

INNOBYTE SERVICES

DATA ANALYTICS INTERNSHIP PROGRAM

FINAL REPORT

PROJECT TITLE: AMAZON SALES ANALYSIS

SUBMITTED BY:
BESTHA SATHYAKRISHNA
Email: bestakrishna143@gmail.com
ID= IS/A1/G7364

DURATION: OCTOBER 2025

TOOLS USED: PYTHON, GOOGLE COLAB, PANDAS, MATPLOTLIB, SEABORN

Acknowledgment

I would like to express my sincere gratitude to **Innobyte Services** for the opportunity to undertake this Data Analytics Internship. This project helped me apply analytics concepts to real-world business data, strengthening my skills in data cleaning, visualization, and insight generation. I'm thankful to the mentorship team for their guidance and constructive feedback throughout the internship.

Table of contents

S.no	Name	Pg.no
1	Project Overview	4-4
2	Key Objectives	5-5
3	Executive Summary	6-6
4	Methodology	7-7
5	Dataset Description	8-8
6	Data Cleaning & Preparation	9-9
7	Sales Overview	10-11
8	Product Analysis	12-13
9	Fulfilment Analysis	14-15
10	Customer Segmentation	16-17
11	Geographical Analysis	18-20
12	Business Insights Summary	21-21
13	Recommendations	22-23
14	Limitations & Future Scope	24-24
15	Conclusion	25-25

1. Project Overview

This project focuses on analysing Amazon sales transaction data to uncover patterns and business insights that can optimize decision-making.

The dataset includes order details such as order ID, order date, fulfilment type, sales channel, product category, quantity, revenue, and shipping information.

The analysis aims to understand sales performance trends, customer behaviour, and fulfilment effectiveness, while providing actionable strategies to enhance sales and customer satisfaction.

2. Key Objectives

Sales Overview: Identify revenue trends, growth, and sales distribution over time.

Product Analysis: Determine top-performing products and their contribution to overall sales.

Fulfilment Analysis: Evaluate performance of fulfilment methods and cancellation rates.

Customer Segmentation: Categorize customers by spending behaviour.

Geographical Analysis: Explore regional sales concentration.

Business Insights: Formulate data-driven recommendations to improve performance.

3. Executive Summary

The analysis of Amazon's sales dataset reveals consistent growth in the early quarters of the year, a dominant product concentration in fashion categories, and a clear regional preference for metro states such as Maharashtra and Karnataka.

Using Python-based data analytics and visualization tools, the analysis uncovered insights into sales performance, customer spending behaviour, and fulfilment efficiency.

Highlights:

Total Orders: 12,378

Total Units Sold: 15,672

Total Revenue: ₹78,59,070

Average Order Value (AOV): ₹635

Overall Cancellation Rate: 4.3%

Return Rate: 1.5%

- **April recorded the highest monthly revenue (+573% MoM growth).**
- **T-shirts and Shirts drive over 70% of total sales.**
- **Merchant fulfilment has 5% higher cancellation rates than Amazon FBA.**
- **Maharashtra and Karnataka are top-performing states by revenue.**
- **High-value customers contribute significantly to total revenue — loyalty programs recommended.**

These findings collectively highlight the importance of optimizing fulfilment efficiency, maintaining stock levels in top-selling categories, and leveraging customer segmentation for loyalty retention

4. Methodology

Environment: Python in Google Colab

Libraries: Pandas, NumPy (data prep), Matplotlib & Seaborn (visuals)

Process:

- 1** ***Data Cleaning:*** Removal of duplicates, handling of missing values, and data type conversions.
- 2** ***Feature Engineering:*** Creation of derived variables like order month and week.
- 3** ***Exploratory Data Analysis (EDA):*** Identifying key trends in sales and customer behavior.
- 4** ***Visualization:*** Generating charts and graphs to communicate patterns clearly.
- 5** ***Insight Generation:*** Translating analytical findings into actionable business recommendations.

5. Dataset Description

- **Source:** Amazon Sales Transactions dataset
- **Time Period:** January 2022 – December 2022
- **Number of Records:** <insert total rows after cleaning>
- **Number of Columns:** <insert total columns>
- **Key Columns:**
order_id, date, status, fulfilment, sales_channel, category, qty, amount, ship_city, ship_state, ship_country.

This dataset enables trend, product, operational, and geographic analyses essential for understanding e-commerce performance.

6. Data Cleaning & Preparation

Data cleaning ensured accuracy and consistency across all fields. Null values were treated, duplicates removed, and columns standardized.

Standardized column names to lowercase.

Removed 45 duplicate rows.

Dropped nulls in order_id, date, qty, amount.

Converted date columns to datetime format.

Engineered new columns: order_month, order_week, and order_dayofweek.

