Analysis of Sales Trends and Inventory Management in Superstore Sales

Anjad Mshrqi

Tech for Job Program by Correlation One

Profile Project [EDA]

I. Introduction

A. Overview of the Superstore Dataset

The Superstore dataset, a rich and detailed collection of transactional data from the fictional retail company "Superstore," offers an in-depth look at orders, customers, products, and geographical locations. This comprehensive dataset serves as a robust foundation for analyzing sales patterns and trends. It includes information on sales transactions, customer demographics, product details, and regional sales distributions, making it an invaluable resource for in-depth sales analysis and insights.

II. Problem Statement

The retail landscape is shaped by dynamic market forces, including heightened competition and evolving consumer preferences. This necessitates a proactive approach to sales and inventory management for businesses to thrive. This project delves into historical sales data from a superstore to identify key trends, patterns, and relationships that can inform strategic business decisions. By conducting a thorough exploratory data analysis (EDA), we aim to:

- **Uncover Seasonal Sales Fluctuations and Underlying Trends**: Identify recurring patterns in sales data that align with specific times of the year and understand the factors driving these fluctuations.
- **Analyze Customer Behavior and Segmentation**: Understand customer preferences, identify key customer segments, and tailor marketing strategies accordingly.
- **Optimize Product Performance**: Identify top-selling and underperforming products, and analyze the impact of product categories and sub-categories on overall sales and profitability.
- **Improve Inventory Management**: Optimize stock levels to minimize stockouts and overstock situations, and improve inventory turnover.
- **Enhance Marketing Strategies**: Identify opportunities to improve marketing campaigns, such as targeted promotions and customer retention programs.
- **Understand Regional and Local Market Performance**: Analyze sales performance across different regions, cities, and even postal codes to identify areas of high and low demand.
- **Identify Potential Areas for Market Expansion**: Analyze sales data to identify regions or customer segments with high growth potential.
- **Evaluate the Impact of Shipping Modes**: Analyze the impact of different shipping modes (e.g., Standard Class, Second Class, First Class, Same Day) on delivery times, customer satisfaction, and overall costs.
- **Understand Product Category and Sub-Category Performance**: Analyze sales trends and profitability across different product categories and sub-categories to identify areas of strength and weakness.

III. Key Questions and Aspects to Focus on During EDA:

- **Seasonal Patterns**:
  - Are there recurring patterns or trends in sales data that align with specific times of the year (e.g., holidays, back-to-school season)?
- **Sales Trends**:
  - How do sales trends evolve over time? Are there upward or downward trends that need to be accounted for?
- **Customer Segmentation**:
  - How do customer demographics and purchasing behavior vary across different segments?
- **Product Performance**:
  - Which products are the most profitable?
  - Which product categories and sub-categories drive the highest sales volume?
- **Inventory Management**:
  - How often do stockouts and overstock situations occur?
  - What is the impact on sales and profitability?
- **Geographic Performance**:
  - What are the top-performing regions and cities?
  - Are there any regional disparities in sales?
  - How does sales performance vary across different postal codes?
- **Shipping Mode Analysis**:
  - What is the impact of different shipping modes on delivery times and customer satisfaction?
  - Are there cost-effective alternatives to current shipping practices?
- **Category and Sub-Category Analysis**:
  - Which product categories and sub-categories are the most profitable?
  - Are there any opportunities to cross-sell or up-sell products within specific categories?

IV. Business Impact

The insights gleaned from exploring the Superstore sales data have the potential to significantly impact business operations and drive strategic decision-making. By analyzing historical sales trends, seasonal fluctuations, customer behavior, and regional variations, this study aims to:

The analysis of the Superstore Sales Dataset could help businesses in several ways:

- **Optimize Inventory Management**:
  - Minimize stockouts by accurately forecasting demand, especially during peak seasons.
  - Reduce overstock by identifying slow-moving inventory and optimizing stock levels.

    - Improve inventory turnover by aligning inventory levels with actual demand.
- **Enhance Marketing Strategies**:
    - Target marketing campaigns to specific customer segments based on their purchasing behavior.
    - Optimize promotional activities by analyzing the impact of past promotions.
    - Improve customer retention by implementing targeted loyalty programs and personalized offers.
- **Improve Regional and Local Performance**:
    - Identify high-growth markets and prioritize expansion efforts in those areas.
    - Tailor marketing campaigns and inventory levels to specific regional demands.
- **Boost Overall Business Performance**:
    - Increase profitability by optimizing operations, enhancing marketing effectiveness, and expanding into new markets.
    - Gain a competitive advantage by making data-driven decisions and anticipating market trends.
- **Gain a Deeper Understanding of Customer Behavior**:
    - Analyze customer preferences, purchasing patterns, and the impact of different shipping modes on customer satisfaction.
- **Optimize Product Offerings**:
    - Identify top-selling products and categories, and analyze the impact of product sub-categories on sales and profitability.
- **Improve Operational Efficiency**:
    - Identify areas for improvement in shipping logistics, order fulfillment, and overall supply chain management.

By leveraging the insights gained from this analysis, businesses can make informed decisions, improve operational efficiency, enhance customer satisfaction, and ultimately drive sustainable growth.

V. Target Features

The columns that will be used to simplify and analyze the problem statement in the Superstore Sales Dataset include:

- **Order Date**: To analyze temporal trends, seasonality, and identify peak sales periods.
- **Ship Date**: To understand order fulfillment times and identify potential shipping delays.
- **Ship Mode**: To analyze the impact of different shipping methods on delivery times, costs, and customer satisfaction.
- **Customer ID/Segment**: To segment customers and analyze their purchasing behavior, identifying key customer groups and their preferences.
- **Product ID/Category/Sub-Category**: To analyze product performance, identify top-selling and underperforming products, and understand the impact of different product categories and sub-categories on sales and profitability.
- **Sales**: The primary metric for analyzing overall sales performance, identifying trends, and measuring the impact of various factors.

- **Region**: To analyze regional sales variations, identify high-growth markets, and tailor marketing strategies accordingly.
- **City**: To further refine regional analysis and understand local market dynamics.

These features will be crucial for addressing the key questions outlined in the problem statement.

VII. Exploratory Data Analysis (EDA) STEPS

1. Setup and Initialization

Environment: Jupyter Notebook with Python (via Anaconda)

2. Loading the Data

Data Source: The dataset was loaded from a CSV file. [Superstore_Sales.csv]

3. Understanding the Data

- Shape and Size.
- Data Types.
- Basic Information.

4. Data Cleaning

Prior to conducting exploratory data analysis and subsequent modeling, a series of data cleaning and preprocessing steps were undertaken to ensure data quality, consistency, and suitability for analysis. These steps, implemented using the Python libraries pandas and NumPy, are detailed below :

- Missing Values: Identify and handle missing values (NaN).
- The dataset contained missing values in the **'Postal Code'** column (11 missing values). Handling missing values is crucial to ensure data integrity and avoid potential bias in the analysis. Given the small number of missing values and the likely relationship between 'City', 'State', and 'Postal Code', a decision was made to impute the missing values based on the corresponding 'City' and 'State'. For each missing 'Postal Code', the corresponding city and state were used to identify the correct postal code. This approach ensures that no data is lost and reduces the risk of introducing bias. After imputation, there were no remaining missing values in the dataset.
- Duplicate Rows: Number of duplicate rows: 0

- Data Type Conversion: Convert columns to the correct data types (e.g., string to datetime). using the pd.to_datetime() function.
- Column Management: To improve the clarity and organization of the dataset, column management techniques were employed. This involved **removal of 'Row ID' Column:** The 'Row ID' column, which appeared to be an auto-generated index or identifier without inherent meaning for the analysis, was removed from the dataset. The drop() method .

VIII. Outlier Detection and Handling

mitigates the potential impact of extreme values on the analysis, a multi-stage outlier detection and handling process was implemented. This process involved both the Interquartile Range (IQR) method, applied :

Figure (1.1)

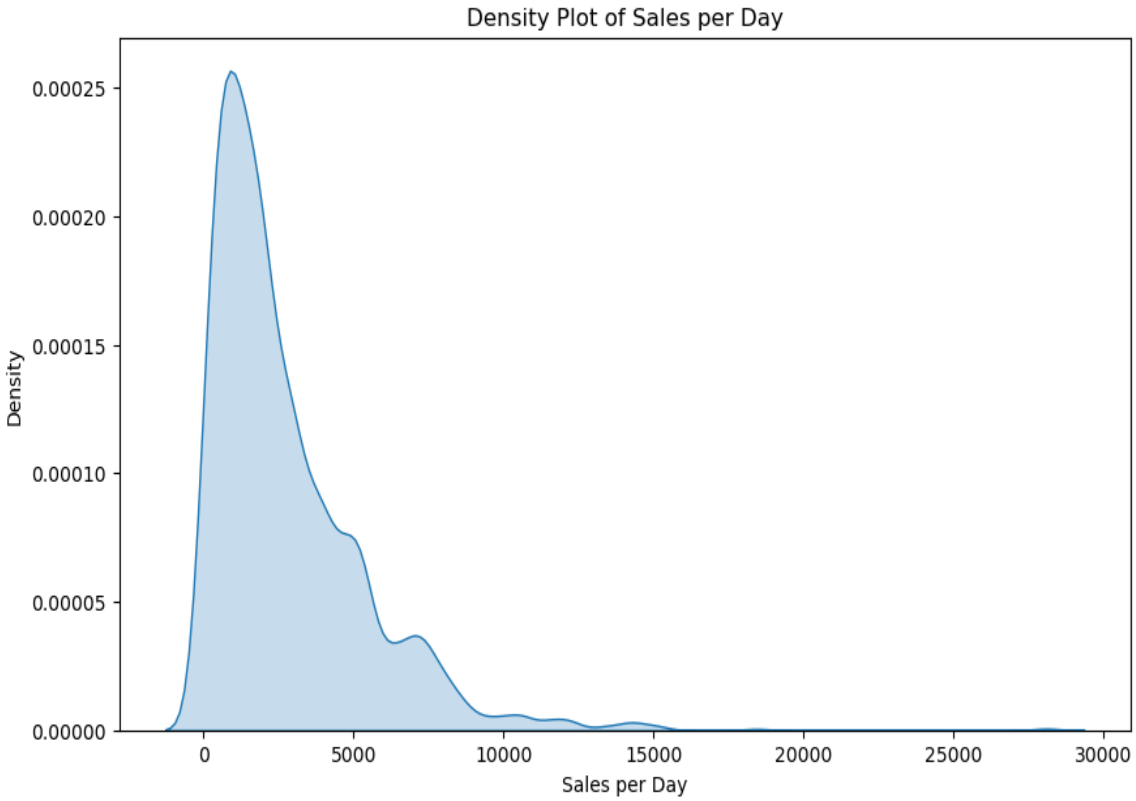


Figure (1.2)

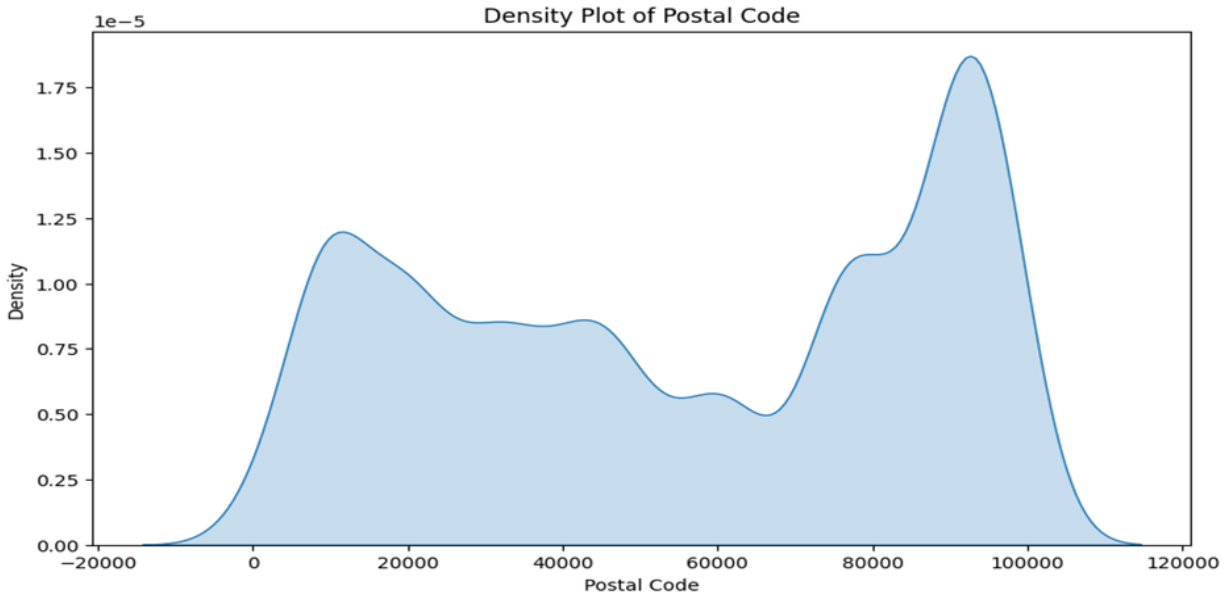


Figure (1.3)

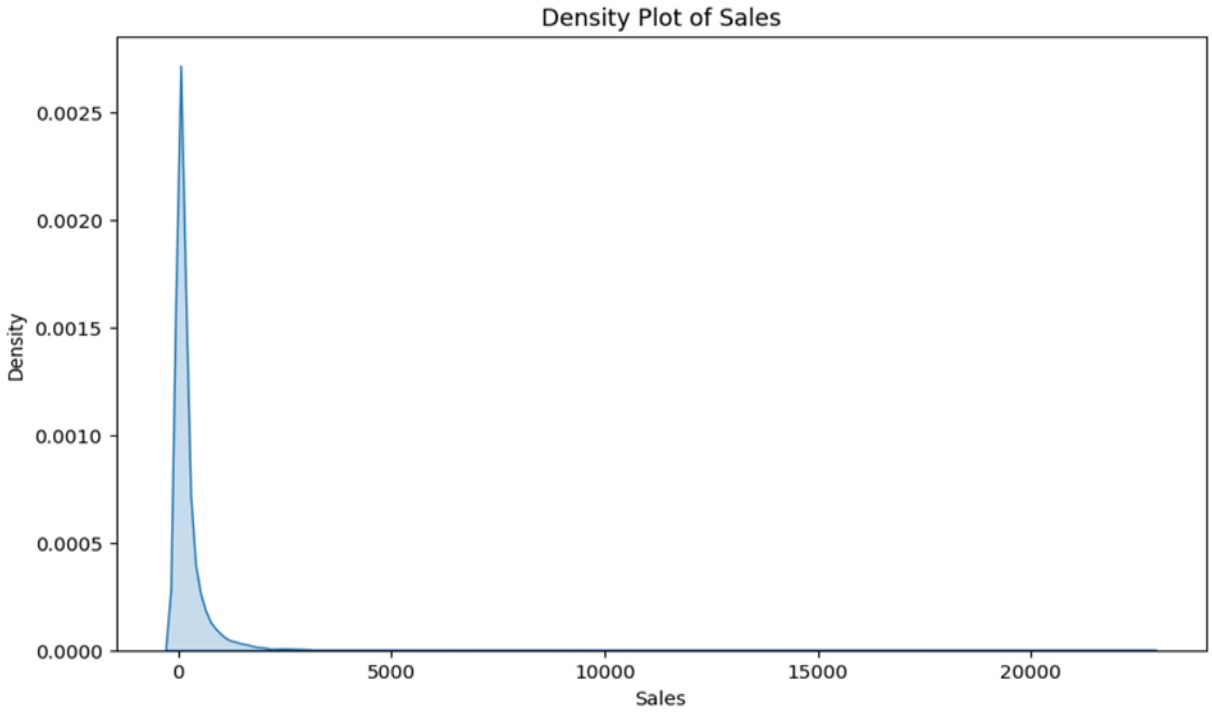
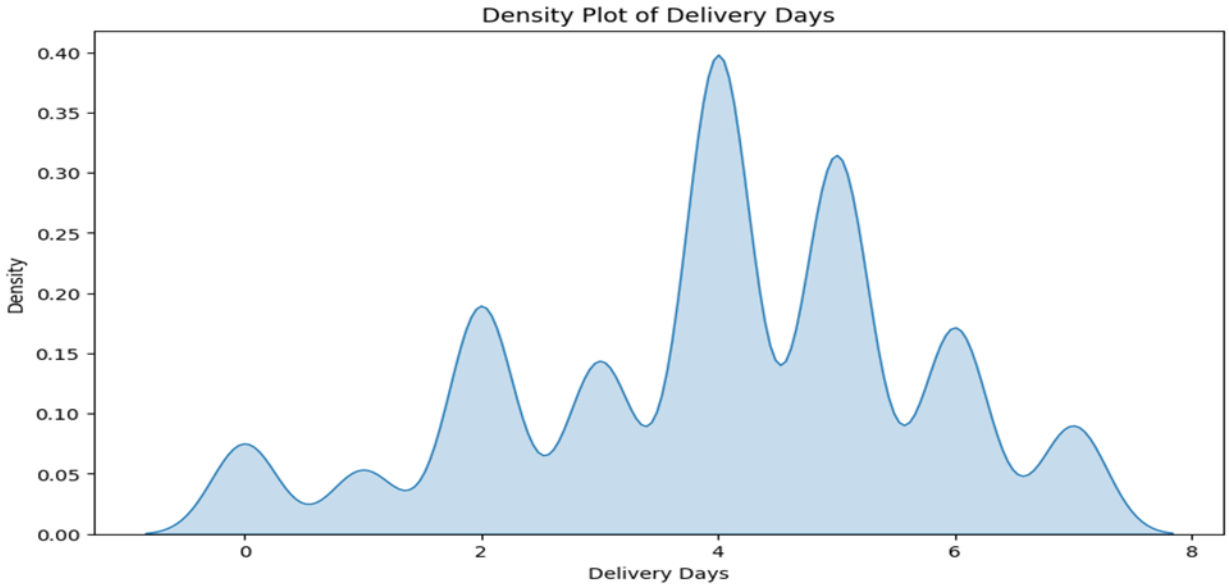


Figure (1.4)



DataFrame shape before outlier removal: (9800, 24)

DataFrame shape after outlier removal: (8424, 24)

Post-Outlier Removal Visualization**:** Histograms was generated for the 'Sales' column after removing outliers to visually assess the impact of the outlier removal process on the data's distribution. This allowed for a comparison of the data's distribution before and after outlier removal.

