

Data Analyst Internship

Title : Analyze and Provide insights on Amazon Sales Report

Internship period : 5/8/24 - 29/8/24

Company : Innobyte Services

Submission Date : 29/8/24

Name : Shaik Mastan Vali

Ph No : 6301145597

Introduction :

In today's competitive e-commerce landscape, gaining a deep understanding of sales performance is crucial for businesses to thrive. This project aims to analyze a comprehensive dataset of Amazon sales transactions to extract valuable insights and inform strategic decision-making.

Key Objectives:

1. **Sales Overview:** Understand the overall sales performance, trends, and patterns over time.
2. **Product Analysis:** Analyze the distribution of product categories, sizes, and quantities sold to identify popular products.
3. **Fulfillment Analysis:** Investigate the fulfillment methods used and their effectiveness in delivering orders.
4. **Customer Segmentation:** Segment customers based on their buying behavior, location, and other relevant factors.
5. **Geographical Analysis:** Explore the geographical distribution of sales, focusing on states and cities.
6. **Business Insights:** Provide actionable insights and recommendations based on the analysis to optimize sales strategies, improve customer satisfaction, and enhance overall business performance.

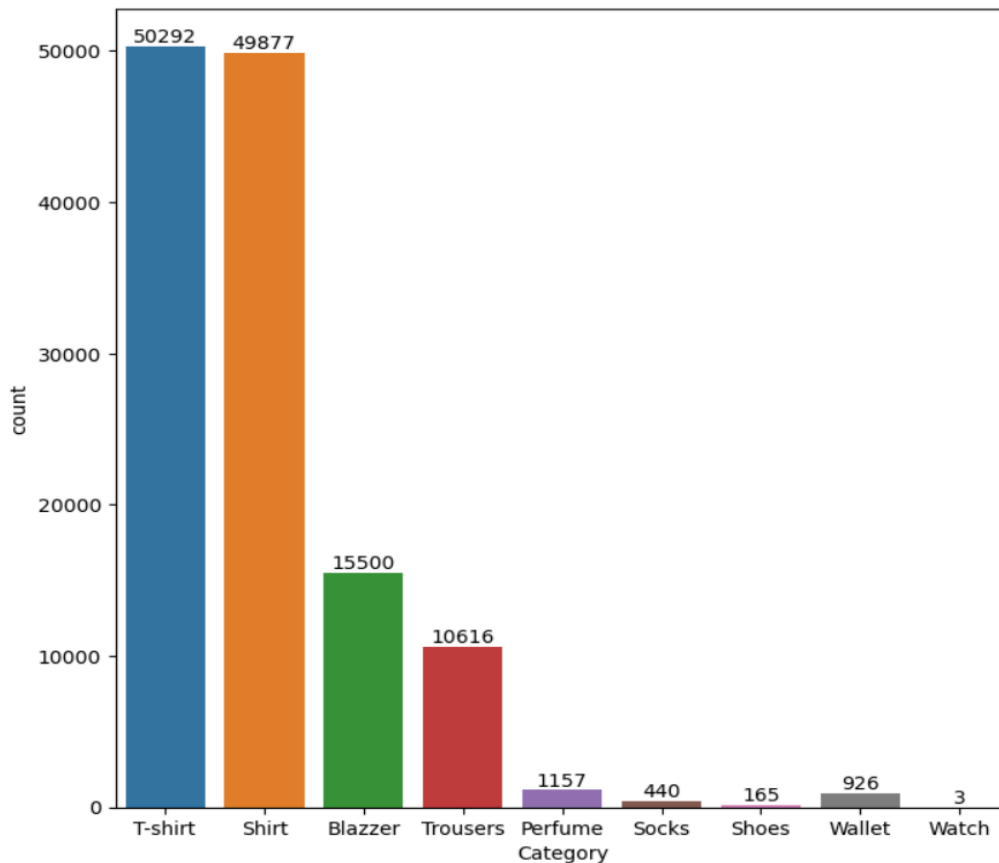
Methodology:

This project will employ a combination of descriptive and exploratory data analysis techniques to achieve the stated objectives. Data cleaning and preprocessing will be performed to ensure data accuracy and consistency. Key performance indicators (KPIs) will be calculated to measure sales performance and identify trends. Statistical analysis and data visualization will be used to uncover patterns and relationships within the data. Finally, actionable insights and recommendations will be presented to support informed decision-making.