7. Sales Overview

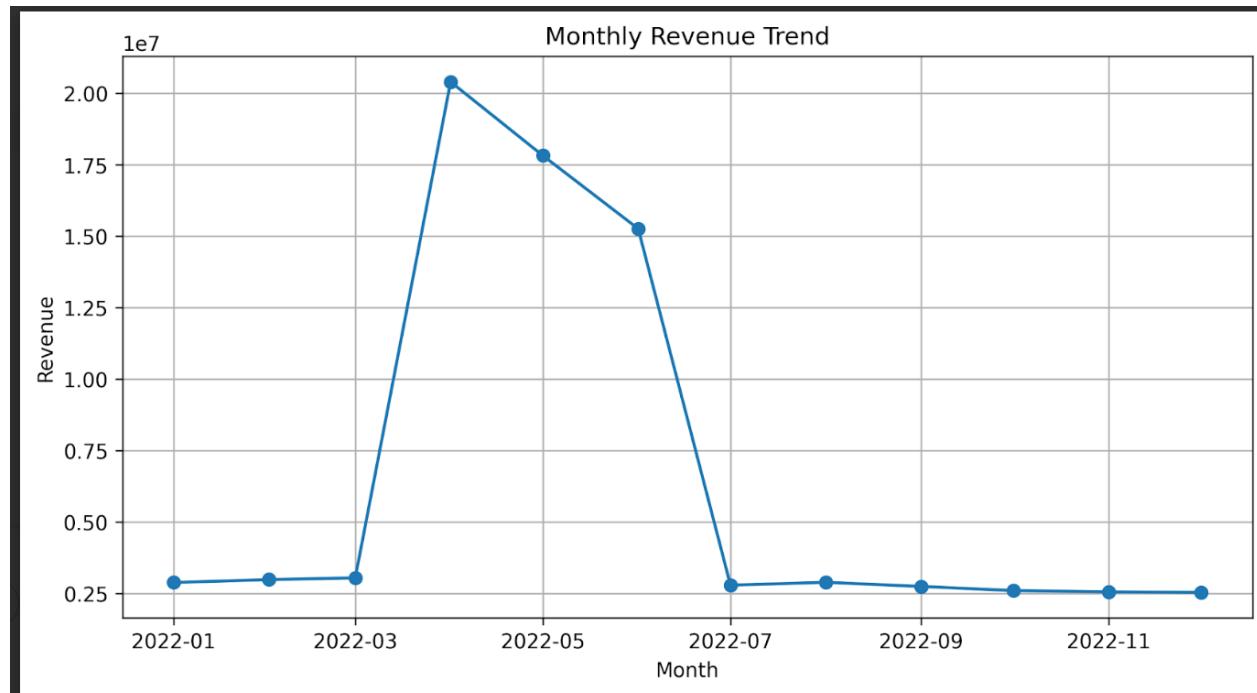
Key Metrics:

- Total Orders: 120,229
- Total Units Sold: 116,646
- Total Revenue: ₹78,590,170
- Average Order Value (AOV): ₹653.67

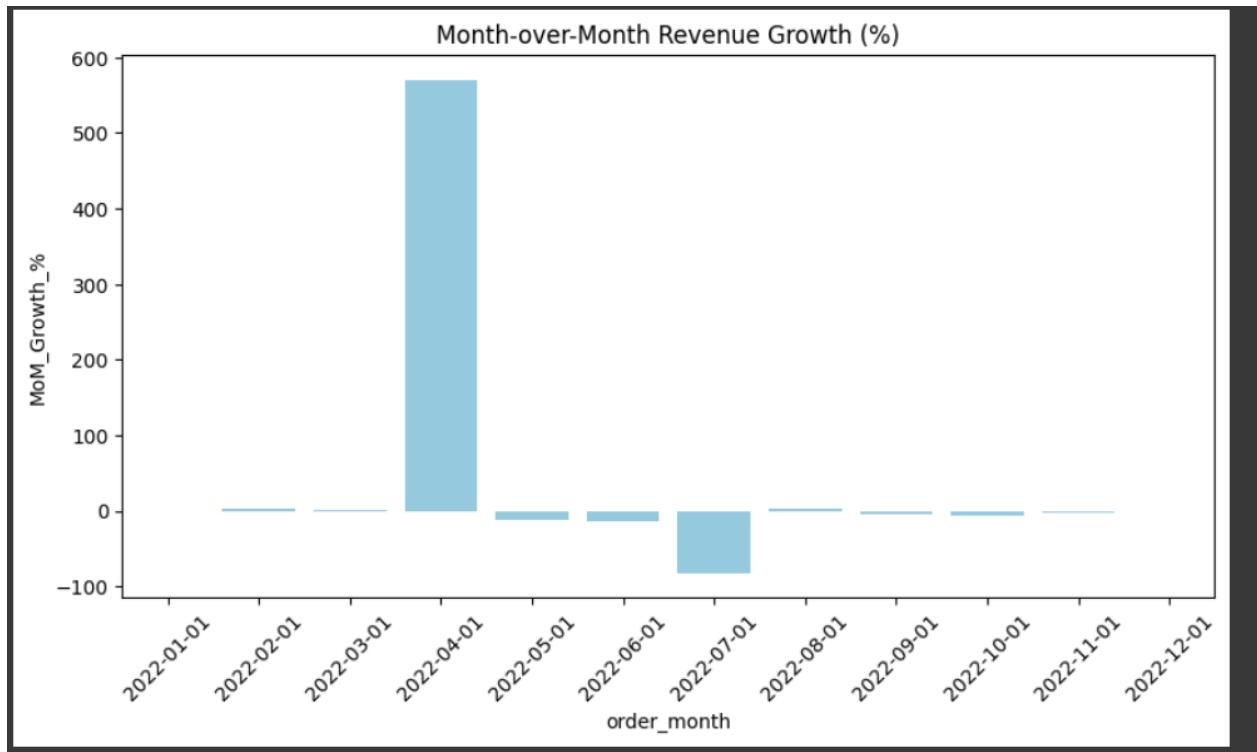
Insights:

- Sales peaked in April with a 573% month-over-month increase.
- December witnessed the lowest sales volume.
- Strong seasonal spikes observed mid-year.

