, and Labels.
* State: Geographic location of each transaction.
* Customer Segment: Classification of customers (e.g., Corporate, Consumer, or Home Office).

The dataset spans multiple years, providing a robust foundation for identifying trends and patterns over time.

## 2.2 Data Cleaning and Preparation

The initial step involved cleaning and preparing the data to ensure it was suitable for analysis. The following actions were taken:

1. Handling Missing Values: Missing entries in critical fields such as sales or profit were carefully examined and removed if they could not be accurately imputed.
2. Removing Duplicates: Duplicate rows were identified and removed to prevent the overrepresentation of specific transactions.
3. Filtering Data: Only transactions within the United States were retained, as the analysis focuses on domestic expansion.
4. Standardizing Data: Columns with text entries, such as state names or product categories, were standardized for consistency.
5. Aggregating Data: To streamline the analysis, transaction-level data was aggregated by state and product subcategory. This allowed for a higher-level view of profitability and sales trends.
6. We created some calculated fields in Tableau so we can expand our data and reach better visualization and charts. Such as Profit margin, average sales Per Order, Rank by Profit, Rank by Sales, and Shipping time.

For Avg sales per order: SUM[Sales] / CountD [OrederID]

For Profit Margin: [Profit] / [Sales]

For Rank by Profit: Rank (SUM [Profit])

For Rank by Sales: Rank (SUM [Sales])

Shipping Time: DateDIFF(‘day’ ,[Order date] , [Ship date])

These steps ensured the dataset was clean, consistent, and ready for visualization and analysis (Zwilling, M. (2022).

## 2.3 Analytical Approach

The analysis was conducted using a combination of Excel for data preparation and Tableau for visualization. The overall approach involved three main steps:

Step 1: Profitability Analysis by State

* The first focus area was state-level profitability. By aggregating sales and profit data by state, we were able to identify regions where the company consistently achieved high-profit margins.
* A bar chart in Tableau was used to compare the total profits across states, which clearly highlighted the top-performing regions.
* Based on this analysis, California, New York, Washington, Virginia, and Michigan emerged as the most profitable states, making them ideal candidates for store expansion (Murphy, S. A. 2013).

Step 2: Product Performance Analysis

* The second area of focus was product performance, particularly within the Office Supplies category. Subcategories such as Paper, Art, Labels, and Envelopes were analyzed to determine their contribution to overall profitability.
* Using a combination of bar charts and trend lines, we compared the sales and profit performance of these subcategories across the top-performing states. This revealed that these four subcategories consistently generated high-profit margins, especially in the identified states (Murphy, S. A. (2013).

Step 3: Customer Segment Analysis

* To further refine our recommendations, we segmented the data by customer type (Corporate, Consumer, and Home Office). This helped us understand which customer groups were driving demand for specific products.
* Tableau visualizations showed that corporate customers had a higher demand for Office Supplies, particularly Paper, and Labels, suggesting a strong business-to-business (B2B) opportunity.

## 2.4 Tools and Techniques

The analysis relied heavily on the following tools:

* Excel: Used for initial data exploration, cleaning, and aggregation. Its simplicity and flexibility made it ideal for preparing the dataset for visualization.
* Tableau: Chosen for its ability to create clear and interactive visualizations. Tableau enabled us to identify trends and patterns that were not immediately evident in the raw data (Batt, S. et al. 2020).

Key Tableau visualizations included:

* Bar charts comparing state-wise profits.
* Line charts depicting sales and profit trends over time.
* Product-wise profitability comparisons within the Office Supplies category.
* maps Visualization: A geographical map was created in Tableau to visualize state-wise profitability. The map provided an intuitive way to compare states based on their total profits (Batt, S. et al. 2020).

## 2.5 Insights and Interpretation

The methodology was designed to yield actionable insights that would directly inform the company’s expansion strategy. By focusing on profitability and product performance, we were able to identify the following:

* Optimal Store Locations: The top five states—California, New York, Washington, Virginia, and Michigan—offer the highest profitability and customer demand, making them ideal for expansion.
* Profitable Product Categories: Within the Office Supplies category, subcategories such as Paper, Art, Labels, and Envelopes consistently performed well across all metrics.
* Customer Preferences: Corporate customers were identified as a key demographic, particularly for Office Supplies, indicating opportunities for targeted marketing and inventory planning.