Results And Analysis :

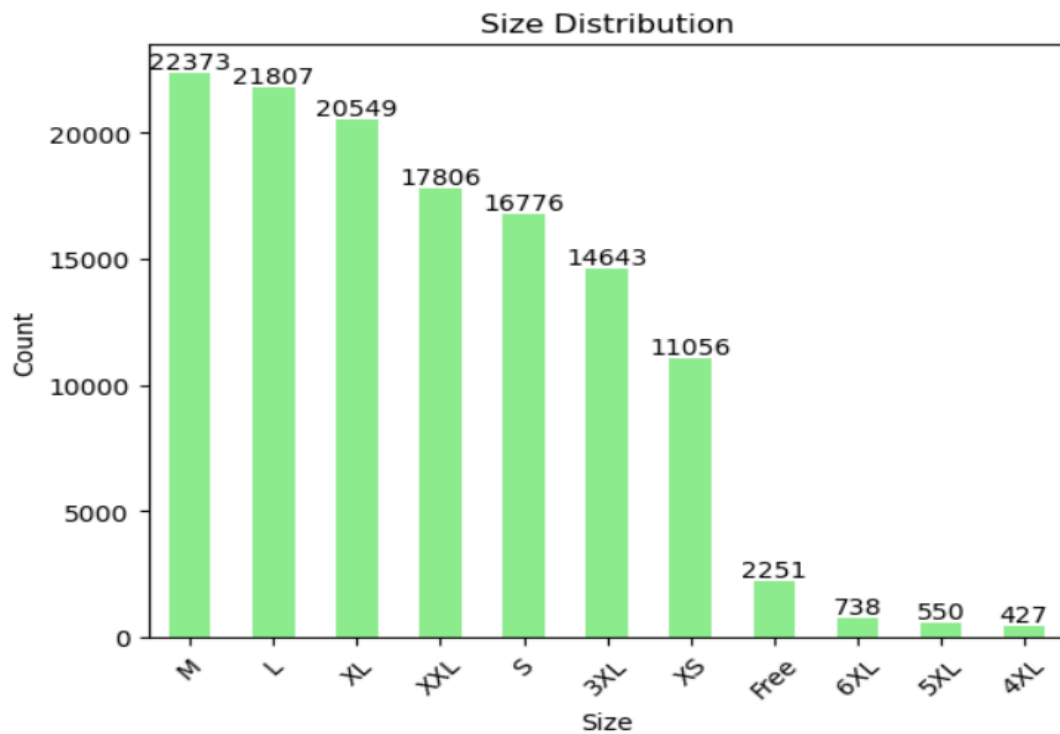
- The dataset contains information about Amazon sales, including order details, product information, and shipping information.
- There are 19 columns, each representing different aspects of the sales data.
- The data types are appropriate for each column, with numeric data stored as int64 or float64, and categorical data stored as objects.
- The 'Date' column is currently stored as an object, which we might want to convert to a datetime format for easier analysis.
- The 'B2B' column is a boolean, indicating whether the sale was a business-to-business transaction.

