**The Amazon Sales dataset** appears to be a detailed report of sales transactions on Amazon, including various attributes related to the orders.

1. **Index**: A sequential index number for each record in the dataset.
2. **Order ID**: A unique identifier for each order, which can be used to track individual transactions.
3. **Date**: The date the order was placed or the transaction occurred.
4. **Status**: Indicates the current state of the order.
5. **Fulfilment**: The entity responsible for fulfilling the order. It could be Merchant (fulfilled by the seller) or Amazon (fulfilled by Amazon).
6. **Sales Channel**: The platform through which the sale was made, in this case, Amazon.in.
7. **Ship-Service-Level**: The shipping service level selected for the order, such as Standard or Expedited.
8. **Category**: The category of the product sold, such as T-shirt, Shirt, Blazer, etc.
9. **Size**: The size of the product sold, which is particularly relevant for clothing items (e.g., S, XL, 3XL).
10. **Courier Status**: The current status of the shipping process, such as On the Way or Shipped.
11. **Qty**: The quantity of items ordered in that particular transaction.
12. **Currency**: The currency used for the transaction, here it is INR (Indian Rupees).
13. **Amount**: The total amount charged for the order.
14. **Ship-City**: The city to which the order is being shipped.
15. **Ship-State**: The state to which the order is being shipped.
16. **Ship-Postal-Code**: The postal code of the shipping address.
17. **Ship-Country**: The country to which the order is being shipped, here it is IN (India).
18. **B2B**: A boolean indicating whether the order is a B2B (Business-to-Business) transaction. TRUE indicates a B2B sale, while FALSE indicates a regular sale.
19. **Fulfilled-By**: Specifies how the order was fulfilled, such as Easy Ship, which is a service where Amazon facilitates the shipping.
20. **New**: This column seems to be empty in the sample data, possibly meant to indicate if the order is new or recently updated.
21. **PendingS**: This column is empty in the provided sample data, and its purpose is unclear without further context.

The **key objectives** are as follows:

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 behaviour, 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.

**Step 1:** First the data is cleaned using various python libraries. The steps are elaborated as follows.

The following data cleaning steps were followed during the cleaning of the Amazon sales dataset:

**Importing necessary libraries:**

* numpy for numerical operations
* pandas for data manipulation and analysis

**Loading the data:**

* The dataset was loaded into a pandas DataFrame.

**Displaying the data:**

* The first few rows of the DataFrame were displayed to get a general overview of the data.

**Handling missing values:**

* **Dropping null columns:** Columns with a high percentage of missing values were dropped.
* **Filling missing currency values:** Missing currency values were filled with the most frequent value.
* **Filling missing amount values:** Missing amount values were filled with zero.
* **Dropping missing values in shipping information:** Missing values in ship-city, ship-state, ship-postal-code, and ship-country were dropped.

**Converting the date column:**

* The date column was converted to a pandas datetime object in the format dd-mm-yyyy.

