# Exploratory Data Analysis Report

Internship Project

**Sahil Chakraborty**

**22 August 2025**

# Amazon Sales Report Analysis

# Overview

This project involves an **Exploratory Data Analysis (EDA)** of Amazon sales transactions. The dataset contains details such as order ID, date, order status, fulfilment method, sales channel, product category, size, quantity, amount, and shipping information. The primary goal is to extract **actionable insights** that can help optimise sales strategies, improve operational efficiency, and enhance customer satisfaction.

## Dataset

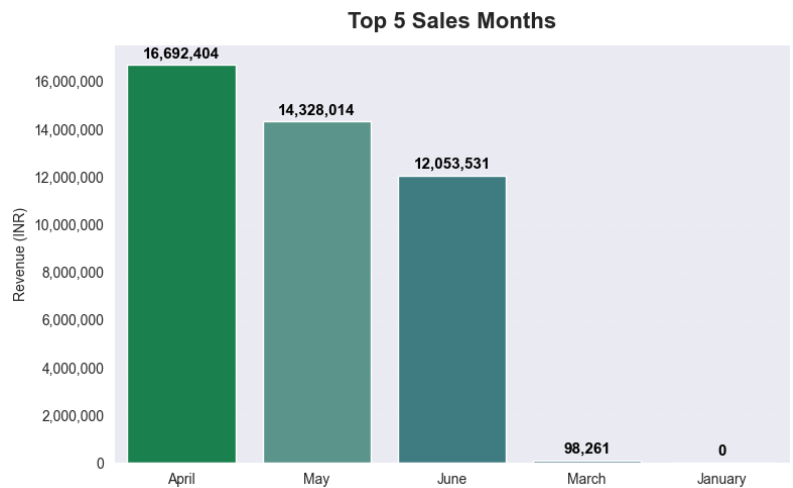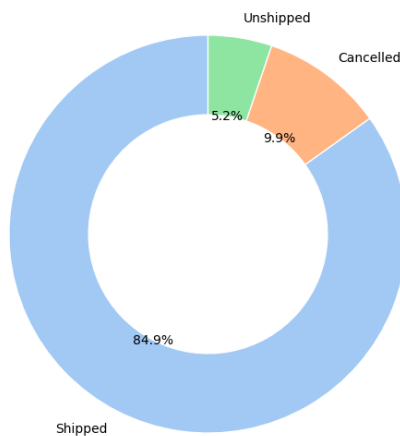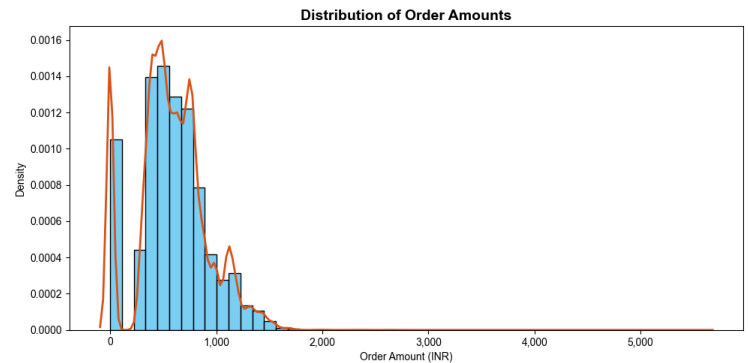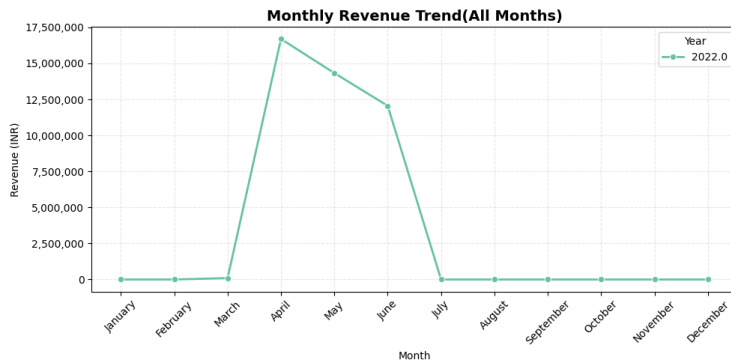The dataset consists of sales records with the following key features:

- **Order Information:** Order ID, Date, Status, Sales Channel, Fulfilment Method

- **Product Details:** Product Category, Size, Quantity, Amount

- **Customer/Shipping Details:** City, State, Pin Code, Ship Service Level

## Key Objectives

1. **Sales Overview** – Understand overall sales performance, trends, and patterns over time.