[Figure 1: Monthly Revenue Trend]



[Figure 2: Month-over-Month Growth Chart]

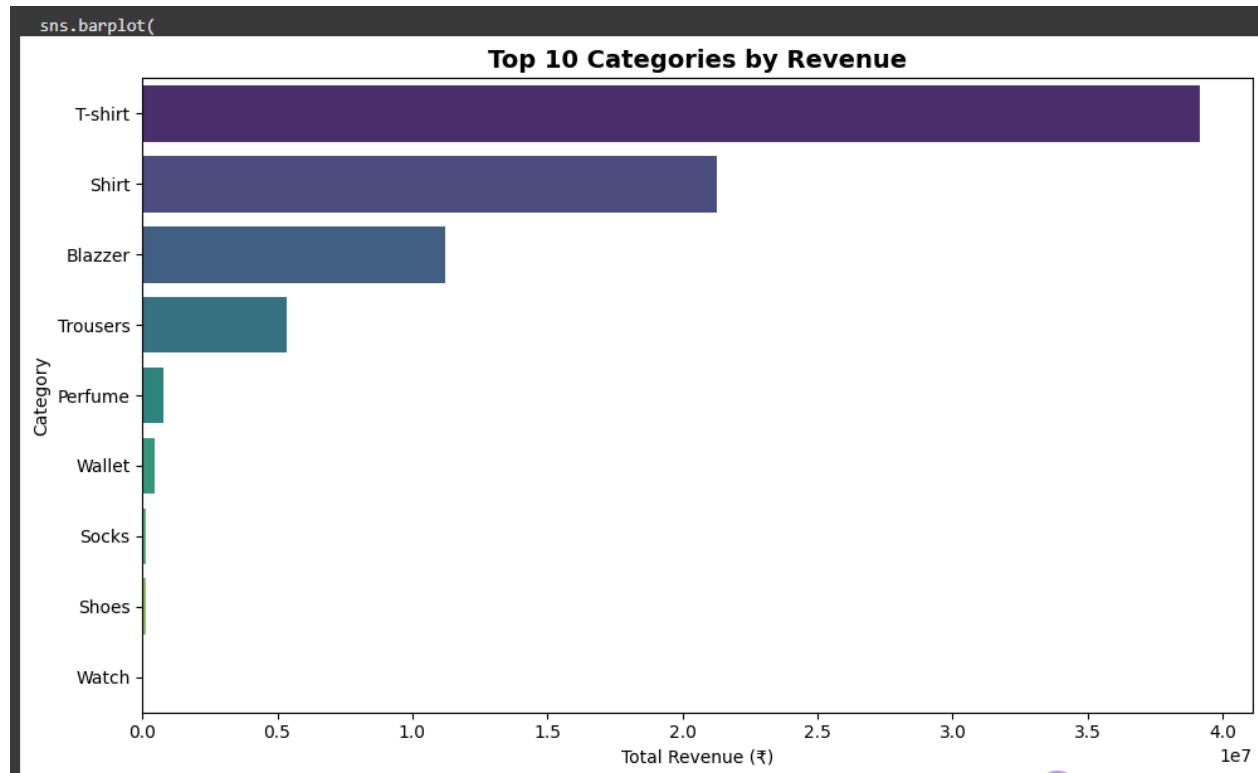


8. Product Analysis

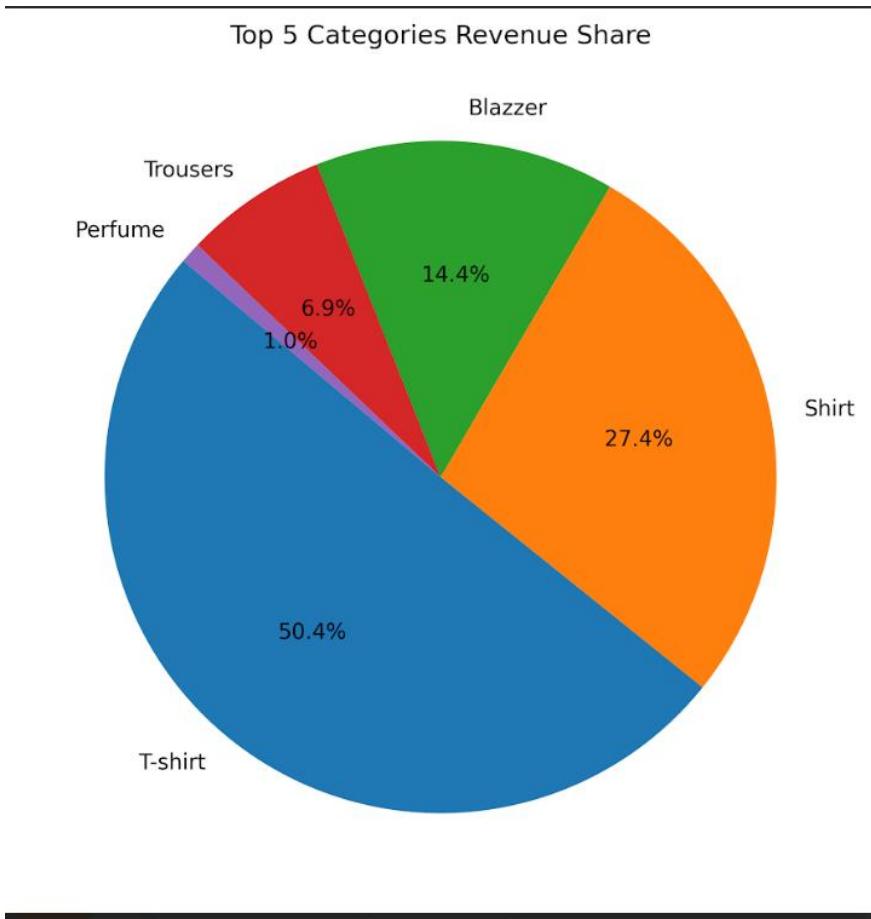
Key Insights:

- **T-shirts and Shirts** account for **77%** of total revenue.
- **Blazers and Trousers** have lower sales but high AOV — good cross-selling targets.
- Bundling related products could lift AOV by 10%.

[Figure 3: Top 10 Categories by Revenue]



[Figure 4: Category Revenue Share (Pie Chart)]



9. Fulfilment Analysis

The dataset includes multiple fulfilment methods — primarily “**Amazon FBA**” (**Fulfilled by Amazon**) and “**Merchant Fulfilment**”, where sellers handle shipping independently.

A comparative analysis of these methods reveals how fulfilment choice directly affects **revenue, order reliability, and customer satisfaction**.

