

FINAL REPORT BASED ON AMAZON SALES DATA—

A. Data Overview

The dataset initially contained **128,976 entries** across **21 columns**. After cleaning, the dataset size used for analysis was **128,941 rows**. The Date column was parsed as a datetime object.

B. Data Cleaning and Quality Issues

1. **Missing Values:** Significant missing values were initially identified:

- currency and Amount: 7,800 missing values.
- Shipping information (ship-city, ship-state, ship-postal-code, ship-country): 35 missing values.
- fulfilled-by: **89,713** missing values.
- New and PendingS: **Completely null** (128,976 missing values initially; 128,941 after dropping shipping nulls).

2. **Null Handling:**

- Rows missing shipping information were dropped.

- Missing values in the **fulfilled-by** column were imputed with the value **'Amazon'**.

- For cancelled orders, the **qty** and **amount** were set to **0**.

3. **Amount Imputation:** Missing **amount** values (or amounts of 0 for non-cancelled orders) were filled using the **mean amount** calculated from non-cancelled orders with amounts greater than 0.

4. **Data Standardisation:** Status values were grouped into fewer, standardized categories like 'Delivered', 'Shipped', 'Pending', 'Cancelled', and 'Returned/Damaged'. The boolean **b2b** column was converted to **'Yes'** or **'No'** strings for transactions within India.

C. Financial Summary and Outliers

- **Average Quantity (Qty):** The mean quantity is approximately **0.90**.

- **Average Amount:** The mean transaction amount is approximately **661.41** (inferred to be INR, as INR is the only currency present).

- **Amount Range:** Amounts range from a minimum of **199.00** to a maximum of **5,584.00**.

- **Outlier Rate:** Outlier detection using the Interquartile Range (IQR) method revealed an **outlier rate of 16.6%**

in the **amount** column, corresponding to 3,747 outlier entries.

II. Univariate and Bivariate Analytical Insights

A. Fulfillment and Status

- **Fulfillment Type:** The majority of orders were fulfilled by **Amazon** (89,691 counts), significantly outpacing Merchant fulfillment. This difference is also visible when viewing orders categorized by **status** and **fulfilment**.

- **Order Status:** The most frequent order status is **'Shipped'** (77,797 occurrences).

- **Service Level:** The most common shipping service level utilized is **'Expedited'** (88,608 occurrences).

- **B2B Status:** Only a small minority of orders are **B2B** (872 True counts) compared to False counts (128,069).

B. Product Performance (Category and Size)

1. By Order Count (Quantity):

- **Top Category:** **T-shirt** is the most frequently ordered category (50,280 counts).

- **Top Size:** The most popular size is **'M'** (22,366 orders).

2. By Sales Amount:

- **Top Category by Sales:** **T-shirt** generates the highest total sales amount, followed by **Shirt**.

- **Top Size by Sales:** The data groups sales by size.

C. Geographic Performance

- **Shipping Country:** All recorded shipments are within **IN (India)**.

- **Top States/Cities by Frequency:** The state/city combination that appears most frequently in the dataset is **Maharashtra/maharashtra** (22,272 counts)