2. **Product Analysis** – Analyse product categories, sizes, and quantities to identify popular items.

3. **Fulfilment Analysis** – Examine fulfilment methods and their effectiveness.

4. **Customer Segmentation** – Segment customers based on purchase behaviour and location.

5. **Geographical Analysis** – Explore sales distribution across states and cities.

6. **Business Insights** – Provide actionable recommendations for improving sales and customer satisfaction.

# Analysis & Visualisations

## 1. Sales Overview

**Monthly Revenue Trend(All Months)**

**Distribution of Order Amounts**

**Top 5 Sales Months**

16,692,404  14,328,014  12,053,531  98,261  0

April  May  June  March  January

Unshipped  Cancelled
5.2%  9.9%
84.9%
Shipped

## Key Insights :

1. **Highly Seasonal Sales** – Revenue is heavily concentrated in **April–June**, with **April (~₹16.7M)** being the peak. This shows that the business experiences **strong seasonal demand**, and marketing campaigns or product launches should focus on this period.

2. **Majority of Orders in the Core Spending Segment** – Most orders fall in the ₹**200–₹1,000 range**, representing the **core customer base** driving consistent revenue. Targeted promotions here can boost overall sales volume efficiently.

3. **High-Value Orders Drive Revenue Despite Low Frequency** – Rare **premium orders (>₹1,500)** and mid-to-high orders significantly contribute to total revenue. **Upselling and retention strategies** for these high-value customers can disproportionately increase profits.

4. **Fulfilment Performance is Strong but Needs Attention**
   – About **84.9% of orders are shipped successfully**, but **~10% cancellations** and **~5.2% unshipped orders** suggest **operational bottlenecks** and areas for improvement in logistics and order management.

5. **Post-Peak Decline Indicates Need for Retention Strategies**
   – After the April peak, revenue drops **~14–17% in May–June**, highlighting the importance of **discounts, retention offers, or marketing initiatives** to maintain sales momentum during declining months.