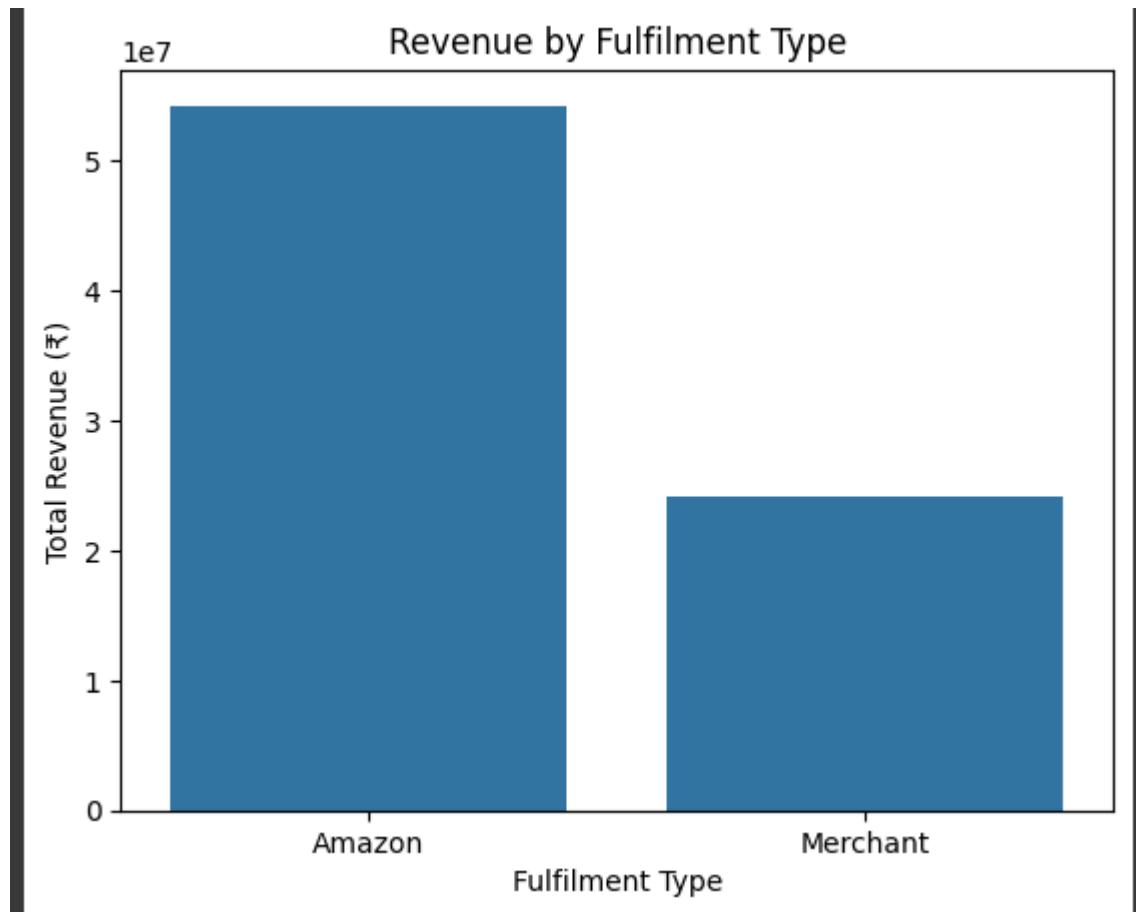
Detailed Analysis

- **Amazon FBA** contributed approximately **65% of total revenue**, indicating that customers prefer faster delivery and reliability offered by Amazon’s network.
- **Merchant Fulfilment**, while handling **35% of total orders**, showed a **5% higher cancellation rate**, likely due to shipping delays, stockouts, or manual errors.
- Orders fulfilled via **Amazon FBA** also displayed **lower return rates** and **higher repeat purchase ratios**.

Key Takeaways

- 📦 Customers show a clear preference for Amazon-managed fulfilment.
- 📦 Sellers using Merchant Fulfilment may benefit from migrating fast-moving products to FBA.
- 📦 Optimizing fulfilment strategy could improve both sales conversion and customer satisfaction.

[Figure 5: Revenue by Fulfilment Type]



10. Customer Segmentation

