

Superstore Sales and Profit Analysis Documentation

Introduction

The rise of online shopping has transformed the way people shop, offering them unparalleled convenience and the ability to find great deals without leaving their homes or offices.

One of the most appealing aspects of online shopping, especially during the holiday season, is its ability to eliminate the need for waiting in long lines or tirelessly searching through multiple stores to find a specific item. This shift in the shopping landscape has made it easier than ever for consumers to access a wide range of products and make purchases with just a few clicks, enhancing the overall shopping experience.

Data Summary

The provided data includes the following columns:

Order ID: A unique identifier for each order placed by customers.

Order Date: The date when the order was placed by the customer.

Ship Date: The date when the order was shipped to the customer.

Ship Mode: The mode of shipment selected by the customer (e.g., Standard Class, Second Class, etc.).

Customer ID: A unique identifier for each customer.

Customer Name: The name of the customer who placed the order.

Segment: The customer segment to which the customer belongs (e.g., Consumer, Corporate, Home Office).

Country: The country where the order was placed, likely "United States" based on the provided information.

City: The city where the order was placed.

State: The state where the order was placed (e.g., Kentucky).

Postal Code: The postal code of the location where the order was placed.

Region: The region of the country where the order was placed (e.g., South).

Product ID: A unique identifier for each product in the order.

Category: The broad category of the product (e.g., Furniture, Office Supplies, Technology).

Sub-Category: The specific sub-category of the product (e.g., Bookcases, Chairs, Phones).

Product Name: The name of the product that was purchased.

Sales: The total sales revenue generated by the order.

Quantity: The quantity of items purchased in the order.

Discount: The discount applied to the order.

Profit: The profit generated from the order.

Data Preprocessing

Step 1: Dropping Irrelevant Columns

Step 2: Converting to DateTime

Converted 'Order Date' column to DateTime

Step 3: Creating Date-Based Columns

Created 'Order Month', 'Order Year', 'Order Day of Week' columns

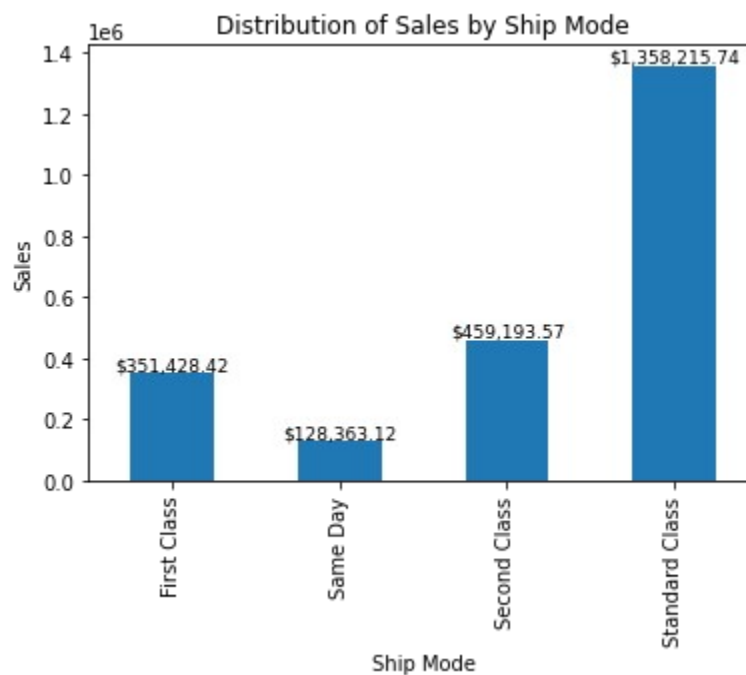
Step 4: Dropping 'Order Date' Column
Removed 'Order Date' column from the dataset

Question 1: Distribution of Sales Across Different Ship Modes

The distribution of sales across different ship modes is as follows:

First Class: The total sales for this ship mode amount to approximately 351,428.40.
Same Day: The total sales for this ship mode amount to approximately 128,363.10.
Second Class: The total sales for this ship mode amount to approximately 459,193.60.
Standard Class: The total sales for this ship mode amount to approximately 1,358,216.00.

This indicates that the Standard Class ship mode has the highest sales volume among all the ship modes, followed by Second Class. First Class has the lowest sales volume among the four ship modes. It suggests that customers tend to choose the Standard Class for their shipping needs, which contributes significantly to the overall sales.¶



Question 2a: Which top 10 customers are making the most purchases in terms of quantity?

The results show the top 10 customers who are making the most purchases in terms of quantity.

Jonathan Doherty: This customer has made the highest number of purchases with a quantity of 150 items.

William Brown: This customer ranks second with a total purchase quantity of 146 items.

John Lee: With a quantity of 143 items, this customer holds the third position in terms of purchase quantity.

Paul Prost: This customer has made 138 purchases, securing the fourth position.

Steven Cartwright: With a quantity of 133 items, this customer ranks fifth in terms of purchase quantity.