Summary of Product Distribution :



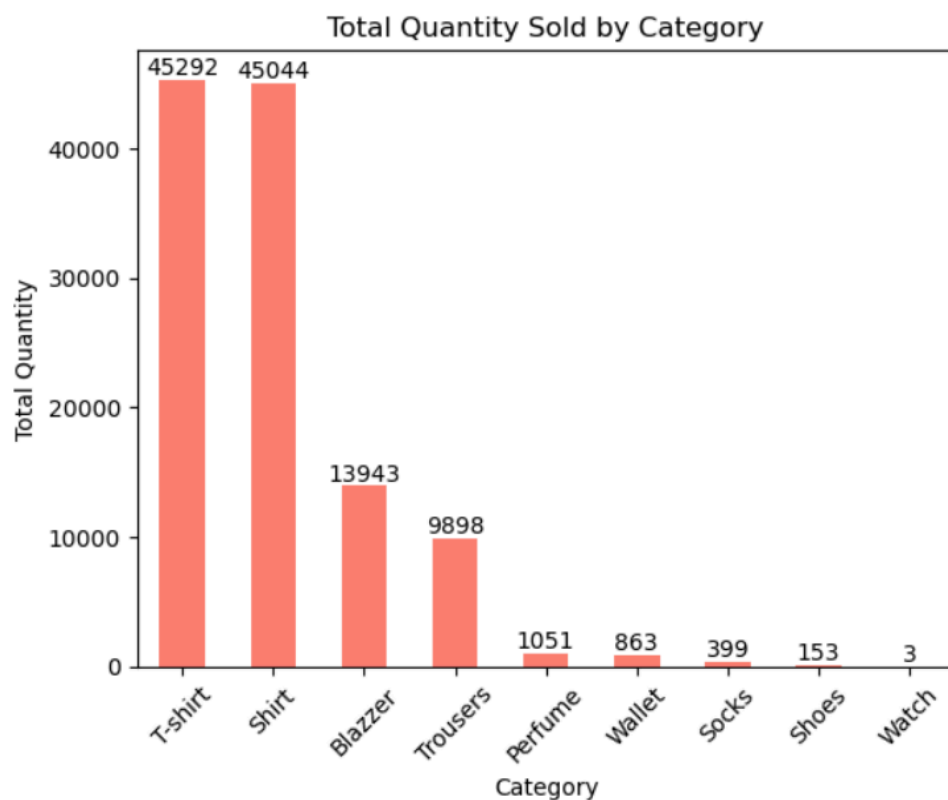
- T-shirts have the highest count, followed by Shirts, Blazers, and Trousers.
- Perfume, Socks, Shoes, Wallet, and Watch have significantly lower counts compared to the top categories.
- The count for Watch is extremely low, with only 3 items.
- There is a significant drop in count after Trousers, with Perfume having a count of 1157.

Summary of size Distribution :



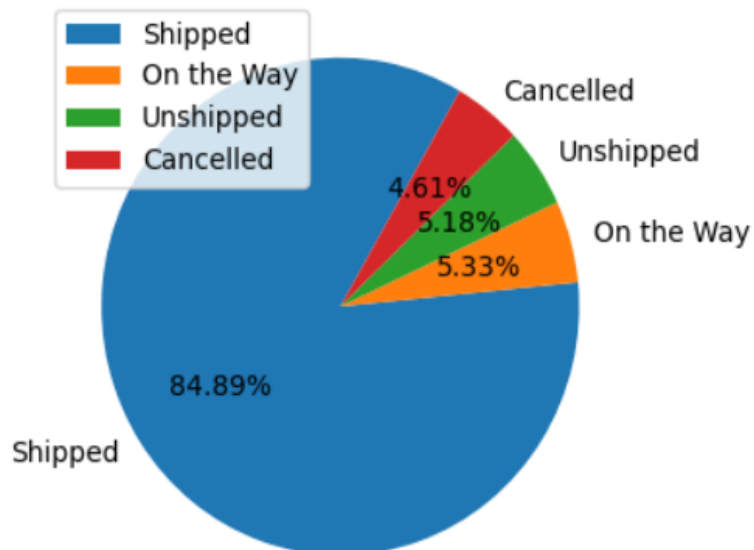
- Size M has the highest count, followed by L, XL, and XXL.
- There is a significant drop in count after XXL, with S having a count of 16776.
- The sizes 6XL, 5XL, and 4XL have extremely low counts.
- The count for Free size is 2251.

