

Amazon Sales Report Analysis

Date: 25-11-2024

Prepared by: Dipti Aher

1. Introduction

The dataset under analysis encompasses detailed information about sales transactions conducted on Amazon. It includes key attributes such as order IDs, transaction dates, fulfillment methods, sales channels, product categories, quantities, amounts, shipping details, and more. This comprehensive dataset serves as the foundation for exploring various dimensions of sales performance, customer behavior, and geographical trends. By conducting an in-depth analysis, the goal is to derive actionable insights that can guide strategic business decisions, optimize operational efficiency, and enhance customer satisfaction.

2. Objective

1. **Sales Overview:** Analyze overall sales performance to identify trends, seasonal variations, and opportunities for growth over time.
2. **Product Analysis:** Assess product categories, sizes, and quantities sold to identify popular items and underperforming categories.
3. **Fulfillment Analysis:** Evaluate the efficiency of different fulfillment methods to ensure timely, cost-effective deliveries.
4. **Customer Segmentation:** Segment customers based on behavior, location, and demographics to enhance marketing strategies and product offerings.
5. **Geographical Analysis:** Examine sales distribution by region to uncover high-performing areas and identify potential markets for expansion.

3. Data Overview

3.1 Data Source

The dataset was sourced from Innobyte, where I am currently completing an online internship. The data provides information about customer orders, including details about products, sales, shipping, and fulfillment. The dataset contains a total of 128,976 rows and 21 columns. The date range of the data is not explicitly mentioned, but the analysis focuses on sales and order data, which spans over a specific period.

3.2 Data Description

The dataset includes the following key variables:

- **Order ID:** Unique identifier for each order.
- **Date:** Date of the order.
- **Status:** Order status (e.g., Shipped, Pending).
- **Fulfillment:** Indicates whether the order is fulfilled by Amazon or another merchant.
- **Sales Channel:** Channel through which the order was placed.
- **Ship-City, Ship-State, Ship-Country, Ship-Postal-Code:** Shipping details for the orders.
- **Category:** Product category of the ordered item.
- **Amount:** Total sales amount for each order.
- **Quantity (Qty):** Quantity of items in the order.
- **Currency:** Currency used for the transaction.

Data Cleaning Steps:

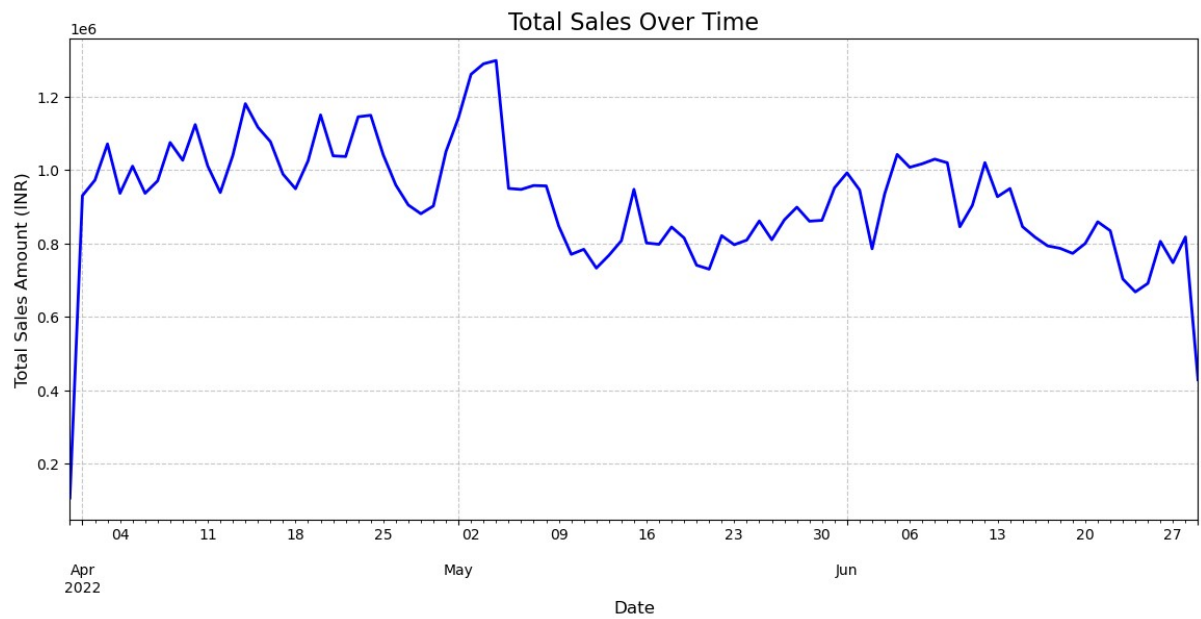
1. **Missing Values:**
 - The "**Fulfilled-by**" column had 69% missing values, which were replaced with "Easy Ship" as the only value in the column.
 - The **Amount** and **Currency** columns had 6% missing values, which were replaced with the median (for Amount) and mode (for Currency).
 - Missing values in the **ship-city**, **ship-state**, **ship-postal-code**, and **ship-country** columns were replaced with the mode of each respective column.
2. **Data Type Conversion:** The **Date** column was originally of the object data type, and it was converted to the **datetime** type for proper time-based analysis.