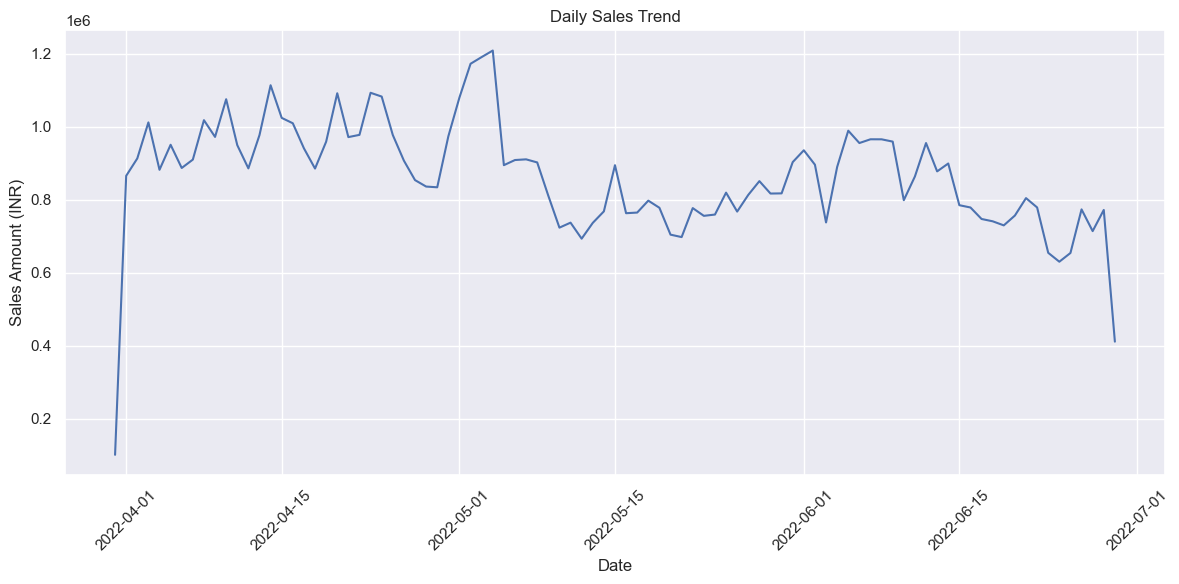
**Exporting the data:**

* The cleaned dataset was exported as a CSV file for further analysis.

**Step 2**: A dashboard in tableau is made using the cleaned dataset. It can be accessed through this link: <https://public.tableau.com/shared/7SMM2HGNB?:display_count=n&:origin=viz_share_link>

From the viz. of Amount week and category and Quantity by week and category the following insights can be determined:

**Sales Overview**

****

**From Amount by week and category the following insights can be determined.**

* **Overall sale** is at **peak** in the **first week of May** and a **steep decline** is seen at the **first week of July**.
* **T-shirts** (in green) appear to be the top-performing product category, consistently generating the highest sales amounts each week. There is a noticeable peak **in late April and early May**, followed by fluctuations, but it remains the top category.
* **Shirts** (in orange) also show strong sales, second to T-shirts, with a noticeable peak **around mid-April**, followed by a gradual decline.
* **Blazer** (in blue) and **Trousers** (in red) are mid-tier performers. Their sales appear to be relatively stable but significantly lower than those of T-shirts and Shirts.
* Other categories such as **Socks**, **Perfumes**, **Shoes**, **Wallets**, and **Watches** contribute to a smaller portion of the sales. They show minimal fluctuations and maintain low sales volumes throughout the period.

