Retail Sales and Supply Chain Analysis Using Business Intelligence Tools

Fatih Batuhan Özatar

5001230011

Faculty of Computer and Information Technologies

Istanbul University

May 25, 2025

Contents

1 Introduction

1.1 Project Aim and Scope

This project aims to analyze the "Retail Supply Chain Dataset" using BI and reporting tools. The main goal is to explore key aspects such as sales performance, customer behaviors, the efficiency of supply chain to provide actionable insights for strategic decision-making. This project covers data preparation, exploratory data analysis (EDA), and developing dynamic reports, and visualizations using Power BI.

1.2 Dataset Used

This analysis uses the "Retail Supply Chain Dataset" publicly presented in Kaggle. The data set covers data including the basic performance metrics like order details, customer information, product features and sales, amount, discount and profit. These data show a comprehensive base to examine sales patterns, customer segments, and supply chain operations.

1.3 Tools Utilized

This project utilized a combination of tools to perform the data analysis and visualization. Python, with its libraries Pandas for data manipulation and Matplotlib/Seaborn for initial visualizations, was employed for data preparation and exploratory data analysis (EDA). Microsoft Excel was used for basic data handling and preliminary visualizations as per project requirements. The primary Business Intelligence tool for developing interactive dashboards and visual analysis was Microsoft Power BI.

2 Data Preparation and Cleaning

2.1 Data Loading and Initial Inspection

The "Retail Supply Chain Sales Dataset" (CSV format) was loaded into a Pandas DataFrame using Python for initial processing. An initial inspection confirmed the dataset includes 9994 rows and 23 columns. A check for missing values revealed that the dataset has no missing entries in any of the columns, which simplified the subsequent cleaning phase.

2.2 Data Cleaning Steps

The data cleaning process was relatively straightforward due to the initial quality of the dataset. Key steps included:

- **Whitespace Trimming in Column Names:** Leading and trailing whitespaces were removed from column names to ensure consistency and prevent potential errors in data referencing.

- **Duplicate Row Check:** The dataset was checked for duplicate entries. No duplicate rows were found, indicating unique records for each transaction.

2.3 Data Transformation and Feature Engineering

Several data transformations were applied to prepare the dataset for analysis:

- The 'Order Date' and 'Ship Date' columns were converted to datetime format for time-based analysis.

- New date-related columns ('Order Year', 'Order Month', 'Order Quarter', 'Order Day Name') were created from 'Order Date'.

- A 'ReturnedIndicator' column (0 for 'Not' returned, 1 for 'Yes' returned) was made from the 'Returned' column.

- A 'Shipping Time (Days)' column was calculated from 'Order Date' and 'Ship Date'.

- The 'Postal Code' column was changed to a string type.

These steps improved the dataset for BI tool analysis. The 'Profit Margin' column was created using DAX in Power BI.

3 Exploratory Data Analysis (EDA) and Key Findings

Exploratory Data Analysis (EDA) was performed on the prepared dataset to uncover initial patterns, trends, and relationships. Key numerical variables such as 'Sales' and 'Profit' exhibited right-skewed distributions, indicating that a smaller number of transactions contribute disproportionately to higher values; median 'Sales' was $54.49 while the mean was $229.86, and median 'Profit' was $8.67 against a mean of $28.66. 'Profit' also showed a wide range, including significant losses (minimum approx. -$6,600). 'Discount' rates varied up to 80%, with a median of 20%. 'Shipping Time (Days)' was also right-skewed, with a median of 4 days but an average of 34.6 days, suggesting some lengthy fulfillment times.

Analysis of categorical variables showed 'Standard Class' as the most common shipping mode (60%) and 'Consumer' as the largest customer segment (52%). Geographically, the 'West' region had the most transactions. 'Office Supplies' was the dominant product category (60%), with 'Binders' and 'Paper' being frequent sub-categories. Approximately 8% of orders were marked as returned. Order frequency peaked on Mondays.

Initial correlation analysis highlighted a notable negative relationship between 'Discount' and 'Profit'. Conversely, 'Sales' and 'Profit' showed a moderate positive correlation. Time series visualization of sales indicated potential seasonal patterns, with variations across months and quarters. These initial findings from EDA guided the subsequent development of interactive dashboards in Power BI for deeper investigation.