4. Methodology

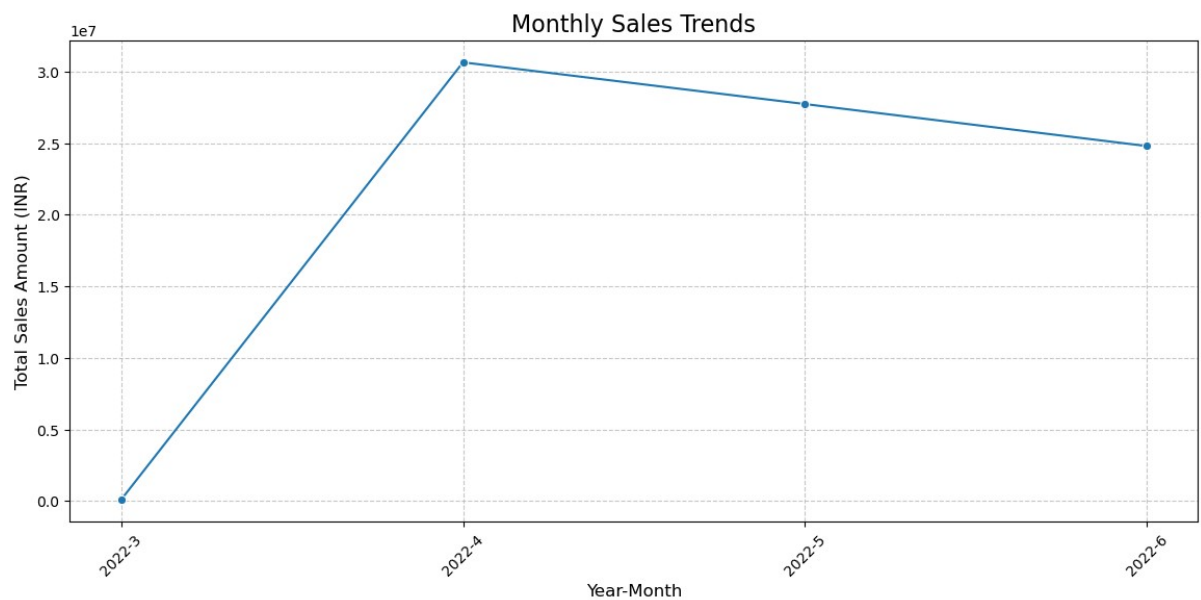
- **Tools Used:** The analysis was conducted using **Python** and its libraries (Pandas for data manipulation, Matplotlib and Seaborn for visualization). Data was cleaned and processed in Python, with aggregation and filtering operations performed for identifying trends.
- **Approach:** The approach focused on time series analysis (monthly sales), categorical analysis (sales by category, state, and city), and fulfillment analysis (order status and shipping insights). Descriptive statistics and visualizations were used to highlight key trends.
- **Actions and Decisions:** Focus on seasonal demand in April, enhance size availability in T-shirts and Shirts, and prioritize regions like Maharashtra and Bengaluru for marketing campaigns.
- **Considerations:** Consider possible biases in shipping or fulfillment data, and review market conditions regularly to adjust strategies as needed. Potential risks include overstocking in specific categories or underestimating demand shifts outside of peak months.

5. Analysis

I] What are the total sales over time?

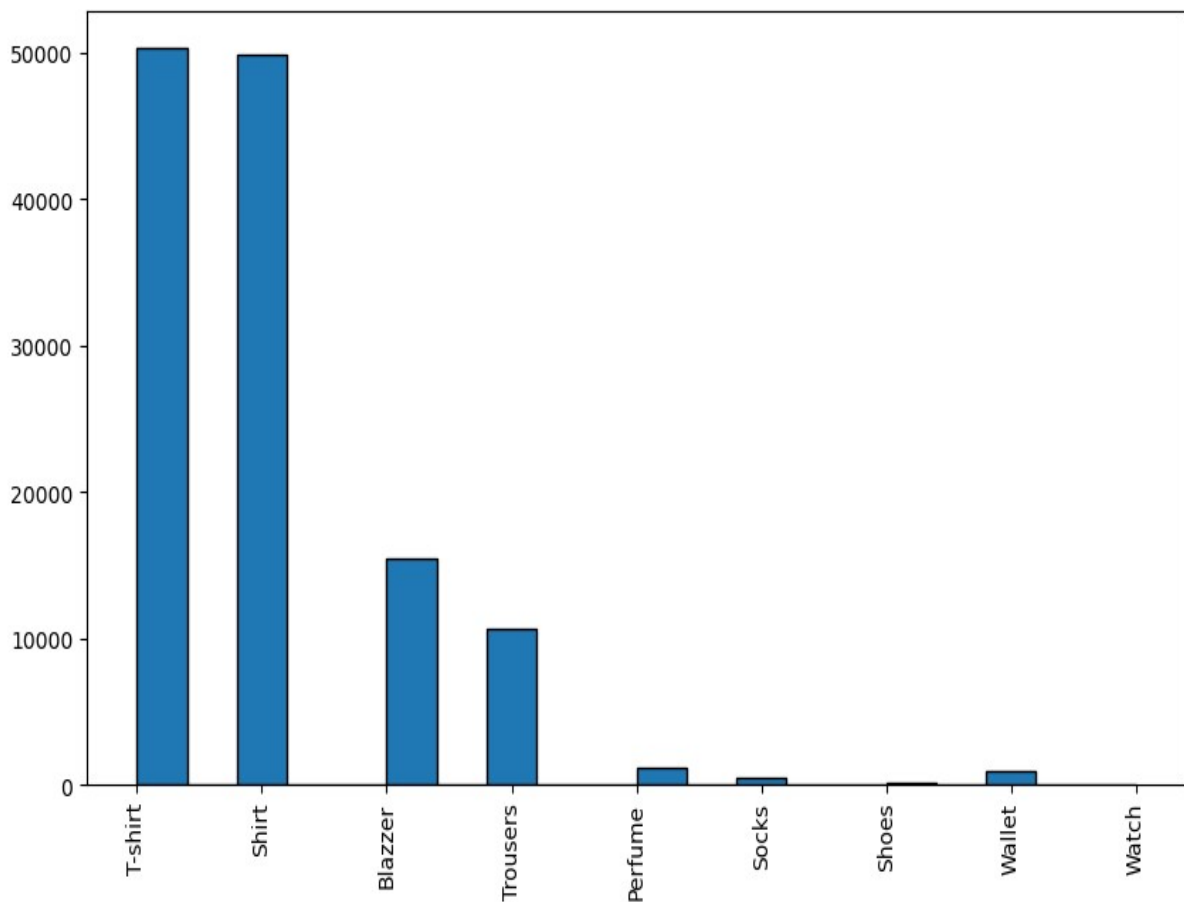


II]