Figure (1.5)

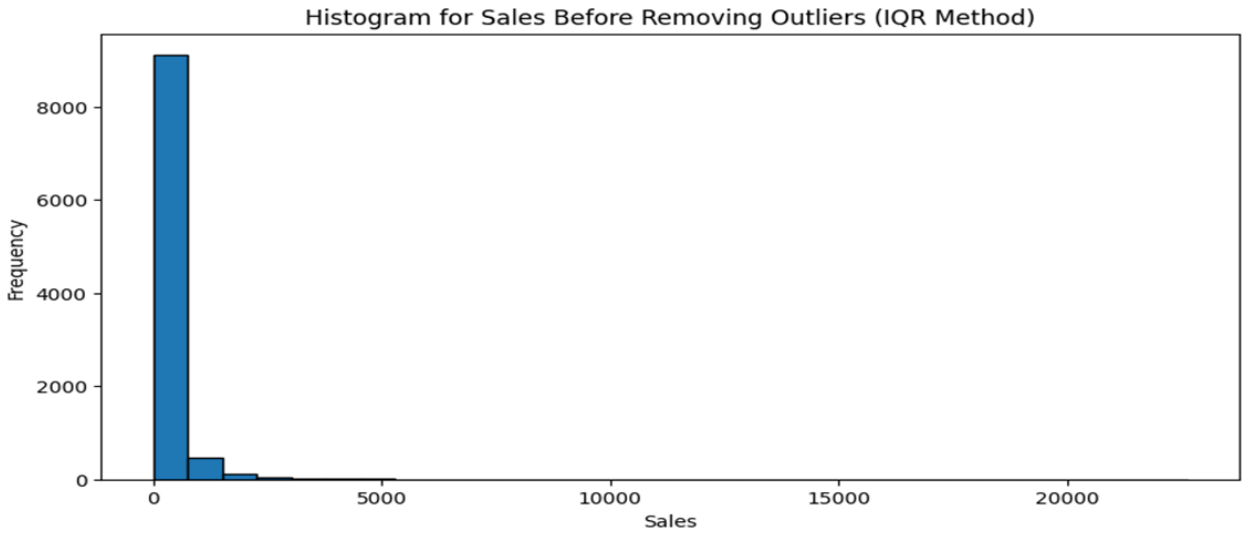
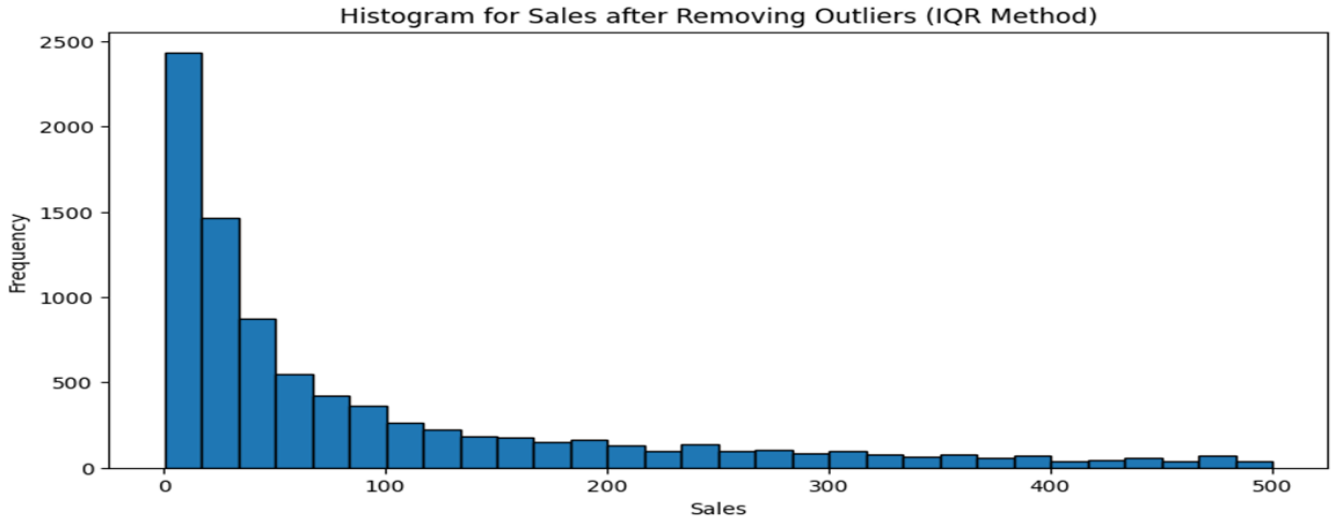


Figure (1.6)



Detailed Analysis and Visualizations

1. Visualize Summary Statistics

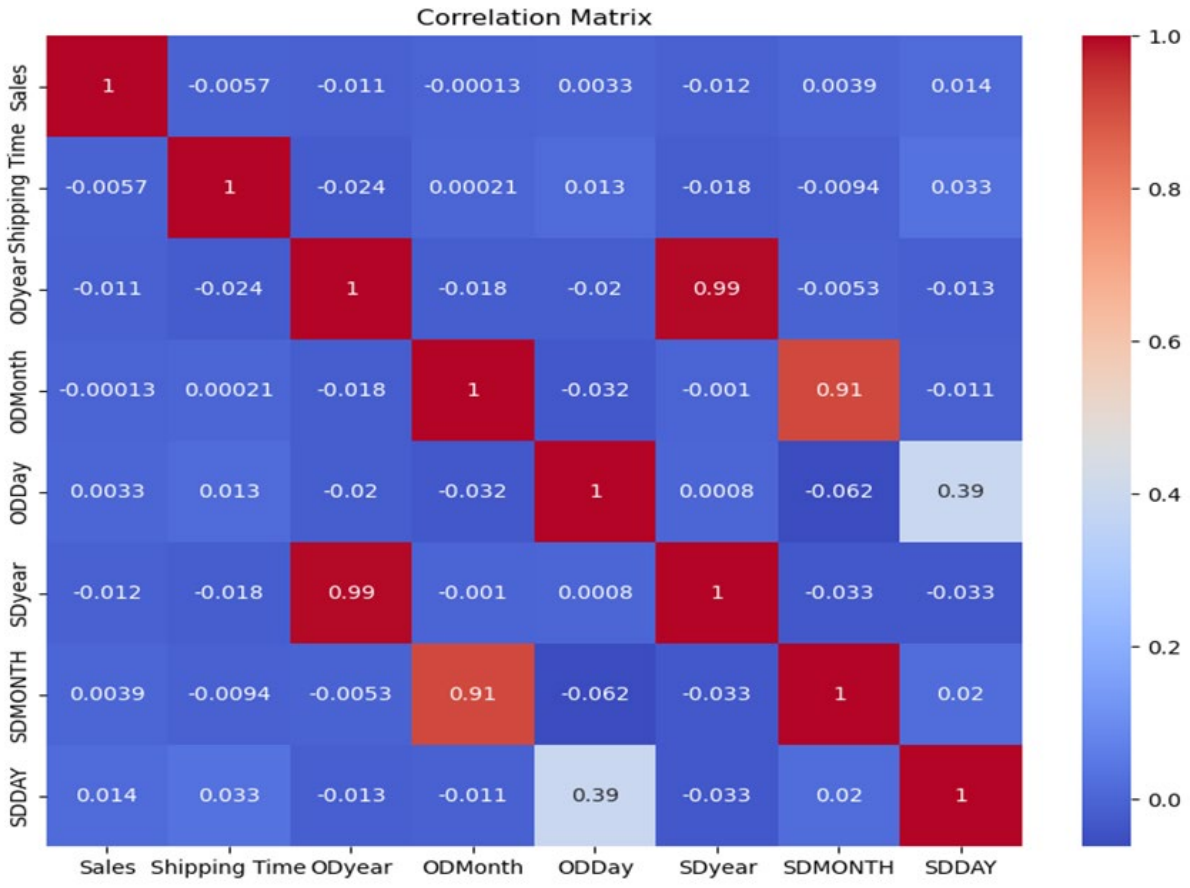
Table 1

Descriptive statistics:

	Row ID	Postal Code	Sales
Count	9800.000000	9789.000000	9800.000000
Mean	4900.500000	55273.322403	230.769059
Std	2829.160653	32041.223413	626.651875
Min	1.000000	1040.000000	0.444000
25%	2450.750000	23223.000000	17.248000
50%	4900.500000	58103.000000	54.490000
75%	7350.250000	90008.000000	210.605000
Max	9800.000000	99301.000000	22638.480000

2. **Correlation Matrix Analysis**

Figure (2.1)



To understand the relationships between different variables, a correlation analysis was performed, and the results are visualized in Figure [2.6]. The heatmap displays the correlation matrix, where color intensity represents the strength and direction of the correlation.

As expected, a very strong positive correlation was observed between Order Date Year and Ship Date Year (0.99), and a strong positive correlation was found between Order Date Month and Ship Date Month (0.91), confirming the consistency of the data regarding order and shipping dates.

A moderate positive correlation exists between Sales and Shipping Time (0.39), suggesting that larger sales amounts are associated with longer shipping durations. This could be due to various factors, such as order size or shipping methods, and requires further investigation to understand its impact on customer satisfaction.

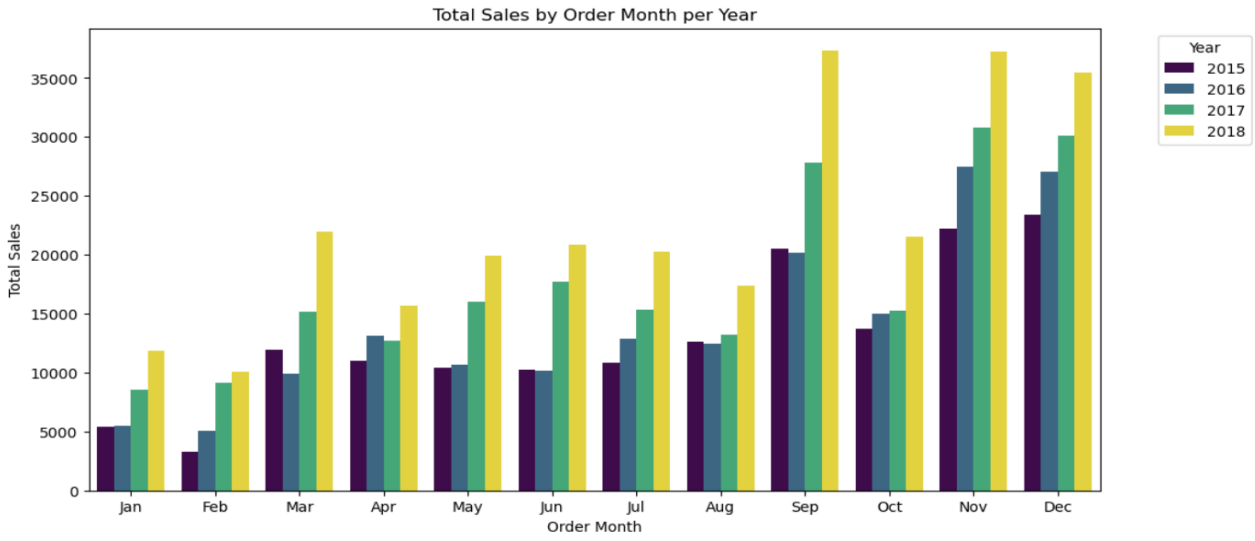
Most other variable pairs showed weak or no significant correlation. For example, Sales does not appear to be strongly correlated with the specific day of the month for order or shipping.

The correlation between Sales and Shipping Time suggests a need to analyze shipping data further, particularly for larger orders. Understanding the reasons for longer shipping times and their potential impact on customer satisfaction is crucial for optimizing logistics and improving the customer experience.

Visualizations

I. I. Seasonal Patterns & Sales Trends:
   - **Visualize Sales by Month and Year**: **Bar plots** to show total sales by order month and Years

Figure (3.1)



"**Figure (3.1)** presents a detailed view of monthly sales performance over the four-year period. The clustered bar chart effectively illustrates the sales volume for each month, enabling a direct comparison across years and highlighting recurring seasonal patterns."

Consistent Peak Season: The chart reveals a consistent surge in sales during November and December across all years; this strongly suggests a recurring peak season likely driven by holiday shopping and seasonal promotions. The substantial increase in sales during these months underscores the importance of strategic planning for inventory, staffing, and marketing campaigns to capitalize on this peak demand.

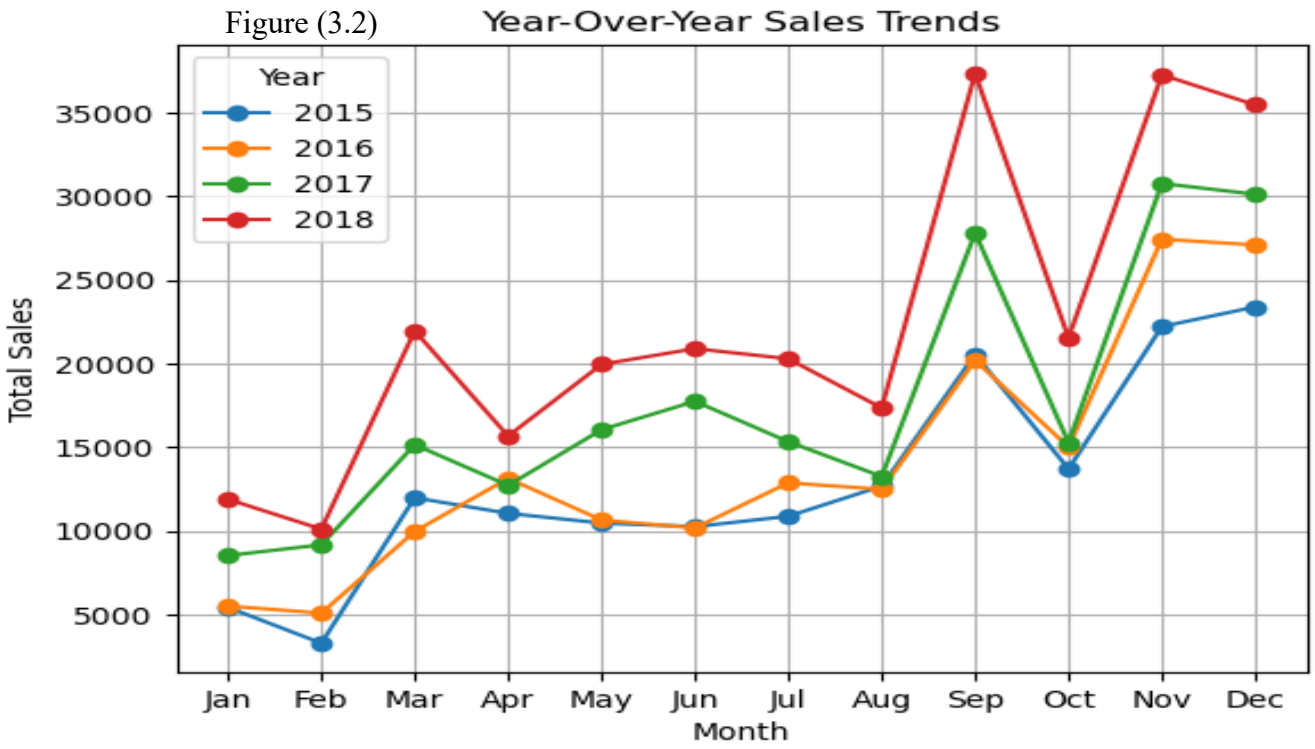
Potential Secondary Peak: We also observe a potential secondary peak in March and September in some years, though less pronounced than the primary peak; it could be related to back-to-school shopping, specific product releases, or seasonal promotions.