-----------------------------------------------------------------------
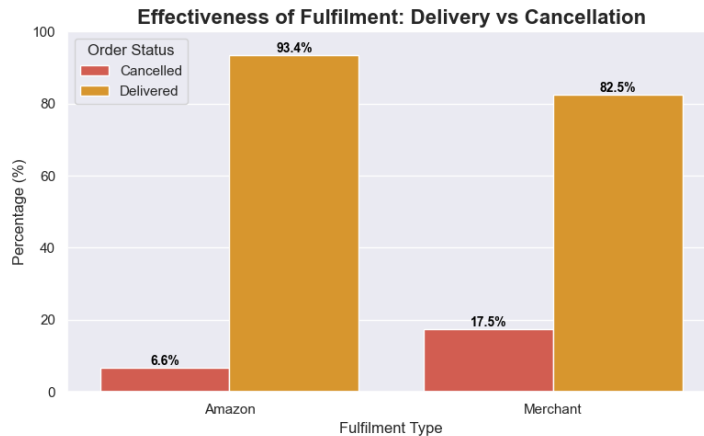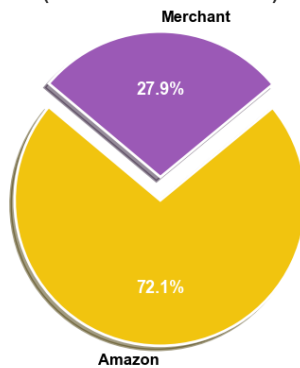
# 2. Product Analysis
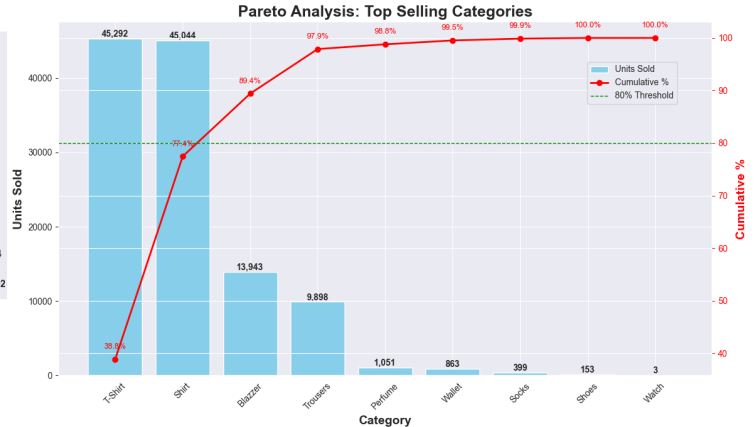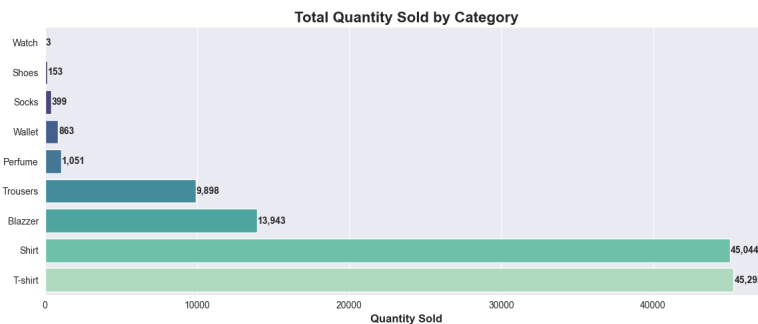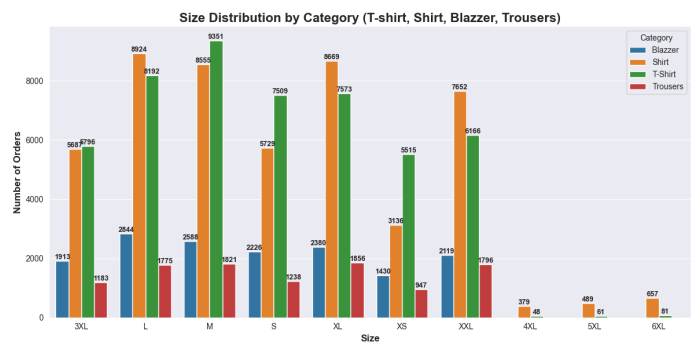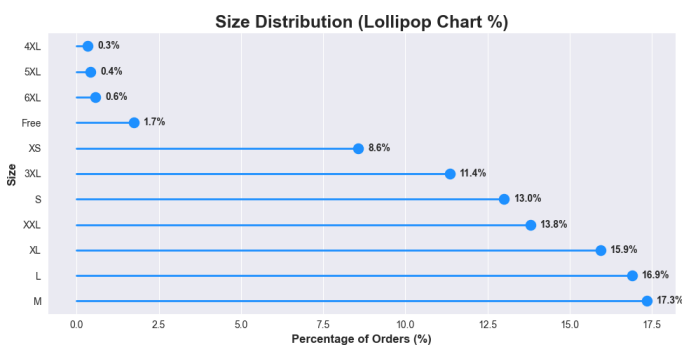


### Key Insights

1. **Top Products Dominate Sales (Pareto Principle)– T-Shirt and Shirt** together account for over **83–85% of total units sold**, showing that a **small fraction of products drives the majority of sales**. This emphasises the importance of prioritising inventory, marketing, and promotions for these top-performing categories.

2. **Significant Drop-Off Beyond Top Products**– After the leading two categories, sales drop sharply. **Blazer and Trousers** form a second tier with moderate sales, while **Perfume, Wallet, Socks, Shoes, and Watch** contribute very little. This highlights the **lack of diversity in high-performing products**, suggesting potential discontinuation or reevaluation of low-demand items.

3. **Extreme Underperformance in Certain Products**– The **Watch** category sold only **3 units**, signalling either **very low demand, poor visibility, or inventory issues**. Such severe under performers need strategic decisions—either phasing out, relaunching with better marketing, or limited stock models.

4. **Demand Concentrated in Popular Sizes**
   – Medium to XL sizes (M, L, XL, XXL, S) account for over **78% of orders**, while XS and extreme sizes (4XL–6XL) contribute less than 1% each. This insight guides **inventory optimisation**, focusing on high-demand sizes while minimising overstock for fringe sizes.

5. **Category-Size Nuances Influence Stocking** – While T-Shirts dominate small and medium sizes, Shirts lead in larger sizes (L–XXL). Blazer and Trousers lag consistently. This suggests **size-specific stocking strategies**, ensuring top categories are fully supplied across their preferred sizes, while niche products are stocked conservatively.