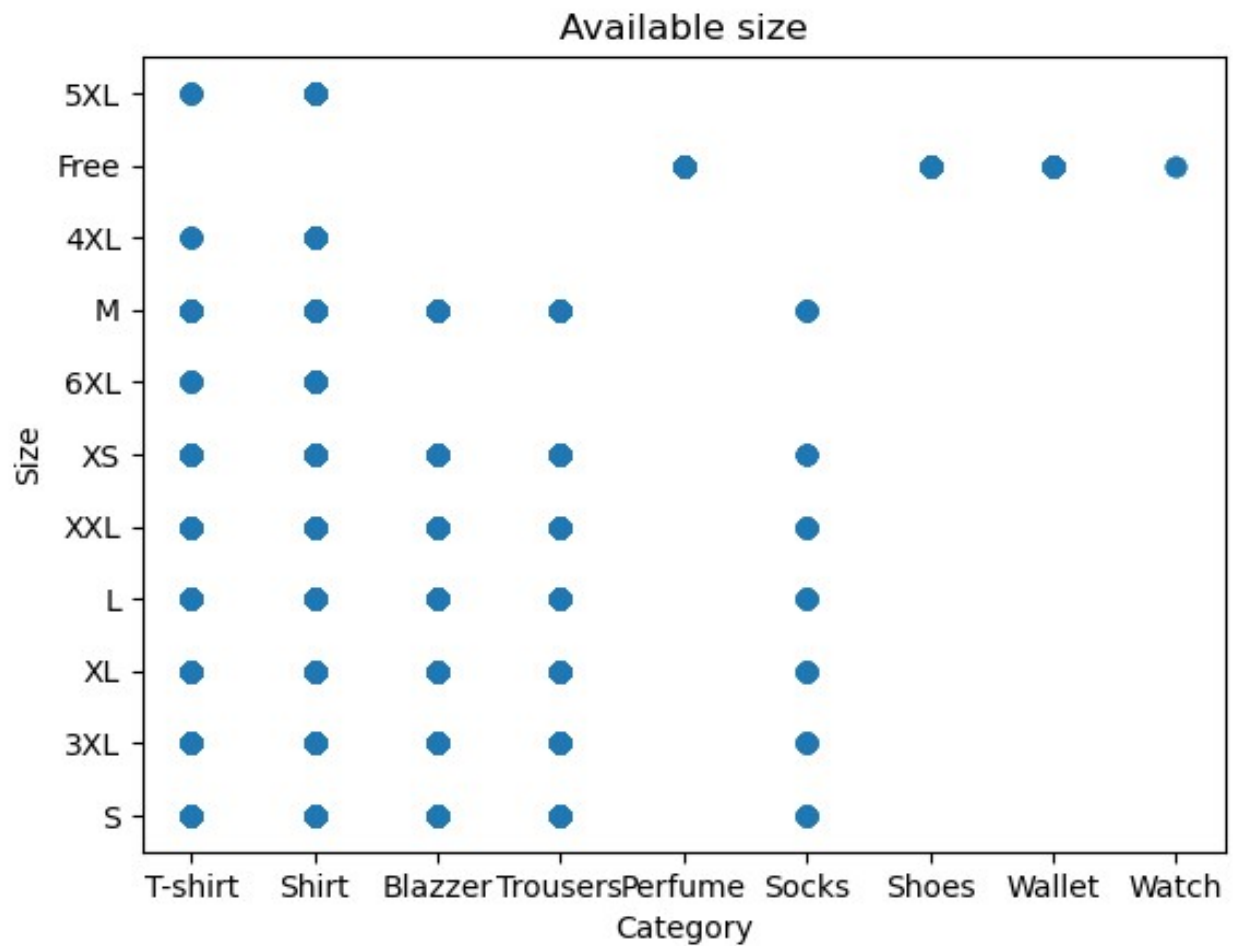
I began by analyzing the total sales over time, specifically focusing on the sales amounts for each month. This showed a clear peak in April, which I identified as the month with the highest sales.

III]What are the top-selling product categories?



