Introduction

In today's highly competitive business environment, data has become a strategic asset for organizations seeking to improve performance and sustain growth. Advances in data collection, storage, and analytics have enabled businesses to transform large volumes of operational data into actionable insights that support informed decision-making. Organizations that effectively leverage analytics are better positioned to optimize operations, understand customer behavior, and drive profitability.

Retail businesses, in particular, operate in data-rich environments where insights derived from sales and customer data can directly influence pricing strategies, inventory management, and marketing effectiveness. This report presents an exploratory data analysis (EDA) of Superstore sales data, a fictitious retail dataset designed to closely reflect real-world retail operations. The analysis focuses on identifying key trends, performance drivers, and profitability patterns to support strategic planning and enhance business decision-making.

This analysis seeks to address the following key business questions:

Sales Performance: What are the overall sales trends over time, and are there notable fluctuations that require management attention?

Product Category Performance: Which product categories contribute most to total sales, and which categories are underperforming?

Geographic Insights: How does sales performance vary across regions, and which markets present growth opportunities or require targeted marketing efforts?

Profitability: Which products generate the highest profitability, which are underperforming, and what factors influence overall profit performance during the period analysed?

In addition to answering these questions, the analysis aims to uncover additional insights that may inform strategic and operational decisions. The ultimate objective is to translate data-driven findings into actionable recommendations that support improved decision-making and enhance overall business performance.
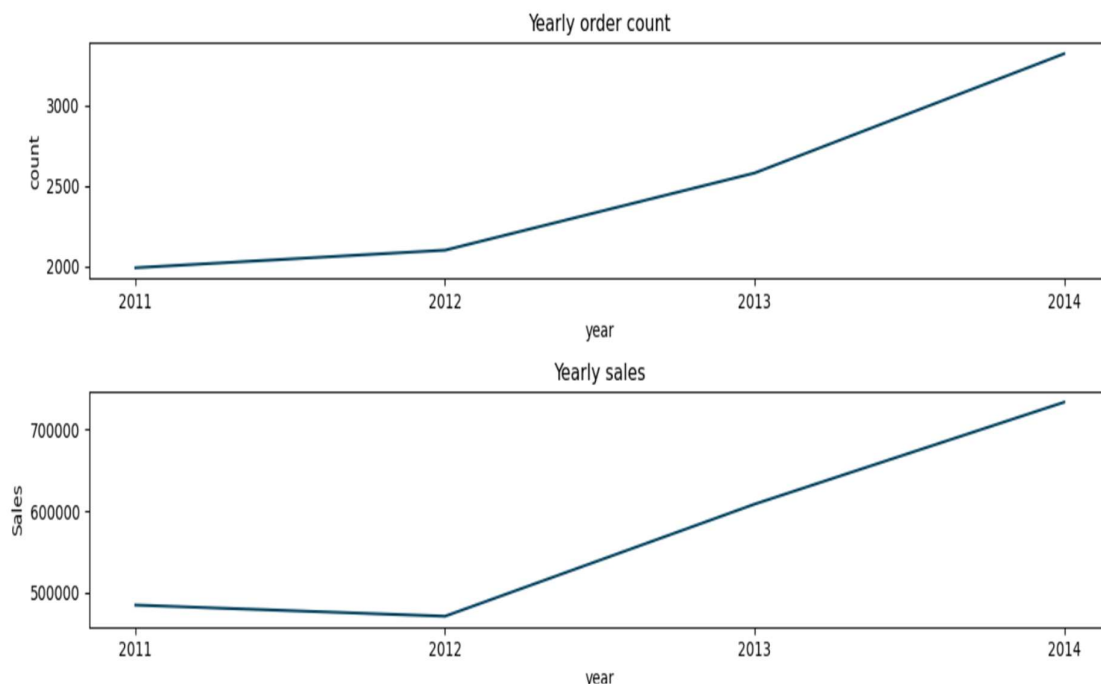
The dataset contains 9994 records (rows) and 21 features (columns). Among the features, 2 have datetime data type (date), 3 are floating point (decimals), 3 are integers (whole numbers), and 13 are object (strings) data types. It also has no missing values.
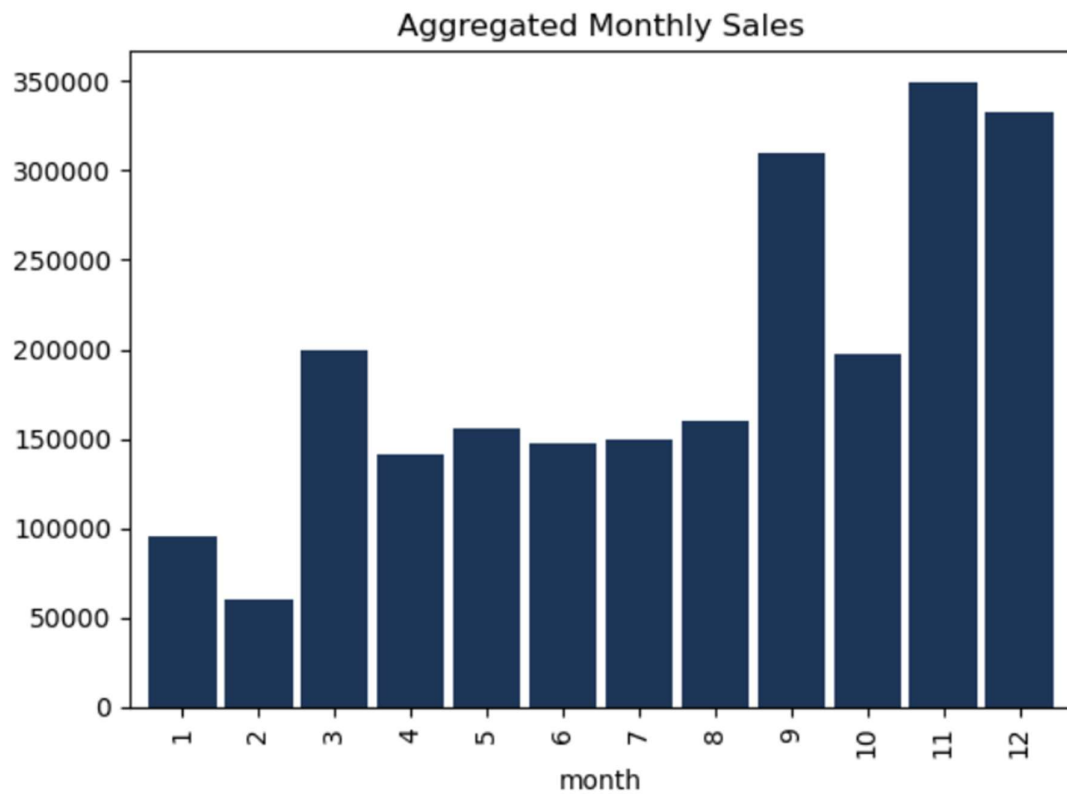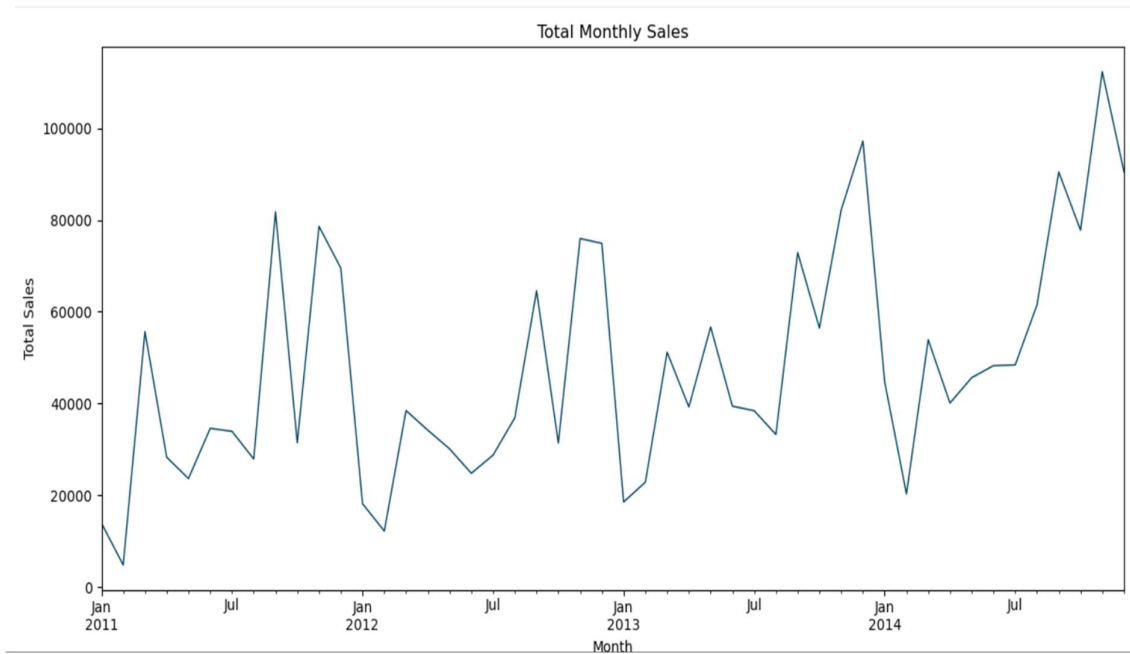
The dataset is now ready for analysis. In real-world business scenarios, data cleaning and transformation are essential steps to ensure the accuracy and reliability of analytical results. These processes typically include handling missing values, converting data to appropriate data types, standardizing or normalizing variables, performing feature engineering, and transforming date-time and string fields.

For this dataset, only minimal data cleaning was required, primarily involving data type conversions and feature creation. With these necessary transformations completed, the dataset is now well-prepared for exploratory data analysis and subsequent business insights.