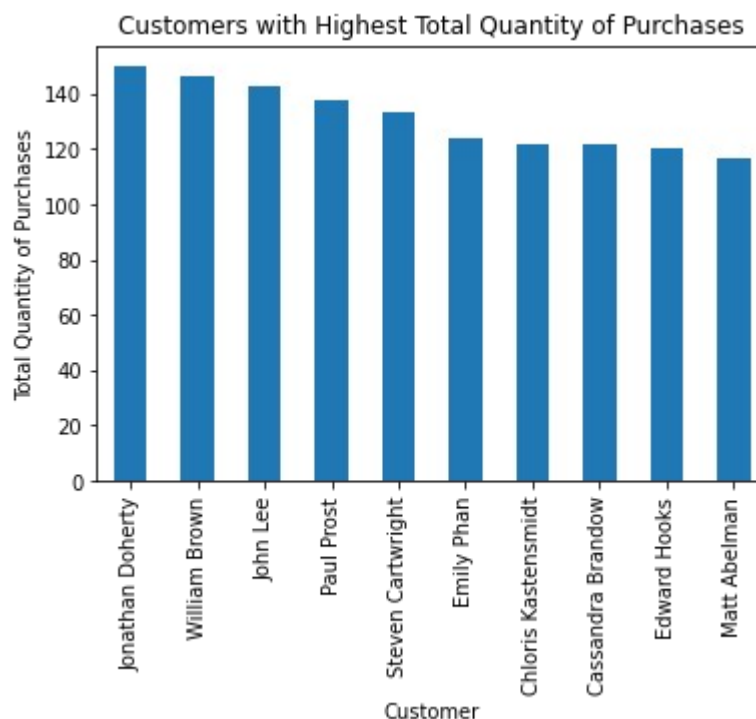
Emily Phan: This customer holds the sixth position with a total purchase quantity of 124 items.

Chloris Kastensmidt: Tied with Cassandra Brandow, both customers have made 122 purchases, securing the seventh position.

Edward Hooks: This customer ranks eighth with a total purchase quantity of 120 items.

Matt Abelman: With a quantity of 117 items, this customer holds the ninth position.

These top 10 customers have made a significant number of purchases, indicating their high level of engagement and potential value to the business in terms of quantity-based sales.



Question 2b: Which top 10 customers are making the most Purchases in terms of Dollars?

The results show the top 10 customers who are making the most purchases in terms of dollars.

Sean Miller: This customer has made the highest purchases in terms of dollars, with a total expenditure of 25,043.05.

Tamara Chand: With a total expenditure of 19,052.22, this customer ranks second in terms of purchase value.

Raymond Buch: This customer holds the third position with purchases amounting to 15,117.34.

Tom Ashbrook: With a total expenditure of 14,595.62, this customer secures the fourth position.

Adrian Barton: This customer ranks fifth with purchases totaling 14,473.57.

Ken Lonsdale: With a total expenditure of 14,175.23, this customer holds the sixth position.

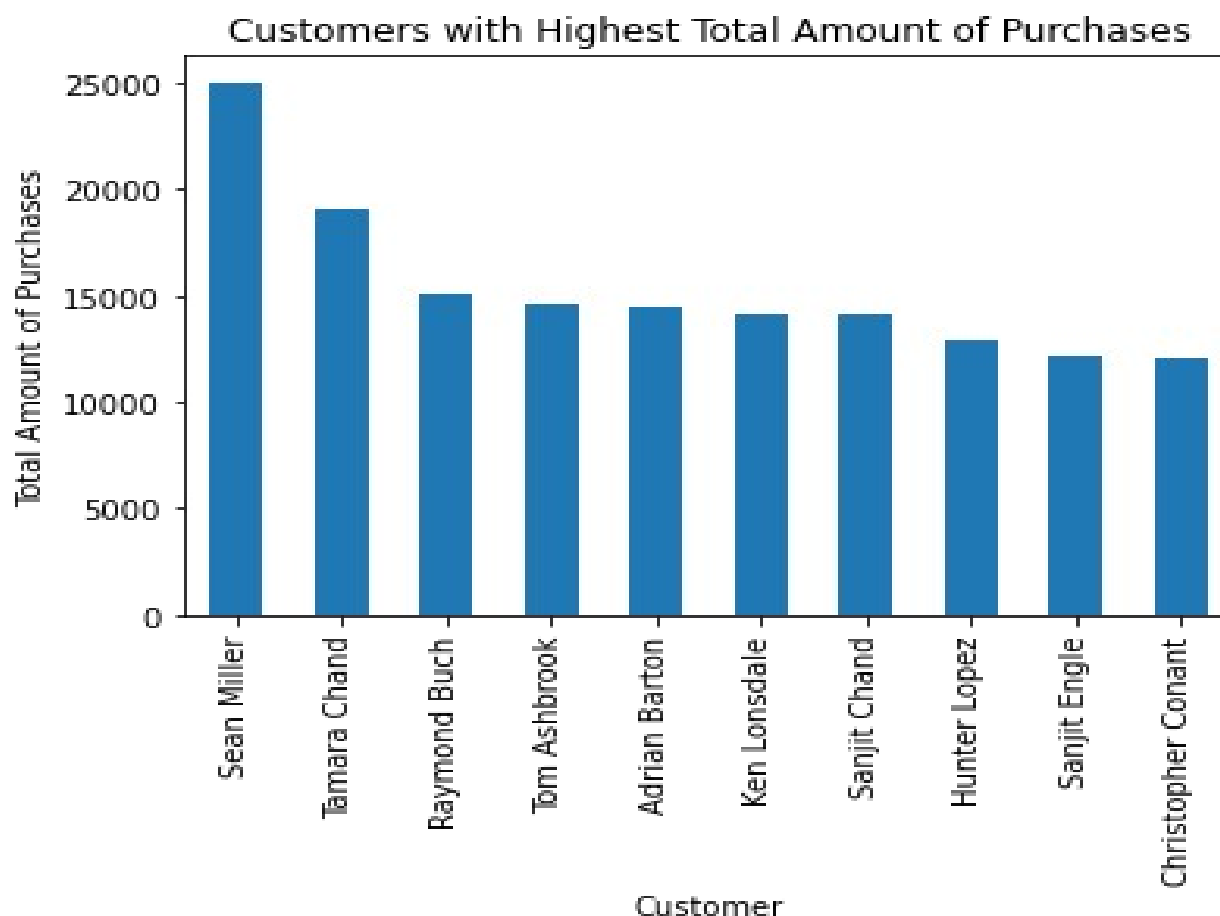
Sanjit Chand: Tied with Hunter Lopez, both customers have made purchases totaling 14,142.33, securing the seventh position.