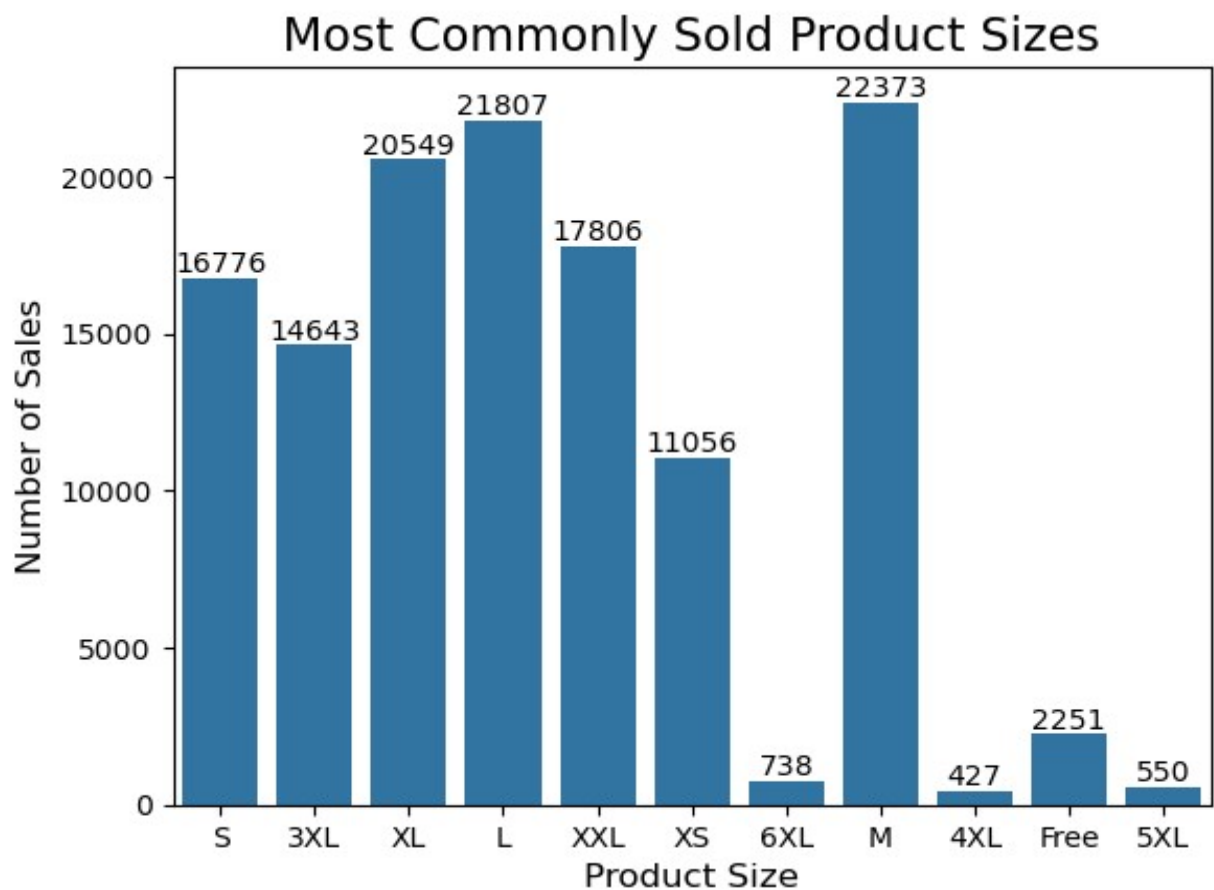
I analyzed the top-selling categories and identified **T-shirts** as the highest-selling category, **Shirts** and **Blazers** followed respectively. Categories like **Shoes** and **Socks** had minimal sales, indicating low demand. This highlights T-shirts' popularity, likely driven by wide availability and customer preference.

IV] sizes available for each category



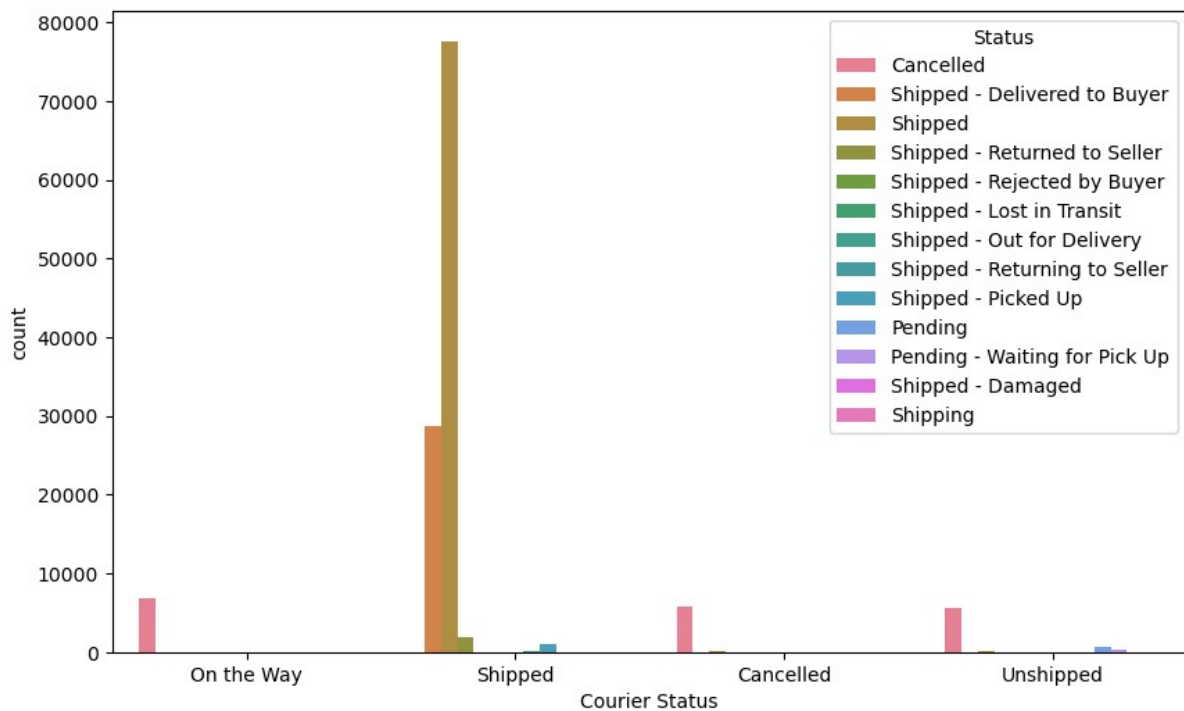
I analyzed the sizes for each category and found that **T-shirts** and **Shirts** have all sizes available except "Free." This variety in sizes makes them more appealing to customers and likely explains their high sales. These categories cater to all size preferences, which is a big reason for their success. Having a wide range of sizes clearly helps them sell better

V]Which product sizes are most commonly sold?