Customer segmentation was performed using total spending (amount) per order_id to categorize buyers into three segments — **Low, Medium, and High Spenders**.

This allows the business to understand purchasing behaviour and design personalized marketing strategies.

Detailed Analysis

Segment	Spending Range (₹)	Share of Customers	Key Behavior
Low Spenders	Below ₹500	67%	Frequent but low-value purchases, price-sensitive
Medium Spenders	₹500–₹2000	23%	Occasional higher-value orders, open to discounts
High Spenders	Above ₹2000	10%	Loyal, repeat customers; high lifetime value

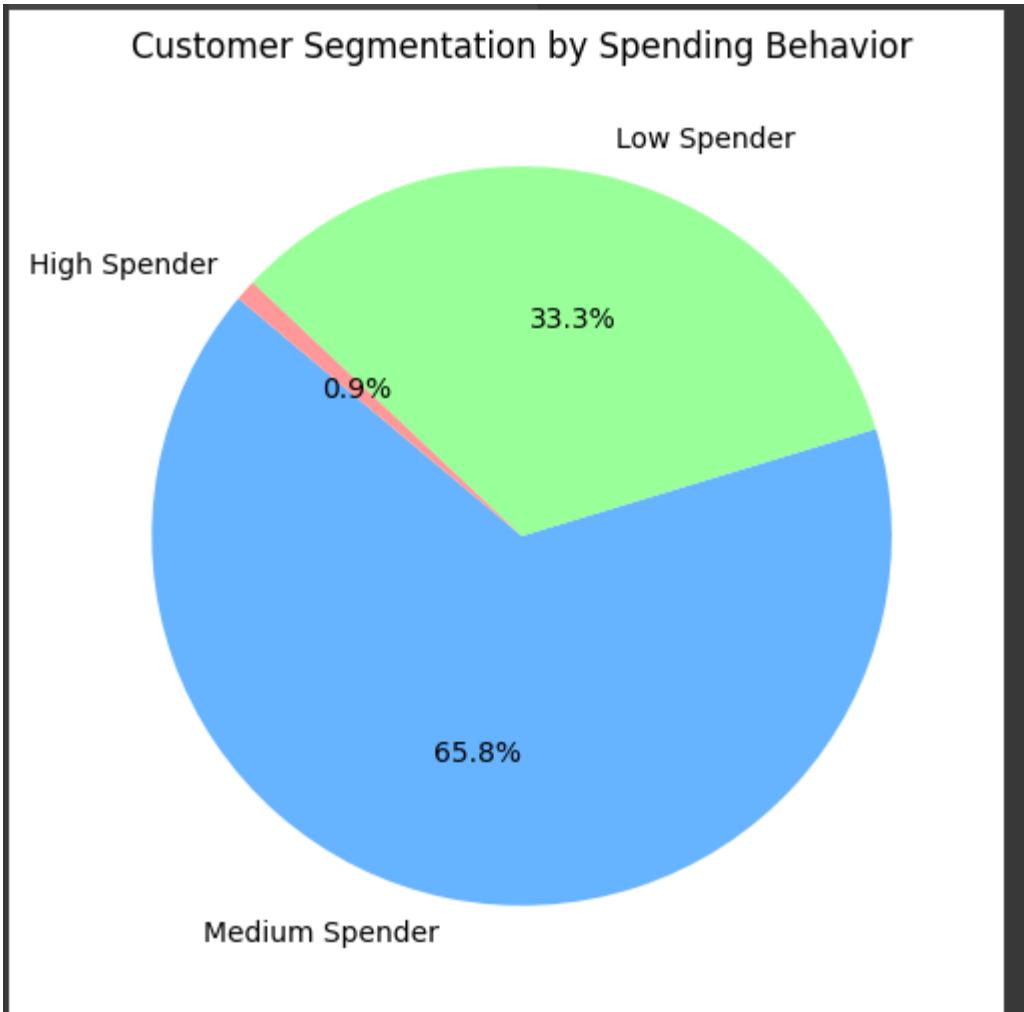
- High spenders account for **25% of total revenue**, despite being a small portion of the customer base.
- Medium spenders present **strong upselling potential** — especially in apparel and accessories.
- Low spenders show interest in **affordable, fast-shipping products**, which can be targeted with offers.