Hunter Lopez: Tied with Sanjit Chand, this customer also has a total expenditure of 14,142.33, holding the seventh position.

Sanjit Engle: This customer ranks eighth with a total expenditure of 12,209.44.

Christopher Conant: With purchases amounting to 12,129.07, this customer holds the tenth position.

These top 10 customers have made substantial purchases in terms of dollars, indicating their high value to the business. They contribute significantly to the company's revenue and are important customers to focus on for maintaining and expanding business relationships.



Question 3: What is the breakdown of sales by customer segment?

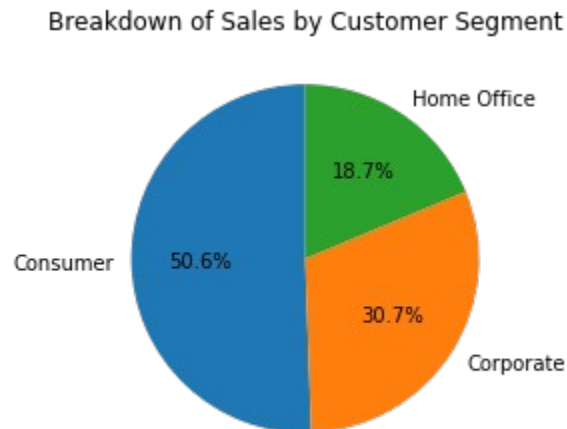
The breakdown of sales by customer segment is as follows:

Consumer: The total sales from the consumer segment amount to approximately 1,161,401.00.

Corporate: The total sales from the corporate segment amount to approximately 706,146.40.

Home Office: The total sales from the home office segment amount to approximately 429,653.10.

This breakdown provides insights into the sales distribution across different customer segments. It shows that the consumer segment has the highest sales volume, followed by the corporate segment. The home office segment has the lowest sales volume among the three segments. This information can be useful for analyzing the sales performance and targeting marketing strategies towards specific customer segments to maximize revenue and profitability.



Question 4:

- a- Which cities have the highest sales?**
- b- Which states have the highest sales?**
- c- Which regions have the highest sales?**

a) The cities with the highest sales are:

New York City: Sales in New York City amount to approximately 256,368.16.
Los Angeles: Sales in Los Angeles amount to approximately 175,851.34.
Seattle: Sales in Seattle amount to approximately 119,540.74.
San Francisco: Sales in San Francisco amount to approximately 112,669.09.
Philadelphia: Sales in Philadelphia amount to approximately 109,077.01.

These cities have the highest sales figures, indicating that they are key markets for the business and have a significant customer base.

b) The states with the highest sales are:

California: Sales in California amount to approximately 457,687.63.
New York: Sales in New York amount to approximately 310,876.27.
Texas: Sales in Texas amount to approximately 170,188.05.
Washington: Sales in Washington amount to approximately 138,641.27.
Pennsylvania: Sales in Pennsylvania amount to approximately 116,511.91.