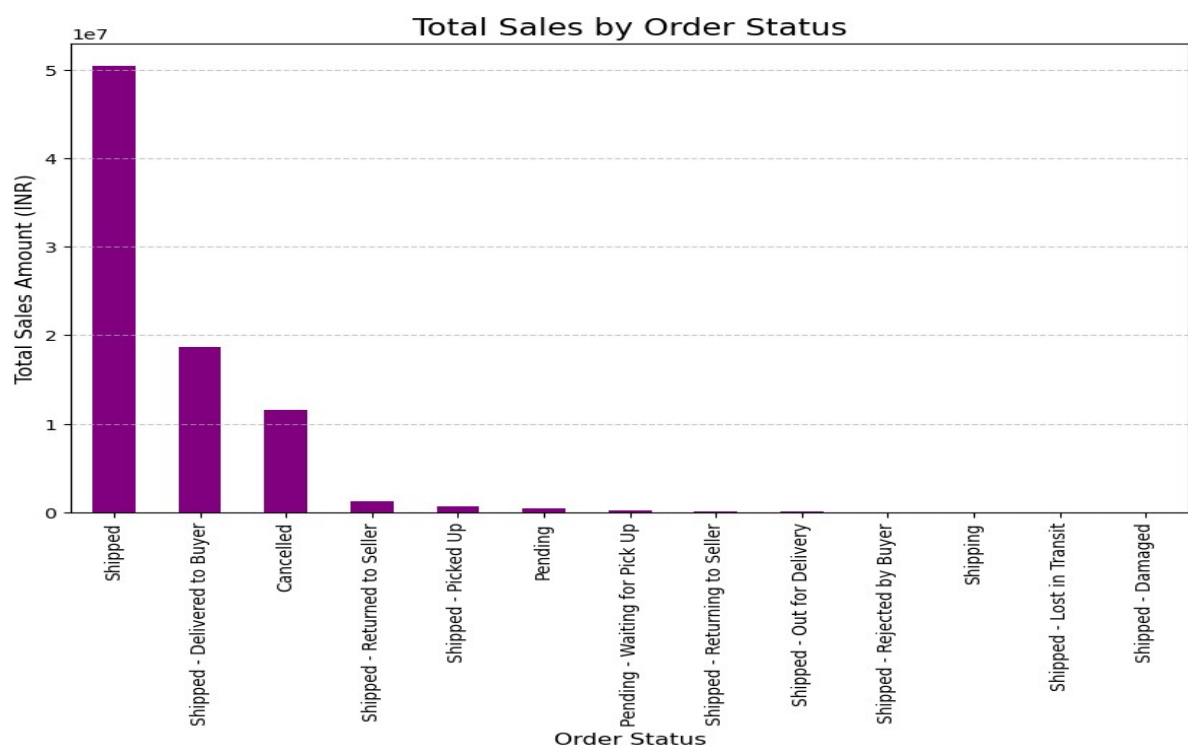
I analyzed which product sizes are most commonly sold and found that **M size** has the highest sales. This suggests that **M size** is the most preferred by customers across categories, making it an important size to keep well-stocked to meet demand.

VI]Courier Status



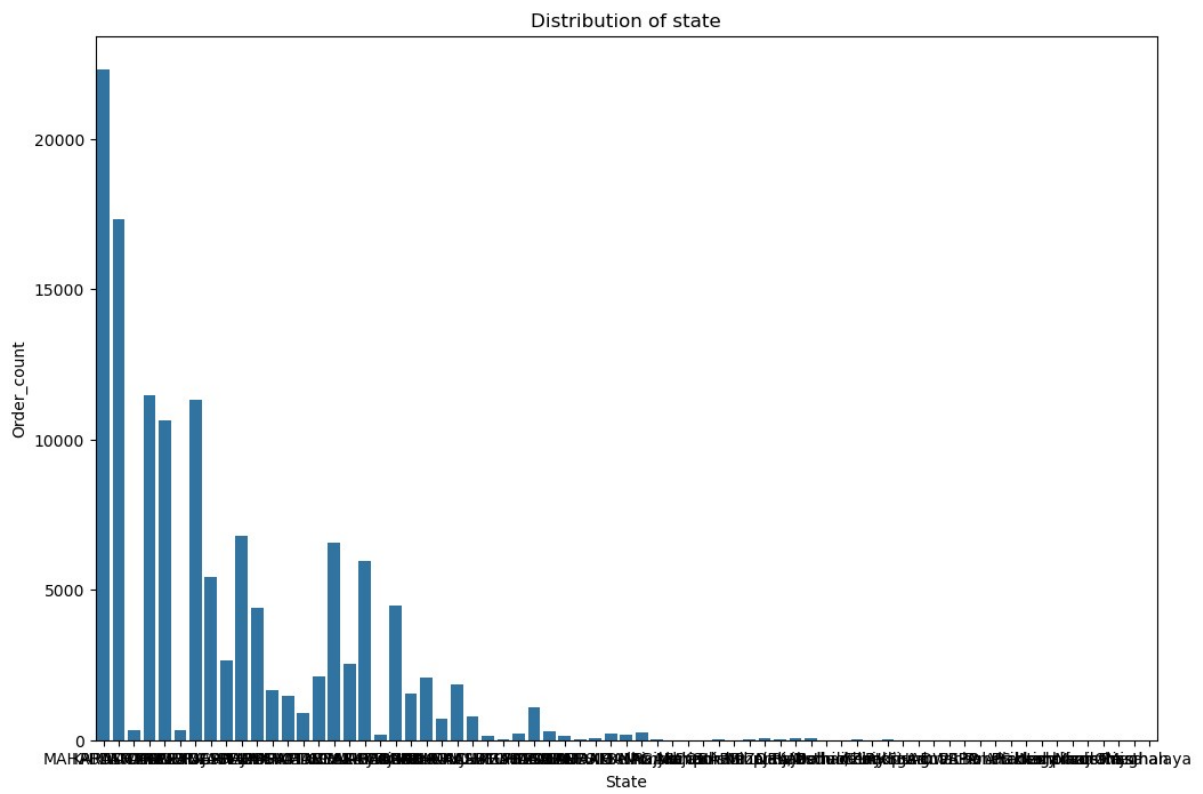
I analyzed the **Courier Status** and found that **Shipped** has the highest count. This indicates that most orders are successfully processed and dispatched, reflecting efficient order management and fulfillment processes.