The dataset contains sales data from 2011-01-04 to 2014-12-31. Earliest Ship Date information was in 2011-01-08, while the latest was in 2015-01-06. No apparent errors or anomalies can be observed with the Sales, Quantity, and Discount columns. With net_profit, (net)_profit_before_discount, and net_profit_per_unit_sold, there are negative values as the min, which may mean actual negative profit or possible error. This requires further investigation.
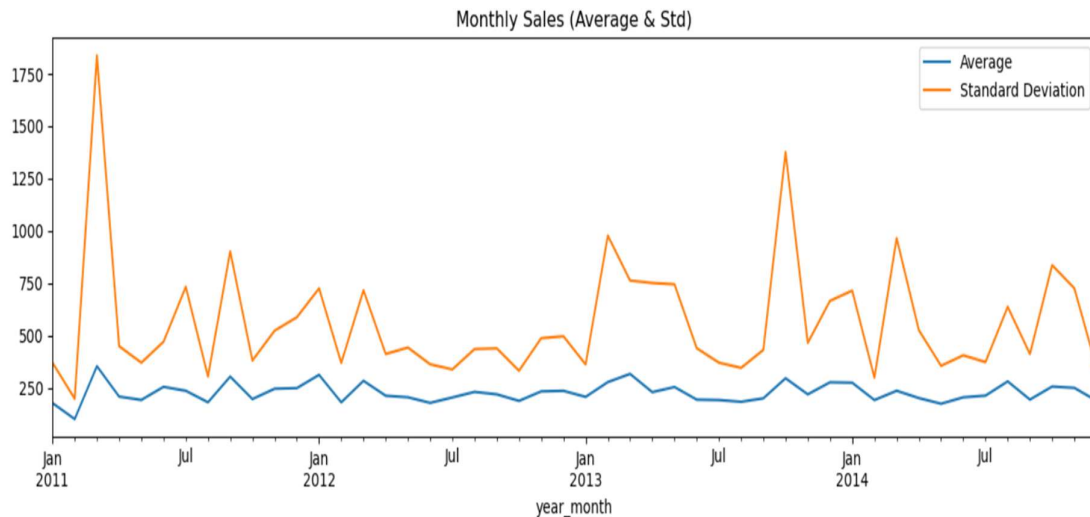
What are the overall sales trends, and how have they evolved over time? Are there any significant fluctuations that need to be addressed?

Yearly order count

Yearly sales

Total Monthly Sales

Aggregated Monthly Sales

The visualization above illustrates total monthly sales aggregated over a four-year period. Overall, sales volumes are higher during the holiday season, reflecting increased consumer spending toward the end of the year. Additionally, the academic calendar in the United States typically begins in late August or early September, which may help explain the noticeable
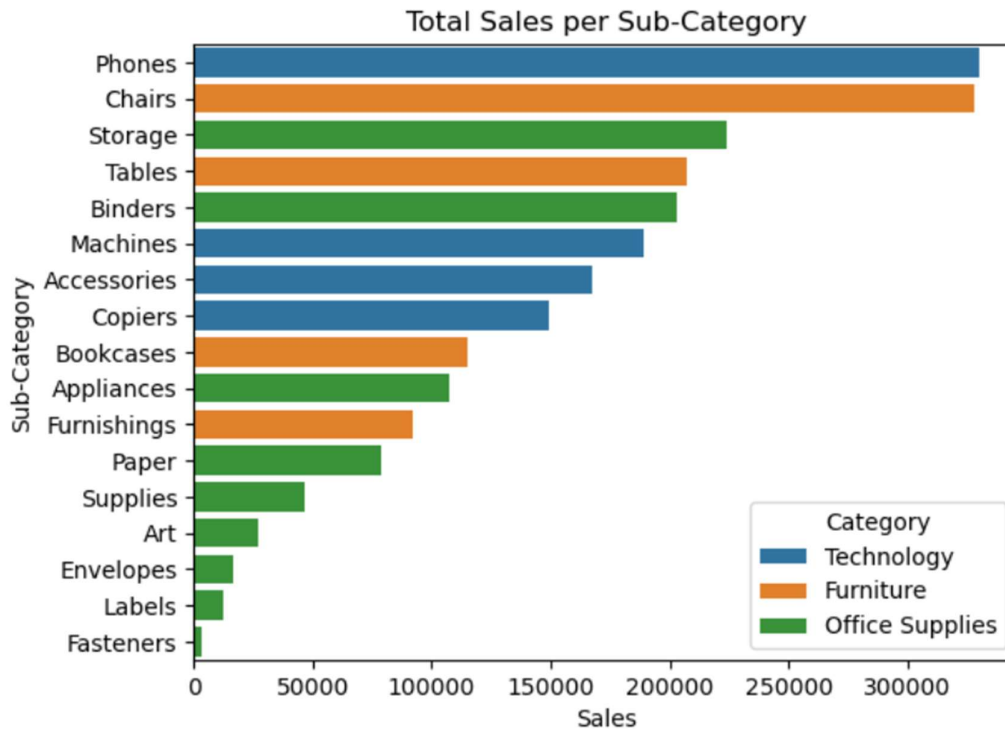
increase in sales during September. This rise is likely driven by demand for school-related products such as binders, paper, home and office supplies, bookcases, and related accessories.

Key findings:

- Substantial intra-month sales variability is evident across the observed period, as quantified by the monthly sales standard deviation. This variability is not random but exhibits recurring temporal patterns. Specifically, sales volatility is elevated in March and again during September and October, suggesting periods of heightened demand uncertainty or promotional activity.

- In contrast, from April 2012 through year-end, sales display relatively low variability, indicating a more stable revenue environment. This period aligns with a modest decline in overall annual sales performance, as total yearly sales decreased slightly from 2011 to 2012. Conversely, higher sales volatility is observed in 2011, 2013, and 2014, reflecting greater fluctuations in customer demand and sales performance during those years.