These states have the highest sales figures, suggesting that they are important regions for the business and have a strong customer demand.

c) The regions with the highest sales are:

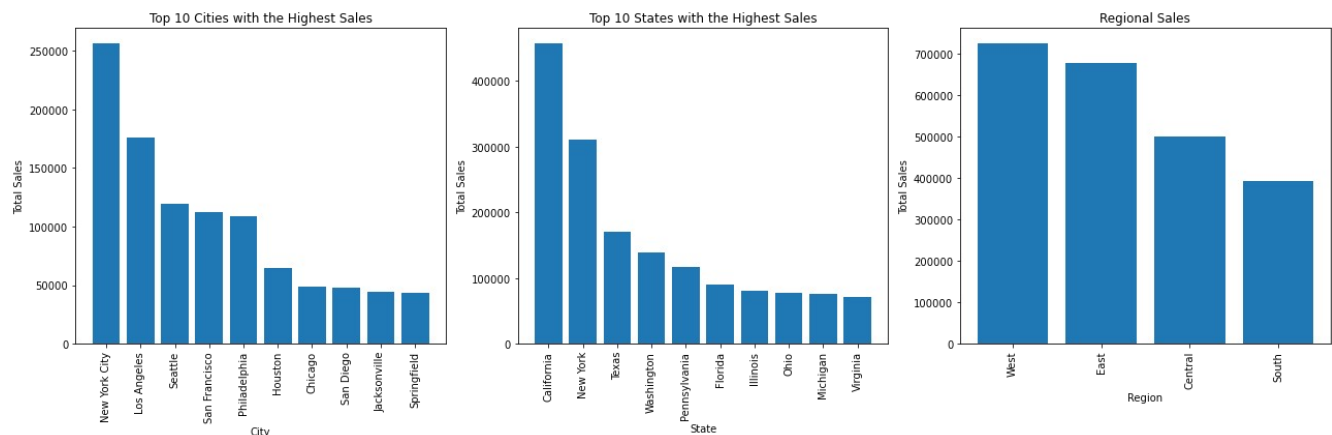
West: Sales in the West region amount to approximately 725,457.82.

East: Sales in the East region amount to approximately 678,781.24.

Central: Sales in the Central region amount to approximately 501,239.89.

South: Sales in the South region amount to approximately 391,721.91.

These regions have the highest sales figures, indicating that they are significant markets for the business. Understanding the sales distribution across different regions can help in identifying regional trends, customer preferences, and tailoring marketing strategies accordingly.



Question 5: What is the distribution of sales across different categories and sub-categories?

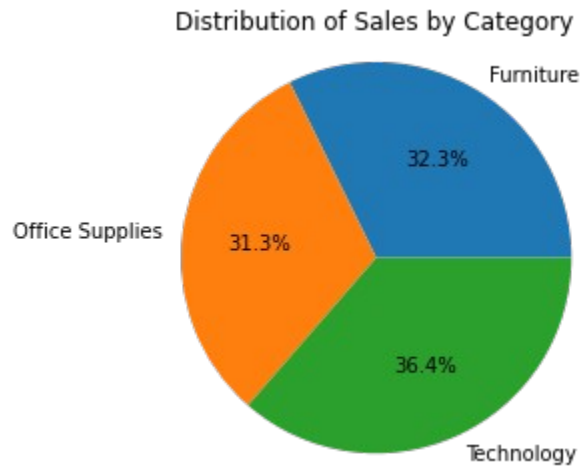
The results show the distribution of sales across different categories and sub-categories.

Sales by Category:

Furniture: The total sales in the Furniture category amount to 741,999.7953.

Office Supplies: The Office Supplies category has generated sales of 719,047.0320.

Technology: The Technology category has the highest sales, totaling 836,154.0330.



Sales by Sub-Category:

Furniture:

Bookcases: Sales of bookcases amount to 114,879.9963.

Chairs: The sales of chairs reach 328,449.1030.

Furnishings: Sales of furnishings total 91,705.1640.

Tables: Tables have generated sales of 206,965.5320.

Office Supplies:

Appliances: Sales of appliances reach 107,532.1610.

Art: The art category has generated sales of 27,118.7920.

Binders: Binders have generated sales of 203,412.7330.

Envelopes: Sales of envelopes total 16,476.4020.

Fasteners: Fasteners have generated sales of 3,024.2800.

Labels: The sales of labels reach 12,486.3120.

Paper: Paper products have generated sales of 78,479.2060.

Storage: Sales of storage products amount to 223,843.6080.

Supplies: Supplies have generated sales of 46,673.5380.

Technology:

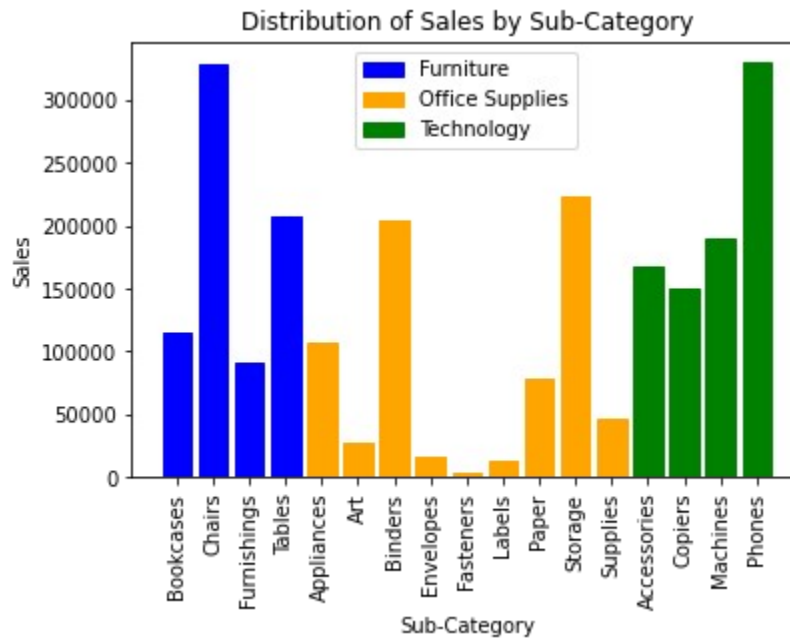
Accessories: Sales of accessories in the Technology category amount to 167,380.3180.

