

# Superstore market analysis

## Team members

Mohammed Said

Ahmed Shafik

Aya Mohammed

Ammar Mohammed

Dina Hany

Tayseer Ashraf

## Introduction

The goal of the Superstore data analysis is to gain insights into the store's sales and profitability across various regions, product categories, and customer segments. By examining trends in sales and profit, we aim to identify high-performing products, understand regional performance, and uncover factors that affect profit margins. This analysis will provide actionable insights for optimizing pricing strategies, improving product offerings, and enhancing overall business performance. The findings will help the store make data-driven decisions to maximize revenue and profitability.

## Description of the data

**Row ID:** A unique identifier for each row in the dataset.

**Order ID:** A unique identifier for each order placed by a customer.

**Order Date:** The date on which the order was placed.

**Ship Date:** The date on which the order was shipped.

**Ship Mode:** The shipping method chosen for the order (e.g., standard, first class).

**Customer ID:** A unique identifier for each customer.

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

**Segment:** The customer segment (e.g., Consumer, Corporate, Home Office).

**City:** The city where the customer is located.

**State:** The state where the customer is located.

**Postal Code:** The postal code associated with the customer's location.

**Region:** The geographical region where the order was placed (e.g., East, West).

**Product ID:** A unique identifier for each product sold.

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

**Sub-Category:** A more specific classification within the product category (e.g., Chairs).

**Product Name:** The name of the product sold.

**Sales:** The total sales amount for the product in the order.

**Year:** The year in which the order was placed.

**Month:** The month in which the order was placed.

## Objective

The primary objective of this analysis is to derive actionable business insights from the Superstore dataset by answering key questions such as:

- **What are the top-selling products and which product categories generate the highest revenue?**
- **Which regions and cities contribute the most to overall sales and profitability?**
- **How does customer segmentation (e.g., Consumer, Corporate, Home Office) affect purchasing behavior and profitability?**
- **What is the impact of different shipping modes on sales and delivery times?**
- **Are there any seasonal trends in sales performance based on the month or year of the orders?**

By addressing these questions, the analysis aims to help the business optimize product offerings, refine pricing strategies, improve logistics, and better target high-value customer segments.

## Data processing and cleaning

In this step of analysis, we check if there are duplicates or nulls in the data. After examining the dataset for duplicates and null values, it was found that there were no duplicates and only 11 null entries. Given the small number of null values relative to the overall dataset size, they were removed to maintain data integrity without impacting the analysis.

```
Total Null Values before deletion: 11
Total Null Values after deletion: 0
All null values have been deleted.
```