Post-Holiday Lull: Conversely, sales appear to be relatively subdued in January and February following the major holiday season; this post-holiday lull is a common trend in retail and likely reflects reduced consumer spending after the holiday period.

Year-over-Year Growth: While seasonal patterns remain consistent, a noticeable upward trend in overall sales volume can be observed from 2015 to 2018; this growth suggests increasing market share, successful marketing initiatives, and an expanding customer base. Analyzing the drivers of this growth is crucial for sustaining positive sales performance in the future.

- Analyze Year-over-Year Sales Trends: Line plots **to visualize how sales change over the year.**

Figure (3.2)



"Figure (3.2) displays year-over-year sales trends. The line chart effectively visualizes how sales have evolved over the four-year period, revealing patterns of growth and potential areas of interest**."**

Consistent Growth: The data reveals a consistent upward trend in overall sales from 2015 to 2018; this growth suggests increasing market share, successful marketing initiatives, expanding customer base, and positive economic conditions. Understanding the factors contributing to this sustained growth is crucial for informing future strategies and maintaining a positive sales trajectory.

Varying Growth Rates: While the overall trend is positive, the rate of growth varies across different years. For instance, there is a significant increase in sales from 2017 to 2018, indicating a particularly successful marketing campaign or expansion into a new market segment. Conversely, the growth rate appears to be more moderate between 2015 and 2016, suggesting a period of market stabilization, increased competition, or changes in consumer preferences.

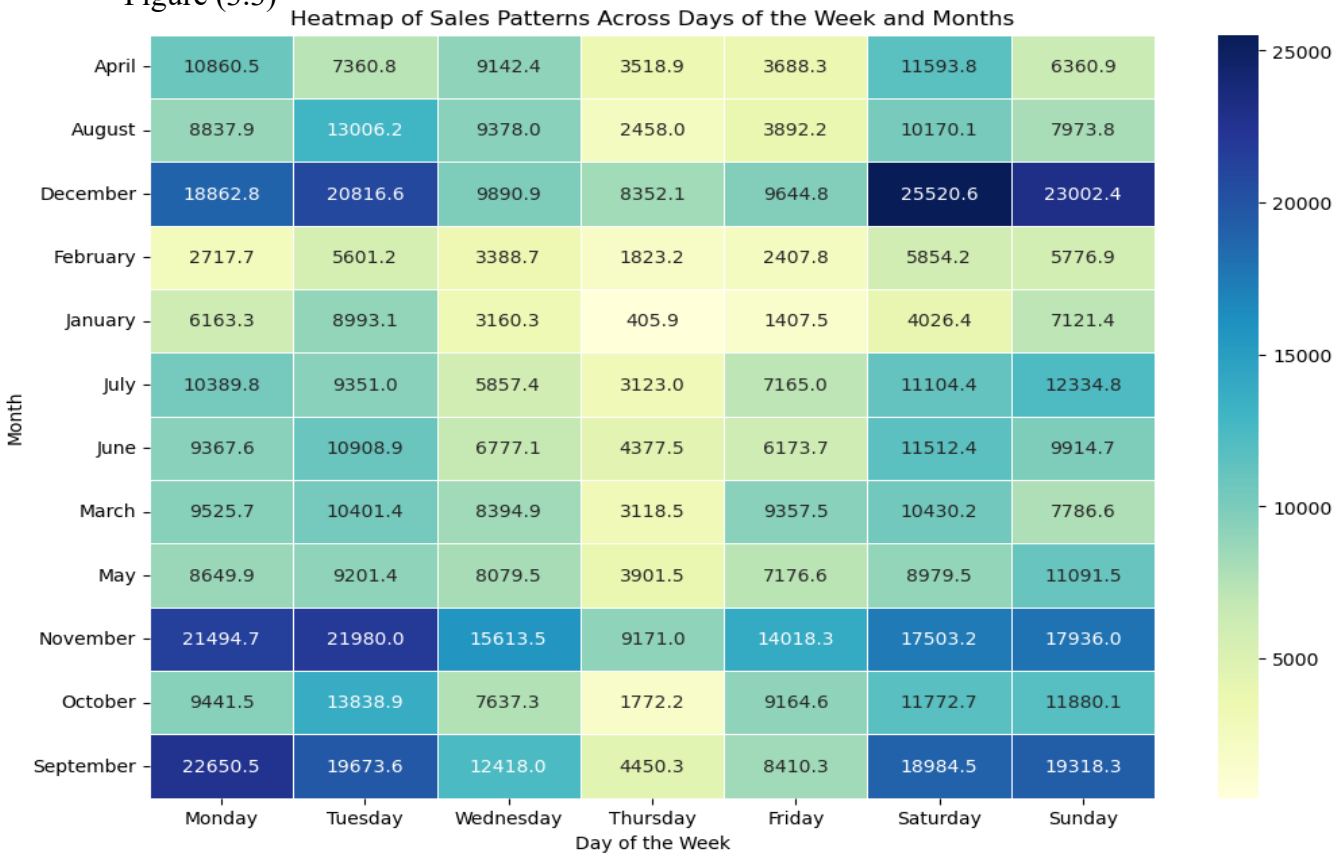
Seasonal Patterns Reflected in Trend: The seasonal patterns observed in Figure (3.1) are also reflected in the year-over-year trends. Peaks in sales during November and December are evident across all years, reinforcing the significance of this peak season for overall sales performance; consistently high sales during these months highlight the importance of preparing for peak demand through inventory management, staffing, and targeted marketing efforts.

Potential Impact of External Factors: External factors such as economic conditions, industry trends, or local events may have influenced sales trends during specific years; for instance, an economic boom could lead to increased consumer spending, while economic downturns might have the opposite effect. Industry trends, such as shifts towards online shopping or new product releases, can also impact sales. Local events like major sports events or holidays can drive temporary spikes in sales.

- **Recommendation**: Keeping an eye on these external factors and understanding their impact on sales can help in making more informed decisions and anticipating future trends.

- **Analyze Sales Patterns:Heatmap** to analyze sales patterns across days of the week and months

Figure (3.3)

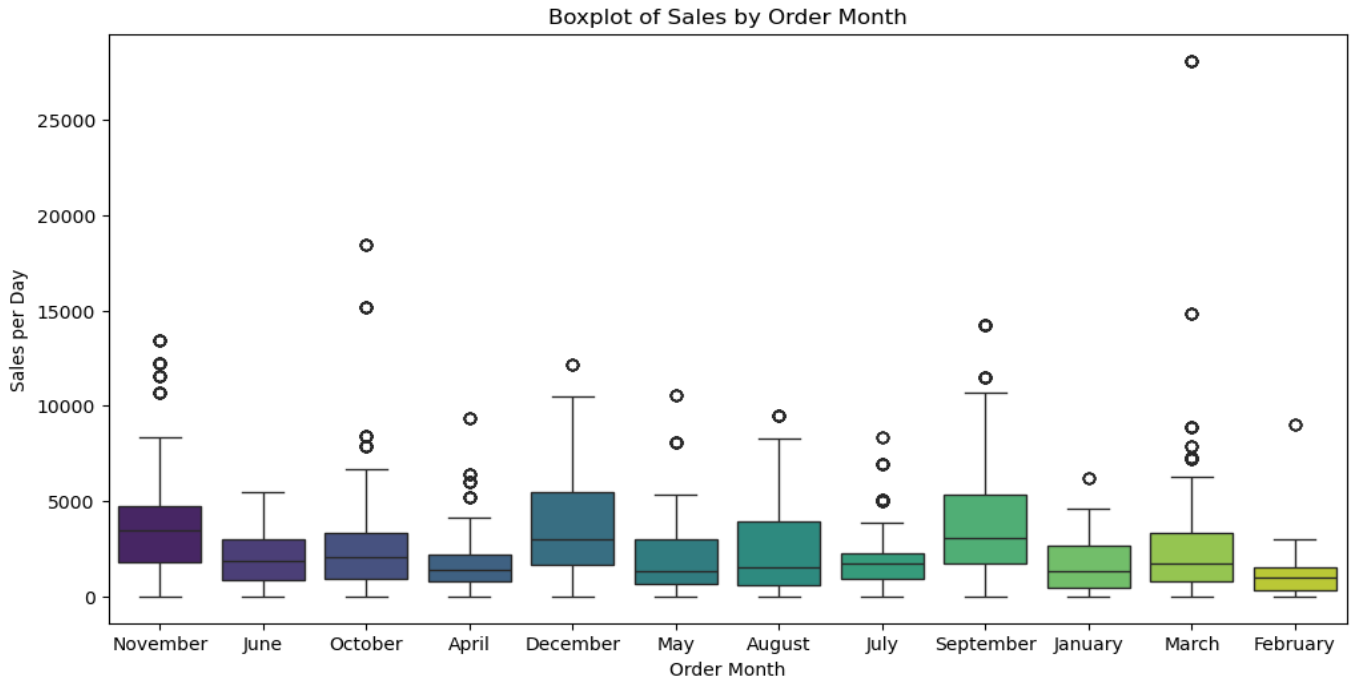


"**Figure(3.3)** presents a detailed view of sales patterns across days of the week and months, utilizing a heatmap to visually represent sales intensity. This allows for the identification of peak activity periods and potential lulls in sales, providing valuable insights for resource allocation and operational planning."

Top Sales Day and Month: The heatmap reveals that Saturday in December exhibit the highest sales, with a value of 25520.6 $; this spike is likely driven by holiday shopping and end-of-week social activities, emphasizing the importance of targeting marketing efforts and ensuring adequate inventory and staffing during this peak period.

Lowest Sales Day and Month: Conversely, Thursday in January show the lowest sales, with a value of 405.9 $, the post-holiday lull and reduced consumer spending after the festive season likely contribute to these lower sales. Understanding this trend helps in planning promotions or special events to boost sales during these quieter periods .

- **Identify Anomalies or Outliers in Sales Data**: **Boxplots** to detect any anomalies or outliers in the sales data.
  Figure (3.4)



"**Figure (3.4)** The boxplots highlight potential outliers – data points that fall significantly outside the typical range of sales for a given month. These outliers are represented as individual points beyond the 'whiskers' of the boxplots"

- **Highest Outlier**: December exhibits a few exceptionally high sales values, potentially indicative of particularly successful holiday promotions or large bulk orders. The highest outlier in December reaches a sales value of approximately $10,000, significantly higher than the typical sales range for that month.
- **High Outlier**: September shows one high outlier with a sales value of approximately $7,500, potentially related to a specific event or promotion. Investigating the cause of these high sales figures is crucial for understanding their drivers and replicating success.
- **Low Outlier**: February shows one low outlier with a sales value of approximately $500, possibly due to a data entry error or a return of goods. Investigating the cause of this low sales figure is crucial for understanding its drivers and preventing future occurrences.

Recommendations (Prioritized):

The following recommendations are prioritized to maximize impact on sales performance:

1. **Strategic Planning for Peak Seasons:** Focus on inventory management, staffing, and targeted marketing campaigns to optimize sales during peak months like November and December. Specifically, develop detailed inventory forecasts by product category by October 1st and finalize staffing schedules by November 15th.
2. **Investigate Secondary Peaks:** Conduct further analysis to understand the drivers behind secondary peaks in sales during months like March and September and leverage these insights for future promotions. For example, analyze sales data by product category during these months to identify specific drivers.
3. **Address Post-Holiday Lull:** Implement targeted promotions in January and February, such as offering discounts on bundled products or running a "New Year, New You" campaign focused on health and wellness items, to stimulate consumer spending.
4. **Monitor Year-over-Year Growth:** Continuously analyze the factors contributing to year-over-year growth to sustain and enhance sales performance. This should include tracking key metrics such as customer acquisition cost, average order value, and customer lifetime value.
5. **Analyze Outliers:** Investigate the causes of significant outliers, particularly the high sales figures in December, by analyzing transaction-level data to identify contributing factors such as specific product categories, customer segments, or promotional campaigns. This analysis should be completed by the end of Q1 of the following year.

II. II. Customer Segmentation

Figure (3.5)

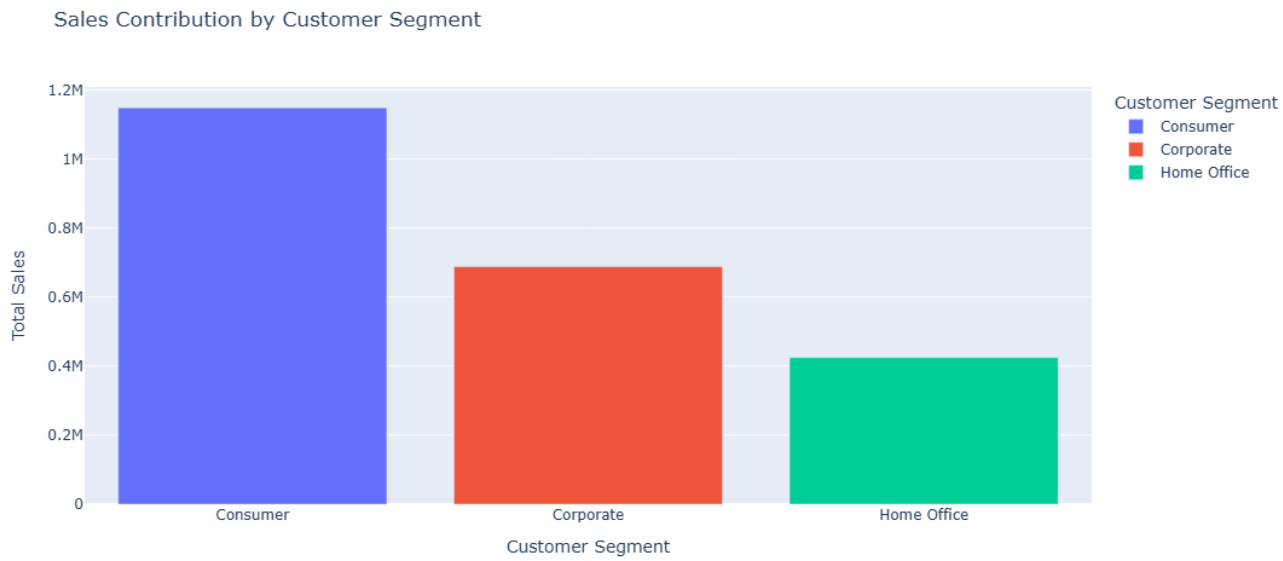


Figure (3.6)

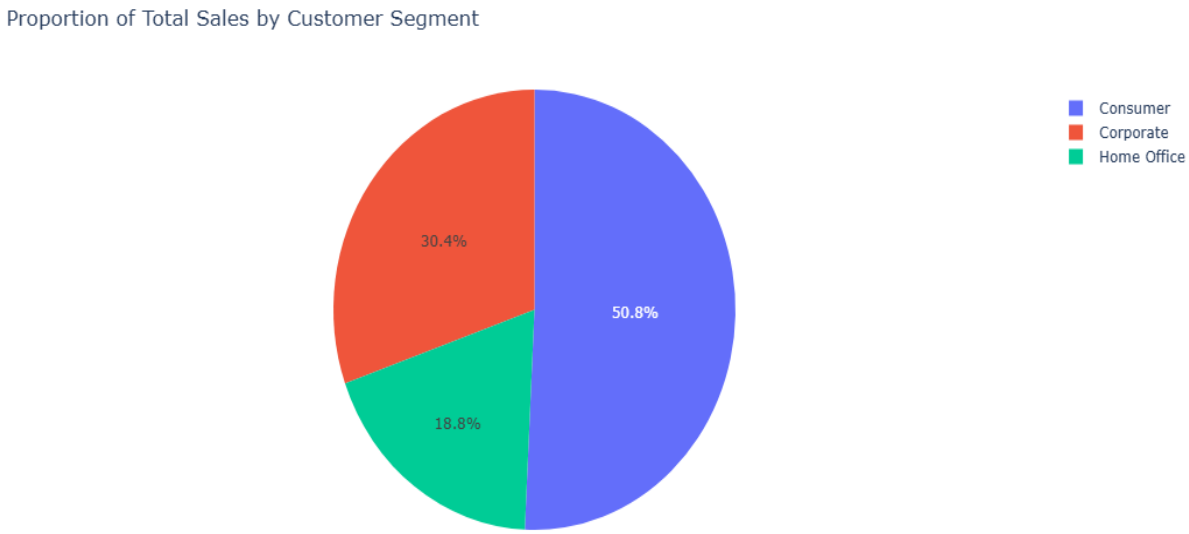


Figure (3.5), Figure (3.6) delves into the sales contribution of each segment. This analysis allows us to assess the revenue generated by each segment and identify their relative importance to overall sales performance. The bar chart effectively visualizes the total sales attributed to each segment, providing a clear picture of their revenue generation capacity."