VII]Which order statuses contribute most to total sales?

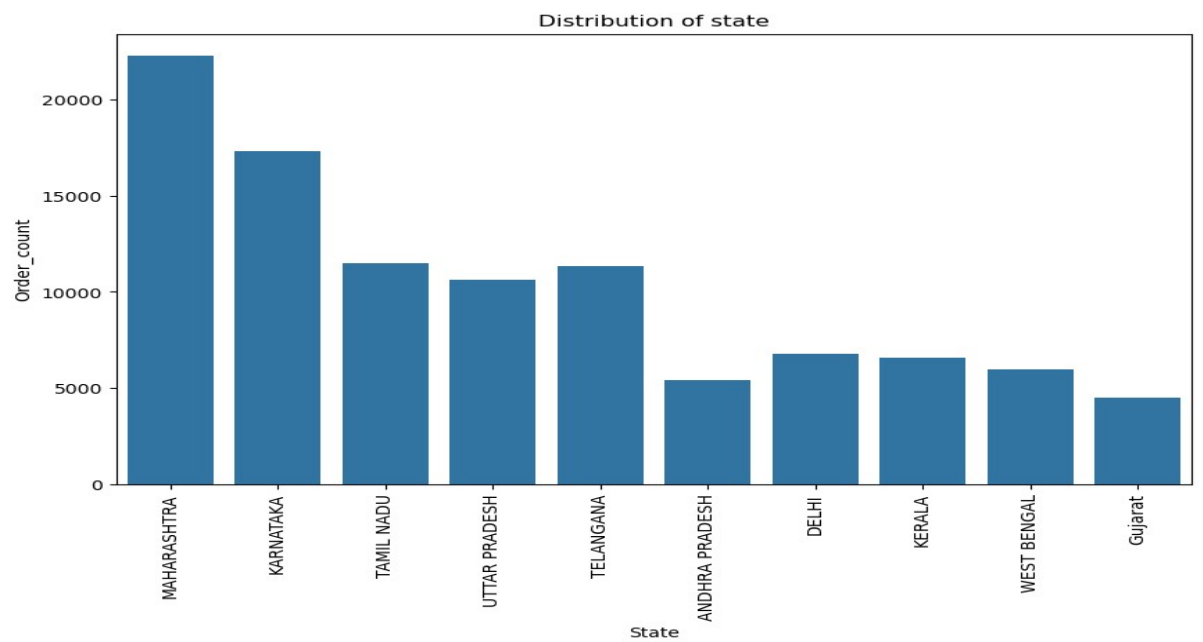


I analyzed which order statuses contribute the most to total sales and found that **Shipped** has the highest sales amount. This highlights that orders successfully shipped contribute significantly to overall revenue, indicating smooth operations in the fulfillment process.