Summary of Quantity Sold By Category :



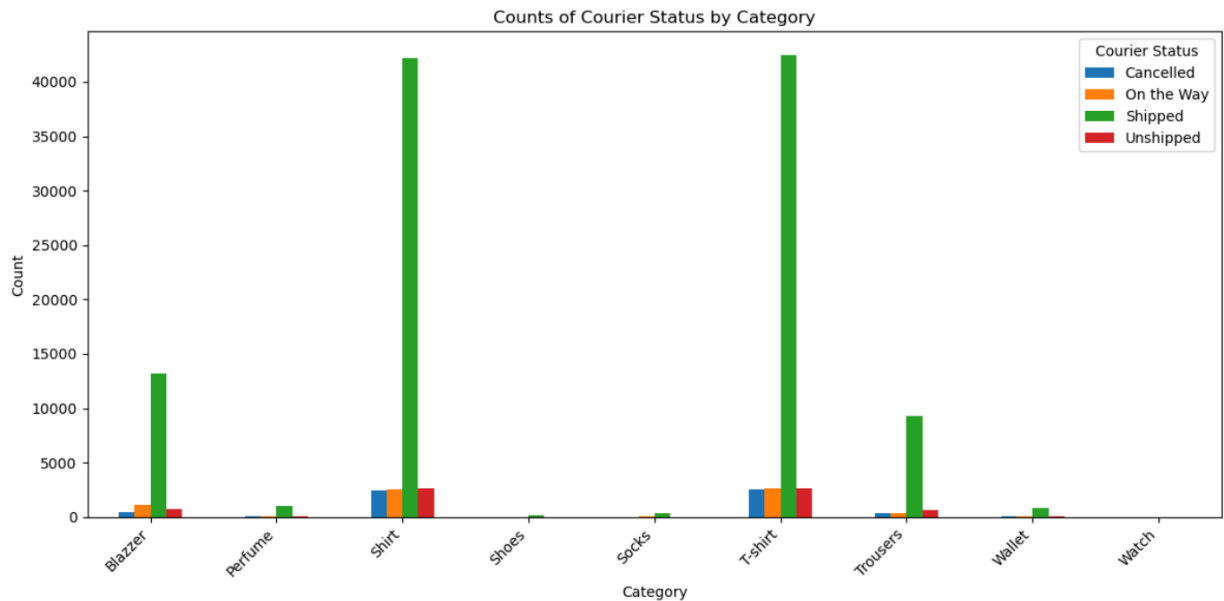
- T-shirts have the highest total quantity sold, followed by Shirts, Blazers, and Trousers.
- Perfume, Wallet, Socks, Shoes, and Watch have significantly lower total quantities sold compared to the top categories.
- The total quantity sold for Watch is extremely low, with only 3 items.
- There is a significant drop in total quantity sold after Trousers, with Perfume having a total quantity of 1051.

Summary of Courier Status :