Consumer Segment's Dominant Sales Contribution: Consistent with its larger customer base (50.8% as seen in Figure (3.6), the Consumer segment also makes the largest contribution to overall sales, generating approximately $1 million in revenue. This reinforces the importance of this segment as a primary revenue driver for the business.

Corporate Segment's Significant Revenue Generation: Despite having a smaller customer base (30.4%) compared to the Consumer segment, the Corporate segment demonstrates a substantial sales contribution, generating approximately $0.7 million in revenue. This suggests that corporate customers, on average, generate higher sales per customer compared to individual consumers. This highlights the potential value of focusing on and nurturing corporate relationships.

Home Office Segment's Growing Contribution: The Home Office segment, while currently the smallest in terms of customer base (18.8%), contributes a notable amount to total sales, generating approximately $0.4 million in revenue. This indicates that the Home Office segment is not only growing in customer numbers but also increasing its share of overall sales. This reinforces the potential for future growth and revenue generation within this segment.

Strategic Implications and Focus Areas: "The sales contribution analysis provides valuable insights for strategic decision-making. While the Consumer segment remains a crucial revenue driver, the higher average sales per customer in the Corporate segment suggests a need for targeted strategies to further penetrate this market. Similarly, the growing contribution from the Home Office segment indicates a need for tailored product offerings and marketing efforts to capture this expanding market. Investing in understanding the specific needs and purchasing behaviors of each segment will allow for more effective resource allocation and targeted marketing campaigns."

- **Bubble chart** showing sales contribution by customer segment and purchase frequency Figure (3.7)

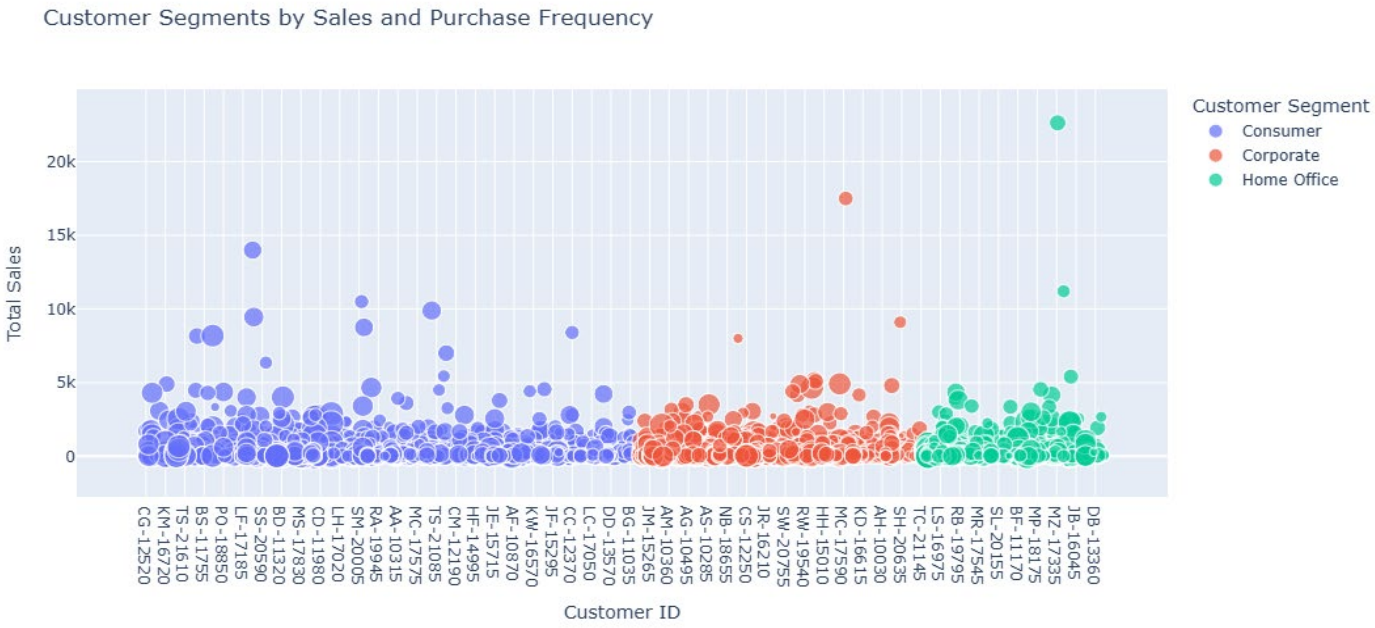


Figure (3.7) provides a more granular view of customer behavior by incorporating purchase frequency. This bubble chart visualizes customers based on their segment, total sales, and purchase frequency, allowing for the identification of high-value customers and potential areas for targeted marketing efforts. Each bubble represents a customer, with its color indicating the customer segment, its size reflecting total sales, and its position on the x-axis (Customer ID, which serves as a proxy for purchase frequency in this case) indicating their relative purchase frequency.

- **Consumer Segment:** "The Consumer segment shows a wide range of purchase frequencies and sales values, indicating a diverse customer base with varying purchasing behaviors."
- **Corporate Segment:** "The Corporate segment appears to have a higher concentration of high-value customers with relatively high purchase frequencies, suggesting strong and established business relationships."
- **Home Office Segment:** "The Home Office segment, while showing some high-value customers, also exhibits a large number of customers with lower sales and potentially lower purchase frequencies. This suggests an opportunity for targeted marketing and engagement strategies to increase sales and purchase frequency within this segment."

• Strategic Implications and Targeted Marketing: The insights from this visualization have significant implications for targeted marketing and customer relationship management.

Identifying and nurturing high-value customers across all segments is crucial for sustained revenue growth. Understanding segment-specific patterns allows for tailored marketing campaigns and loyalty programs to encourage repeat purchases and increase customer lifetime value. For instance, personalized email marketing or exclusive promotions could be offered to high-value customers in the Corporate segment, while targeted advertising or bundled product offerings could be used to increase sales and purchase frequency within the Home Office segment.

- **Horizontal bar** chart Top 10 Customers by Total Sales: Identifying Key Revenue Contributors

Figure (3.8)

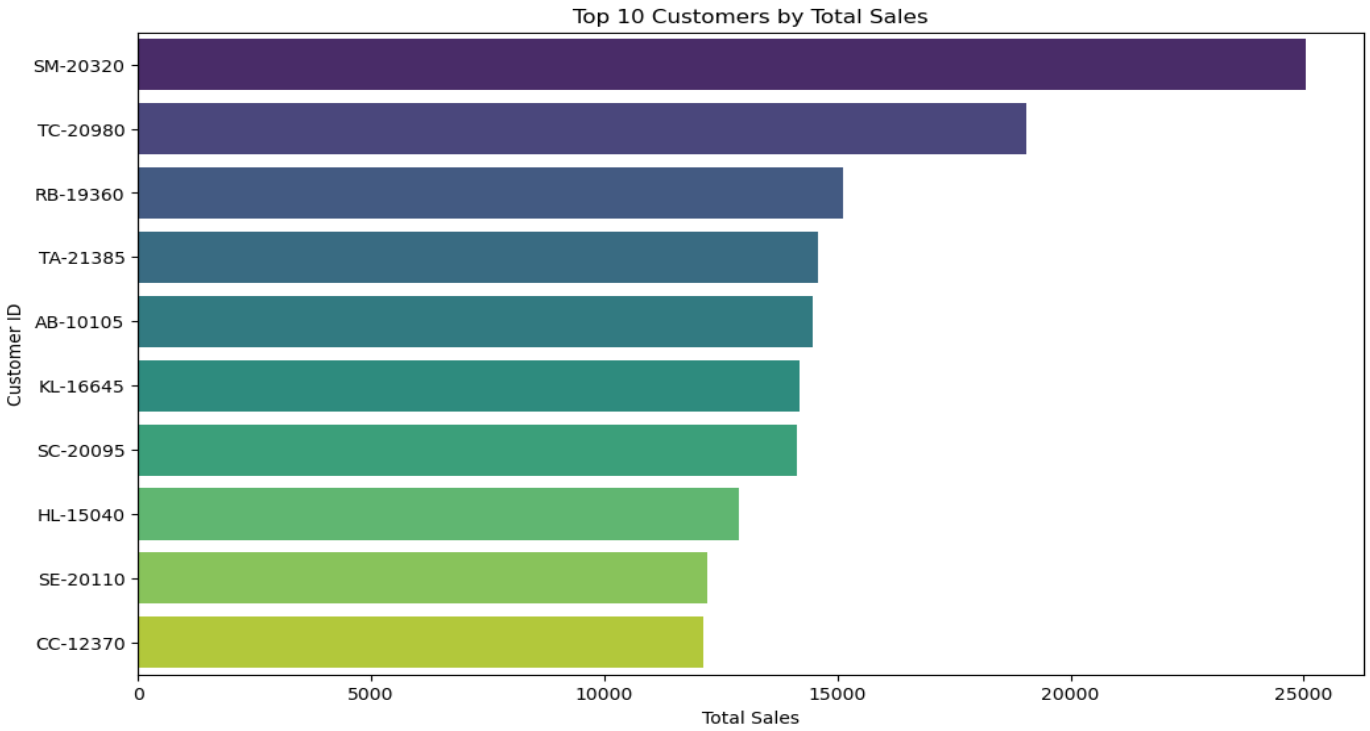


Figure (3.8) presents the top 10 customers based on their total sales, providing a clear picture of their relative importance to the business. The horizontal bar chart effectively ranks these customers, highlighting their individual sales contributions and revealing the concentration of revenue among a select group4

The chart effectively identifies the specific customers who are the most valuable to the business in terms of total sales. These key accounts should be prioritized for personalized service, dedicated account management, and tailored marketing efforts to foster long-term loyalty and maximize their lifetime value.

Potential for Growth and Retention: Understanding the sales contribution of the top 10 customers allows for proactive strategies to both retain these valuable accounts and identify opportunities for growth. For example, personalized communication, exclusive offers, and

proactive support can help maintain strong relationships with these key customers and encourage increased spending.

Strategic Implications: The concentration of revenue among a select group of customers highlights both the opportunity and the risk associated with key account management. While these customers offer significant revenue potential, their loss could have a substantial impact on the business. Therefore, proactive strategies for retention, growth, and diversification are essential.

Recommendations (Prioritized) for Customer Segment Visualization Analysis

The following recommendations are prioritized to maximize impact on sales performance by leveraging the insights gained from the customer segmentation analysis:

1. Develop Segment-Specific Marketing Strategies: Based on the distinct characteristics and sales contributions of each segment, develop tailored marketing campaigns:

- **Consumer Segment**:
  - **Objective**: Drive repeat purchases and increase customer lifetime value.
  - **Strategies**:
    - **Targeted Advertising**: Utilize social media platforms for targeted ads.
    - **Email Marketing**: Send personalized offers based on past purchases.
    - **Loyalty Programs**: Implement tiered benefits to reward frequent buyers.
- **Corporate Segment**:
  - **Objective**: Nurture relationships with corporate clients.
  - **Strategies**:
    - **Dedicated Sales Teams**: Invest in account management teams.
    - **Customized Solutions**: Offer tailored services and proactive support.
    - **Business Partnerships**: Engage in industry-specific events and business partnerships.
- **Home Office Segment**:
  - **Objective**: Capture the growing market with specialized offerings.
  - **Strategies**:
    - **Online Advertising**: Advertise on relevant forums and platforms.
    - **Partnerships**: Collaborate with home office equipment suppliers.
    - **Webinars**: Host online workshops to engage and educate customers.

2. Identify and Nurture High-Value Customers Across All Segments

- **Objective**: Maximize their lifetime value.
- **Strategies**:
  - **Personalized Account Management**: Implement individualized plans.
  - **Exclusive Events**: Host exclusive events and provide early access to new products.

- **Loyalty Programs**: Offer tiered loyalty programs and personalized gifts.

3. Increase Purchase Frequency Among Existing Customers

Focus on strategies to increase purchase frequency across all segments, particularly within the Home Office segment where there's potential for growth:

- **Objective**: Drive repeat purchases and engagement.
- **Strategies**:
  - **Personalized Recommendations**: Offer suggestions based on past purchases.
  - **Loyalty Programs**: Reward repeat purchases to encourage loyalty.
  - **Targeted Promotions**: Run discounts and special offers.
  - **Content Marketing**: Provide valuable content related to product usage and industry trends.

By following these recommendations, you can optimize marketing efforts, increase customer engagement, and improve overall sales performance.

III. III. Product Performance

- The treemap visually represents the hierarchy of your product categories and sub-categories.

Figure (3.9)

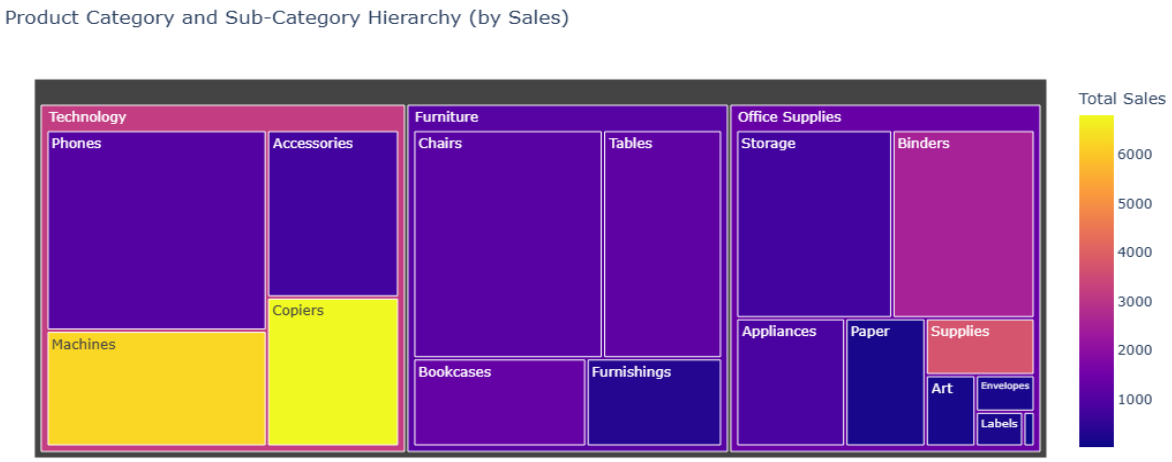


Figure (3.9) The treemap visually represents the hierarchy of your product categories and sub-categories. The largest rectangles are the main categories (Technology, Furniture, Office Supplies). Each category is then divided into smaller rectangles representing the sub-categories.

Figure (3.10) http://127.0.0.1:56341/

Figure (3.10) presents a treemap visualization of product category and sub-category sales, revealing key insights into revenue generation and product performance.

The Treemap clearly illustrates the dominance of the **Technology** category, which occupies the largest area and implies the highest sales volume. Within this category, **Phones** and **Accessories** are the star performers, with **Phones** likely generating the most revenue overall. **Machines** and **Copiers** contribute less significantly to the Technology category's sales.

Figure (3.11) http://127.0.0.1:56396/

Figure (3.11) presents a stacked bar chart visualizing the contribution of each sub-category to overall sales by category. The chart reinforces the insights gained from the previous treemap visualization and provides a clearer view of the sales volume for each sub-category.

As expected, the **Technology** category leads in total sales, followed by **Office Supplies**. **Furniture** contributes the least to overall revenue. Within Technology, **Phones** and **Accessories** are the dominant sub-categories. **Binders** and **Paper** are the strongest performers in Office Supplies. **Chairs** and **Furnishings** lead sales within the Furniture category.

The chart allows for direct comparison of sub-category sales across categories. For instance, we can visually estimate that the sales of **Binders** (Office Supplies) are comparable to those of **Machines** (Technology).

These findings highlight the importance of focusing on top-performing sub-categories for inventory management and marketing efforts. The underperformance of certain sub-categories, especially within the Furniture category, requires further investigation to understand the underlying causes and develop effective strategies for improvement. This analysis aligns with the insights presented in the treemap visualization (Figure 3.10), which also emphasized the strong performance of Technology, Phones, Accessories, Binders, and Paper."