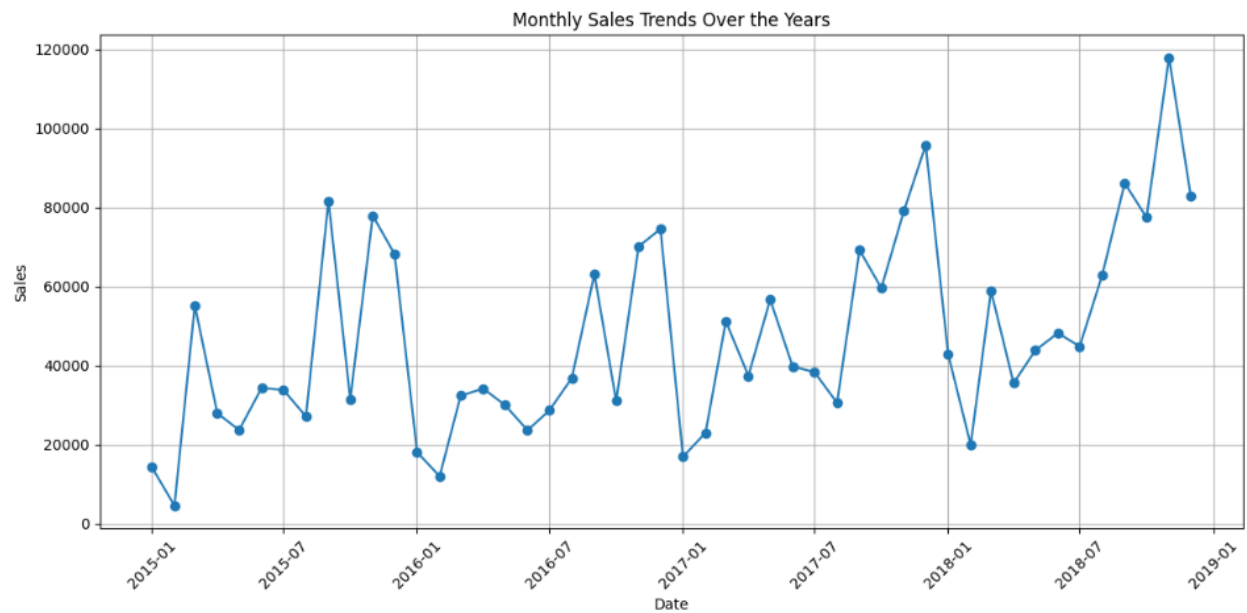
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9800 entries, 0 to 9799
Data columns (total 18 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Row ID                9800 non-null   int64
1   Order ID              9800 non-null   object
2   Order Date            9800 non-null   object
3   Ship Date             9800 non-null   object
4   Ship Mode             9800 non-null   object
5   Customer ID           9800 non-null   object
6   Customer Name         9800 non-null   object
7   Segment              9800 non-null   object
8   Country               9800 non-null   object
9   City                 9800 non-null   object
10  State                9800 non-null   object
11  Postal Code          9789 non-null   float64
12  Region              9800 non-null   object
13  Product ID          9800 non-null   object
14  Category            9800 non-null   object
15  Sub-Category        9800 non-null   object
16  Product Name        9800 non-null   object
17  Sales               9800 non-null   float64
dtypes: float64(2), int64(1), object(15)
memory usage: 1.3+ MB
None
```

We conducted a summary of the dataset and performed basic statistical analysis to gain initial insights into key metrics such as mean, median, standard deviation, and distribution patterns across relevant fields. This helped in identifying trends, potential outliers, and overall data characteristics for further analysis.