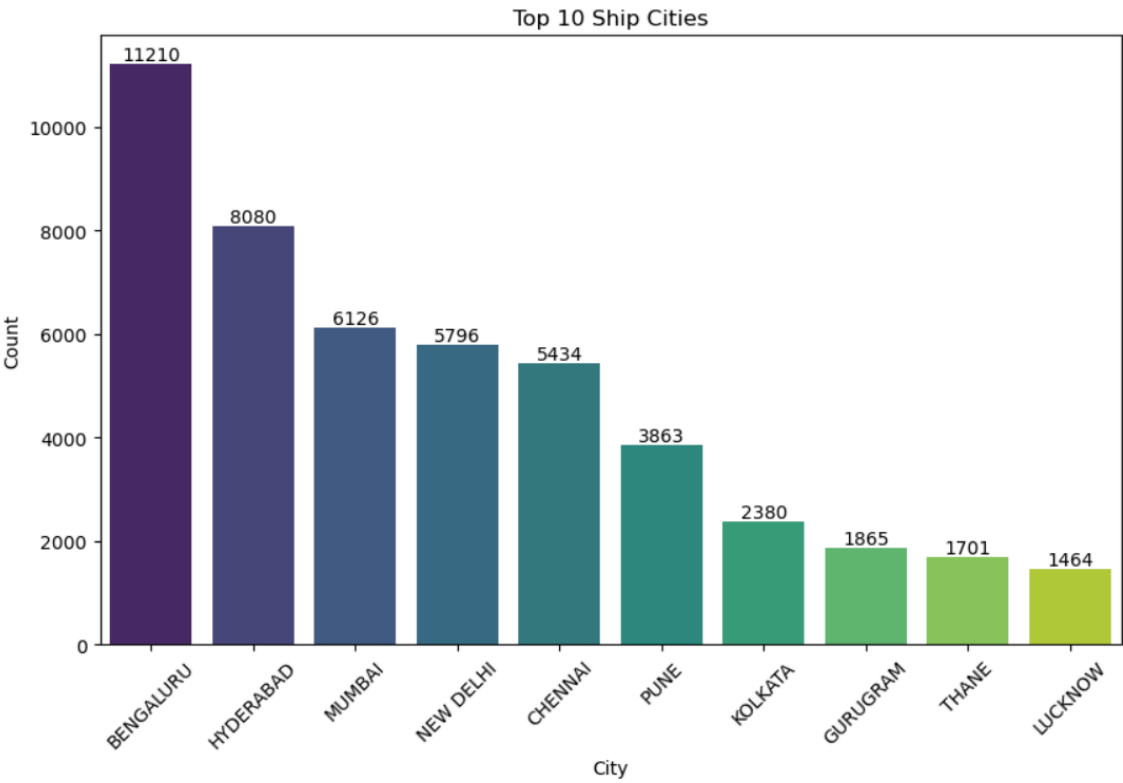
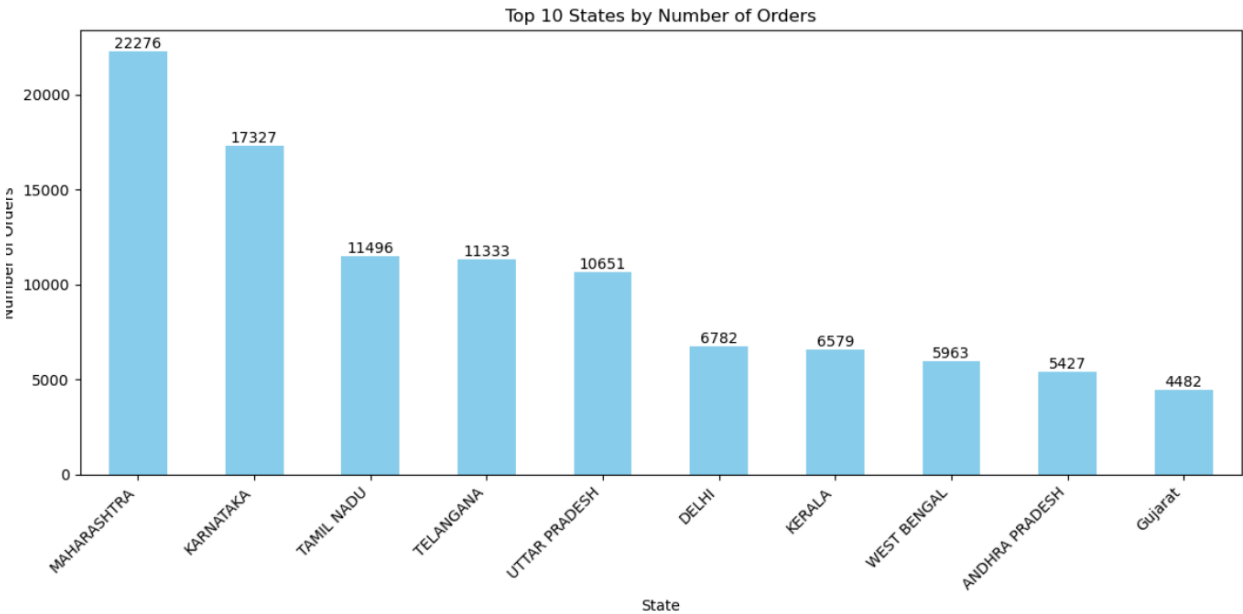
- The majority of orders have been "Shipped," accounting for 84.89% of the total.
- "On the Way" orders represent 5.33% of the total.
- "Unshipped" orders make up 5.18% of the total.
- "Cancelled" orders constitute 4.61% of the total.