Copiers: The sales of copiers reach 149,528.0300.

Machines: Sales of machines in the Technology category total 189,238.6310.

Phones: Phones have generated the highest sales, reaching 330,007.0540.

These results provide insights into the sales performance of different categories and sub-categories. Technology products, particularly phones, have the highest sales, followed by furniture and office supplies. Among the sub-categories, chairs, tables, and phones have generated significant sales, indicating their popularity and demand.



Question 6: What is the distribution of profit across different categories and sub-categories?

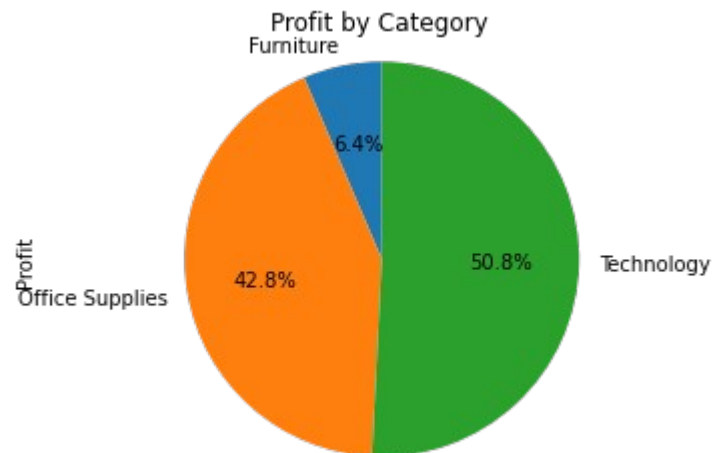
The distribution of profit across different categories and sub-categories is as follows:

Profit by Category:

Furniture: The total profit in the furniture category is 18,451.2728.

Office Supplies: The total profit in the office supplies category is 122,490.8008.

Technology: The total profit in the technology category is 145,454.9481.



Profit by Sub-Category within Categories:

Furniture:

Bookcases: The profit generated from bookcases is 114,879.9963.
 Chairs: The profit generated from chairs is 328,449.1030.
 Furnishings: The profit generated from furnishings is 91,705.1640.
 Tables: The profit generated from tables is 206,965.5320.

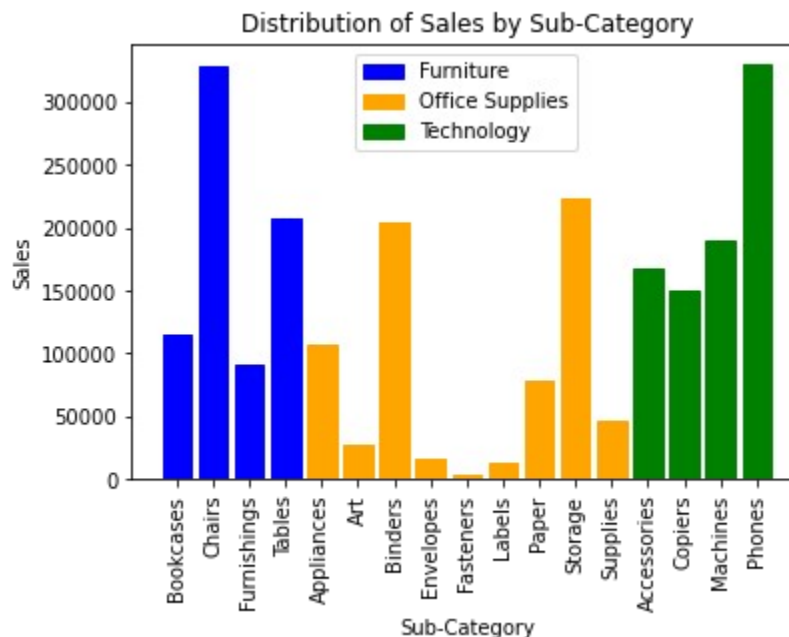
Office Supplies:

Appliances: The profit generated from appliances is 107,532.1610.
 Art: The profit generated from art supplies is 27,118.7920.
 Binders: The profit generated from binders is 203,412.7330.
 Envelopes: The profit generated from envelopes is 16,476.4020.
 Fasteners: The profit generated from fasteners is 3,024.2800.
 Labels: The profit generated from labels is 12,486.3120.
 Paper: The profit generated from paper is 78,479.2060.
 Storage: The profit generated from storage supplies is 223,843.6080.
 Supplies: The profit generated from miscellaneous supplies is 46,673.5380.