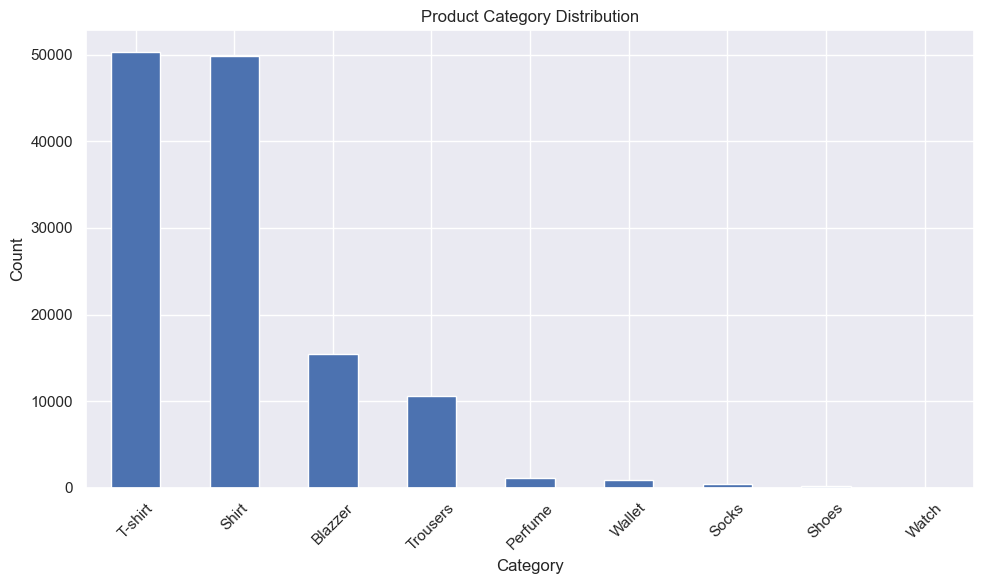
**Trends and Patterns:**

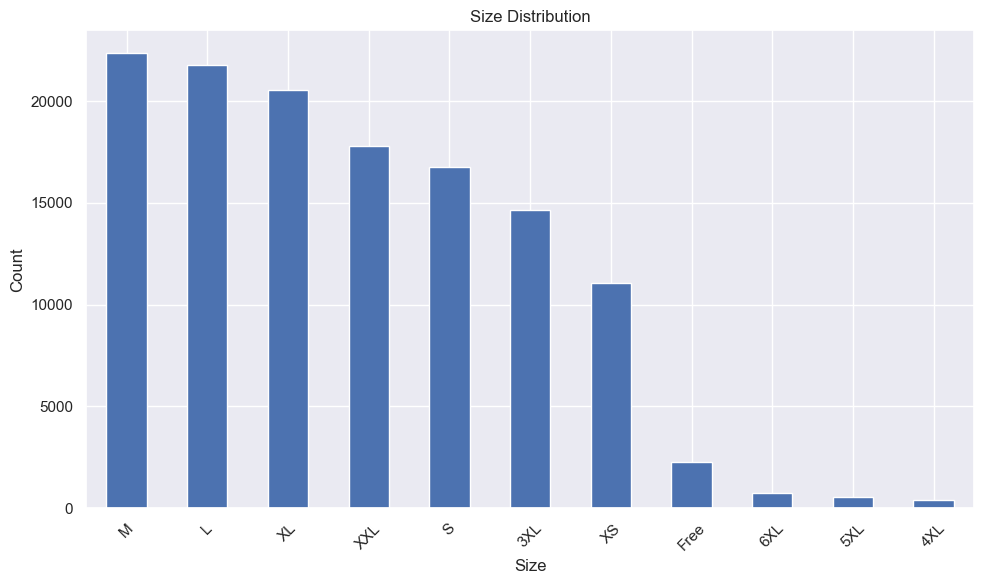
* **May 2022** shows a significant peak across multiple categories, particularly T-shirts and Shirts, which could indicate a successful promotion or seasonal trend. It can also be due to the fact of summer season that the sales of T-Shirts and Shirts increases during this time period.
* **June 2022** indicates a general decline in sales across most categories, which might suggest a post-peak slowdown as the rainy season begins.

**Recommendations:**

* **Focus on T-shirts and Shirts**: These categories are consistently high performers, suggesting they are popular among customers. Consider expanding inventory or introducing new styles within these categories.
* **Investigate May Promotions**: The spike in May sales might be due to a specific event or promotion or due to advent of summer season. Analysing the cause can help replicate this success in future months.
* **Monitor Declining Categories**: The decline in June sales across categories, especially high performers, should be monitored. If the trend continues, it may require action such as new marketing strategies or inventory adjustments.

**Product Analysis**





From Quantity by week and category the following insights or recommendations can be given (**refer to tableau dashboard**)

**Quantity Sold by Category:**

* **T-shirts (in green)** consistently represent the largest portion of total items sold each week, confirming their popularity.
* **Shirts (in orange)** are the second most sold category, showing significant contribution across all weeks.
* **Blazers (in blue)** also account for a noticeable quantity, though their share is smaller compared to T-shirts and Shirts.
* **Trousers (in red)** maintain a consistent but smaller portion of sales quantity, indicating steady demand.
* Other categories like Socks, Perfumes, Shoes, Wallets, and Watches contribute minimally to the overall sales quantities.

**Trends and Patterns:**

* **Month of April, 2022**: This month shows the highest quantity of products sold, particularly with T-shirts and Shirts driving most of the sales.
* **Week of June 26, 2022**: There is a noticeable drop in the quantity sold, indicating a decline in sales activity by the end of June.

**Insights:**

* **Product Focus**: T-shirts and Shirts should remain a focus for marketing and inventory due to their consistent high volume.
* **Sales Fluctuations**: The fluctuations in sales quantities suggest possible external factors such as promotions, seasonality, or changing customer preferences.

**Quantity by size and category**

* **Distribution of Product Categories:** The categories T-shirt (green) and Shirt (orange) dominate in terms of quantities sold across almost all sizes. This indicates that these two categories are likely the most popular.
* **Size Popularity:** Medium (M), Large (L), and Extra Large (XL) sizes seem to be the most popular, with the highest quantities sold. Smaller sizes like XS and less common sizes like 4XL to 6XL have much lower sales volumes.
* **Category-Size Insights:** The T-shirt and Shirt categories have strong sales across multiple sizes, especially in M, L, and XL, suggesting these products are in high demand across these sizes.