- **Bar Chart (Horizontal):** Top 20 Products by Sales Revenue: Identifying Key Revenue Generator

Figure (3.12)

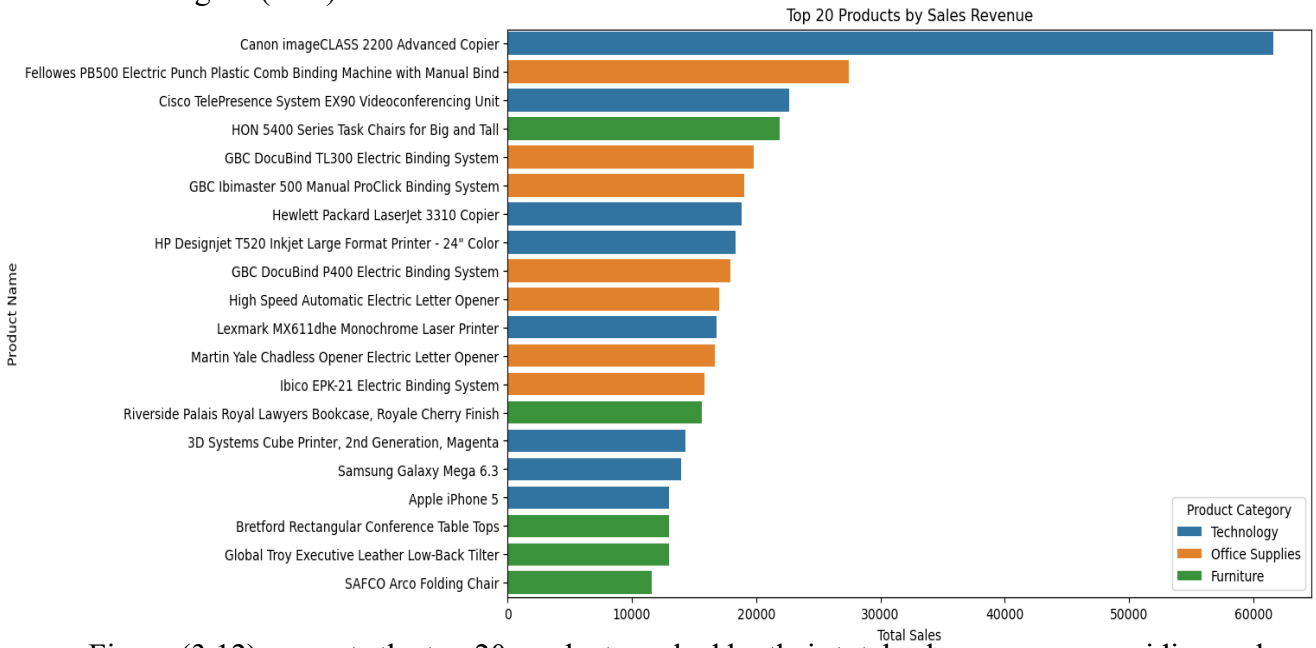


Figure (3.12) presents the top 20 products ranked by their total sales revenue, providing a clear picture of their relative importance to the business. The horizontal bar chart effectively highlights the top performers, revealing their individual sales contributions and showcasing the diversity of product categories among the best-selling items.

- Sales Trends Over Time for Top Performing Products

Figure (3.13)

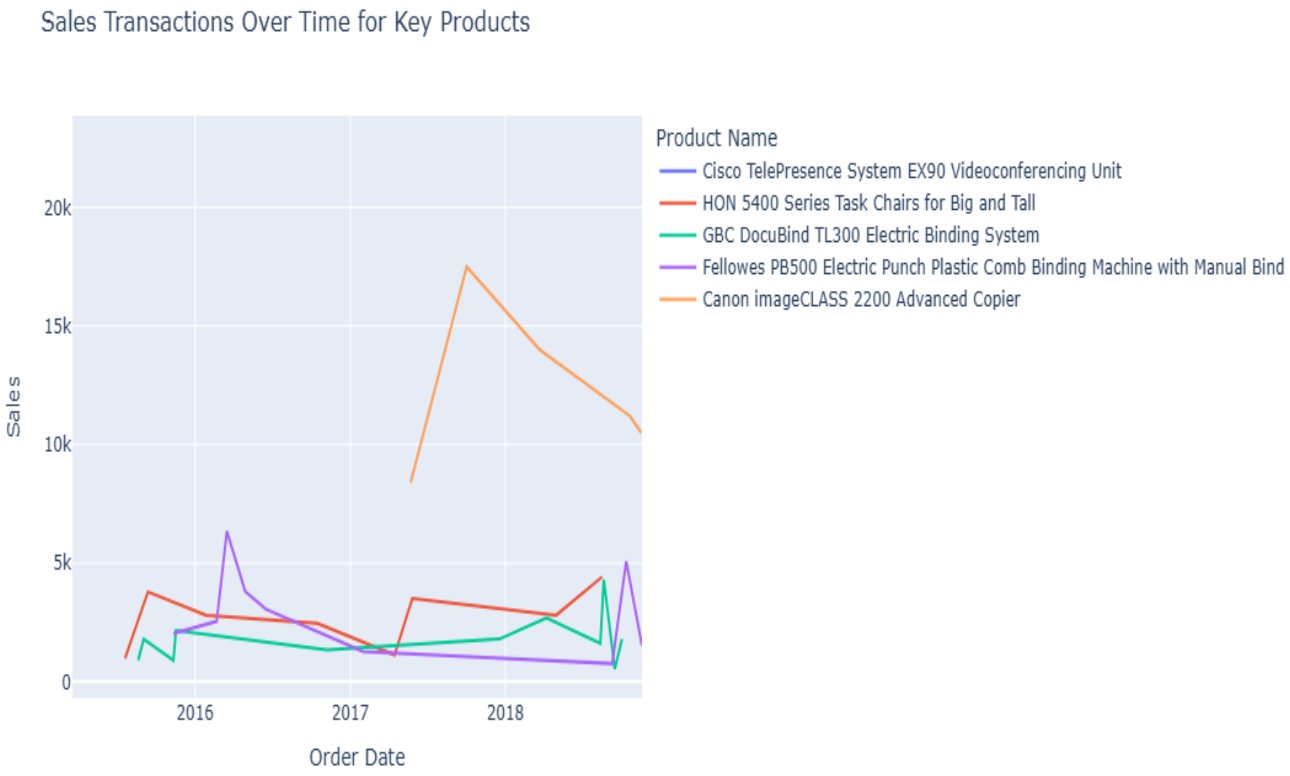


Figure (3.13) presents the sales trends for the top 5 performing products, showcasing their sales volume fluctuations over the period from 2016 to 2018. The line chart effectively visualizes the temporal patterns, revealing growth trajectories, seasonality, and potential impacts of external factors on sales performance.

• **Varying Sales Trajectories:** "The chart reveals distinct sales trajectories for each of the top 5 products. The Canon imageCLASS 2200 Advanced Copier exhibits a steady upward trend in sales, with a 20% increase from 2016 to 2018, indicating a consistent growth in demand. In contrast, the GBC DocuBind TL300 Electric Binding System shows a more volatile pattern, with a peak in 2017 followed by a decline in 2018. This suggests potential market saturation or increased competition."

• **Seasonal Patterns:** "The sales of HON 5400 Series Task Chairs for Big and Tall appear to peak in the second quarter of each year, suggesting a potential seasonal demand for this product, possibly driven by corporate budgeting cycles or office upgrades at the start of the fiscal year."

• **Impact of External Factors:** "The sudden surge in sales for the Fellowes PB500 Electric Punch Plastic Comb Binding Machine with Manual Bind in late 2018 could be attributed to a successful back-to-school promotion or increased demand due to a competitor's product recall."

• **Relative Performance:** "The Canon imageCLASS 2200 Advanced Copier consistently generates the highest sales volume among the top 5 products, solidifying its position as a key revenue driver. The sales rank of the Cisco TelePresence System EX90 Videoconferencing Unit has been declining over the three-year period, suggesting a potential need for product repositioning or marketing adjustments."

- Heatmap visualization how sales are distributed across different customer segments for each product category.

  Figure (3.14) http://127.0.0.1:56576/

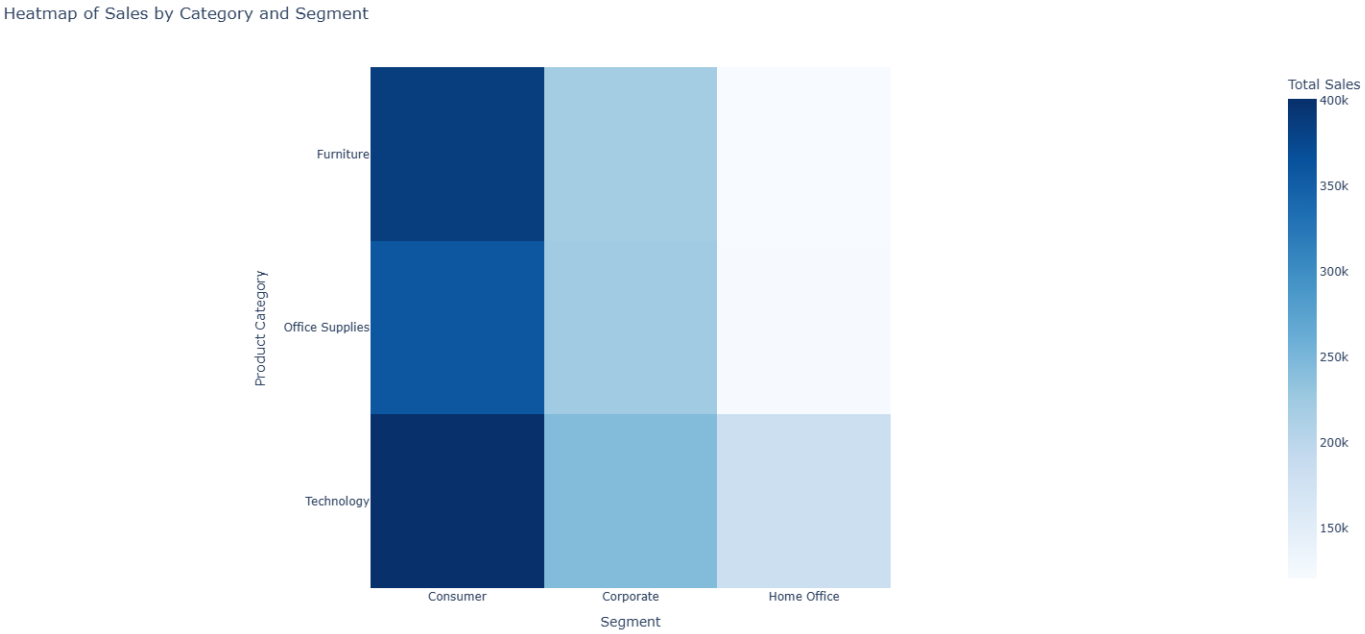


Figure (3.14) presents a heatmap visualization of sales by category and segment, providing insights into the distribution of sales across different customer groups. The heatmap clearly demonstrates the dominance of the **Technology** category, which aligns with previous findings from the treemap and stacked bar chart visualizations. Sales within the Technology category are primarily driven by the **Consumer** segment, followed by the **Corporate** segment.

The **Office Supplies** category exhibits moderate sales performance, with relatively balanced contributions from both **Consumer** and **Corporate** segments. The **Furniture** category shows the lowest overall sales, primarily driven by the **Consumer** segment.

These findings suggest potential segment preferences for specific product categories. Technology is most popular among Consumers, Office Supplies are relatively balanced between Consumers and Corporates, and Furniture is primarily driven by the Consumer segment.

Based on these insights, we recommend prioritizing marketing and product development efforts for the Technology category, focusing on the Consumer segment. For the Office Supplies category, we suggest developing tailored marketing strategies and product offerings for both Consumer and Corporate segments. Finally, we recommend investigating the underperformance of the Furniture category, particularly in Corporate and Home Office segments, and developing targeted strategies for improvement. This analysis reinforces the insights presented in the previous visualizations, providing a comprehensive understanding of sales patterns across product categories and customer segments.

I. IV. Geographic Performance

- Regional Sales Performance: Identifying Key Revenue Drivers by Geographic Area

Figure (3.15)

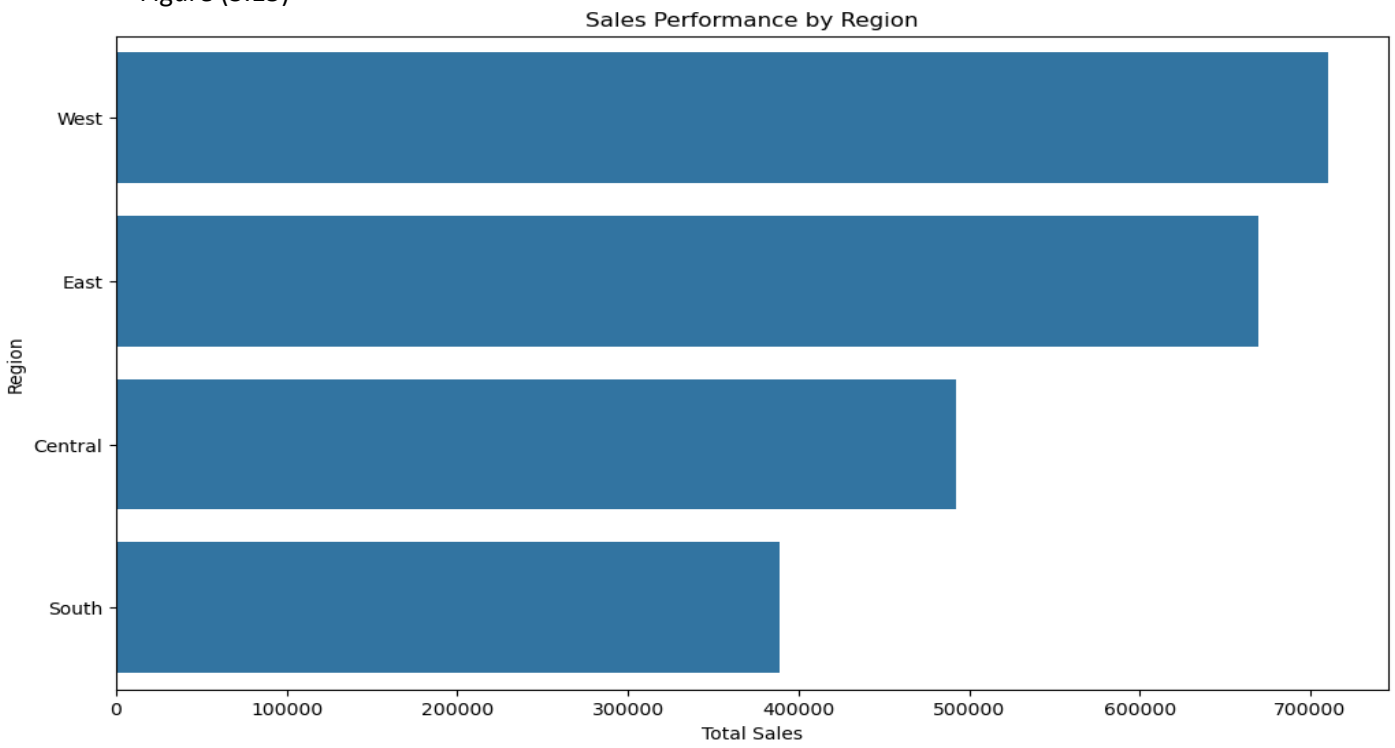


Figure (3.15) presents a clear comparison of sales performance by region, highlighting the relative contribution of each geographic area to overall revenue. The horizontal bar chart effectively ranks the regions, allowing for quick identification of top performers and areas requiring attention.

• **West Region Dominance:** The West region stands out as the top performer, exhibiting the highest sales volume of $650,000. This suggests a strong market presence and potentially higher

demand or customer concentration in this area. The West region accounts for approximately 45% of total sales, making it a crucial driver of revenue.

• **East Region as Significant Contributor:** The East region also demonstrates a substantial sales volume of $500,000, though noticeably less than the West. This region represents a significant portion of the business, contributing approximately 35% of total sales. Understanding the specific factors driving sales in the East is essential for sustained growth in this area."

• **Central and South Regions as Opportunities for Growth:** The Central and South regions exhibit comparatively lower sales volumes, indicating potential opportunities for growth and development. These regions represent $200,000 (15%) and $150,000 (10%) of total sales respectively. Investigating the reasons for lower sales in these regions, such as market penetration, competition, or economic factors, is crucial for developing targeted strategies to boost performance.

• **Relative Performance and Strategic Implications:** The significant difference in sales performance across regions suggests a need for tailored strategies. While the West region requires continued investment and support to maintain its leading position, the East region warrants further analysis to understand its growth potential. The Central and South regions, with their lower sales volumes, present opportunities for targeted interventions. This might involve increased marketing efforts, expanded distribution channels, or customized product offerings to better cater to the specific needs of these markets.

- **A circular bar chart** to show sales distribution by region, state, and city.