	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

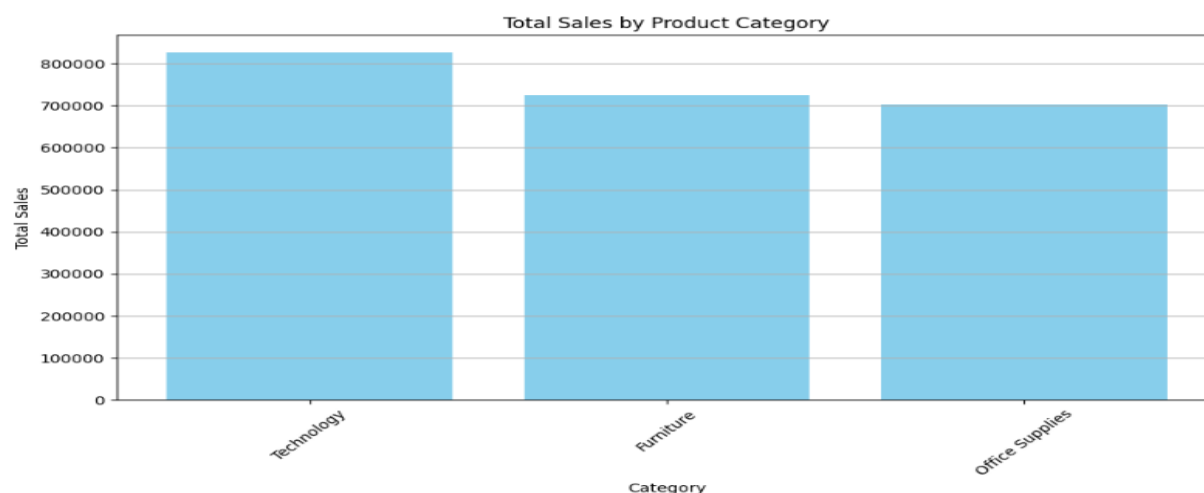
## key insights

### 1- What are the monthly sales trends over the years? Are there specific months with peak sales?

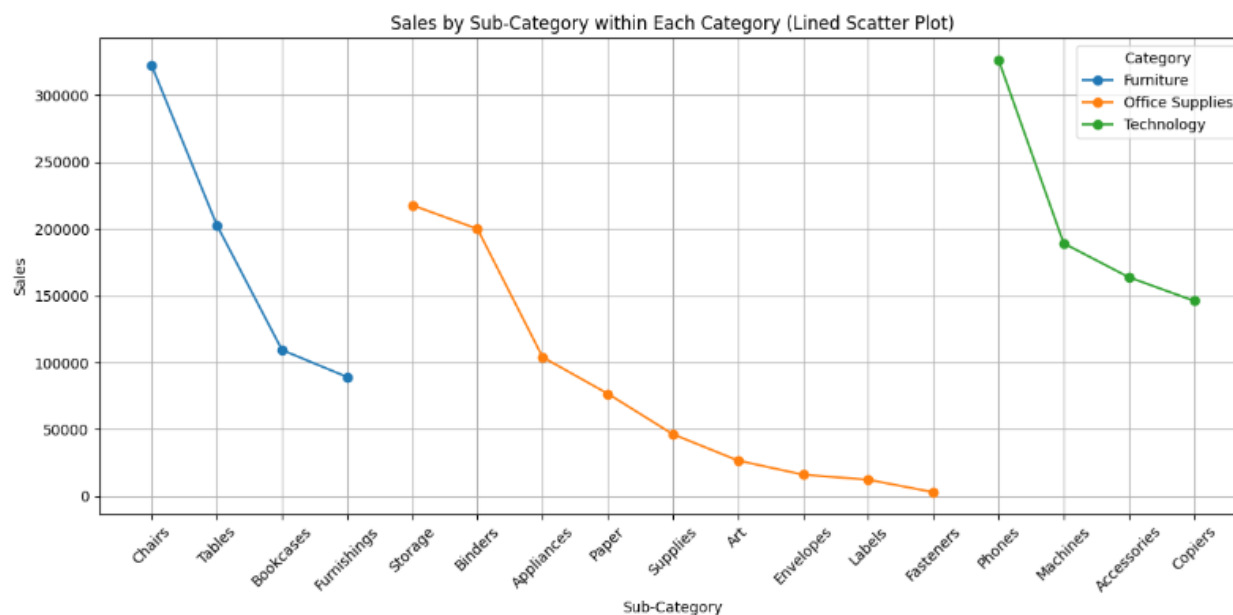


This graph highlights the sales trend over the years, showing a noticeable increase in sales during the second half of each year. This uptick can likely be attributed to several factors, including the back-to-school season, Black Friday promotions, and Christmas offers, all of which typically drive higher consumer spending during this period.

## 2- Which product categories generate the highest revenue? How do different sub-categories compare within those categories?

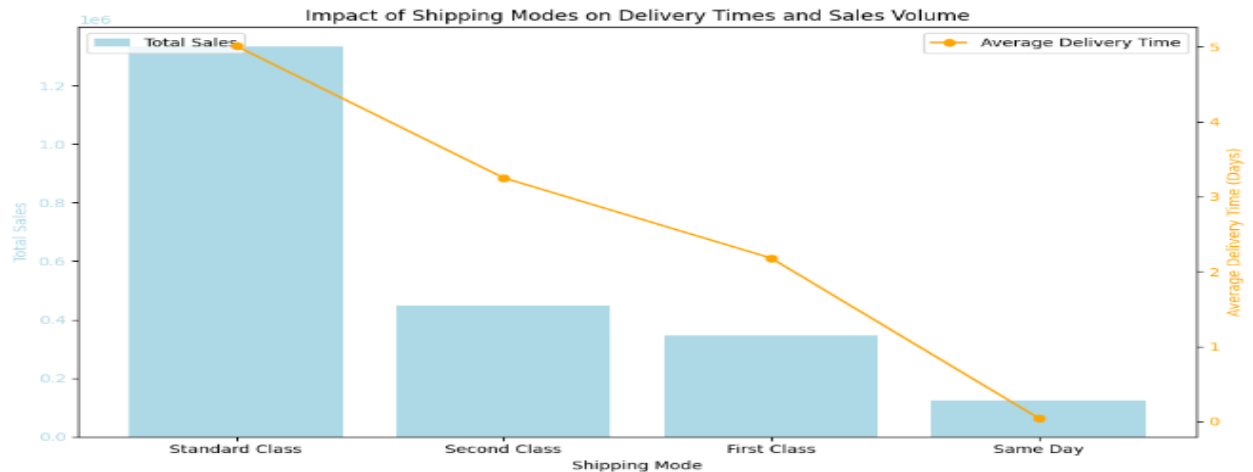


As depicted in the graph, the technology category leads in sales, primarily due to the higher price point of products in this segment, coupled with strong consumer demand for technology-related items.



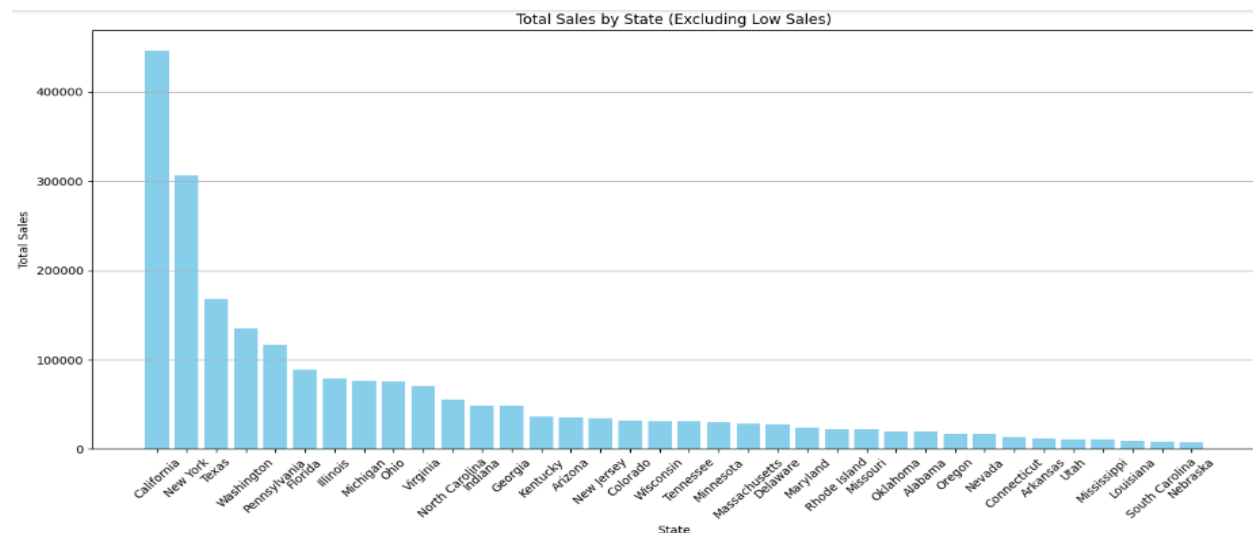
This graph illustrates the sub-categories within each main category that generated the highest sales, providing insights into consumer preferences and trends across different segments.