---

# 3. Fulfilment Analysis



**Key Insights**

1. **Amazon Dominates Fulfilment** – Handles **72.1%** of non-cancelled orders (~84K), proving it's the primary logistics channel.

2. **Higher Reliability** – **93.4% of Amazon orders** are delivered successfully, with only **6.6% cancellations**, showing strong trust and efficiency.

3. **Merchant Weakness** – Merchants fulfil **27.9% of orders (~32K)** but face a **17.5% cancellation rate**, nearly **3x higher** than Amazon.

4. **Customer Experience Gap** – Amazon buyers have a **9 in 10 chance** of delivery, while Merchant buyers risk almost **1 in 6 orders being cancelled**.

5. **Strategic Implication** – To compete, Merchants must **improve stock management and logistics** or shift to Amazon FBA to reduce cancellations.

# 4. Customer Segmentation
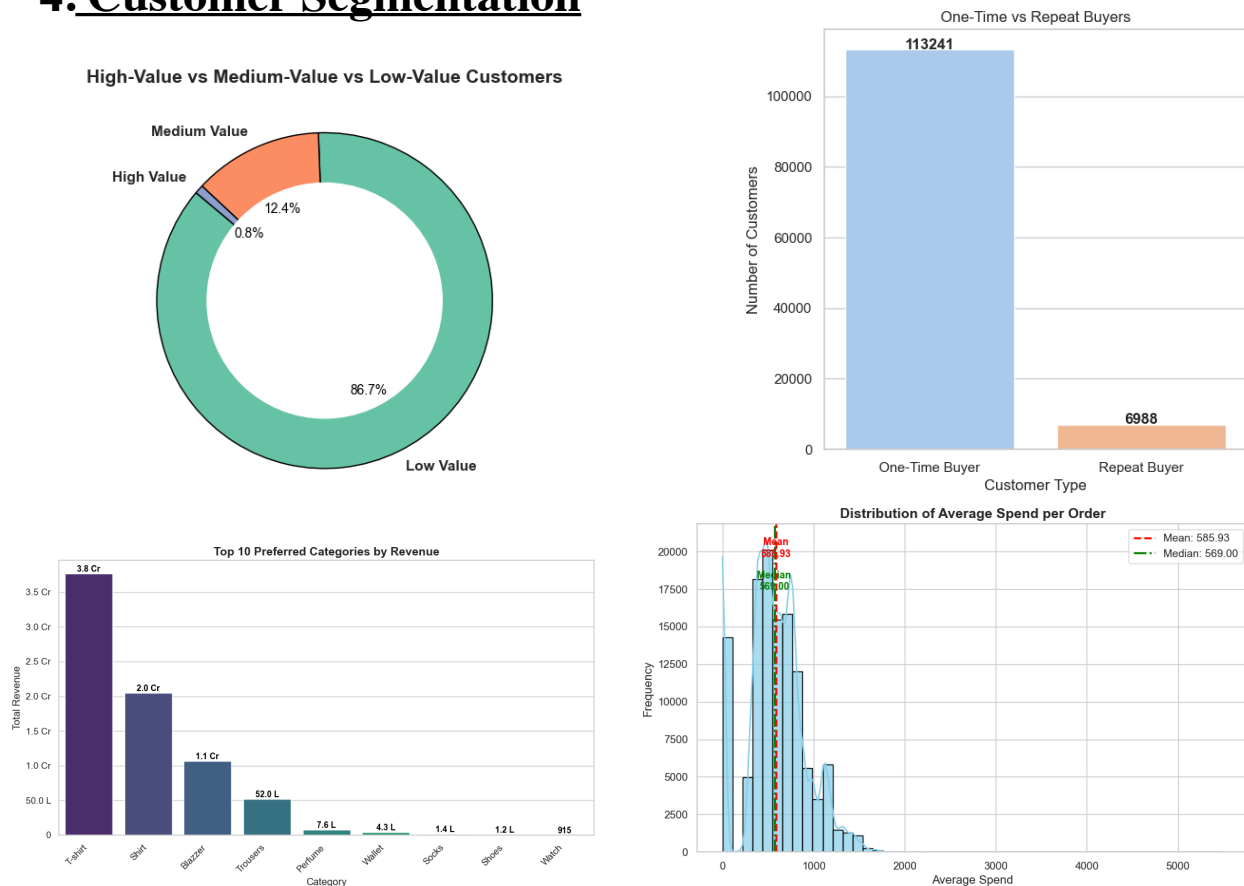


## Key Insights

1. **Customer Base is Low-Value Heavy** – Nearly **87% of customers spend <₹1000**, showing dependence on **volume, not high spenders**. This makes customer acquisition cost-sensitive.