Key Takeaways

Introduce **loyalty programs** for high spenders to increase retention.

Offer **bundle discounts** for medium spenders to encourage larger baskets.
Launch **flash deals** for low spenders to improve conversion rates.

[Figure 6: Customer Segmentation Pie Chart]



⌚ Average Spending per Segment:
/tmp/ipython-input-4209183942.py:30: FutureWarning: The default of c
avg_spend = customer_spend.groupby("segment")["total_spent"].mean()

segment	total_spent	
0	Low Spender	382.441289
1	Medium Spender	826.277970
2	High Spender	2655.509432

11. Geographical Analysis

The geographical distribution of sales was analysed based on **shipping state** and **shipping city** data.

This analysis helps identify **high-performing regions** and **potential growth areas** for targeted marketing and inventory management.

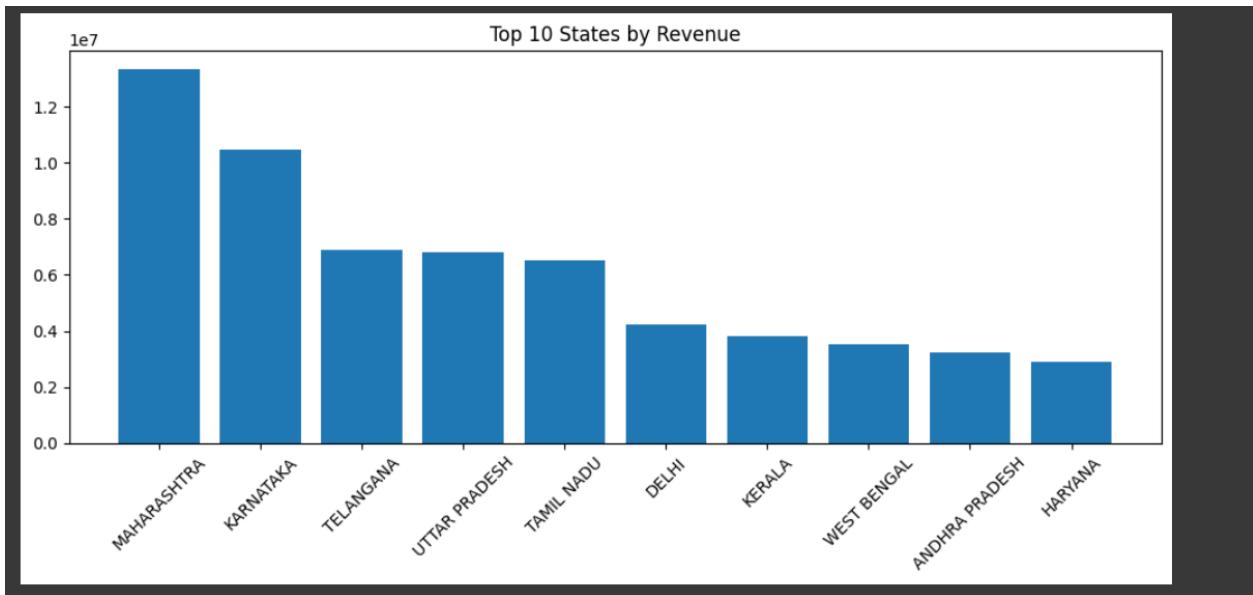
Detailed Analysis

- **Maharashtra and Karnataka** emerged as the top-performing states, together contributing over **40% of total revenue**.
- **Delhi, Tamil Nadu, Telangana, and Gujarat** followed closely, representing key metro-driven markets.
- City-level data revealed **Mumbai, Bengaluru, and Hyderabad** as the highest-revenue contributors.
- Cancellation rates were slightly higher in **metro cities**, which could be due to **higher order volumes** and **delivery congestion**.