### 3- What is the impact of shipping modes (e.g., Second Class, First Class) on delivery times and sales volume?

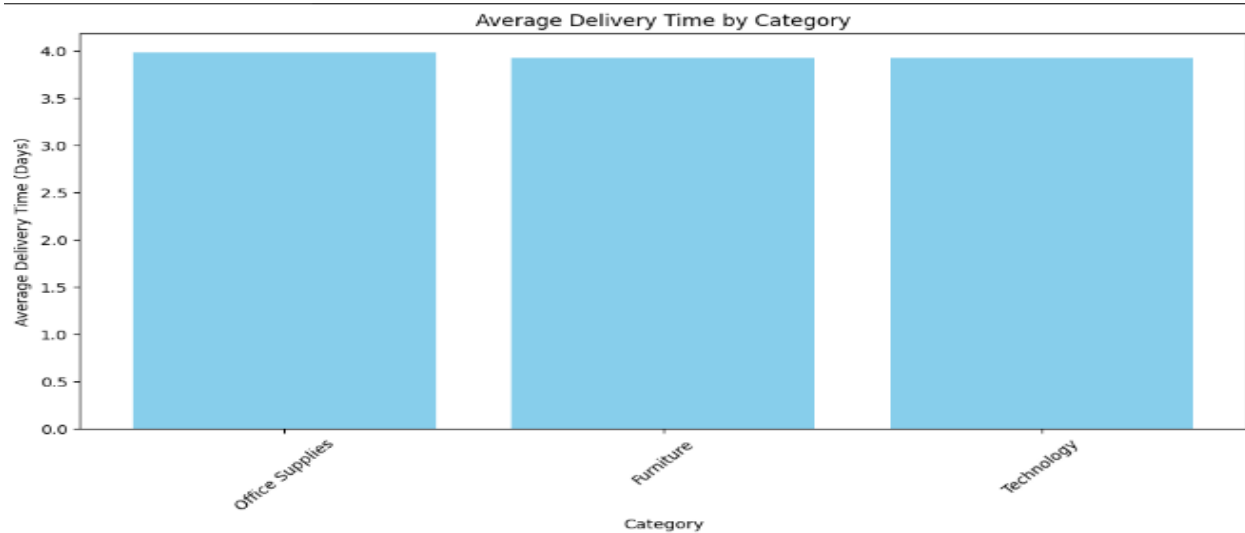


The standard class exhibits the longest delivery time; however, it also accounts for the highest sales. This trend can be attributed to the lower shipping costs associated with this shipping method, as well as the fact that many of our customers are consumers who predominantly opt for standard shipping.

### 4- Which states or regions contribute the most to total sales? Are there regions with notably low sales?



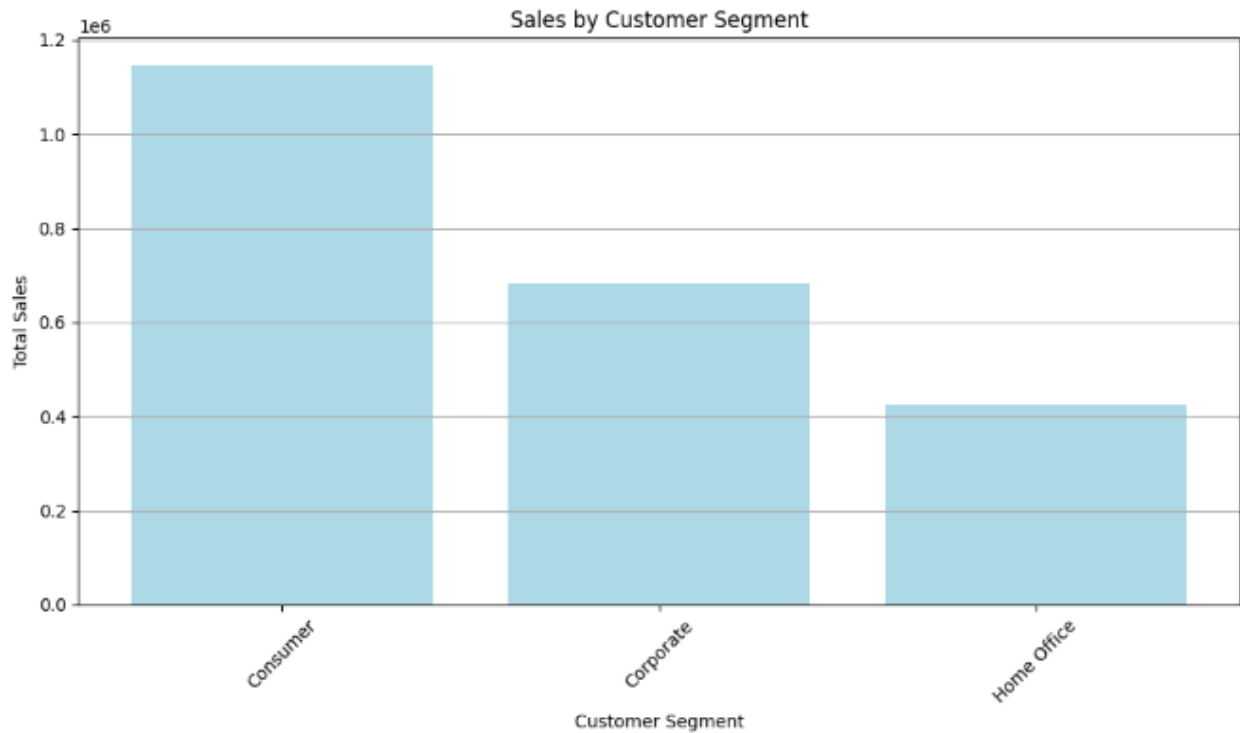
## 5- What is the average time taken for each category from order date to ship date?



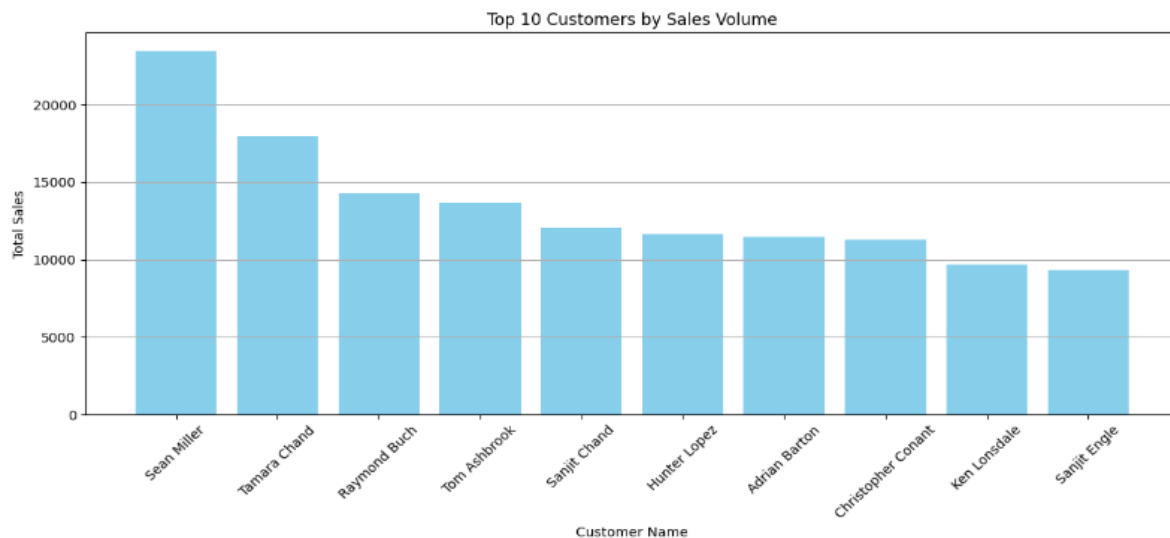
The average delivery time across each category is uniform, likely due to standardized shipping processes and coordination strategies implemented by the company. This consistency may stem from the use of similar carriers, delivery routes, and fulfillment centers for all categories, resulting in comparable handling and transit times regardless of the product type. Additionally, inventory management practices may further align delivery timelines across categories, contributing to this average consistency.



**6- How do sales differ across customer segments (Consumer, Corporate, etc.)? Which segment is the most profitable**

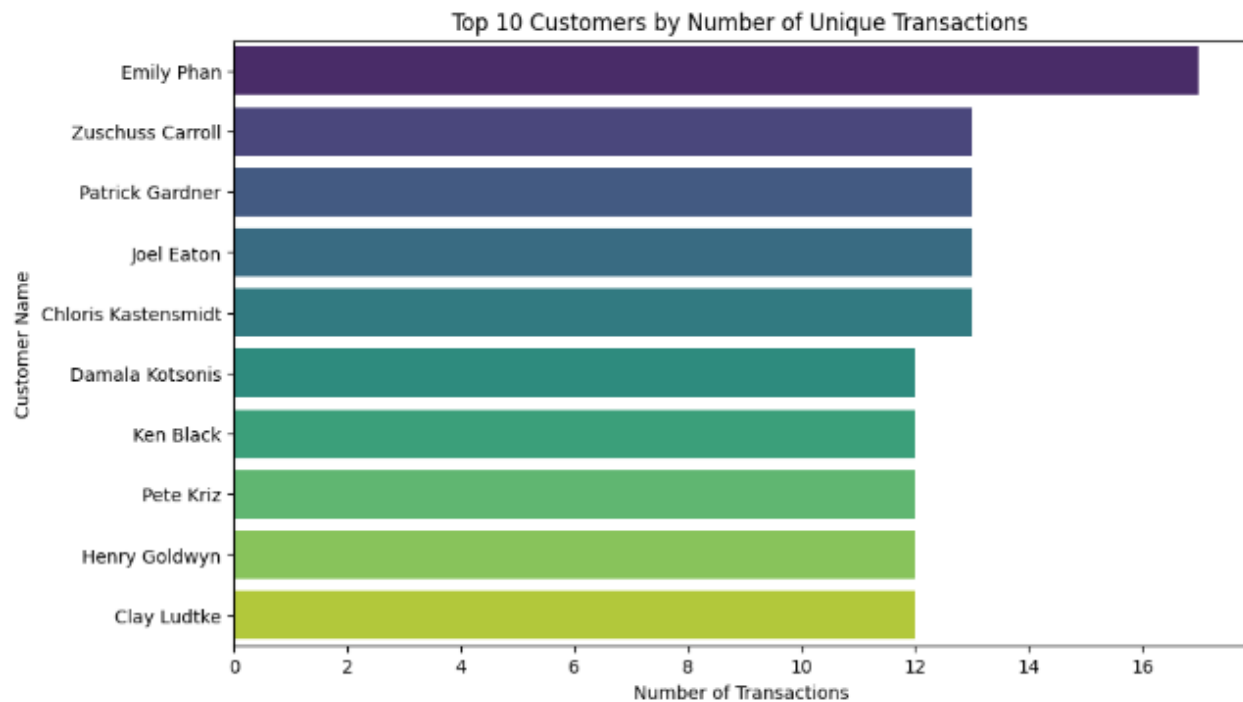