4 Insights from Power BI Dashboards and Support for Business Decisions

4.1 Executive Summary: Overall Performance Dashboard

The "Overall Performance" dashboard gives a quick look at key business numbers. Total Sales were about $2.3 Million, and Total Profit was around $286,000. The Average Profit Margin was 12%. There were 5,009 total orders, and 8% of these were returns. Orders took an average of 34.61 days to ship. These numbers show the company's current general status.

The sales and profit trend chart shows that sales and profit generally increased over the years, with some ups and downs. The forecast for the next two years also shows a similar pattern with slight growth. For product categories, 'Technology' had the most sales, then 'Furniture', and then 'Office Supplies'. The sales amounts for these categories were quite similar.

Regional performance showed big differences. The 'West' region had the highest sales ($725K) and profit ($108K). It also had the fastest shipping (32.59 days) but the most returns (15%). The 'East' and 'South' regions had good profit margins (17% and 16%) and low return rates. The 'Central' region was notable because it made a loss, with a -10% profit margin, despite good sales ($501K). It also had the longest shipping times (36.98 days). The 'Consumer' segment made up most of the sales.

4.2 Product Performance & Profitability Analysis Dashboard

The "Product Performance & Profitability Analysis" dashboard details how different products impact the company's financial results. This view helps identify which products and categories contribute most to profits or cause losses.

The ranking of sub-categories by total sales reveals that 'Phones' is leader in sales, followed by 'Chairs', 'Storage', and 'Binders'. Conversely, sub-categories like 'Fasteners', 'Labels', 'Art', and 'Envelopes' show very low sales volumes. This highlights the key revenue drivers for the company.

Analysis of average profit margins by sub-category presents an interesting contrast. Sub-categories with the lowest sales volumes, such as 'Labels', 'Paper', 'Envelopes', 'Fasteners', and 'Art', exhibit some of the highest average profit margins. This suggests that while these items do not sell in large quantities, each sale is relatively profitable. This situation requires checking if sales of these high-margin items can be increased.

Critically, several sub-categories consistently show negative average profit margins. Specifically, 'Tables', 'Binders', 'Bookcases', and 'Machines' operate at a loss on average. This indicates potential issues with pricing, high costs, or excessive discounts for these items, requiring urgent review and strategic adjustments.

The scatter plot shows that as discount increases, profit margin usually drops. This is important for smarter discount strategies. A detailed matrix helps check sales, quantity, discount, and profit per product, and filters let us explore this data by year, region, and segment.

4.3 Customer Insights & Operational Efficiency Dashboard

The "Customer Insights Operational Efficiency" dashboard provides a closer look at customer behavior and key aspects of supply chain performance.

Customer segment analysis showed different patterns across regions. For example, in the 'West' region, the 'Consumer' segment brings the most profit, even though their average order value and order frequency per customer are the lowest among segments in that region. Conversely, 'Home Office' customers in the West order most frequently but are the least profitable. In the 'Central' region, the 'Corporate' segment was the most valuable, showing high frequency, order value, and profit, while the 'Consumer' segment was least profitable. The 'East' region's 'Consumer' segment showed good profitability with higher order frequency and value compared to other segments in that region. These regional differences in segment performance shows that customer strategies may need to be considered by region.

The decomposition tree analysis of total sales showed that 'Technology' and 'Furniture' are the leading main categories, closely followed by 'Office Supplies'. Within 'Technology', 'Phones' are by far the top-selling sub-category. For 'Furniture', 'Chairs' lead in sales, and in 'Office Supplies', 'Storage' and 'Binders' are prominent. This helps identify key product drivers within each category.

Regarding shipping operations, 'Standard Class' is prominently the most used ship mode across all regions. However, its average shipping time is also generally the highest. The 'Same Day' ship mode has very low usage, particularly in the 'South' and 'Central' regions. The trend analysis of average shipping times showed regional variations: shipping times decreased from 2014 to 2015/2016 in most regions but then started to increase again in 2017, with a notable exception in the 'Central' region where shipping times began to decrease rapidly after 2016. This highlights inconsistencies and areas for operational review.