**Recommendations:**

* **Focus on Best Sellers:** Consider focusing inventory and marketing efforts on the most popular sizes (M, L, XL) and categories (T-shirts and Shirts).

**Fulfilment Analysis**

**Amazon Fulfilment:**

(**refer to tableau dashboard**)

**72.10%** of orders are fulfilled by Amazon. This suggests that the majority of sales are being handled directly by Amazon, likely through their fulfilment centers.

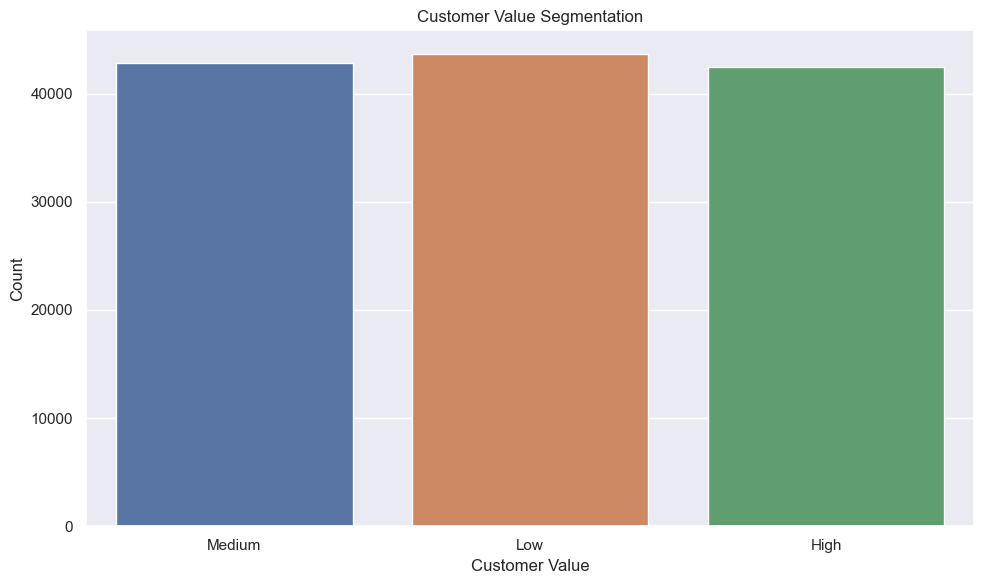
**Effectiveness:** Using Amazon for fulfilment can be advantageous due to their established logistics network, which might result in faster delivery times and potentially higher customer satisfaction. It also indicates a strong reliance on Amazon's fulfillment services, which could provide benefits like better storage, handling, and delivery.

**Merchant Fulfillment:**

**27.90%** of orders are fulfilled by merchants. This represents a smaller portion of total sales but still a significant percentage.

**Effectiveness:** Merchant fulfillment allows for more control over the shipping process, potentially lower costs for sellers who can manage their logistics efficiently, and flexibility in handling special requests or unique products. However, it may also lead to variability in delivery times and customer experience depending on the merchant's resources and capabilities.

**Customer Segmentation**



As it can be seen there is a slightly more number of low value customers placing orders in the site.

 **Low Value**: These customers might be occasional buyers with lower total spending.

 **Medium Value**: These customers are more engaged, possibly with moderate spending and frequency.

 **High Value**: These are likely the most valuable customers, possibly frequent buyers with higher total spending.

The fact that the segments are balanced indicates a healthy customer base, but it also highlights the need to maintain or increase engagement across all segments to ensure growth.

A graph of different colored bars

Description automatically generated

It can be inferred that the state of Maharashtra and Karnataka has a strong customer base and efforts should be made to retain this customer base.

**Insights:**

**Maharashtra Dominates:**

Maharashtra has the highest number of customers, with over 20,000, making it a key market.

**Karnataka as a Major Market:**

Karnataka also has a significant customer base, though slightly less than Maharashtra, suggesting it is another important region for sales.

**Moderate Customer Bases:**

States like **Tamil Nadu, Uttar Pradesh and Telangana** have a moderate number of customers, indicating steady but not dominant markets.