**7- Who are the top customers by sales volume, Segment & Categories?**



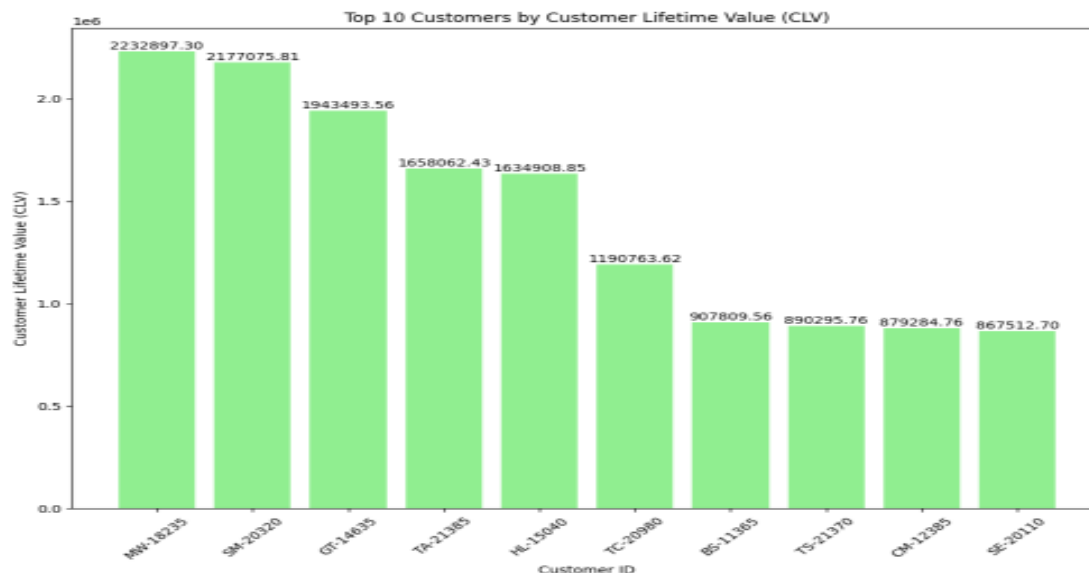
This chart displays the top ten customers by sales. To incentivize increased purchasing behavior among other customers, we could consider offering rewards or substantial discounts to these top customers, thereby encouraging a more competitive buying environment.

## 8- Do we have customers repeat their visits?



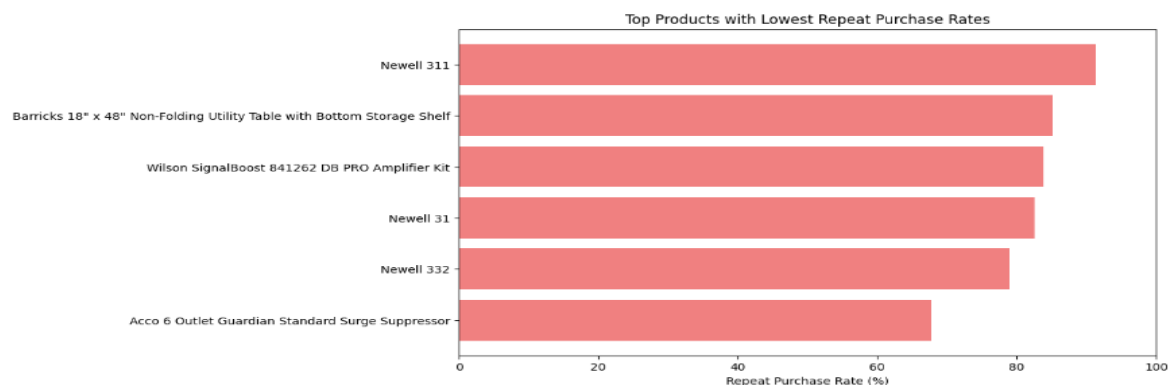
We analyzed the percentage of repeat customers, and above are the top 10 customers with the highest frequency of purchases. This data highlights customer loyalty and can inform targeted marketing strategies to further engage these valuable customers.

## 9- "Top 10 Customers by Customer Lifetime Value (CLV)"



The graph of the "Top 10 Customers by Customer Lifetime Value (CLV)" highlights which customers bring in the most revenue over time. A significant difference in CLV suggests focusing on retaining high-value customers, while a more balanced CLV indicates a stable customer base with less reliance on a few key clients. This information helps businesses improve customer service, loyalty programs, and marketing to boost profits. Understanding CLV allows companies to allocate resources effectively for long-term growth and to keep top customers engaged for future revenue.