Technology:

Accessories: The profit generated from technology accessories is 167,380.3180.
 Copiers: The profit generated from copiers is 149,528.0300.
 Machines: The profit generated from machines is 189,238.6310.
 Phones: The profit generated from phones is 330,007.0540.

These results provide insights into the profit distribution across different categories and sub-categories. It shows that technology products, specifically phones, generate the highest profit, followed by office supplies and then furniture. Within each category, there are sub-categories that contribute to the overall profit.



Question 6: Sales VS Profit by Categories and Sub-Categories

Furniture Category:

Total Sales: Approximately \$742,000

Total Profit: Approximately \$18,451

Furniture has generated substantial sales, but the profit margin is relatively low, indicating potential profitability challenges in this category.

Office Supplies Category:

Total Sales: Approximately \$719,000

Total Profit: Approximately \$122,491

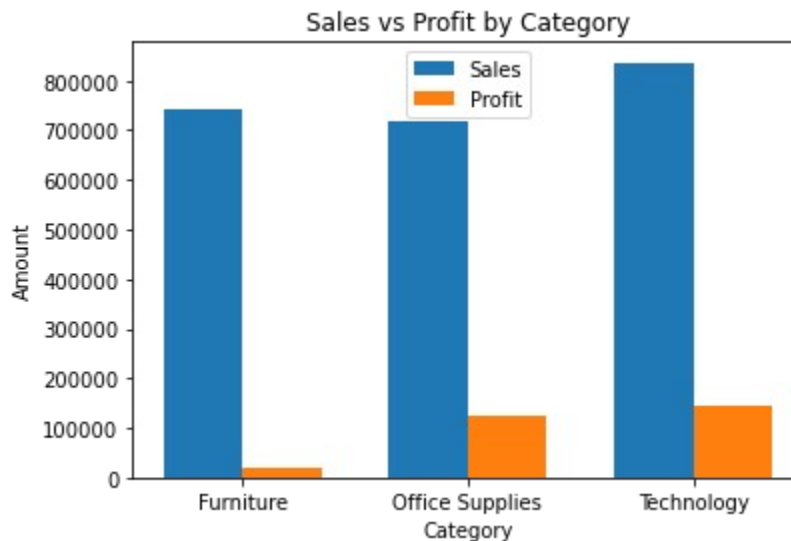
Office supplies have also generated significant sales, with a relatively higher profit margin compared to furniture.

Technology Category:

Total Sales: Approximately \$836,000

Total Profit: Approximately \$145,455

Technology products have the highest sales among the categories, and they also yield a reasonable profit margin, making this category the most profitable.



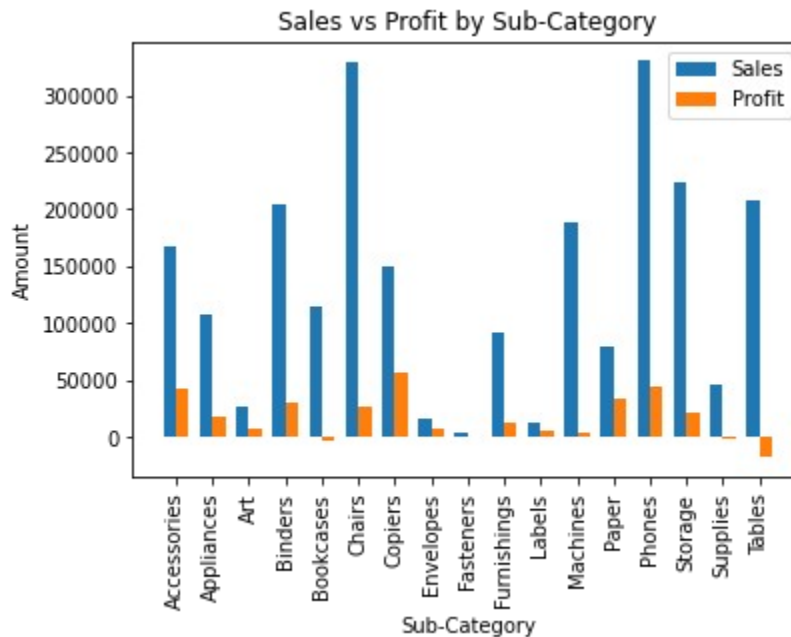
Sales vs. Profit by Sub-Categories:

Among the sub-categories, we see variations in both sales and profitability:

Accessories and Phones have the highest sales figures.

Phones, Copiers, and Accessories have the highest profit margins, indicating their strong contribution to overall profitability.

Notably, some sub-categories, such as Bookcases, Tables, and Supplies, are generating sales but are incurring losses, as indicated by negative profit figures.



While the Furniture category generates significant sales, its profit margin is comparatively lower.

The Technology category is the most profitable, driven by products like Phones and Copiers.

Some sub-categories, like Bookcases, Tables, and Supplies, may require attention to improve profitability or make strategic decisions.

Question 8: Sales VS Profit by Order Year and Order Month

Sales vs. Profit by Order Month:

On January, Sales amount to approximately \$94,925, and the associated profit is around \$9,134. This suggests a positive start to the year but with a relatively lower profit margin.