- Overall, retail sales variability is influenced by multiple structural and market-driven factors, including seasonality, shifts in consumer purchasing behavior, promotional cycles, and competitive dynamics, all of which have direct implications for demand forecasting, inventory planning, and revenue management.
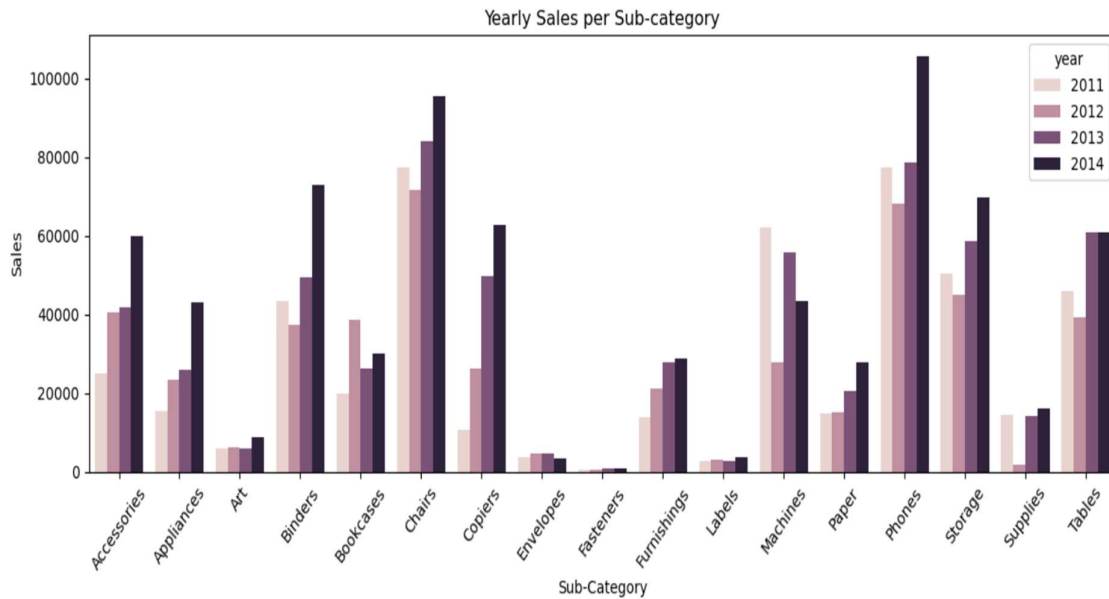
Which product categories contributed the most to the company's sales? Which categories are underperforming, if any?

Total Sales per Sub-Category

The visualization provides an aggregate view of sales magnitude across product sub-categories over the four-year period from 2011 to 2014. Within the Technology category, Phones are the highest revenue-generating sub-category. Similarly, Chairs dominate sales within the Furniture category, while Storage leads in the Office Supplies category. Collectively, Phones, Chairs, and Storage consistently rank as the top three revenue-generating sub-categories throughout the period, followed by Tables, Binders, and Machines.

Conversely, Copiers represent the lowest-performing sub-category within Technology, while Furnishings and Fasteners generate the least sales within the Furniture and Office Supplies categories, respectively.
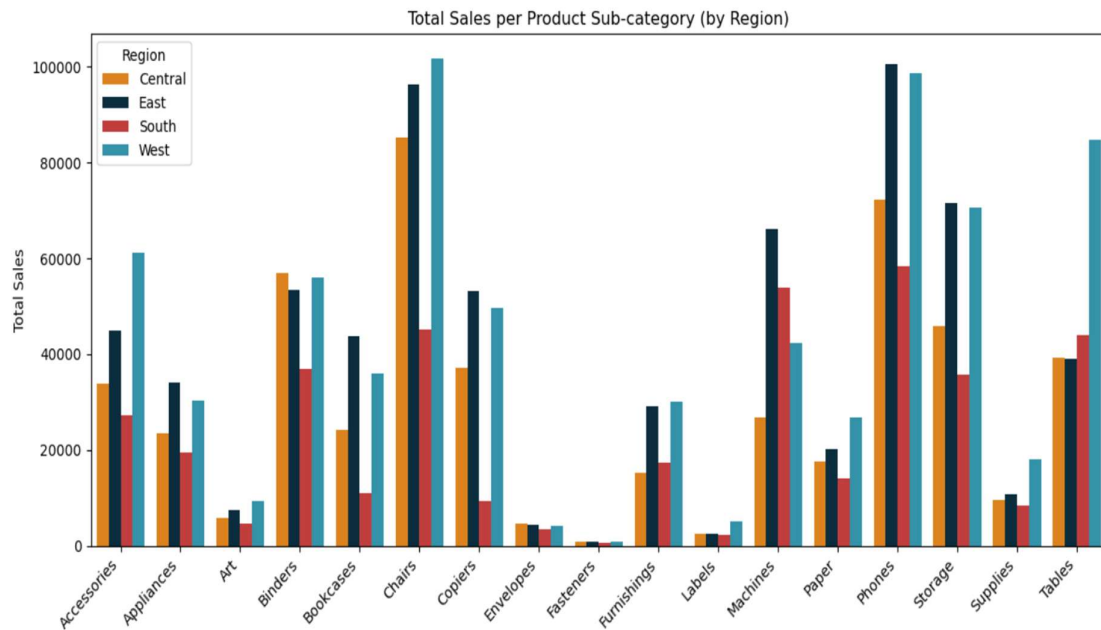
Notably, the sub-categories with the highest sales—Phones and Chairs—belong to the Technology and Furniture categories, which typically comprise higher-priced items. This suggests that overall revenue is significantly driven by high-ticket products, highlighting the importance of pricing strategy, demand management, and inventory optimization within these categories.