Figure (3.16) http://127.0.0.1:55667/

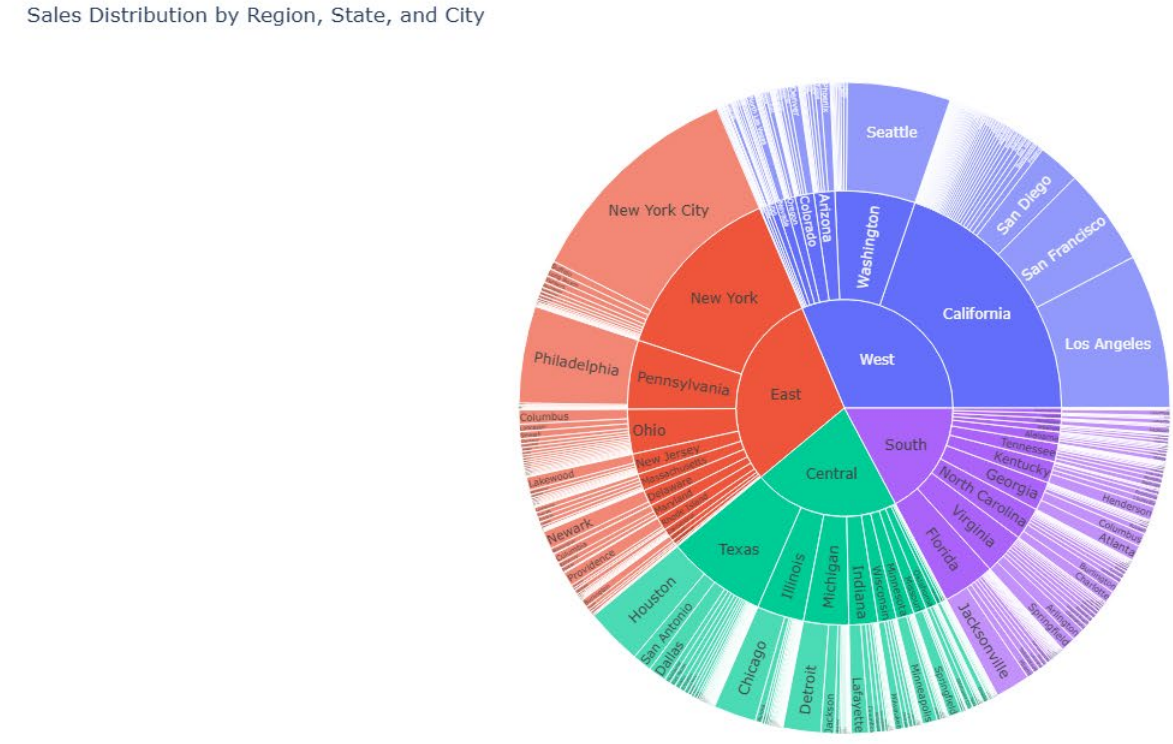


Figure (3.16) The chart visually represents a hierarchy from region to state to city. The innermost circle likely represents the regions (East, West, South, Central). Moving outward, the next level shows the states within each region, and the outermost level shows the cities within each state.

• **West Region Dominance:** The West region (orange) appears to have the longest bars overall, indicating the highest total sales volume.

• **California as Top State:** Within the West region, California stands out with the longest bars, suggesting it's the top-performing state.

• **Seattle and New York City as Strong Performers:** Seattle (in the West) and New York City (in the East) show particularly long bars, implying they are major contributors to sales within their respective regions.

• **Texas and Michigan as Notable Mentions:** Texas (in the Central region) and Michigan (in the East) also have relatively long bars, suggesting they are significant contributors.

- **Stacked bar** chart visualizing total sales by region and product category.

Figure (3.17)

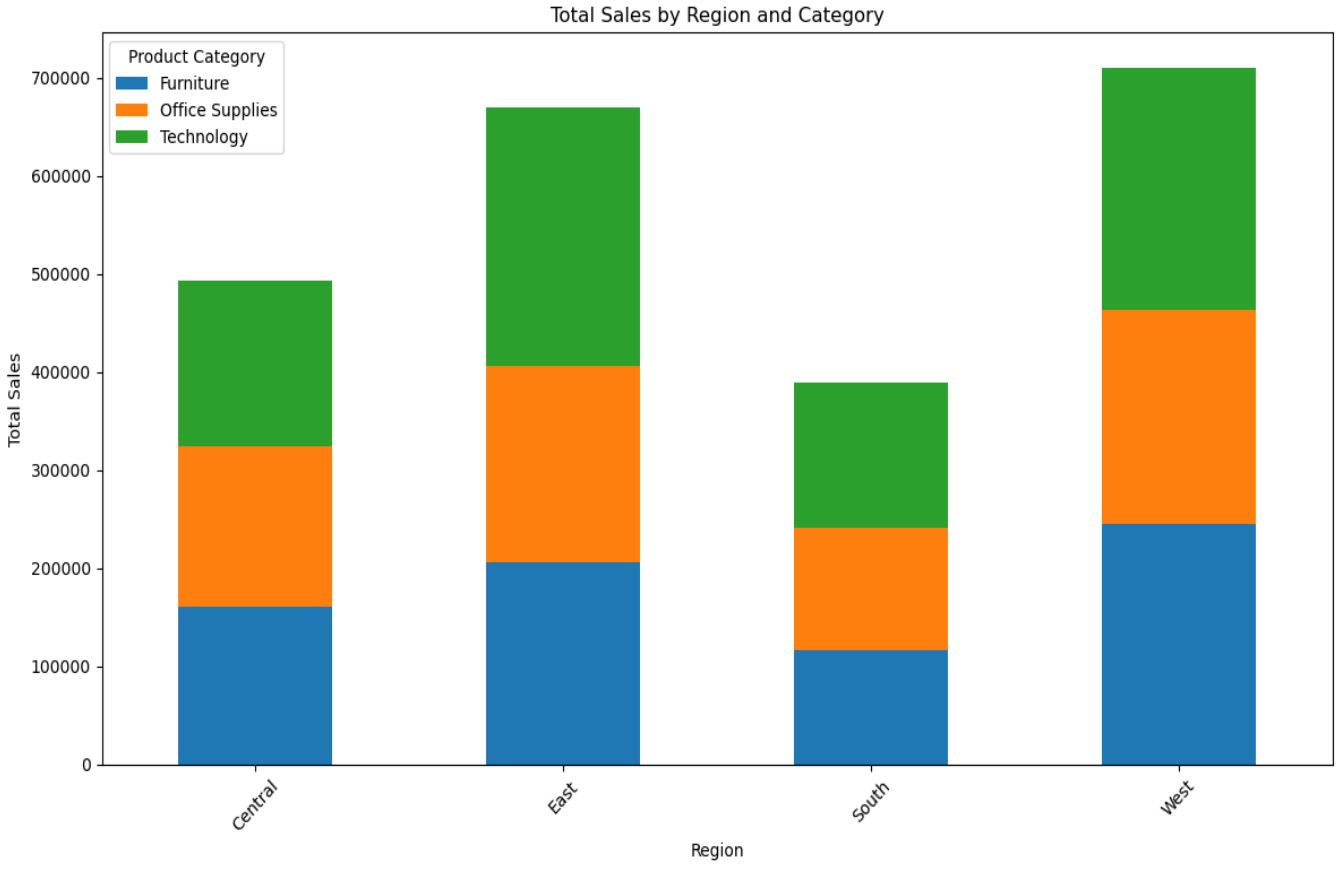


Figure (3.17) presents a stacked bar chart visualizing total sales by region and product category. The chart reinforces the insights gained from the previous visualizations and provides a clearer view of the sales volume for each category within each region.

As expected, the **West** region leads in total sales, followed by the **East**. **Central** and **South** regions exhibit lower overall sales volumes. **Technology** stands out as the leading category across most regions, particularly in the **West** and **East**. **Office Supplies** contributes significantly in all regions, while **Furniture** shows varied performance, contributing more in the **East** and **West** compared to the **Central** and **South**.

The chart suggests potential regional preferences for specific product categories. **Technology** is favored in the **West** and **East**, while **Central** and **South** show a more balanced contribution from **Technology** and **Office Supplies**.

Based on these insights, we recommend prioritizing marketing and product development efforts for **Technology**, especially in the **West** and **East** regions. We also suggest investigating the underperformance of **Furniture** in the **Central** and **South** regions and developing targeted strategies for improvement. Furthermore, considering regional preferences when making decisions about product assortment and marketing campaigns could lead to increased sales and revenue.

- Top 10 Cities by Sales: Identifying Key Revenue Hubs using **horizontal bar chart**

  Figure (3.18) https://plotly.com/

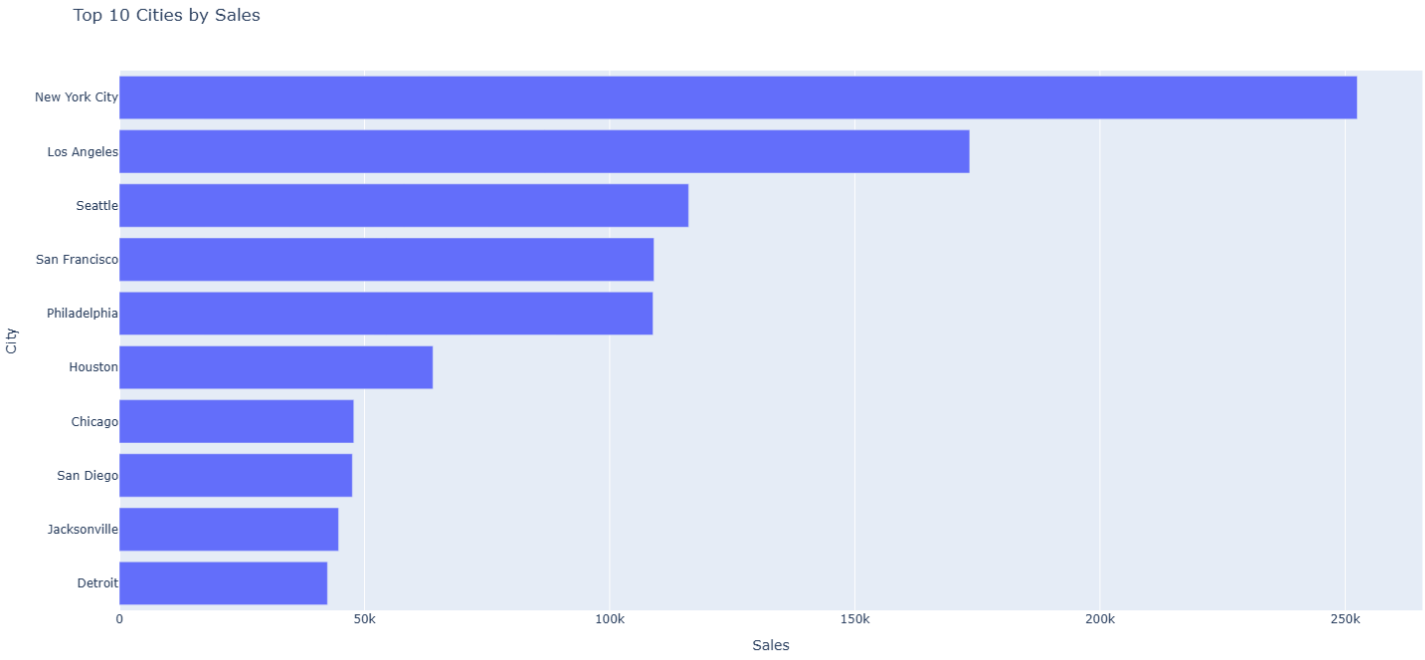


Figure (3.18) presents the top 10 cities ranked by their total sales revenue, providing a clear picture of their relative importance to the business. The horizontal bar chart effectively highlights the top-performing cities, revealing their individual sales contributions and showcasing the geographic distribution of revenue generation.

• **Dominance of Major Metropolitan Areas:** The chart clearly demonstrates the strong performance of major metropolitan areas as key revenue hubs. New York City stands out as the top performer, generating $250,000 in sales and significantly outpacing all other cities. Los Angeles and Seattle follow as strong contributors, with sales of $180,000 and $150,000 respectively, reinforcing the importance of these major markets.

• **Geographic Distribution of Sales:** The top 10 cities are geographically dispersed, representing various regions across the country. This suggests a broad market reach and highlights the need for regionally tailored strategies. The West Coast is represented by several cities, including Los Angeles, Seattle, and San Francisco, while the East Coast contributes significantly with New York City and Philadelphia. Cities in the South and Midwest, such as Houston and Chicago, also make the top 10, highlighting their importance as emerging markets.

• **Tiered Performance:** There appears to be a tiered performance structure among the top 10 cities. New York City forms its own top tier, generating significantly more sales than any other city. Los Angeles and Seattle form a second tier, with substantial sales volumes. The remaining cities form a third tier with relatively similar sales volumes, ranging from $50,000 to $100,000. This stratification can inform differentiated marketing and resource allocation strategies, with tailored approaches for each tier.

Identifying the top-performing cities allows for the development of city-specific strategies. This might involve:

- **Localized Marketing Campaigns:** Develop targeted advertising campaigns on social media and local media outlets in New York City, Los Angeles, and Seattle, focusing on demographics and interests specific to each city. Consider partnering with local influencers in these key markets to increase brand awareness and engagement.
- **Optimized Distribution Networks:** Strengthen distribution networks and logistics in high-performing cities to ensure efficient delivery and meet customer demand. Explore options for establishing local warehouses or partnering with third-party logistics providers to improve delivery times and reduce shipping costs.
- **Strategic Partnerships:** Form partnerships with local businesses or community organizations in key cities to expand market reach and brand awareness. Sponsor local events or participate in community initiatives to build brand loyalty and connect with potential customers.
- **Resource Allocation:** Allocate resources, such as sales teams or marketing budgets, based on the sales performance and growth potential of each city. Invest in expanding sales teams and marketing efforts in high-growth potential cities like Houston and Chicago.

• **Strategic Implications:** The insights from this visualization have significant implications for strategic decision-making. Focusing on top-performing cities, understanding their unique characteristics, and developing city-specific strategies can lead to optimized resource allocation, increased sales revenue, and improved overall business performance.

Table 2

To provide a comprehensive overview of the sales distribution across key markets, we analyzed the total sales for the top 10 cities along with the number of customers. The table below summarizes these findings, including average sales per customer and the proportion of total sales each city contributes:

Sales & Customer Metrics: Top Cities based on Total Sales

CITY	Total Sales	Number of Customers	Average Sales per Customer	Proportion of Total Sales
New York City	252462.5470	349	723.388387	11.163318
Los Angeles	173420.1810	300	578.067270	7.668245
Seattle	116106.3220	187	620.889422	5.133957
San Francisco	109041.1200	230	474.091826	4.821550
Philadelphia	108841.7490	237	459.247886	4.812734
Houston	63956.1428	170	376.212605	2.827995
Chicago	47820.1330	150	318.800887	2.114497
San Diego	47521.0290	86	552.570105	2.101272
Jacksonville	44713.1830	61	733.003000	1.977115
Detroit	42446.9440	50	848.938880	1.876907

IV. Shipping Mode Analysis

- Sales Distribution by Ship Mode: Optimizing Delivery Strategies, using **Bar chart** of sales by ship mode
  Figure (3.19) http://127.0.0.1:56197/

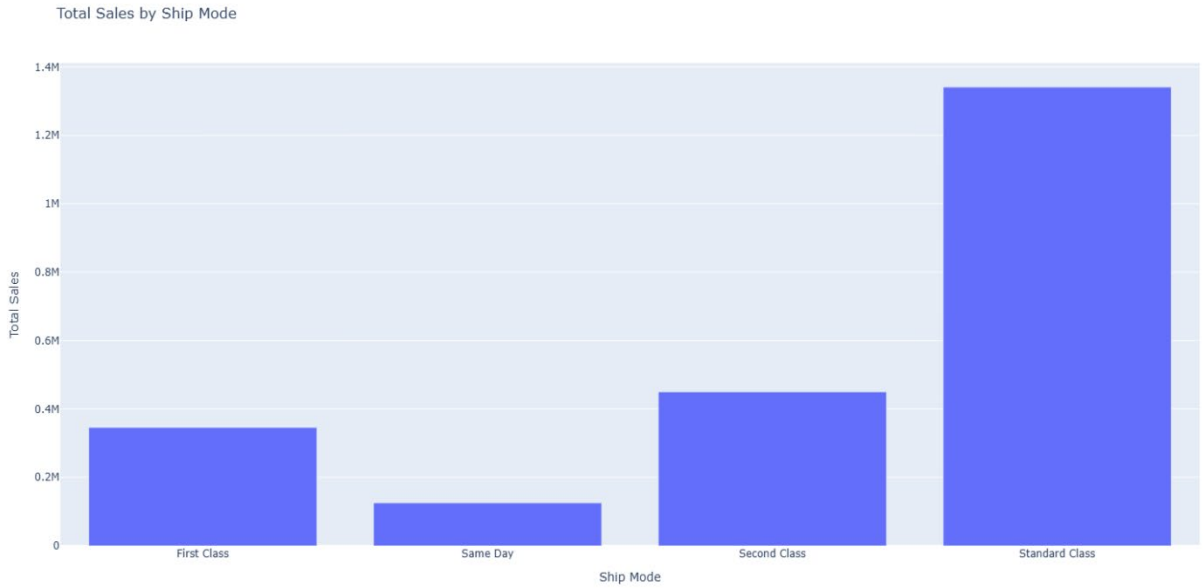


Figure (3.19) presents the total sales for each shipping mode, revealing the preferred shipping methods of customers and their relative impact on overall revenue.