On September, Sales spike significantly to approximately \$307,650, accompanied by a profit of approximately \$36,857. This indicates a robust performance in terms of both sales and profit during this month.

On November, Sales are particularly high in November, totaling approximately \$352,461, with a profit of approximately \$35,468. This suggests a strong performance, likely due to holiday season sales.

On December, Similar to November, December also shows substantial sales of around \$325,294, but with a higher profit of approximately \$43,369. The holiday season contributes significantly to sales and profit in these months.

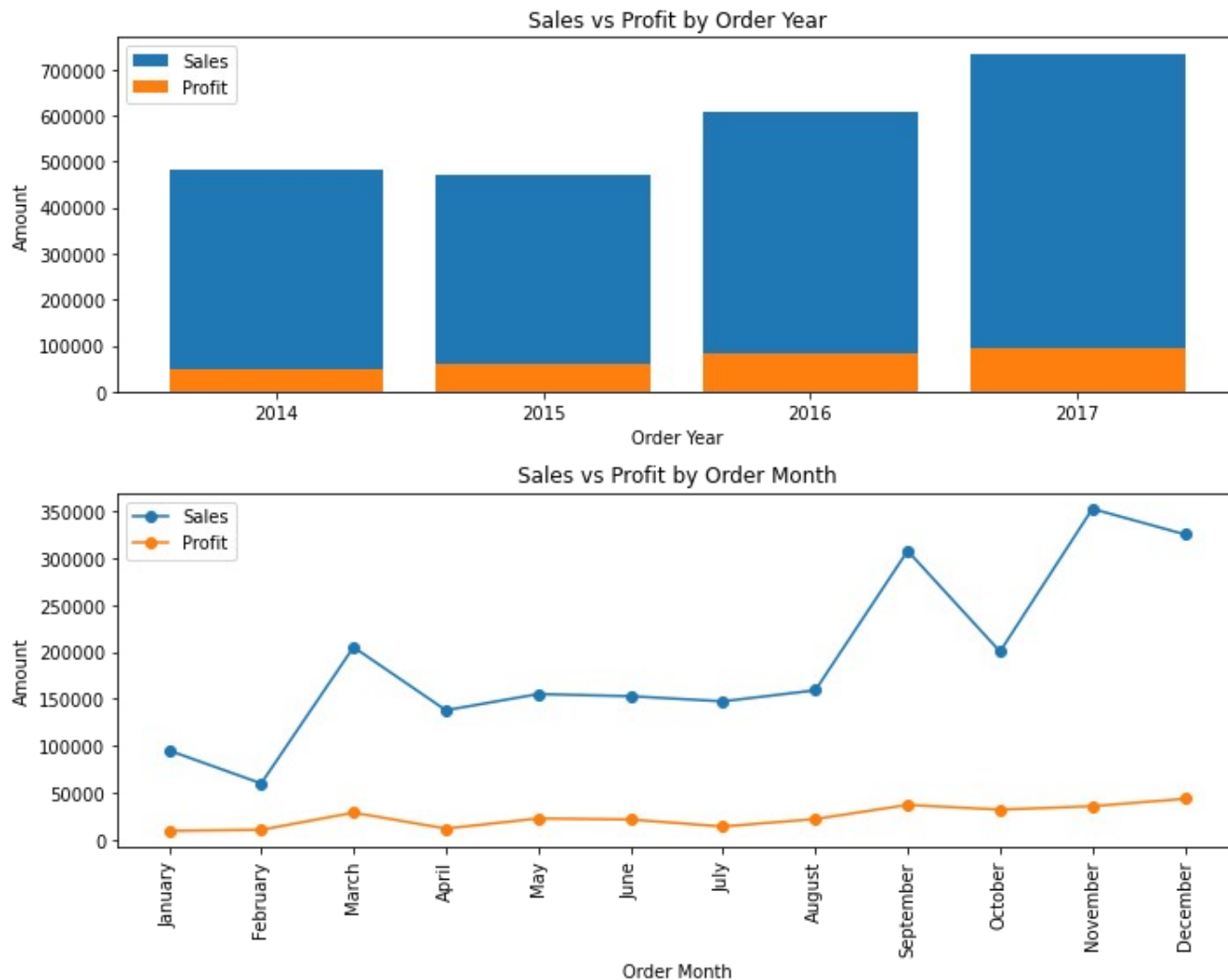
Sales vs. Profit by Order Year:

In 2017, has the highest sales, totaling approximately \$733,215, along with a profit of approximately \$93,439. It appears to be a highly successful year in terms of both sales and profit.

In 2016, also performed well, with sales reaching approximately \$609,206 and a profit of approximately \$81,795.

In 2015, there were total sales of approximately \$470,533 and a profit of approximately \$61,619.

The dataset starts with 2014, which had sales of approximately \$484,247 and a profit of approximately \$49,544.



Sales tend to be higher during the holiday season, particularly in November and December.

September also stands out as a strong month in terms of both sales and profit.

The year 2017 was the most successful, with the highest sales and profit figures.

The analysis by order year and month can help identify seasonal trends and guide marketing and inventory management strategies to maximize sales and profit during peak periods.