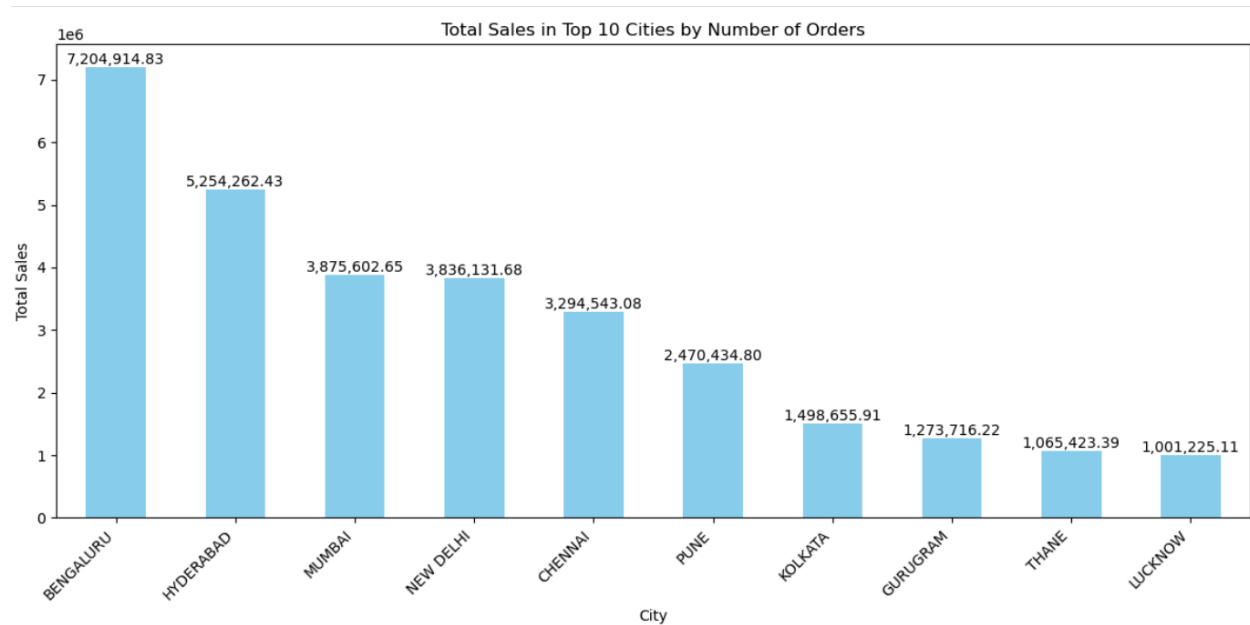
Courier status by Category :



- The "Shipped" status has the highest count for most categories, particularly for T-shirts and Shirts.
- The "Unshipped" status has a significant count for T-shirts and Trousers.
- The "Cancelled" and "On the Way" statuses have relatively lower counts across all categories.
- There are noticeable variations in the distribution of courier statuses among different categories.