**Lower Representation**:

States like Bihar, Punjab, Odisha, Rajasthan and Madhya Pradesh have relatively lower customer counts, suggesting these are less significant markets for the business.

**Recommendations**:

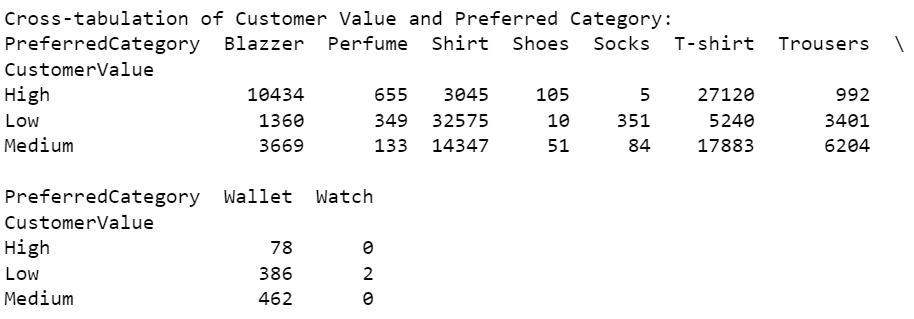
* **Targeted Marketing:**

Maharashtra and Karnataka could be prioritized for targeted marketing efforts or product launches.

States with moderate customer bases could benefit from enhanced engagement strategies to boost sales.

* **Growth Opportunities**:

The states with lower customer numbers might represent growth opportunities where increased marketing or localized strategies could expand the customer base.



**Insights**:

**High-Value Customers**:

* T-shirt: The most popular category with 27,120 high-value customers.
* Blazer: The second most popular category with 10,434 high-value customers.
* Shirt: 3,045 high-value customers prefer shirts.
* Trousers and Wallets also have notable high-value customer segments.

**Low-Value Customers**:

* Shirt: Dominates the low-value customer category with 32,575 customers.
* T-shirt: Significantly lower at 5,240 customers compared to shirts.
* Trousers: With 3,401 customers, is another important category.

**Medium-Value Customers:**

* T-shirt: The most popular in this segment as well, with 17,883 customers.
* Shirt: Follows with 14,347 customers.
* Blazer: Is also a preferred category with 3,669 customers.