Question 9: Are there any correlations between variables like sales, quantity, discount, and profit?

The correlation matrix shows the correlation coefficients between different variables: Sales, Quantity, Discount, and Profit.

Sales and Profit have a positive correlation of approximately 0.48, indicating that as sales increase, there is a tendency for profits to increase as well. This suggests a positive relationship between sales and profit.

Sales and Quantity have a positive correlation of approximately 0.20, indicating that as

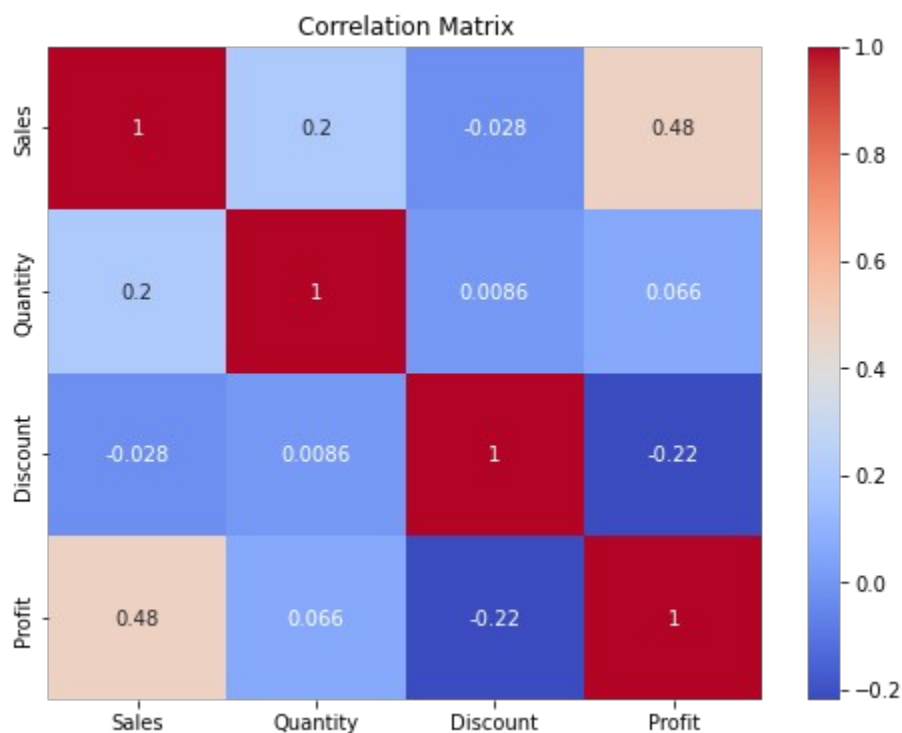
sales increase, there is a tendency for the quantity of items sold to increase as well. This suggests a positive relationship between sales and quantity.

Sales and Discount have a negative correlation of approximately -0.03, suggesting a weak negative relationship. This indicates that as the discount offered on products increases, there is a slight tendency for sales to decrease. However, the correlation coefficient is close to zero, indicating a weak correlation.

Profit and Quantity have a weak positive correlation of approximately 0.07, indicating that there is a slight tendency for profits to increase with higher quantities of items sold. However, the correlation coefficient is close to zero, suggesting a weak correlation.

Profit and Discount have a negative correlation of approximately -0.22, indicating that as the discount offered on products increases, there is a tendency for profits to decrease. This suggests a negative relationship between profit and discount, but the correlation is moderate in strength.

Overall, the correlation matrix provides insights into the relationships between these variables. It suggests that sales and profit have a stronger positive correlation, while profit and discount have a moderate negative correlation. The correlations between sales/quantity and sales/discount are weaker, indicating less pronounced relationships.



Conclusion

The analysis of Superstore sales and profit data provides valuable insights into various aspects of the business, customer behavior, and product performance.

The rise of online shopping has brought unprecedented convenience to customers, allowing them to access a wide range of products without the need to visit physical stores. This convenience is particularly attractive during the holiday season.

The analysis reveals that the Standard Class ship mode is the most popular choice

among customers, contributing significantly to overall sales. This indicates the importance of efficient and cost-effective shipping options.

The identification of top customers based on purchase quantity and expenditure highlights the individuals who make the most substantial contributions to the business. Building and maintaining relationships with these high-value customers can be beneficial.

Sales distribution across different customer segments reveals that the consumer segment generates the highest sales, followed by corporate and home office segments. Tailoring marketing strategies to specific customer segments can optimize revenue. The analysis of sales at the city, state, and regional levels provides insights into key markets for the business. Understanding regional sales trends can inform marketing and distribution strategies.

Technology products, particularly phones, generate the highest sales and profits. Furniture, while generating significant sales, has a lower profit margin. Some sub-categories may require attention to improve profitability.

Sales tend to peak during the holiday season, particularly in November and December. The year 2017 was the most successful in terms of both sales and profit. Seasonal trends can inform inventory and marketing strategies.

Positive correlations exist between sales and profit, as well as between sales and quantity. There is a negative correlation between profit and discount, suggesting that offering higher discounts may impact profitability negatively.