Yearly Sales per Sub-category

Based on average annual sales growth rates, Envelope products exhibited the slowest growth over the period. In contrast, Supplies recorded the highest growth, achieving an average annual growth rate (AAGR) of 185%. This was followed by Copiers and Appliances, which posted strong growth rates of 86% and 43% AAGR, respectively, indicating rapid expansion within these sub-categories.

Key Findings

- Sales Contribution by Sub-Category: Over the four-year period, Phones generated the highest sales within the Technology category, Chairs within Furniture, and Storage within Office Supplies. These leading sub-categories were followed by Tables, Binders, and Machines. Conversely, Copiers, Furnishings, and Fasteners were the lowest revenue-generating sub-categories within the Technology, Furniture, and Office Supplies categories, respectively.

- Year-over-Year Sales Variability: Annual sales performance varied across all product sub-categories, with no consistent growth pattern observed over time. For Phones, Binders, Appliances, and Accessories, the highest sales growth occurred in 2014, whereas Copiers, Machines, and Tables experienced their peak growth in 2013.

- Average Annual Growth Performance: Based on average annual sales growth rates, Supplies, Copiers, and Appliances emerged as the fastest-growing sub-categories. In contrast, Envelopes, Chairs, and Machines recorded the slowest average growth, indicating relatively mature or stagnant demand in these segments.
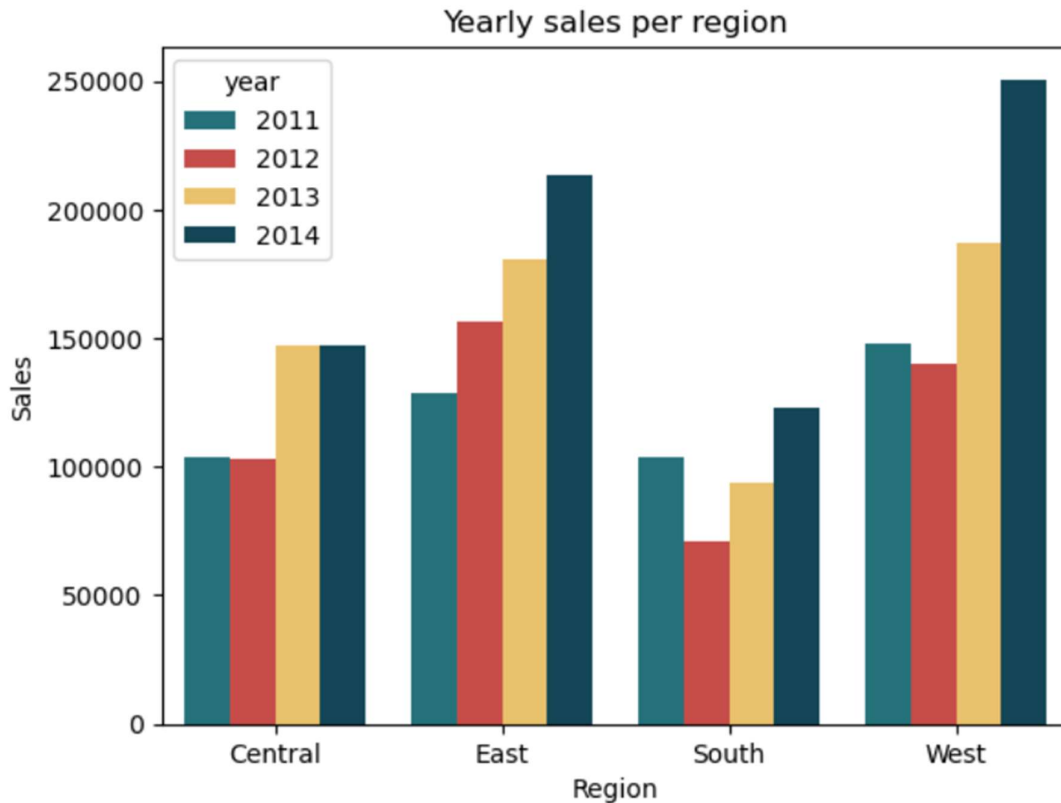
Total Sales per Product Sub-category (by Region)

Sales across most sub-categories are notably lower in the South and consistently higher in the West. For several products, this disparity is particularly pronounced. For example, sales of chairs and copiers in the South are significantly lower compared to all other regions. In contrast, machines and tables exhibit higher sales in the South than in other regions, making them clear exceptions to the general regional trend.

For most sub-categories, sales in the Central region are only marginally higher than those in the South, suggesting relatively similar demand levels between these two regions. Another prominent pattern is that certain product sub-categories record substantially higher sales in the West compared to all remaining regions. These include tables, office supplies, and technology accessories, indicating strong demand concentration in the West.

Although the dataset is fictional and does not provide contextual information about the regions, sales patterns allow for reasonable hypotheses. The higher sales observed in the West and East may suggest a greater presence of offices, corporate environments, or business districts in these regions. Supporting this assumption, sales of office-related products—such as storage, binders, and appliances—are significantly lower in the South and Central regions, while demand for these items is considerably higher in the West and East. Despite this trend, it is worth noting that the South records the second-highest sales for machines, indicating a possible regional specialization or localized demand for this product category.