## 2.6 Challenges and Limitations

While the methodology was comprehensive, a few challenges and limitations were encountered:

1. Historical Data Bias: The dataset reflects past trends, which may not fully predict future market conditions. To mitigate this, we focused on stable patterns that are less likely to fluctuate significantly over time (Carlisle, S. 2018).
2. Lack of External Variables: Factors such as local competition, real estate costs, or macroeconomic trends were not included in the dataset, which could affect the feasibility of expansion in certain states.
3. Aggregation Impact: While aggregating data simplifies analysis, it may obscure transaction-level details that could provide additional insights (Carlisle, S. 2018).

# 3. Data Analysis and Insights

Data analysis is the cornerstone of making informed decisions for business expansion. By carefully examining the Sample Superstore dataset, we derived actionable insights that pinpoint the most profitable states and product categories for opening new retail stores. This section provides a detailed breakdown of the analysis process and the key findings (Becker, L. T. and Gould, E. M. 2019).

## 3.1 Analysis of State-Level Profitability

The first step of our analysis focused on identifying states that consistently exhibited high profitability. Using aggregated profit data, we created bar charts in Tableau to compare the performance of all 49 states in the dataset. This visualization highlighted clear differences in profit margins across the country.

The top five states that emerged as the most profitable were:

California: With its large population and diverse economy, California consistently delivered the highest profits. It is a well-established market with robust demand across all product categories.

New York: As another densely populated state, New York demonstrated strong sales and profit margins, especially in urban areas where retail activity is concentrated.

Washington: This state showed consistent profitability, indicating a steady demand for products in our target categories.

Virginia: While not as dominant as California or New York, Virginia displayed promising growth trends and a stable profit margin, making it an emerging market.

Michigan: Despite its economic challenges in the past, Michigan exhibited substantial profit growth in certain product categories, making it a viable candidate for expansion.

## 3.2 Analysis of Product Categories and Subcategories

After identifying the most profitable states, the next focus was on determining which products would maximize profitability in the new stores. The dataset divides products into three main categories: Office Supplies, Furniture, and Technology. Within these, we analyzed subcategories to identify the top performers (Becker, L. T. and Gould, E. M. 2019).

Our analysis revealed that Office Supplies was the most profitable category, driven by four standout subcategories:

* Paper: Consistently high demand across all five states. This product appeals to both individual consumers and corporate clients, making it a reliable choice for inventory.
* Art: Frequently purchased by creative professionals, students, and schools, particularly in urban areas.
* Labels: Popular among corporate clients for organizational needs, indicating a strong B2B demand.
* Envelopes: A steady performer due to its wide usage in both personal and professional contexts.

The profitability of these subcategories is attributed to their low production costs and frequent purchase cycles, which generate stable revenue streams (Becker, L. T. and Gould, E. M. 2019).

## 3.3 Insights on Customer Segmentation

Customer segmentation further refined our understanding of demand patterns. By dividing customers into three groups—Corporate, Consumer, and Home Office—we identified distinct purchasing behaviors:

Corporate Customers: This segment accounted for a significant proportion of Office Supplies purchases, particularly Paper and Labels. These findings suggest that targeting corporate clients with tailored marketing strategies and bulk purchase discounts could drive higher sales.

Individual Consumers: Consumers showed a preference for Art supplies, which are often associated with educational or recreational use.

Home Office Segment: This group demonstrated moderate demand for products like Envelopes and Labels, driven by small business needs.

The segmentation analysis highlighted the potential for both business-to-business (B2B) and business-to-consumer (B2C) strategies, allowing the company to cater to diverse customer needs.

## 3.4 Time-Based Trends

A line chart analysis of sales and profit trends over time revealed seasonal spikes in demand:

Back-to-School Season: Sales of Office Supplies, particularly Paper and Art products, surged during the months leading up to the academic year.

Holiday Season: A noticeable increase in demand for general Office Supplies was observed during November and December, likely driven by year-end office purchases and holiday activities.

These trends suggest that inventory planning should align with these seasonal fluctuations to capitalize on peak demand periods.