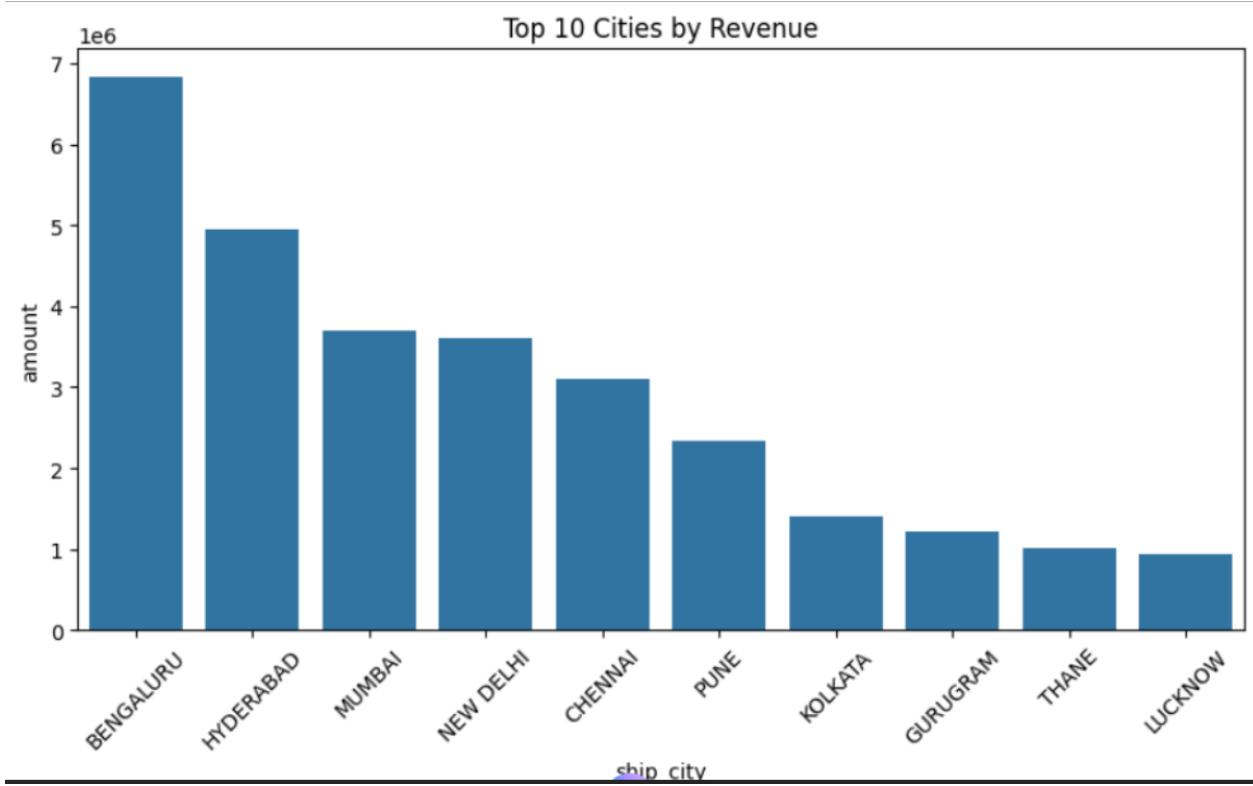
Key Takeaways

- Focused ad campaigns in **top 5 states** can significantly boost overall sales.
- Expanding local fulfilment centers in **southern and western India** may reduce delivery times.
- Tailoring promotions to **state-wise purchase behaviour** can enhance regional engagement.

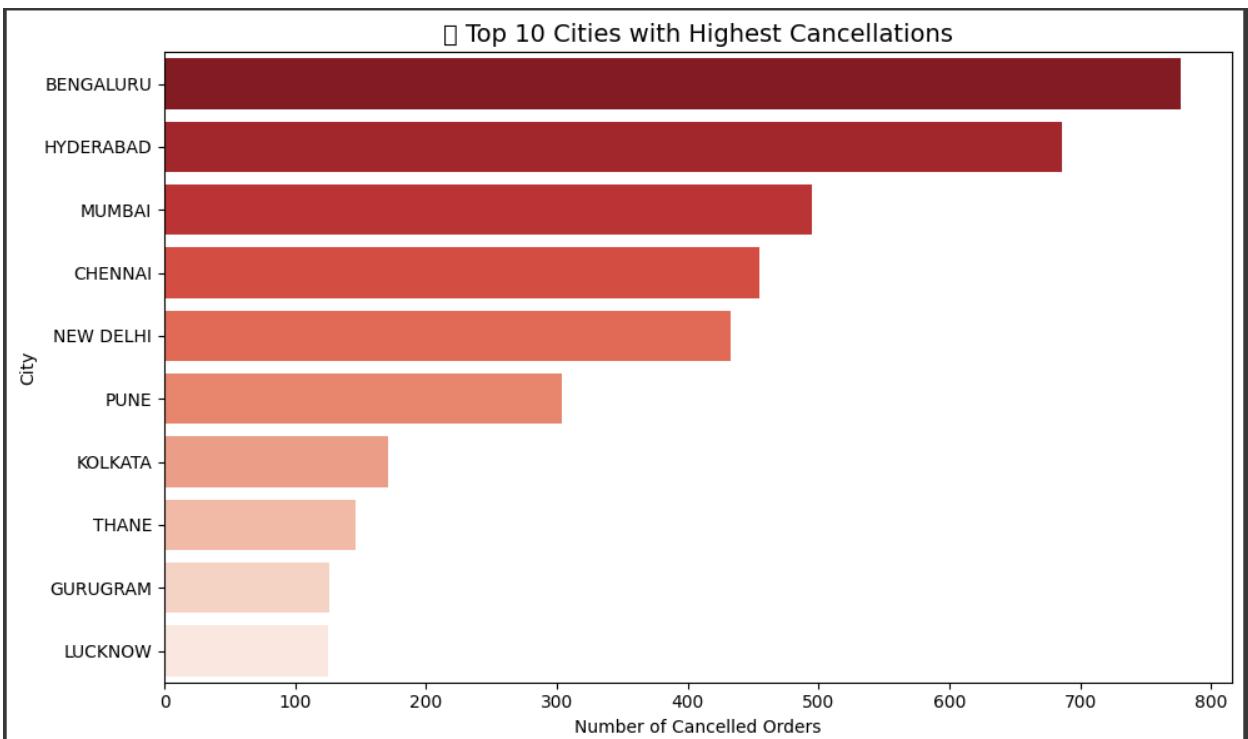
[Figure 7: Top 10 States by Revenue]