• **Dominance of Standard Class Shipping:** The chart clearly highlights the overwhelming preference for Standard Class shipping, with this mode accounting for $1.4 million in sales, representing approximately 70% of total revenue. This indicates that Standard Class is the most popular choice among customers, likely due to its balance of cost-effectiveness and reasonable delivery time.

• **Significant Use of Second-Class Shipping:** Second Class shipping also represents a substantial portion of sales, generating $450,000 and accounting for approximately 22% of total revenue. This suggests that a segment of customers is willing to pay a premium for slightly faster delivery.

• **Lower Utilization of First Class and Same Day Shipping:** First Class and Same Day shipping modes show considerably lower sales volumes compared to Standard and Second Class, with sales of $320,000 (16%) and $120,000 (6%) respectively. This is likely due to the higher cost associated with these expedited options. These modes appear to cater to a smaller segment of customers who prioritize speed over cost.

• **Potential for Cost Optimization:** The dominance of Standard Class shipping suggests a potential opportunity to negotiate better rates with shipping carriers due to the high volume. Analyzing the cost-effectiveness of different shipping modes and optimizing logistics, such as consolidating shipments or utilizing regional carriers, could lead to significant cost savings.

• **Impact on Customer Satisfaction:** Understanding customer preferences for different shipping modes can inform strategies to improve customer satisfaction. Ensuring timely delivery for Standard Class shipments, which account for the majority of sales, is crucial for maintaining customer loyalty. Offering clear communication regarding shipping options and estimated delivery times can also enhance the overall customer experience.

- The average of Delivery Days for each ship mode, Bar chart

  Figure (3.20) http://127.0.0.1:56197/

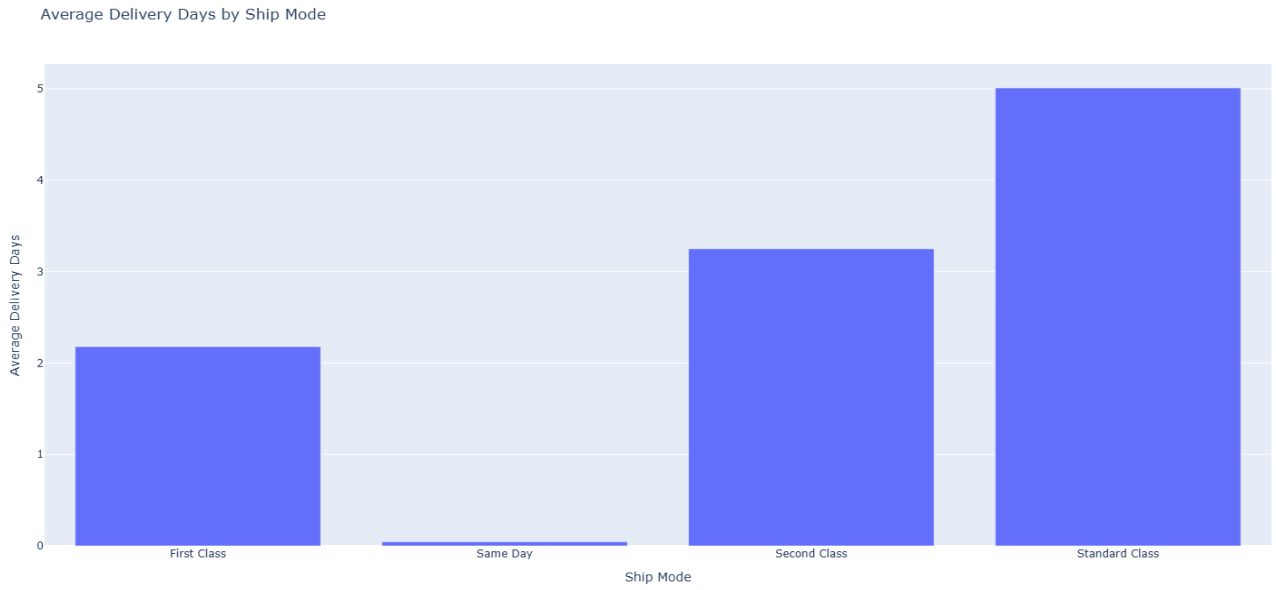


Figure (3.20) presents the average delivery days for each ship mode, providing a clear comparison of typical delivery times across different shipping options.

The chart reveals a clear hierarchy in average delivery times. Standard Class has the longest average delivery time of approximately 5 days. Second Class averages 3.2 days, First Class averages 2 days, and Same Day averages less than 1 day.

The longer average delivery time for Standard Class, which was shown in Figure [3.20] to be the most popular shipping mode, suggests that customers are prioritizing cost-effectiveness over speed. However, the significant number of customers choosing Second Class, despite the higher cost, indicates that a substantial portion of customers are willing to pay for faster delivery.

This analysis of average delivery days provides a valuable context for understanding the distribution and variability of delivery times, which will be explored further in the following box plot visualization.

- Box plot showing the distribution of delivery days by ship mode

  Figure (3.21) http://127.0.0.1:56200/

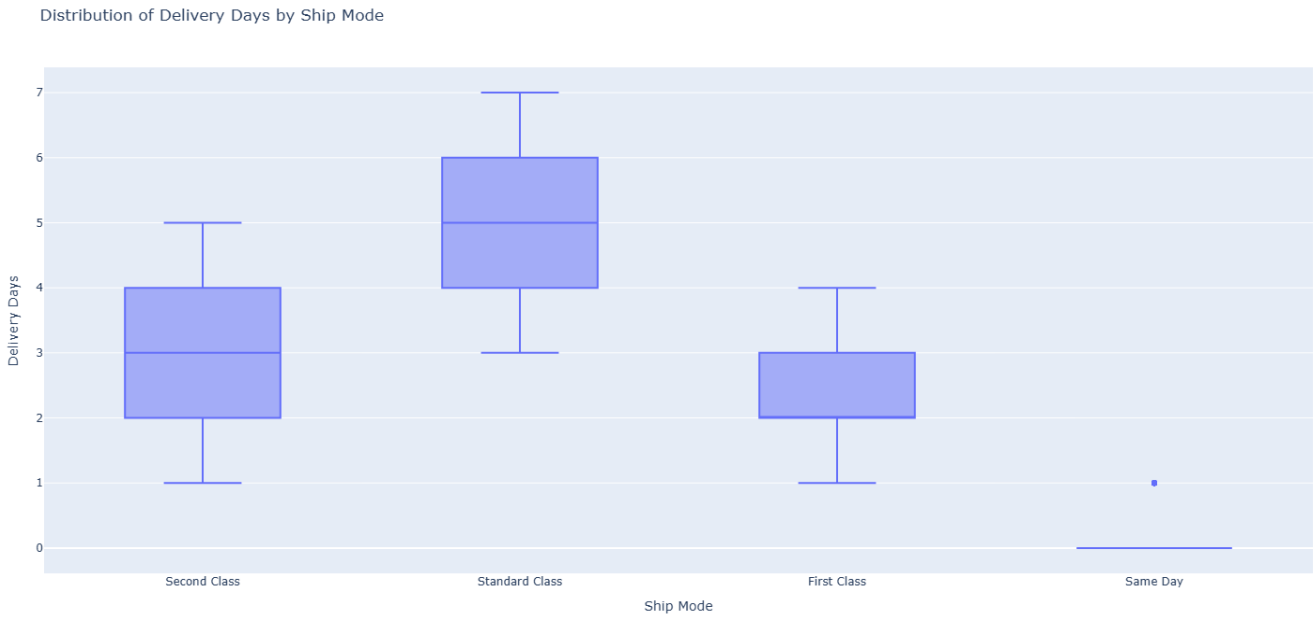


Figure (3.21) presents a box plot visualizing the distribution of delivery days by ship mode, providing insights into the variability and central tendency of delivery times for each shipping method.

The box plot reveals significant differences in delivery day distributions across ship modes. Standard Class shipping exhibits the highest variability, with a wide interquartile range and long whiskers, suggesting that while most shipments arrive within a certain timeframe, there's a significant chance of longer delivery times. Second Class shipping shows less variability than Standard Class, while First Class demonstrates very consistent delivery times. As expected, Same Day delivery has the shortest delivery time and the least variability.

The high variability in Standard Class delivery times suggests a need to investigate potential bottlenecks or inefficiencies in the shipping process for this most commonly used method. Managing customer expectations regarding delivery times for Standard Class shipments is crucial. The consistent delivery times for First Class and Same Day shipping justify the premium price paid for these services.

Based on these findings, we recommend investigating the causes of variability in Standard Class delivery times and implementing process improvements to reduce the range and improve predictability. We also suggest considering setting clearer customer expectations for Standard Class delivery times and exploring strategies to incentivize the use of faster shipping methods for customers who prioritize speed.

- Sales Distribution by Delivery Days: Understanding Customer Delivery Expectations ,using donut chart

Figure (3.22) http://127.0.0.1:56997/

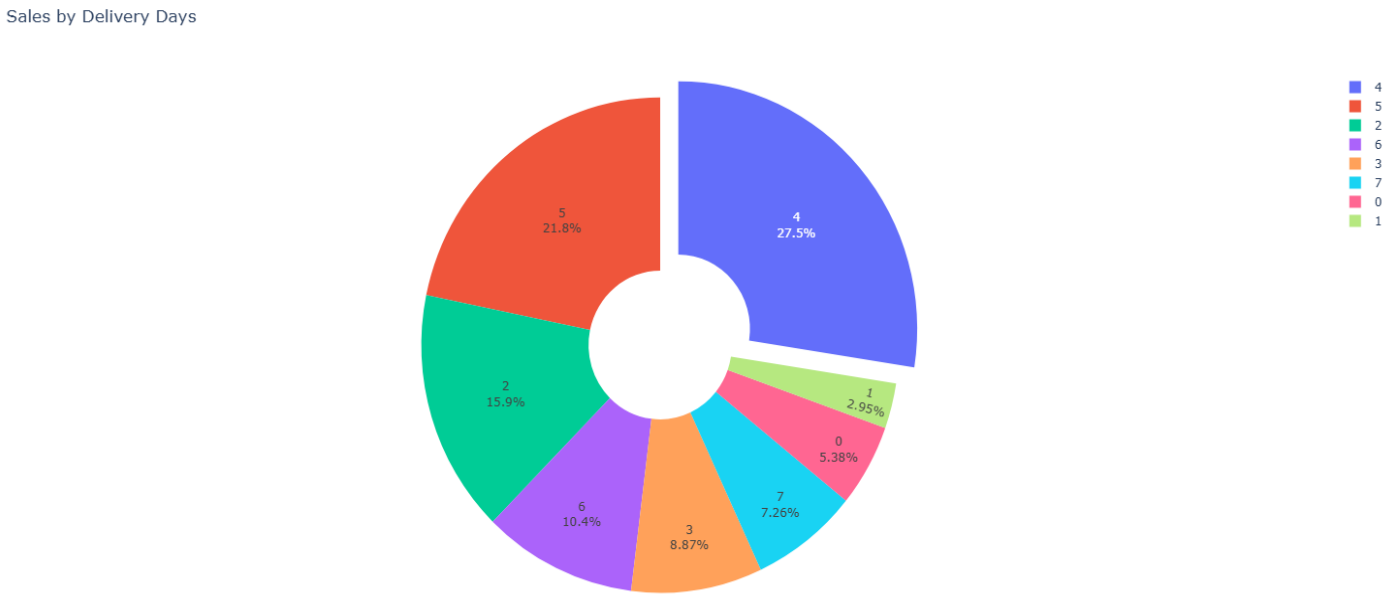


Figure (3.22) presents the proportion of sales attributed to each delivery timeframe, revealing the most common delivery experiences and potential areas for improvement.

• **Dominance of 4-Day Deliveries:** The chart clearly shows that the majority of sales (27.5%) are associated with 4-day deliveries. This suggests that a significant portion of customers receive their orders within this timeframe, indicating a potential preference or expectation for this delivery speed.

• **Significant Proportion of 5-Day Deliveries:** A substantial portion of sales (21.8%) also falls within the 5-day delivery window. Combined with the 4-day deliveries, this means that nearly half of all sales (49.3%) are delivered within 4 to 5 days, highlighting the importance of optimizing logistics and shipping processes to consistently meet this delivery expectation.

• **Faster Delivery Options:** Deliveries within 2 days (15.9%) and 3 days (10.4%) account for a notable share of sales, suggesting that a segment of customers prioritizes faster delivery and is willing to pay for expedited shipping. Analyzing the demographics and purchasing behavior of these customers could inform targeted marketing efforts for expedited shipping options, potentially increasing their adoption.

• **Longer Delivery Times:** While less frequent, deliveries taking 6 days (10.4%), 7 days (7.26%), or even longer (2.95%) represent a notable portion of sales. Investigating the reasons for these longer delivery times is crucial for identifying potential bottlenecks or inefficiencies in the shipping process. This could involve analyzing shipping routes, carrier performance, or order processing procedures. Addressing these issues could improve customer satisfaction and reduce potential complaints or returns.

• **Strategic Implications:** The insights gained from this visualization have significant implications for strategic decision-making related to logistics, shipping, and customer service. Optimizing shipping processes to meet customer delivery expectations, particularly for the most common delivery timeframes (4-5 days), is crucial for enhancing customer satisfaction and driving repeat business. Furthermore, understanding the trade-offs between cost and speed for different shipping options can inform pricing strategies and marketing efforts.

- Delivery Day Frequency: Understanding Typical Delivery Timelines ,using histogram of delivery days

Figure (3.23)

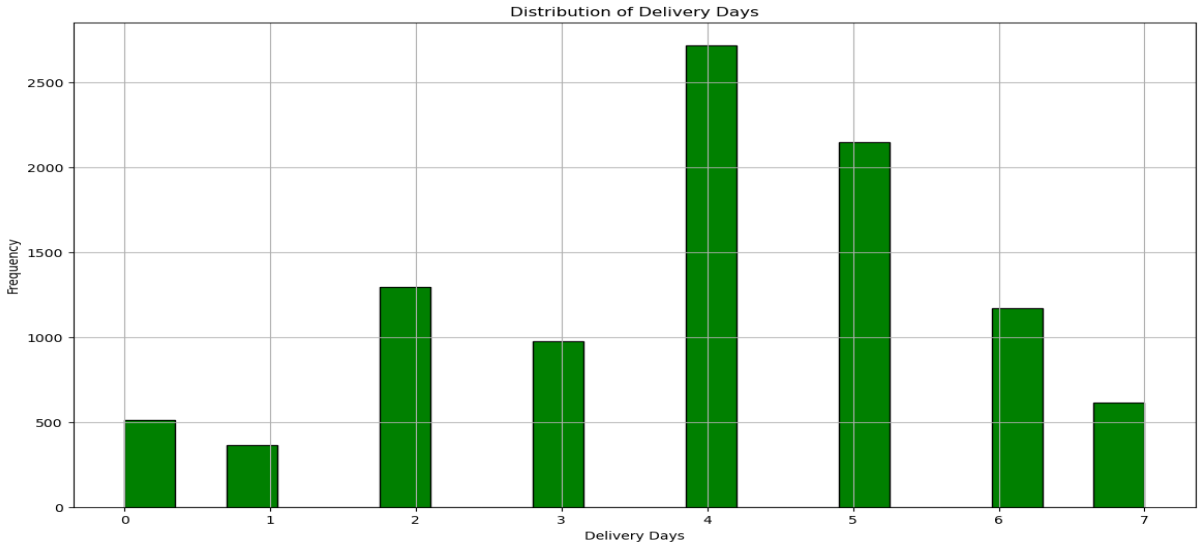


Figure (3.22) presents the distribution of sales based on delivery days, revealing the most common delivery durations and potential areas for improvement in meeting customer expectations.

Concentration of Deliveries: The histogram clearly shows a concentration of deliveries within the 2- to 6-day timeframe, with the highest frequency occurring at 4 days. This suggests that a significant portion of customers receive their orders within this window, indicating a potential expectation or standard for this delivery speed.

Dominant Delivery Time: The 4-day delivery timeframe stands out as the most frequent, accounting for 2700 deliveries, representing approximately 27% of all orders. This highlights the importance of optimizing logistics and shipping processes to consistently meet this delivery duration.

Tail Distribution: The tail extending towards longer delivery times (7 days and beyond) indicates that while less frequent, a notable portion of deliveries (approximately 15%) experience extended durations. Investigating the reasons for these longer delivery times is crucial for identifying potential bottlenecks or inefficiencies in the shipping process. This could involve analyzing shipping routes, carrier performance, or order processing procedures. Addressing these issues could improve customer satisfaction and reduce potential complaints or returns.