## 3.5 Observations on Discounts and Profitability

While discounts can boost sales volumes, their impact on profitability is not always positive. A scatter plot analysis comparing discounts and profit margins showed that excessive discounts often led to reduced overall profits, particularly in non-essential categories like Furniture. However, moderate discounts on Office Supplies, especially for bulk purchases, proved effective in driving sales without significantly eroding profit margins.

## 3.6 Limitations in the Analysis

While the insights derived are robust, certain limitations should be noted:

Lack of Predictive Analysis: The dataset reflects historical trends, which may not fully account for future market dynamics, such as changes in consumer behavior or economic conditions.

External Factors: Variables like local competition, logistical costs, and state-specific regulations were not included in the dataset but could influence the feasibility of opening new stores in the identified locations.

## 3.7 Summary of Insights

From the analysis, several key takeaways emerged:

Optimal Locations: California, New York, Washington, Virginia, and Michigan are the most profitable states, making them ideal for the company’s expansion strategy.

Top Products: Within Office Supplies, Paper, Art, Labels, and Envelopes consistently generate high profits and should be the focus of inventory.

Target Customers: Corporate clients are a major driver of profitability, suggesting the potential for targeted B2B marketing efforts.

Seasonal Demand: Inventory planning should consider seasonal trends, particularly the back-to-school and holiday periods.

Discount Strategy: Discounts should be applied strategically to boost sales without compromising profit margins.

By combining geographical analysis, product performance evaluation, and customer segmentation, this analysis provides a comprehensive roadmap for data-driven business expansion. The recommendations derived from these insights will enable the company to maximize profitability and minimize risks as it enters new markets.

# 4. Recommendations and Conclusion

## 4.1 Store Locations

Based on our analysis, we recommend opening new stores in the following states:

1. California: High sales and profit potential, driven by a large and diverse customer base.
2. New York: Strong market stability and urban customer density.
3. Washington: Consistent profitability across product categories.
4. Virginia and Michigan: Emerging markets with untapped potential.

## 4.2 Product Focus

To maximize profitability, the new stores should prioritize the following products:

* Paper
* Art supplies
* Labels
* Envelopes

These products have demonstrated high demand and profitability, making them ideal for a focused marketing strategy.

## 4.3 Operational Strategies

* Inventory Management: Leverage data-driven tools to optimize stock levels and reduce waste.
* Targeted Marketing: Develop marketing campaigns tailored to corporate and individual customer segments.
* Performance Monitoring: Use dashboards to track sales and profitability in real-time, enabling quick adjustments as needed.

## 4.4 Conclusion

* Our analysis offers a practical and clear roadmap for the company’s expansion, built on reliable data and thoughtful insights. By focusing on the right states and products, the company is well-positioned to capitalize on current trends and build a strong foundation for long-term success.
* California, New York, Washington, Virginia, and Michigan stand out as the most promising locations for new stores. These states have demonstrated consistent profitability and strong demand for the company’s products, offering both stability and growth opportunities. Meanwhile, Office Supplies, particularly Paper, Art, Labels, and Envelopes, emerged as high-margin products with broad appeal to both corporate clients and individual consumers. These dual revenue streams make them an ideal focus for the expansion.

# References

Batt, S. et al. (2020) ‘Learning Tableau: A data visualization tool’, *The Journal of Economic Education*, 51(3–4), pp. 317–328. doi: 10.1080/00220485.2020.1804503.

Becker, L. T. and Gould, E. M. (2019) ‘Microsoft Power BI: Extending Excel to Manipulate, Analyze, and Visualize Diverse Data’, *Serials Review*, 45(3), pp. 184–188. doi: 10.1080/00987913.2019.1644891.

Carlisle, S. (2018) ‘Software: Tableau and Microsoft Power BI’, *Technology and Architecture + Design*, 2(2), pp. 256–259. doi: 10.1080/24751448.2018.1497381.

Murphy, S. A. (2013) ‘Data Visualization and Rapid Analytics: Applying Tableau Desktop to Support Library Decision-Making’, *Journal of Web Librarianship*, 7(4), pp. 465–476. doi: 10.1080/19322909.2013.825148.

Zwilling, M. (2022) ‘Big Data Challenges in Social Sciences: An NLP Analysis’, *Journal of Computer Information Systems*, 63(3), pp. 537–554. doi: 10.1080/08874417.2022.2085211.