It is important to emphasize that these figures represent total aggregated sales from 2011 to 2014 and do not capture year-over-year fluctuations or temporal trends. Consequently, this visualization serves as a high-level comparison of sales magnitude across regions and product sub-categories rather than an analysis of sales performance over time.

Yearly sales per region

The visualization illustrates how regional sales evolved over time. In the Central region, sales remained relatively stable initially before experiencing a sharp upward trend beginning in 2013. The East region demonstrated consistent year-over-year growth throughout the entire period, indicating steady demand and resilience to broader market fluctuations.

In contrast, the South region experienced a decline in sales in 2012, followed by a clear recovery and sustained positive growth in subsequent years. A similar pattern is observed in the West, where sales also dipped in 2012 but rebounded strongly thereafter, maintaining positive growth for the remainder of the period.

Overall, a slight decline in total sales is evident in 2012. From the regional trends, it can be inferred that the South contributed most significantly to this downturn, followed by the West and then the Central region. Notably, despite the overall dip, the East continued to grow positively in 2012, highlighting its relative stability compared to the other regions.

Using the Sales Average Annual Growth Rate (AAGR) of each region for the 4-year period, we see that the West had been the fastest growing Region in terms of sales, followed by the East, and then the Central region, and then South. Fruther investigation can be done to understand possible factors that may affect the differences which include economic factor, market conditions (saturation), consumer preference, among others.

Average Annual Growth Rate (AAGR) is calculated by getting the arithmetic mean of the yearly growth rates
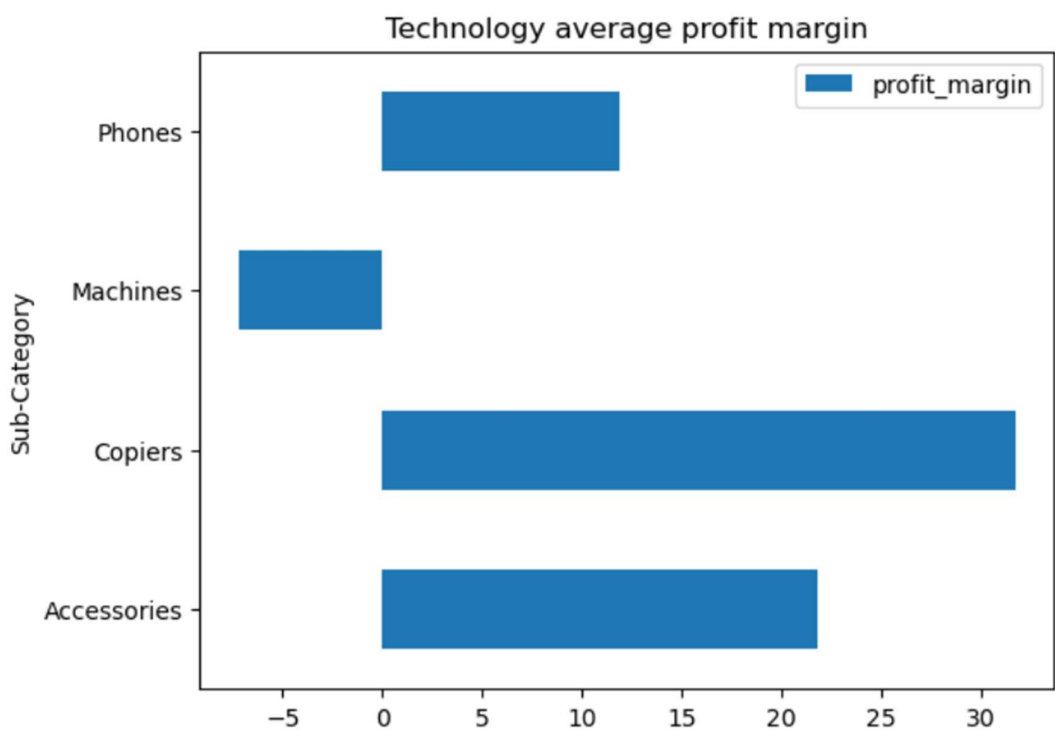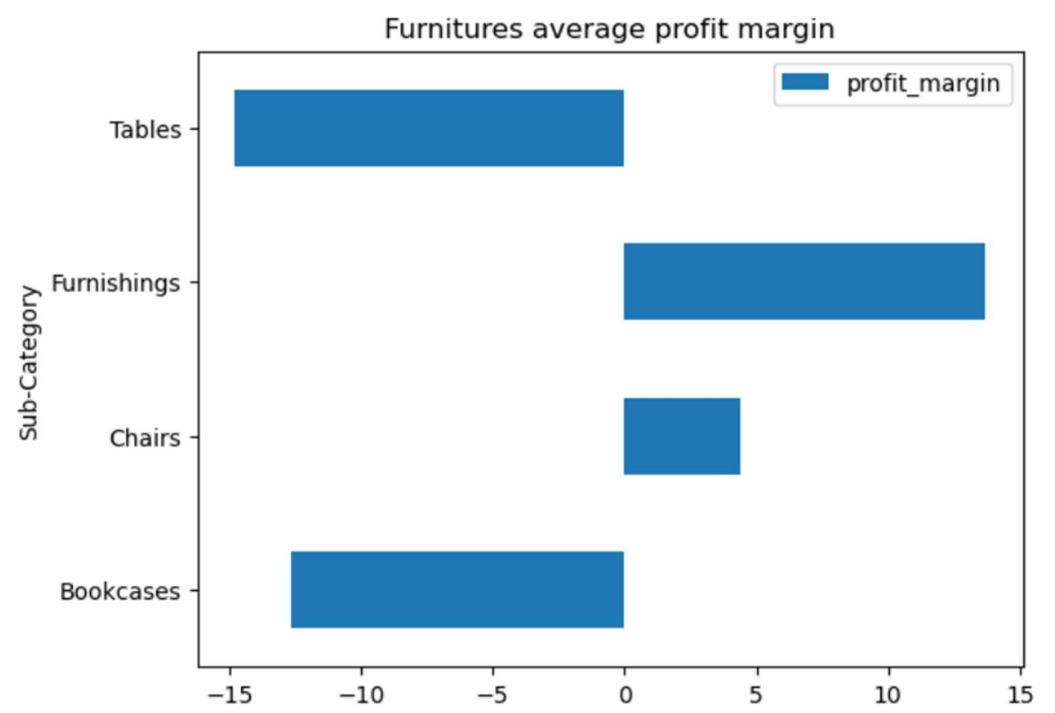
Key Findings:

- A clear seasonal pattern is evident across all regions. Sales generally increase toward the end of the year, with pronounced peaks in November and December. A notable surge also occurs in September, which may be associated with the start of the academic year. This interpretation is supported by increased sales of school-related products and a rise in sales within the consumer segment during this month.

- Across most months, the West consistently records the highest sales, followed by the East, while the Central and South regions trail behind. Sales in the South remain the lowest for most of the year, with one notable exception in March 2011, when South sales were more than three times higher than the next-best performing region, indicating a possible one-time event or anomaly.

- In terms of product sub-categories, sales are generally lower in the South and higher in the West. Specifically, office supplies and technology products show stronger performance in the West and East, whereas tables and machines record higher sales in the South. The Central region exhibits relatively stable and consistent sales across most sub-categories. For certain office supply items—such as art supplies, envelopes, fasteners, and labels—sales levels are broadly comparable across all regions, suggesting uniform demand.

- Regarding growth trends, all regions experienced negative growth in 2012, with the exception of the Central region, which maintained positive growth during that year. Based on Average Annual Growth Rate (AAGR), the West emerges as the fastest-growing region, followed by the East, then the Central, and finally the South. While the Central region demonstrates moderate and stable growth, its performance lags behind the stronger expansion observed in the West and East.
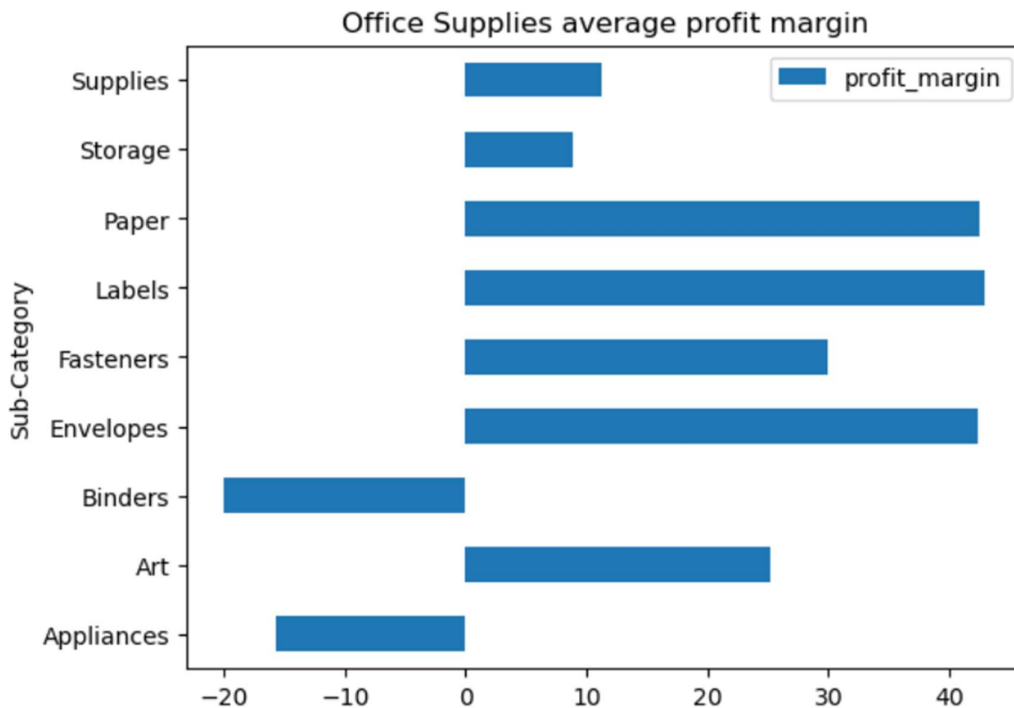
Which products are more profitable and which were not? With the available data, what factors affected the company's profit? How is the company's profitability during the period?