VIII]State wise ship date

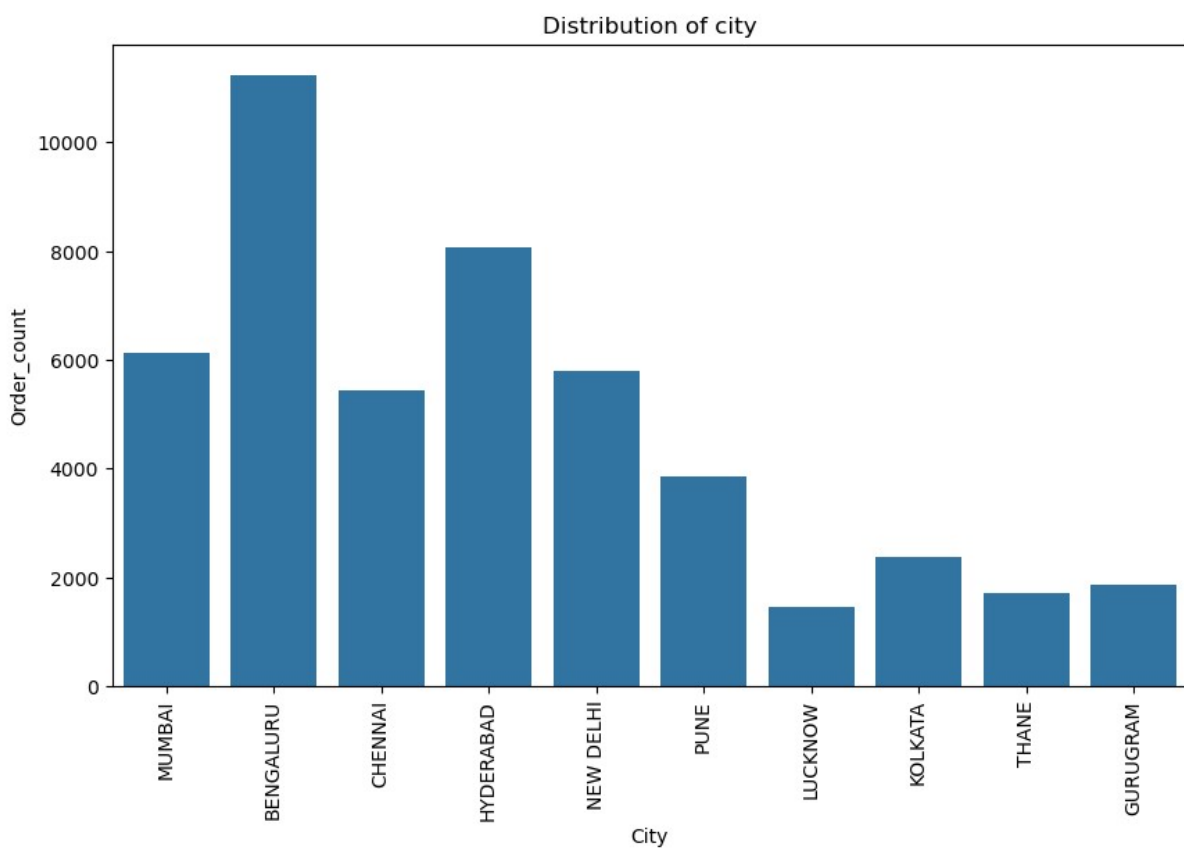


IX]Top 10 states



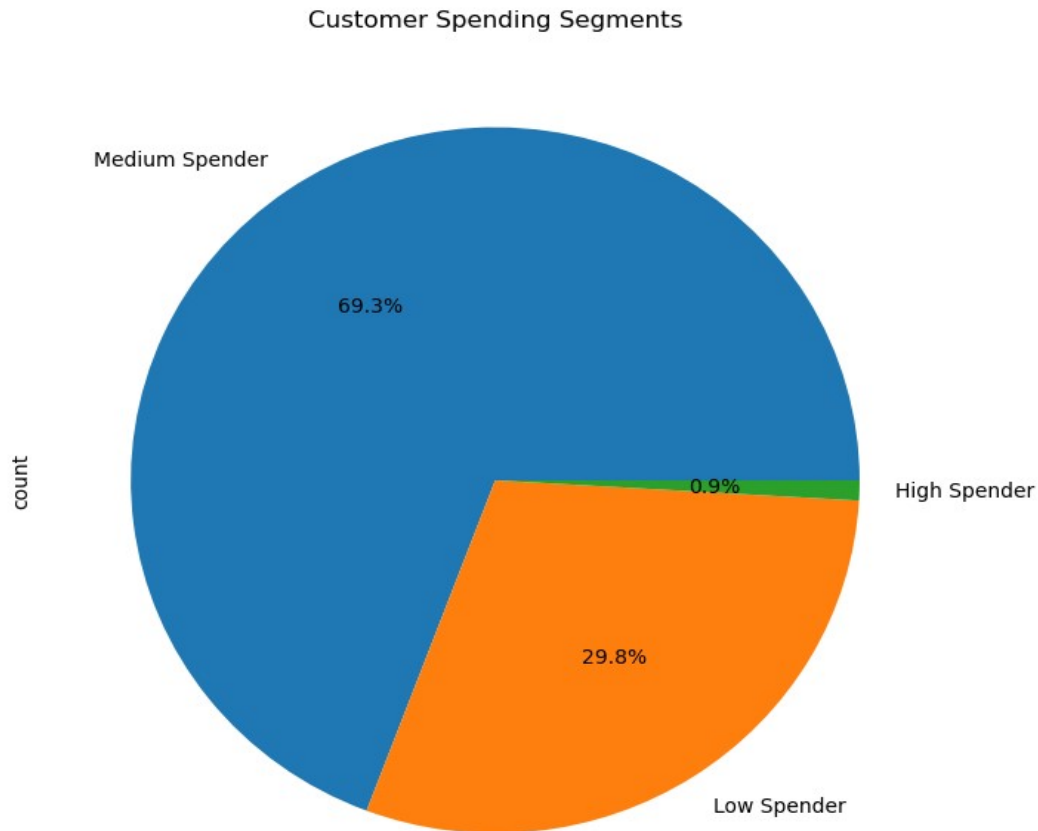
I analyzed the top 10 states based on shipping data and found that **Maharashtra** has the highest order count. This indicates a strong customer base or demand in the region, making it a key market to focus on for sales and marketing strategies.

X]Top 10 Cities



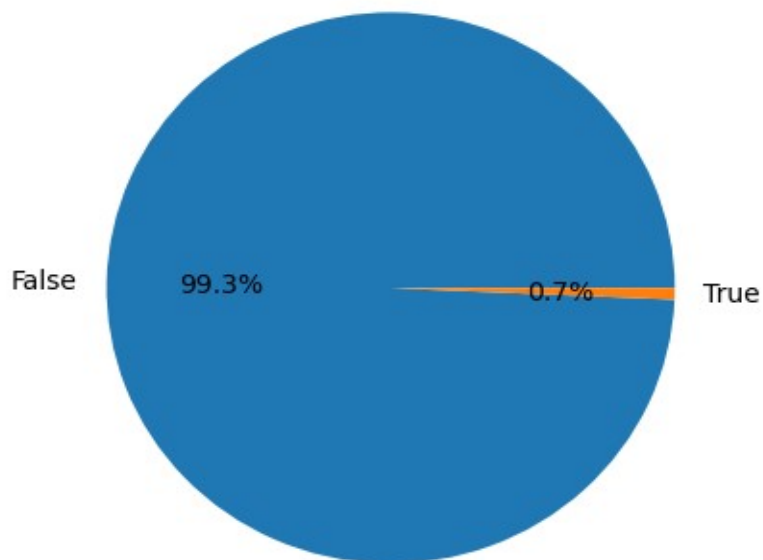
I analyzed the top 10 cities and found that **Bengaluru** has the highest order count. This indicates that Bengaluru is a major market with strong demand, making it an important city to prioritize for inventory and marketing efforts.