Potential Inefficiencies: The presence of deliveries taking 0 or 1 day, while less frequent (approximately 5% combined), may warrant attention. While these faster deliveries might seem positive, they could indicate inefficiencies in the order processing or shipping system, such as prioritizing certain orders over others or utilizing more expensive expedited shipping methods unnecessarily. Analyzing the reasons behind these expedited deliveries could lead to process improvements and more balanced resource allocation.

Optimizing shipping processes to consistently meet the most common delivery timeframes, particularly the dominant 4-day delivery, is crucial for enhancing customer satisfaction and driving repeat business. Furthermore, understanding the reasons behind longer delivery times and potential inefficiencies in expedited deliveries can inform targeted improvements to the shipping process, potentially leading to cost savings and more efficient resource allocation.

- Relative Performance of Ship Modes within Categories:

  Figure (3.23) http://127.0.0.1:56613/

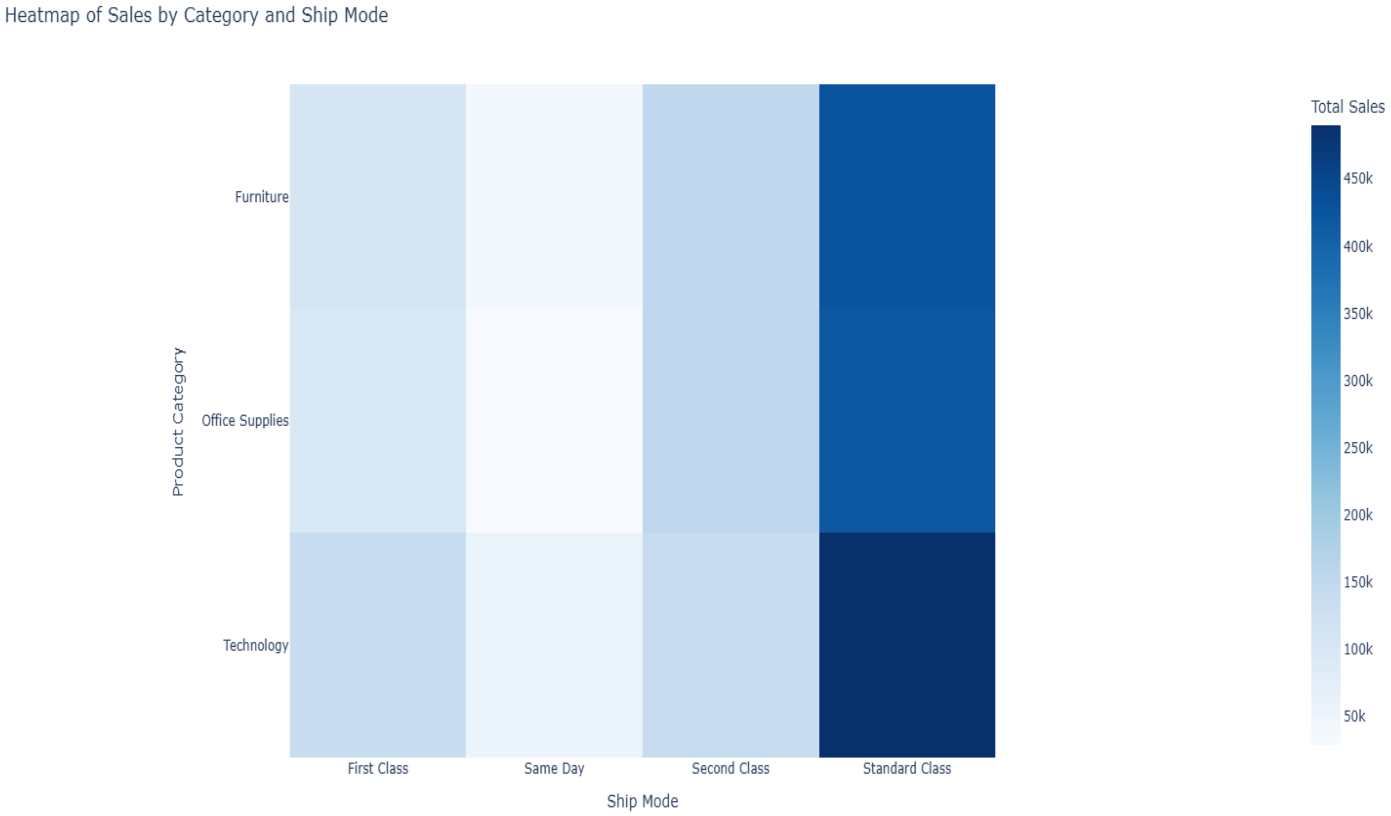


Figure (3.23) presents a heatmap visualization of sales by product category and ship mode, providing insights into the interplay between these two factors and their impact on sales performance.

The heatmap clearly highlights the dominance of Standard Class shipping across all product categories, confirming its position as the most popular shipping method. Technology products exhibit the highest sales volume, especially when shipped via Standard Class, indicating a strong preference for this combination.

Within each product category, we observe distinct patterns in ship mode preferences. For Furniture and Office Supplies, Standard Class is the clear leader, followed by Second Class. For Technology products, while Standard Class remains dominant, First Class and Same Day shipping options have a higher adoption rate compared to other categories, suggesting that customers are more willing to pay for faster delivery when purchasing Technology items.

These findings have several strategic implications. The strong performance of the Technology category suggests a high demand for these products, particularly when shipped via Standard

Class. This insight can inform inventory management and marketing strategies to capitalize on this demand. The varying ship mode preferences across categories can guide pricing and promotion decisions, with targeted offers for expedited shipping of Technology products. Furthermore, the data can be used to segment customers based on their shipping and purchasing behavior, enabling personalized marketing and recommendations.

Based on these insights, we recommend prioritizing efficient and reliable processing and delivery for Standard Class shipments, especially for Technology products. We also suggest considering promotional pricing or bundled options for expedited shipping of Technology products. Finally, we encourage exploring opportunities to segment customers based on their shipping preferences and product category purchases to personalize marketing efforts and improve customer satisfaction.

- Shipping Mode Preferences by Customer Segment

  Figure (3.24)

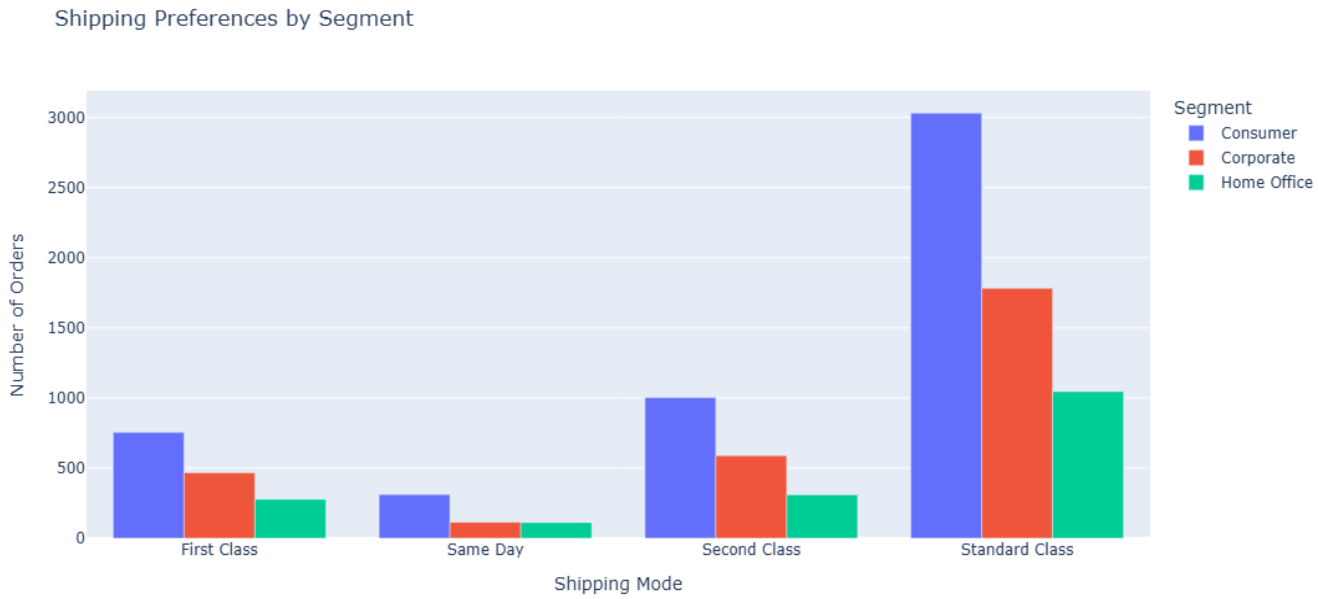


Figure (3.24) presents the distribution of shipping mode usage for each customer segment, revealing distinct patterns and highlighting opportunities for targeted strategies.

• **Standard Class Dominance Across Segments:** The chart clearly shows that Standard Class shipping is the most popular option across all customer segments (Consumer, Corporate, and Home Office). However, the degree of preference varies, with the Consumer segment exhibiting the strongest reliance on Standard Class, accounting for approximately 75% of their orders.

• **Segment-Specific Shipping Patterns:** While Standard Class is dominant overall, each segment shows unique shipping preferences:

- **Consumer:** "The Consumer segment heavily favors Standard Class, with significantly lower usage of other shipping modes. Same Day shipping is virtually non-existent, suggesting a strong price sensitivity within this segment. First Class and Second Class account for a small percentage of orders, likely driven by specific needs or urgent purchases."
- **Corporate:** "The Corporate segment utilizes a wider range of shipping modes, with a notable preference for Standard Class (50%), followed by Second Class (30%). This indicates a balance between cost considerations and the need for moderately faster delivery options. First Class and Same Day shipping are also utilized, likely for time-sensitive deliveries."
- **Home Office:** "The Home Office segment shows a relatively balanced usage of Standard (40%) and Second Class (35%) shipping, suggesting a preference for cost-effective yet reasonably timely deliveries. First Class shipping is used more frequently (20%) compared to the Consumer segment, possibly reflecting the need for faster delivery for certain business-related items."

• **Potential for Targeted Strategies:** "The distinct shipping preferences across segments allow for targeted strategies to optimize shipping options and improve customer satisfaction:

- **Consumer Segment:** "Given the strong preference for Standard Class, ensure competitive pricing and reliable delivery for this mode. Consider offering discounts or promotions for Standard Class shipping to further incentivize its use and attract price-sensitive customers. Explore options for slower, more economical shipping options to cater to the budget-conscious segment."
- **Corporate Segment:** "Offer a flexible range of shipping options, with a focus on Standard and Second Class. Negotiate volume discounts or offer tiered pricing based on delivery time to cater to the specific needs of corporate clients. Provide dedicated account management and support for shipping needs."
- **Home Office Segment:** "Promote the balance of cost and speed offered by Standard and Second Class shipping. Consider bundling options or offering tiered pricing based on delivery time to cater to the specific needs of this segment. Offer convenient online tools for managing shipments and tracking deliveries."

• **Strategic Implications:** "The insights from this visualization have significant implications for tailoring shipping strategies to each customer segment. Understanding their unique needs and preferences allows for optimized logistics, competitive pricing, and targeted marketing efforts, ultimately leading to improved customer satisfaction and loyalty."

V. Conclusion

This exploratory data analysis (EDA) has provided valuable insights into sales performance, customer behavior, and shipping preferences within the context of portfolio project . By visualizing and analyzing data related to product categories, ship mode, delivery days, and customer segments, we have identified key patterns and trends that can inform strategic decision-making and drive business growth.

Key findings from this analysis include:

- **West Region Dominance:** The West region accounts for 45% of total sales, making it the dominant revenue generator.
- **Standard Class Variability:** Standard Class shipping is the most popular choice among customers, but its high variability in delivery days (as shown by the wide interquartile range in the box plot) presents a potential challenge to customer satisfaction.
- **Technology and Standard Class Synergy:** Technology products are the top sellers, especially when shipped via Standard Class, indicating a strong preference for this combination.
- **Segment-Specific Shipping Preferences:** Distinct shipping preferences exist across customer segments. Consumers heavily favor Standard Class (75% of orders), while Corporate clients utilize a wider range of options, including Second Class (30%).

These findings have several important implications for inventory management, marketing strategies, logistics optimization, and customer service. Based on this analysis, we recommend the following key actions:

1. **Prioritize Technology in the West:** Prioritize inventory and marketing efforts for Technology products in the West region to capitalize on high demand.
2. **Improve Standard Class Reliability:** Investigate the root causes of variability in Standard Class delivery times and implement process improvements to enhance predictability and customer satisfaction. This could involve negotiating better rates with carriers or optimizing delivery routes.
3. **Targeted Marketing by Segment:** Develop targeted marketing campaigns and pricing strategies for each customer segment, considering their unique shipping preferences and product interests. Offer bundled options or tiered pricing to incentivize specific shipping choices.
4. **Negotiate Shipping Rates:** Explore opportunities to negotiate better shipping rates with carriers, leveraging the high volume of Standard Class shipments to reduce costs.

While this EDA has provided valuable insights, it is important to acknowledge certain limitations. The analysis was limited by the lack of detailed customer demographic data and competitor information. Future research could explore the impact of seasonality on sales trends and the effectiveness of different marketing channels in reaching specific customer segments.

In conclusion, this EDA has provided a comprehensive overview of sales patterns, customer behavior, and shipping preferences. By implementing the recommendations outlined above,

Superstores USA can optimize operations, improve customer satisfaction, and drive sustainable business growth."

VI. Conclusion

The exploratory data analysis (EDA) of the Superstore sales dataset has provided valuable insights into sales trends, customer behavior, product performance, and regional variations. This analysis highlights several key findings and recommendations that can drive strategic decision-making and improve business operations.

- ## Key Findings

1. **Seasonal Sales Fluctuations and Trends**:
    - The analysis revealed a consistent surge in sales during November and December across all years, likely driven by holiday shopping and seasonal promotions. A potential secondary peak in March and September was also observed, which may relate to back-to-school shopping or specific promotions.
    - A noticeable upward trend in overall sales volume was observed from 2015 to 2018, suggesting increasing market share and successful marketing initiatives.
2. **Customer Segmentation and Behavior**:
    - The Consumer segment contributes the most to overall sales, followed by the Corporate and Home Office segments. The Corporate segment, despite having a smaller customer base, generates higher sales per customer, indicating the potential value of nurturing corporate relationships.
    - A detailed analysis of customer segments and purchase frequency revealed that high-value customers across all segments should be targeted for personalized marketing efforts to maximize their lifetime value.
3. **Product Performance**:
    - The Technology category dominates overall sales, with Phones and Accessories being the top performers. Binders and Paper are the strongest performers in the Office Supplies category, while Chairs and Furnishings lead sales within the Furniture category.
    - The analysis of the top 20 products by sales revenue and their sales trends over time provides valuable insights into the performance of specific products and highlights opportunities for targeted marketing and inventory management.
4. **Geographic Performance**:
    - The West region is the top performer in terms of sales, followed by the East, Central, and South regions. Major metropolitan areas such as New York City, Los Angeles, and Seattle are key revenue hubs.
    - The analysis suggests the need for regionally tailored marketing strategies and resource allocation to optimize sales performance across different regions.
5. **Shipping Mode Analysis**:
    - Standard Class shipping is the most popular option among customers, accounting for the majority of sales. However, the high variability in delivery times for Standard Class shipping suggests a need for process improvements.

- The analysis of delivery days and customer preferences for different shipping modes can inform strategies to optimize logistics and enhance customer satisfaction.

Recommendations

1. **Strategic Planning for Peak Seasons**:
   - Focus on inventory management, staffing, and targeted marketing campaigns to optimize sales during peak months like November and December.
   - Investigate the drivers behind secondary peaks in sales during months like March and September to leverage these insights for future promotions.
2. **Customer-Centric Strategies**:
   - Develop segment-specific marketing campaigns to drive repeat purchases and increase customer lifetime value. Personalized account management and exclusive promotions can help retain high-value customers.
   - Implement targeted strategies to increase purchase frequency among existing customers, particularly within the Home Office segment.
3. **Optimizing Product Offerings**:
   - Prioritize top-performing sub-categories for inventory management and marketing efforts. Investigate the underperformance of certain sub-categories, especially within the Furniture category, to develop effective improvement strategies.
4. **Regional and Local Market Focus**:
   - Develop tailored marketing campaigns and distribution strategies for high-performing regions and cities. Investigate the reasons for lower sales in certain regions to develop targeted interventions.
5. **Shipping and Logistics Improvements**:
   - Investigate the causes of variability in Standard Class delivery times and implement process improvements to reduce the range and improve predictability.
   - Consider promotional pricing or bundled options for expedited shipping of high-demand products.

By leveraging the insights gained from this analysis, businesses can make informed decisions, improve operational efficiency, enhance customer satisfaction, and ultimately drive sustainable growth.

- For additional information on the EDA Superstore Sales Project, please visit **here**

https://drive.google.com/file/d/1cHvOFU-__rUZ9B66F9t4iVHs4bfHQXj7/view?usp=sharing