**Implications**:

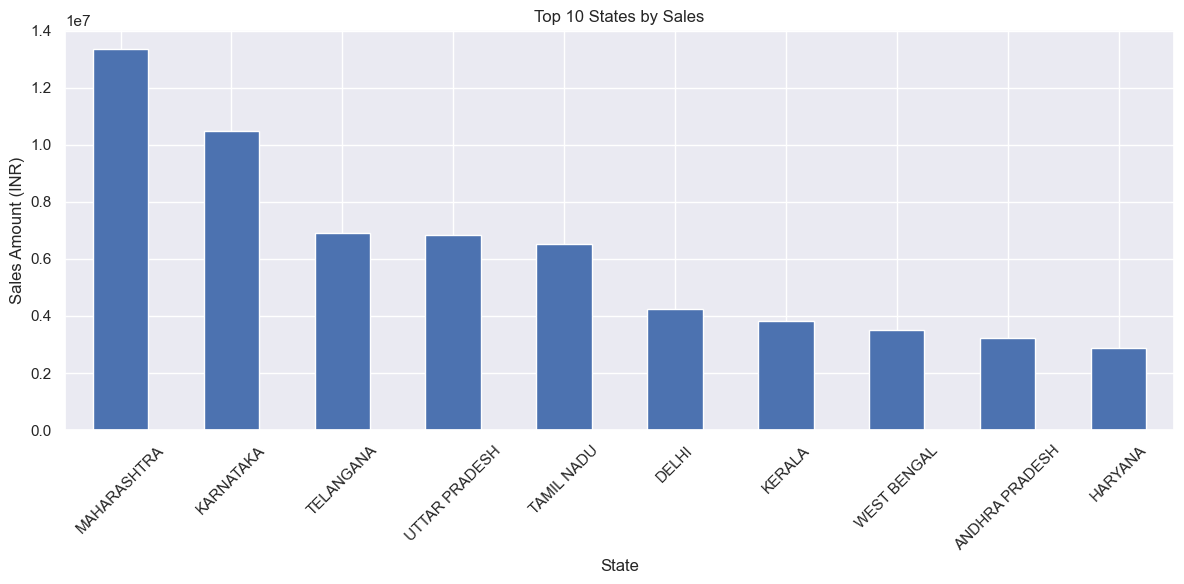
**Product Popularity:**

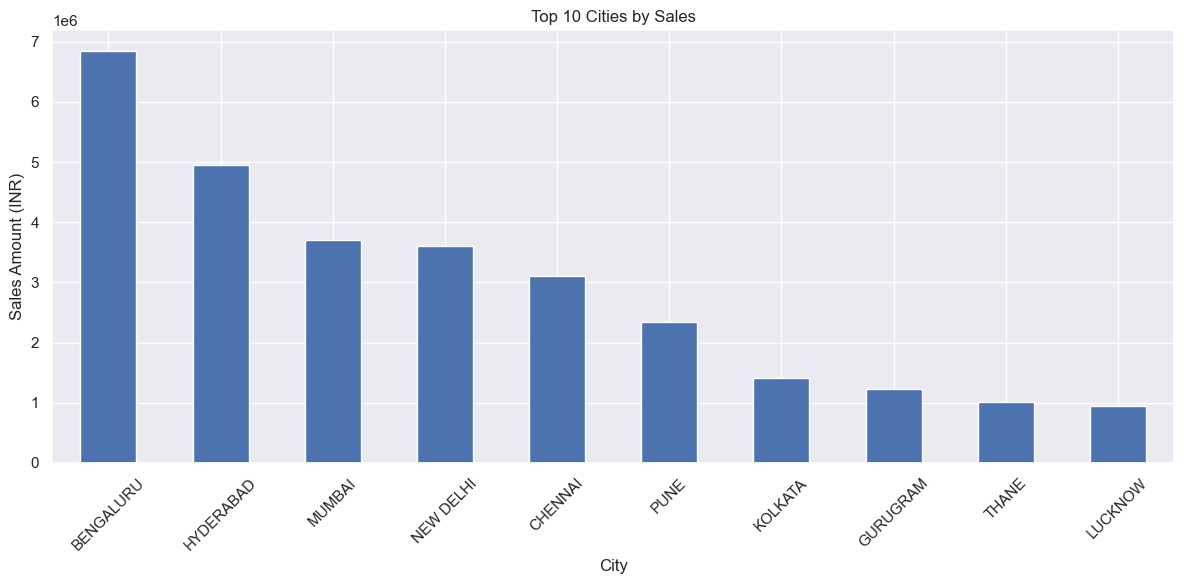
* T-shirts and Shirts are universally popular across all customer value segments.
* Blazers are particularly favored by high and medium-value customers.

**Recommendations**:

* High-value customers: Could be targeted with premium T-shirt, Blazer, and Shirt offerings.
* Low-value customers: Might respond well to promotions or discounts on Shirts.
* Medium-value customers: May prefer quality T-shirts and Blazers, which could be highlighted in marketing campaigns.

**Geographical Analysis**

****



**Key Findings:**

**States:**

* Maharashtra is the top-performing state with sales of around 13.5 million INR.
* Karnataka follows as the second-highest with about 10.5 million INR in sales.
* There's a significant gap between the top two states and the rest.

**Cities:**

* Bengaluru leads in city-wise sales with approximately 6.8 million INR.
* Hyderabad is the second-highest performing city with about 5 million INR in sales.
* The top 5 cities (Bengaluru, Hyderabad, Mumbai, New Delhi, and Chennai) account for a large portion of overall sales.

**Insights:**

* State vs. City Performance: While Maharashtra is the top state, its capital Mumbai is only the third-highest performing city. This indicates that sales in Maharashtra are more evenly distributed across multiple cities.
* Market Penetration: There's a significant drop-off in sales after the top 5-6 states/cities, indicating potential for growth in other regions.

**Recommendations:**

* Focus on High Performers: Continue to strengthen market presence and strategies in high-performing states like Maharashtra, Karnataka, Tamil Nadu and Telangana.
* Expand in Potential Markets: Develop targeted strategies to boost sales in large states with currently lower sales, such as Uttar Pradesh, West Bengal and Odisha. These represent significant growth opportunities.