XI]Customer Spending Segments



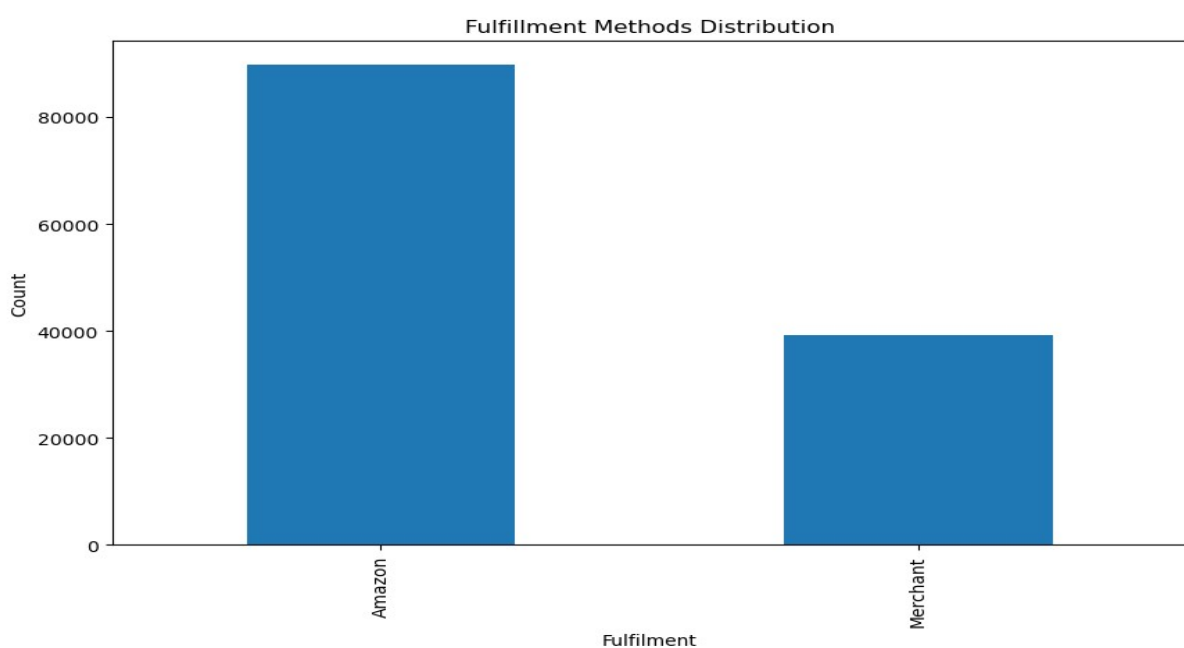
I analyzed **Customer Spending Segments** and found that **Medium Spenders** make up the majority at **69.3%**, followed by **Low Spenders** at **29.8%**, and **High Spenders** at only **0.9%**. This shows that most customers fall into the medium spending range, suggesting a focus on products and pricing that cater to this segment can drive overall sales

XII]check b2b dtata



I analyzed the **B2B data** and found that **99.3% of buyers are retailers**, while only **0.8% are wholesalers**. This indicates that the majority of business sales come from retailers, suggesting that the focus should be on maintaining strong relationships with retailers to drive sales, while wholesalers represent a much smaller portion of the market.

XII]Fullfillment Analysis



I analyzed the **Fulfillment Data** and found that **Amazon** has the highest count, surpassing **Merchants**. This indicates that Amazon handles a larger volume of orders, suggesting that leveraging Amazon's fulfillment network may be more efficient for fulfilling customer orders compared to merchant-based fulfillment.

6. Recommendations

- **Target Peak Sales Period:** Focus marketing efforts in April to capture the seasonal demand and build on previous success in this period.
- **Enhance T-shirt and Shirt Inventory:** Ensure that T-shirts and Shirts are well-stocked across all sizes, particularly M, to meet demand.
- **Leverage Regional Insights:** Allocate more resources to Maharashtra and Bengaluru for targeted promotions, stock replenishment, and regional marketing.
- **Focus on Medium Spenders:** Develop offers or loyalty programs tailored to medium spenders to increase customer retention and sales.
- **Strengthen Retailer Partnerships:** Since most B2B sales come from retailers, consider offering incentives, bulk pricing, or dedicated support to build stronger relationships with them.
- **Optimize Amazon Fulfillment:** Rely more on Amazon's fulfillment services for faster and more efficient order processing, while continuing to maintain relationships with merchants.

7. Conclusion

This analysis provided valuable insights into sales trends, product performance, and customer preferences. Key findings indicate that April is a peak sales month, T-shirts lead in sales due to their variety in sizes, and Maharashtra and Bengaluru are critical markets. By focusing on these areas and optimizing fulfillment processes through Amazon, the business can strengthen its market position.

Moving forward, targeting specific customer segments, improving stock management for popular sizes, and leveraging insights on regional demand can help sustain growth and improve sales performance.