2. **Weak Retention is a Major Risk** – About **94% are one-time buyers**, with only **6% repeat customers**. The business lacks loyalty and long-term engagement, highlighting a **serious retention gap**.

3. **Mid-Range Spending Dominates, Premium Segment Exists** – Average order value is ₹**586**, mostly in the ₹**500–₹1000** range. A small premium group (>₹2000) exists and, if nurtured, can drive disproportionate revenue.

4. **Revenue is Over-Concentrated in Apparel** – T-Shirts (₹**3.8 Cr**) and Shirts (₹**2 Cr)** dominate revenue, while accessories barely contribute. The business is **highly apparel-dependent**, with cross-sell potential in underperforming items.

5. **Growth Strategy Should Balance Retention & Upselling** – To grow sustainably:

   ○ Introduce **loyalty/retargeting** to convert one-time buyers.

- Use **bundling & cross-selling** to push AOV above ₹1000.

- Focus on apparel while leveraging accessories as **add-on boosters**.

––––––––––––––––––––––––––––––––––––––––

# 5. Geographical Analysis
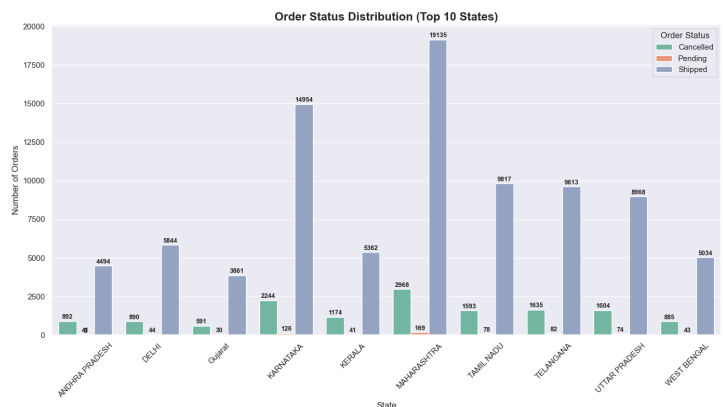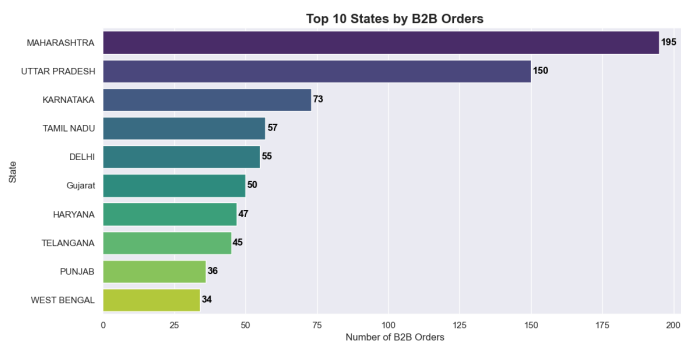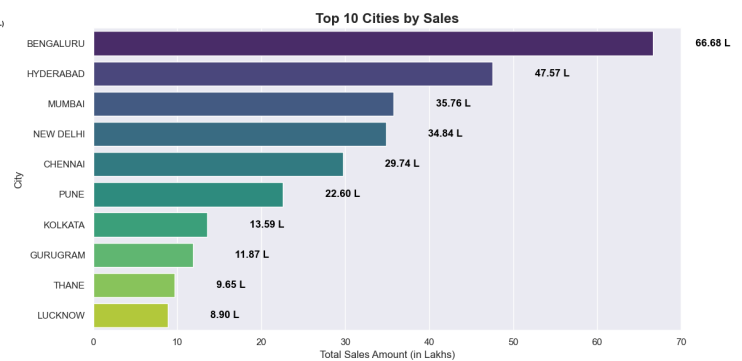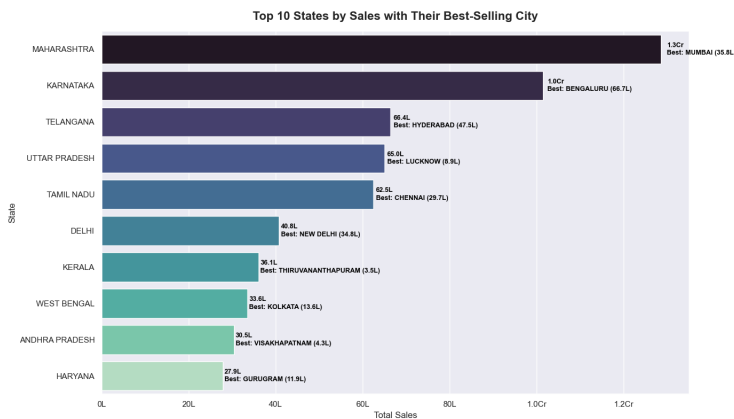