The return rate matrix (Region vs. Category) indicated that while overall return rates are

manageable, there are specific problem areas. The 'Furniture' category, particularly in the 'Central' region, shows a significantly higher return rate (appearing as dark red/orange on the heatmap) compared to other category/region combinations. 'Technology' and 'Office Supplies' generally have lower return rates across all regions. This shows that furniture or delivery problems, especially in the Central region, should be investigated to lower costs and improve customer satisfaction.

5 Sales and Marketing Recommendations

5.1 Fixing Discount Policies to Increase Profit

Our analysis shows that high discounts, especially above 20-30%, often lower profit margins or cause losses. The 'Furniture' category, like 'Tables' and 'Bookcases', is very sensitive to this.

- **What to Do:**
  - Give smaller discounts for sub-categories that often lose money. For example, set a top discount of 20% for 'Furniture' items.
  - Offer promotions for product bundles or specific customer groups. This is better than big general discounts and can protect profit margins.
  - Watch closely how any discount changes affect sales numbers and profit margins.
- **What We Expect:** Fewer sales that lose money. This should increase the overall profit margin and total profit. Sales numbers might decrease in some areas, but overall profit should be better.

5.2 Improving Regional Plans and How Things Work

The 'West' region has the most sales and profit but also a high return rate (15%). The 'Central' region loses money on average (-10% profit margin) even with good sales. It also has long shipping times, though this got better after 2016. The 'East' and 'South' regions are more balanced, make good profit, and have low return rates.

- **What to Do:**
  - **For 'Central' Region:** Find out why it's losing money. Check prices, local costs, and the products sold there. Use the ideas that helped improve shipping times after 2016.
  - **For 'West' Region:** Understand why the return rate is high (15%). Do things to lower returns, like writing better product details or checking product quality.
  - **For 'East' and 'South' Regions:** See what makes these regions do well. Try to use these good ideas in other regions.
  - **General:** Work to make shipping times shorter in all regions where they are long. Use good ideas from regions or times that had better shipping.
- **What We Expect:** More profit in regions like 'Central'. Lower costs from returns, especially in the 'West'. Things working better overall, and happier customers.

5.3 Using Segment-Specific Marketing and Sales Plans

Customer segments act differently and bring in different profits in each region. For example, the 'Corporate' segment is very valuable in the 'Central' region. The 'Consumer' segment brings

most profit in the 'West' and 'East'. The 'Home Office' segment in the 'West' orders often but does not make much profit.

- **What to Do:**
  - Create special marketing for the most profitable segments in each region. For example, focus on business-to-business plans for the 'Corporate' segment in the 'Central' region.
  - For segments that buy a lot but have low average order values (like 'Consumers' in some areas), find ways to make them spend more each time (e.g., sell product groups, suggest extra items).
  - For segments that order often but bring low profit (like 'Home Office' in the 'West'), check what products they buy and their profit margins. Try to increase profit with special deals or price changes.
- **What We Expect:** More sales and profit from important customer segments. Customers staying longer because their specific needs are met. Better use of marketing money.

5.4 Implement Segment-Specific Marketing and Sales Strategies

Customer segments show different behaviors and profitability across regions. For instance, the 'Corporate' segment is highly valuable in the 'Central' region, while the 'Consumer' segment drives most profit in the 'West' and 'East'. The 'Home Office' segment in the 'West' orders frequently but is not very profitable.

- **Actionable Steps:**
  - Develop targeted marketing campaigns for the most profitable segments in each region.
  - For high-volume, lower average order value segments (like 'Consumers' in some areas), explore strategies to increase average order value.
  - For segments with frequent orders but low profit (like 'Home Office' in the 'West'), analyze product choices and margins to improve profitability through specific promotions or pricing.
- **Expected Outcome:** Increased sales and profitability from key customer segments. Better customer retention by addressing specific segment needs. More efficient use of marketing resources.

6 Conclusion

This project analyzed retail data using Python and Power BI. The goal was to find useful insights on sales, customers, and supply chain operations.

Key findings showed that high discounts reduce profit margins, especially for furniture. Some products like 'Tables' and 'Bookcases' often lose money. Regional performance also differed: the 'Central' region made losses, while the 'West' had high sales but also many returns. Customer segments and shipping operations also showed areas for improvement.

The Power BI dashboards clearly showed these trends and issues. This analysis provides valuable information to help the business make better strategic decisions and improve its performance.