Customer Segmentation Based on Buying Behavior, Location, and Key Demographics





		Number of Orders	Total Quantity	Total Sales
ship-state	ship-city			
KARNATAKA	BENGALURU	11210	10419	7204754
TELANGANA	HYDERABAD	8071	7331	5248327
MAHARASHTRA	MUMBAI	6126	5650	3875494
DELHI	NEW DELHI	5786	5243	3830286
TAMIL NADU	CHENNAI	5434	4950	3294427
MAHARASHTRA	PUNE	3863	3539	2470372
WEST BENGAL	KOLKATA	2380	2164	1498615
HARYANA	GURUGRAM	1865	1731	1273691
MAHARASHTRA	THANE	1701	1539	1065387
UTTAR PRADESH	LUCKNOW	1464	1289	1001186

Geographical Focus :

Insight: Sales are concentrated in major metropolitan areas and specific states.

Recommendations:

- Strengthen presence in top-performing cities like Bengaluru, Hyderabad, and Mumbai by increasing marketing efforts and potentially opening more physical stores or distribution centers.
- Develop targeted expansion strategies for high-potential but currently underperforming cities and states.
- Investigate the reasons for success in top-performing areas and try to replicate these factors in other regions.

Product Strategy :

Insight: T-shirts and shirts are the most popular products.

Recommendations:

- Expand the range of t-shirts and shirts, focusing on styles that are selling well.
- Introduce region-specific designs or styles that cater to local preferences in top-selling states.
- Consider bundling less popular items (like trousers) with best-sellers to boost their sales.
- Analyze the performance of Blazers, which are the third most popular item, to see if there's potential for growth in this category.

Supply Chain Optimization :

Insight: Sales are concentrated in specific regions.

Recommendations:

- Optimize inventory management by stocking more products in or near high-selling areas to reduce delivery times and costs.
- Consider setting up regional distribution centers in states like Maharashtra, Karnataka, and Telangana to improve delivery speed and reduce shipping costs.

Marketing and Promotion :

Insights: There's a significant variation in sales across regions.

Recommendations:

- Develop region-specific marketing campaigns that resonate with local cultures and preferences.
- Increase digital marketing efforts in underperforming regions to boost awareness and sales.
- Partner with local influencers or celebrities in top-performing states to further strengthen market position.

Customer Satisfaction and Retention :

Insight: While we have sales data, we need more information on customer satisfaction.

Recommendations:

- Implement a robust customer feedback system to gather insights on product quality, delivery experience, and overall satisfaction.
- Analyze the "Status" column in the dataset to identify any issues with cancellations or returns, and develop strategies to address these.
- Create a post-purchase engagement strategy to encourage repeat purchases and build brand loyalty.

Seasonal Trends :

Recommendation:

Analyze sales data over time to identify any seasonal trends. This can help in planning inventory, promotions, and product launches.

Delivery and Shipping :

Insight: The dataset includes information about shipping methods (Standard, Expedited).

Recommendations:

- Analyze the correlation between shipping methods and customer satisfaction or repeat purchases.
- Consider offering free or discounted expedited shipping for high-value customers or large orders to increase customer satisfaction.

Product Development :

Recommendation:

- Use the sales data to inform future product development. For example, if certain sizes or styles are selling particularly well, consider expanding those lines.

Competitive Analysis

Recommendation:

- Conduct a competitive analysis in the top-performing cities and states to understand what local or national competitors are doing and how you can differentiate your offerings.
- By implementing these recommendations, you can create a more targeted and effective sales strategy, improve customer satisfaction through better product offerings and service, and ultimately enhance overall business performance. Remember to continuously monitor the results of these changes and be prepared to adjust strategies as needed based on new data and market responses.

- **Tech Stack :**

Programming language : Python

Data Cleaning and Analysis :

Pandas: A powerful Python library for data cleaning and analysis, especially for working with structured data (tables).

Data Visualization Tools :

Matplotlib: A basic plotting library in Python for creating static, animated, and interactive visualizations.

Seaborn: A Python library built on top of Matplotlib, providing a higher-level interface for drawing attractive and informative statistical graphics.