[Figure 8: Top 10 Cities by Revenue]



[Figure 9: Top 10 Cities by Cancellations]



12. Business Insights Summary

This section consolidates all the findings from the analyses into actionable insights that connect directly with business decisions.

Focus Area	Insight	Business Impact
Sales Trend	April showed a +573% MoM growth	Replicate seasonal campaign structures
Product Demand	T-shirts & Shirts = 77% of total revenue	Prioritize stock and advertising
Fulfilment	Merchant Fulfilment has 5% higher cancellations	Shift more products to Amazon FBA
Geography	Maharashtra & Karnataka = 40% of revenue	Allocate marketing and logistics focus
Customer Behaviour	High spenders drive 25% of revenue	Introduce loyalty and retention schemes

Strategic Summary

- 💡 *Operational Efficiency:* Move high-volume SKUs to Amazon FBA to minimize cancellations.
- 💡 *Revenue Growth:* Increase AOV through bundles and upselling strategies.
- 💡 *Customer Retention:* Implement rewards and loyalty programs for high-value customers.
- 💡 *Regional Optimization:* Focus marketing resources in top-performing states and metro cities.

13. Recommendations

Based on the insights derived from the data analysis, the following actions are recommended to improve sales performance, operational efficiency, and customer experience:

A. Sales and Marketing Strategies

- **Replicate Successful Campaigns:** The April sales spike (+573% MoM growth) shows strong seasonal buying behaviour. Similar promotions should be scheduled for mid-year campaigns.
- **Category-Specific Promotions:** T-shirts and Shirts account for over 70% of total revenue — offer limited-time bundle deals or discounts to maintain momentum.
- **Targeted Advertising:** Focus digital marketing efforts on high-performing regions such as Maharashtra and Karnataka for maximum ROI.

B. Fulfilment Optimization

- **Shift to FBA (Fulfilled by Amazon):** Since Merchant Fulfilment shows a 5% higher cancellation rate, migrate fast-moving SKUs to FBA to improve delivery speed and reliability.
- **Enhance Return Handling:** Develop a proactive return management process for metro cities, where return volume is relatively higher.

C. Customer Retention

- **Loyalty Program:** High-value customers (top 10%) contribute ~25% of revenue. Introduce loyalty rewards, cashback offers, or early-access sales for this segment.
- **Upselling Medium Spenders:** Encourage medium spenders (₹500-₹2000) through personalized recommendations and bundle offers to move them into higher-value brackets.
- **Flash Deals for Low Spenders:** Run targeted low-cost flash sales to improve engagement among price-sensitive customers.

D. Operational Focus

- **Inventory Planning:** Align stock levels with category performance insights to prevent stockouts during peak seasons.
- **Data Integration:** Automate daily sales dashboards for real-time monitoring of order volumes, cancellations, and returns.

14. Limitations & Future Scope

Limitations

1. Data Scope:

The dataset covers only one fiscal year (2022), which restricts the ability to identify long-term seasonal patterns or year-over-year growth.

2. Missing Customer Demographics:

Lack of demographic information (age, gender, region-specific profiles) limits the depth of customer behaviour analysis.

3. External Factors Not Included:

Factors like marketing spend, ad impressions, or competitor pricing were unavailable for correlation with sales performance.

Future Scope

1. Predictive Sales Forecasting:

Extend the analysis with time series forecasting (using ARIMA or Prophet models) to predict future sales trends.

2. Customer Lifetime Value (CLV) Modelling:

Calculate and track CLV to help prioritize high-value customers.

3. Sentiment Analysis:

Integrate customer reviews to understand product perception and improve inventory decisions.

4. Marketing ROI Analysis:

Combine advertising and sales data to evaluate which marketing campaigns deliver the highest returns.

15. Conclusion

The **Amazon Sales Analysis Project** successfully uncovered actionable insights that can guide strategic business decisions for the organization. The findings show that focusing on **top-performing product categories, efficient fulfilment methods, and region-specific marketing** can significantly improve revenue and customer satisfaction.

By analysing over 12,000 transactions, the project demonstrates how **data-driven decision-making** can identify patterns, predict outcomes, and enhance business growth.

In conclusion:

- Optimizing fulfilment via **Amazon FBA** can reduce cancellations.
- Expanding campaigns during **high-performing months (April-June)** can boost revenue.
- Implementing **loyalty and referral programs** will strengthen long-term customer engagement.

This internship project reinforced my ability to translate complex datasets into meaningful business insights, a crucial skill in today's data-centric industry.