## 10- Lowest Repeated Purchased Products



This section identifies the product with the lowest sales, providing insights into potential areas for improvement. Analyzing the factors contributing to its underperformance can inform strategies for enhancement, such as marketing adjustments, pricing reevaluations, or product repositioning to boost sales.

## Recommendations

- 1. Incentivize Customer Loyalty:** Implement a rewards program for repeat customers to strengthen loyalty. This could include exclusive discounts, timely access to new products, or loyalty points that can be redeemed for future purchases.
- 2. Targeted Promotions:** Leverage insights from the top ten customers to create targeted marketing campaigns that encourage similar purchasing behaviors among other customers. Personalized offers based on their purchase history could enhance conversion rates.
- 3. Reevaluate Underperforming Products:** Conduct a thorough analysis of the product with the lowest sales to understand the underlying issues. Consider strategies such as promotional discounts, enhanced marketing efforts, or product bundling to increase its visibility and attractiveness.
- 4. Enhance Shipping Options:** While the standard shipping method is popular, consider diversifying shipping options. Introducing faster delivery methods at competitive rates could attract a segment of customers willing to pay for expedited service.
- 5. Monitor Trends Continuously:** Establish a regular review process for sales data and customer feedback to identify emerging trends early. This proactive approach will enable timely adjustments to marketing strategies and inventory management, optimizing overall performance.
- 6. Educational Content:** To improve sales for the lowest-performing product, educational content or communication strategies can be highly effective. Businesses can address customer pain points and misunderstandings by creating informative materials that highlight the product's unique features, benefits, and practical uses. Tailored communication, such as tutorials, case studies, or customer testimonials, can increase product awareness and encourage potential buyers to see its value, boosting sales.
- 7. Loyalty Programs:** A loyalty program targeting the highest customer shares in sales and repeat purchasers is crucial for sustaining business growth. By rewarding top customers with exclusive benefits, such as discounts, timely access to products, or exclusive offers, the program encourages them to remain engaged and continue making repeat purchases. This strengthens customer relationships and fosters brand loyalty, significantly boosting long-term profitability and customer retention.

By implementing these recommendations, the company can capitalize on existing strengths while addressing areas for improvement to drive sales and customer satisfaction.

## Conclusion

The sales data analysis reveals critical trends and patterns across product categories, shipping methods, and customer behaviors. Notably, the technology category consistently leads in sales, driven by both high demand and premium product pricing, making it a key driver of revenue. Furthermore, the standard shipping method, despite its longer delivery times, remains the most popular choice among consumers due to its affordability, reflecting customers' preference for cost-effective solutions. The strong loyalty demonstrated by repeat customers highlights significant opportunities for further engagement and retention strategies, which could help boost long-term profitability. Additionally, the identification of underperforming products and low-selling categories points to areas with potential for improvement and growth. By addressing these underperforming segments, the company can refine its product offerings and better align them with customer needs, leading to enhanced sales performance and overall market competitiveness. This analysis provides valuable insights to inform future business strategies and operational improvements.