## Key Insights

1. **Maharashtra & Karnataka are the backbone** – Maharashtra leads both in **order volume (19K+) and sales (~₹1.3 Cr)**, while Karnataka follows with **~₹1.0 Cr**. Together, they dominate overall business performance.

2. **Bengaluru & Mumbai are the crown jewels** – Bengaluru (**~₹66.7L, 67% of Karnataka**) and Mumbai (**~₹35.8L, 27% of Maharashtra**) are the **most critical single markets**, contributing outsized revenue compared to other cities.

3. **Sales are metro-centric & highly concentrated** – Hyderabad, Chennai, Delhi, and Kolkata show the same pattern → **1 city contributes nearly half or more of the state's revenue.** This makes the business heavily reliant on Tier-1 cities.

4. **Cancellation hotspots in Maharashtra & Karnataka**
– Despite strong sales, both states face **high cancellation volumes (~3K in Maharashtra, ~2.2K in Karnataka)**. Reducing cancellations here will directly protect top-line revenue.

5. **B2B demand is concentrated in Maharashtra & UP**
– Together, they account for **~50% of B2B orders**, showing strong industrial/SME dependence. Other states contribute much smaller shares, creating a **two-tier B2B structure**.

---

# Conclusion and Final Insights

The business shows **strong revenue performance (₹75.4M from ~129K orders)**, but sales are heavily skewed towards a few products and regions:

- **Product Concentration:** Following the **80/20 rule**, most revenue comes from **T-shirts and Shirts**. However, the **Average Order Value (AOV)** is only ₹585, with single-item purchases dominating. This highlights a **big upselling opportunity**.

- **Customer Behaviour:** The base is dominated by **one-time, low-value buyers**. Over 86% of customers spend <₹1,000, while only 6% return for repeat purchases—signalling a **retention challenge**.

- **Geographic & Seasonal Dependence:**

  - **Regions:** Sales are concentrated in **Maharashtra and Karnataka**, especially **Bengaluru and Mumbai**.

  - **Seasonality:** A strong spike in **Q2 (Apr–Jun)** with sharp drops elsewhere. Weekly peaks occur on **Saturdays**.

- **Operational Issues:** About **10% of orders are cancelled**, mostly from **Merchant-fulfilled sellers**—triple the rate of Amazon-fulfilled ones. This weakens trust and reliability.

# Strategic Recommendations

### 1. Increase Average Order Value (AOV)

- Introduce **bundled combos** (e.g., *Shirt + Blazer*, *T-shirt + Trousers*) to push multi-item purchases.

- Add **cross-sell prompts** for accessories (Perfume, Wallets, Socks) at checkout.

- Set a **free shipping threshold (₹750+)** to encourage larger carts.

### 2. Improve Customer Retention

- Launch a **loyalty program** with rewards for repeat purchases.

- Use **personalised follow-ups** (emails/SMS with product suggestions + discounts) to bring one-time buyers back.

### 3. Optimise Marketing & Operations

- Focus campaigns on **top categories (T-shirts, Shirts)** and high-performing regions like **Bengaluru, Mumbai**.

- Maximise **Saturday spikes** and plan inventory/ads for **Q2 surge**.

- Reduce **cancellations** by auditing merchant-fulfilled sellers or shifting them to Amazon fulfilment.

By targeting **higher order value, better retention, and operational fixes**, the business can shift from low-value, one-time sales to a **more profitable and sustainable growth model**.

# Future Scope

- Advanced customer segmentation using clustering (K-Means/DBSCAN).

- Predictive modelings of future sales using Machine Learning.

- Interactive dashboards with Tableau/Power BI for business teams.

*This project was done as part of my Internship on Data Analytics. The analysis provides both high-level patterns and actionable insights to support better decision-making for e-commerce growth , Thank you.*

=======================================================================
=======================================================================