Using profit margin as the primary measure of profitability, the company recorded its lowest performance in 2011, with a margin of 10.23%. Profitability improved markedly in 2012, rising to 13.10%, and continued to increase in 2013, when the company achieved its highest profit margin of 13.43%. However, this upward trend did not persist. In 2014, the company experienced a slowdown, with profit margin declining to 12.74%, falling below the 2012 level though remaining higher than in 2011.

Although profit margin is a useful indicator of financial performance, it does not provide a complete picture on its own. Ideally, a more comprehensive analysis would incorporate additional financial metrics such as Cost of Goods Sold (COGS), operating income, and shareholders' equity. However, because the dataset primarily contains sales-related

information and lacks detailed cost and financial structure data, profit margin serves as the most appropriate and consistent metric for evaluating profitability within the scope of this analysis.

Furnitures average profit margin

Technology average profit margin
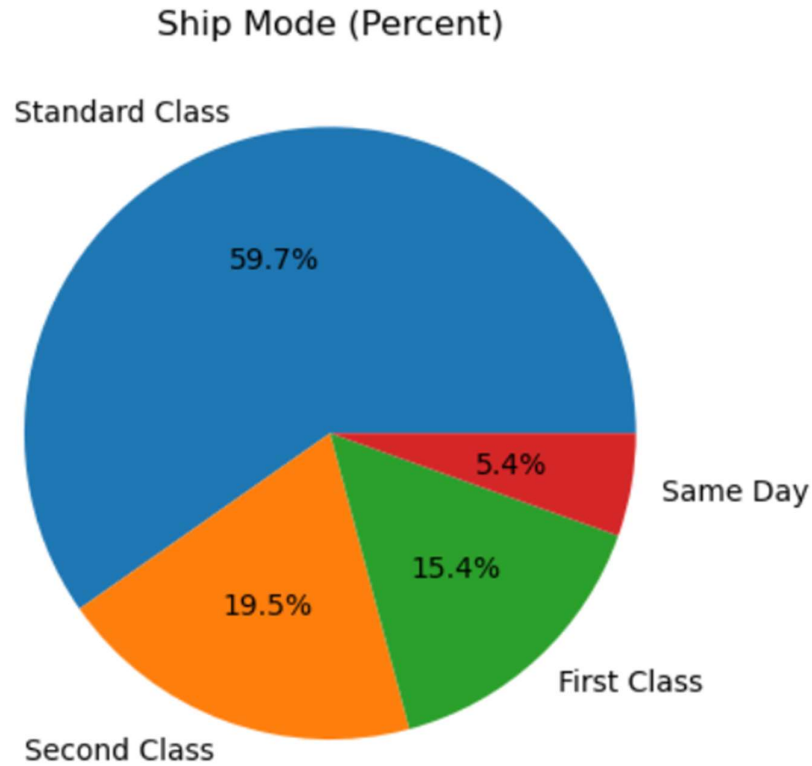
Office Supplies average profit margin

By product sub-category, phones, chairs, and binders generated the highest total profit within the technology, furniture, and office supplies categories, respectively. These three sub-categories also emerged as the top profit contributors overall, highlighting their central role in the company's profitability.

At the opposite end, machines, bookcases, and fasteners recorded the lowest total profit within the technology, furniture, and office supplies categories, respectively. Among all sub-categories, fasteners generated the least profit, with a total of only $949, indicating minimal contribution to overall earnings.

When accounting for discounts, profitability declines substantially across nearly all product sub-categories. Most notably, tables experienced a dramatic reversal, with gross profit falling from 17,725 after discounts—a 167% decrease (1.67×). Bookcases and supplies also shifted from positive to negative profit after discounts were applied. As shown in the accompanying table, profit declined across all sub-categories, with tables showing the largest percentage drop, while labels were the least affected, with an approximate 10% decrease.

These results suggest a need for a comprehensive review of the company's discounting and promotional strategies, particularly for products where aggressive discounts were not offset by sufficient increases in sales volume.

Finally, it is important to note that the values presented in the visualization represent aggregated totals. As such, they may give the misleading impression that all orders were profitable prior to discounts. In reality, some individual orders may have incurred losses, which were offset by higher profits from other transactions within the same sub-category.

Ship Mode (Percent)

Key Findings

Summarized below are the key findings from this analysis. Throughout the 4-year period from 2011 to 2014:

Sales Performance

Superstore sales increased yearly, with the fastest growth in 2013 and the slowest in 2012.

Seasonal sales trends was observed, notably in November, December, and September.

Sales exhibited high variability, particularly in March, September, and October.

Product Categories

Phones, chairs, and storage products led in sales by category, while copiers, furnishings, and fasteners performed least.

No clear sales pattern emerged for product sub-categories.

Supplies (technology), copiers, and appliances experienced the highest average annual sales growth, while envelopes, chairs, and machine products grew the slowest.

Profitability

The company maintained a profit margin above 10%, decreasing slightly in 2014.

Furnishings, copiers, and labels were the most profitable sub-categories, while chairs, phones, and storage products were the least profitable.

Phones, chairs, and binders generated the highest total profit, while machines, bookcases, and fasteners generated the least.

Discounts significantly affected profits, with tables, office supplies, and bookcases experiencing the largest drops.

Chairs were highly profitable in the furniture category, with copiers showing impressive profit growth. Machine products, on the other hand, had stagnant growth.