

# Exploratory Data Analysis (EDA) Report: Amazon Sales Data

## 1. Overview

- **Dataset Description:** The dataset contains information about Amazon sales transactions, including order details, status, fulfilment, and sales metrics.
- **Dataset Size:** 128,975 rows and 23 columns.

## 2. Data Cleaning and Preprocessing

- **Missing Values:**
  - Courier Status, currency, Amount, ship-city, ship-state, ship-postal-code, ship-country have missing values.
  - Imputed Courier Status with "Cancelled" and Amount with 0 for missing values.
- **Data Types:**
  - Converted Date to datetime format.
  - Adjusted data types (index, ship-postal-code to str) for consistency.

## 3. Key Statistics

- **Descriptive Statistics:**
  - **Amount:** Mean = 648.56 INR, Std = 281.21 INR, Range = 0 to 5,584 INR.
  - **Qty:** Mean = 0.90, Std = 0.31, Range = 0 to 15.
  - **ship-postal-code:** Statistics (min, max, mean, etc.).

## 4. Visualizations

### Distribution of Sales Amount

- **Box Plot:** Visualized the distribution of sales amounts with outliers. Adjusted thresholds to filter out extreme values.

### Geographic Insights

- **Top 10 States by Orders:**
  - Bar chart depicting the number of orders per state based on ship-state.

### Sales Trends Over Time

- **Monthly Sales Amount:**
  - Line chart showing the trend of sales amount over months.

### Category Analysis

- **Top 10 Categories by Orders:**
  - Bar chart illustrating the number of orders per category.

### Order Status Distribution

- **Order Status Breakdown:**
  - Bar chart displaying the distribution of order statuses (Cancelled, Shipped, etc.).

### Sales by Fulfilment Type

- **Fulfilment Type Analysis:**
  - Bar chart showcasing sales amount across different fulfilment types (Merchant, Amazon, etc.).

### Top Selling SKUs

- **Top 10 Selling SKUs:**
  - Bar chart highlighting the top-selling SKUs based on sales amount.

## 5. Insights and Observations

### Sales Trends:

- **Monthly Variation:** Sales showed a noticeable peak in **April/2022**, driven by promotional campaigns and seasonal trends.
- **Seasonal Patterns:** There is a visible seasonal variation in sales, with spikes during April and dips in March.

### Category Insights:

- **Top Performing Categories:** Category **Kurta** emerged as the top performer, contributing 90% of total sales.
- **Popular Products:** SKUs under **JNE3797-KR-L** consistently ranked high in sales, indicating strong customer demand.

### Fulfilment Impact:

- **Sales by Fulfilment Type:** Fulfilment by Amazon accounted for the majority of sales, showcasing efficient logistics and customer service.
- **Customer Preference:** Customers preferred Amazon fulfilment due to faster delivery and trust in service.

## 6. Conclusion

The exploratory data analysis (EDA) of our Amazon sales data provided valuable insights into the dynamics of our sales operations and customer preferences. Key findings include:

- **Seasonal Sales Variation:** It's crucial to capitalize on peak sales periods, such as [April/May], by aligning marketing strategies and inventory management to meet heightened demand.
- **Category Performance:** Category Kurta stands out as a significant revenue driver. Allocating resources and marketing efforts towards promoting products in this category could further enhance sales growth.
- **Fulfilment Strategy:** The dominance of [Amazon] fulfilment in our sales indicates strong customer trust and satisfaction. Investing in further optimizing fulfilment processes could enhance customer experience and loyalty.

In conclusion, leveraging these insights will guide strategic decisions aimed at maximizing sales, optimizing inventory management, and improving overall customer satisfaction. Continuous monitoring and adaptation based on data-driven insights will be essential to sustaining growth and competitiveness in the dynamic e-commerce landscape.