Marketing & Customer Analytics in company Half Life

Marketing & Analytics Internship (2026) – Bconic

Sandini Fernando

Final-Year Undergraduate – Business Analytics

University of Moratuwa

Datasets Overview

1.CUSTOMERS DATASET

Records: ~5,000 customers

Key columns:
- Customer ID
- Age
- Email
- Registration date

Purpose: Identifies who the customers are

2.PRODUCTS DATASET

Records: ~200 products

Key columns:
- Product ID
- Product name
- Category
- Subcategory
- Brand
- Price
- Rating

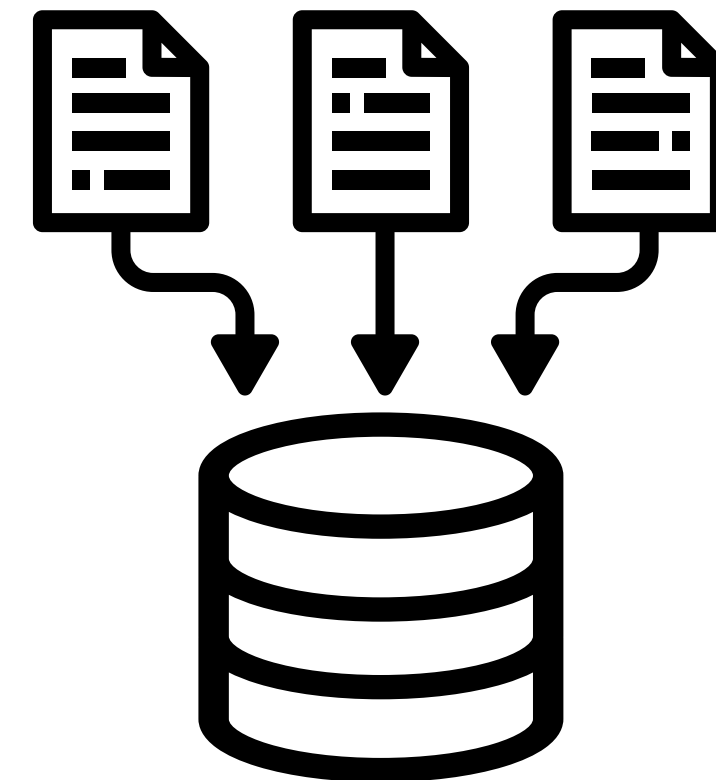Purpose:Provides product and category context for sales analysis

3.TRANSACTIONS DATASET

Records: 23,814 transactions

Key columns:
- Transaction ID
- Customer ID
- Product ID
- Transaction date
- Order status
- Quantity
- Total amount

Purpose: Captures customer purchase behavior

Data Pre-processing & Preparation

Data Exploration
- Reviewed each dataset structure and data types
- Checked for missing values and inconsistencies

MISSING VALUE HANDLING

- Filled missing customer age using the median
- Filled missing product ratings using the median
- Replaced missing customer emails with "Unknown"
- No records were removed to keep all customers in the analysis

Why median?
- Less affected by extreme values
- Only a small number of missing entries
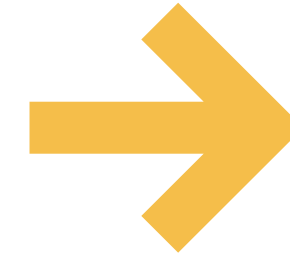
DATA PREPARATION STEPS

1. Converted all date columns to datetime format
2. Filtered only completed transactions to represent real sales
3. Merged datasets step by step:
   Transactions + Customers
   Then added Products

Customer Analytics Creation

	customer_id	total_spend	num_orders	last_purchase_date	days_since_last_purchase	avg_order_value
0	CUST000000	812.22	3	2024-08-16 07:41:45	138	270.740000
1	CUST000002	257.16	3	2024-10-05 10:55:13	88	85.720000
2	CUST000003	835.86	4	2024-08-24 20:42:57	130	208.965000
3	CUST000004	2465.33	12	2024-11-24 20:27:19	38	205.444167
4	CUST000005	414.42	1	2024-08-29 18:13:12	125	414.420000

Customer Segmentation Approach

HOW SEGMENT THRESHOLDS WERE DEFINED → **USED PERCENTILES TO CREATE CLEAR CUT-OFFS**

Threshold Logic

- Top 25% in total spend → Higher-value customers
- Top 25% in number of orders → More frequent buyers
- Top 25% in recency (days since last purchase) → Less recently active
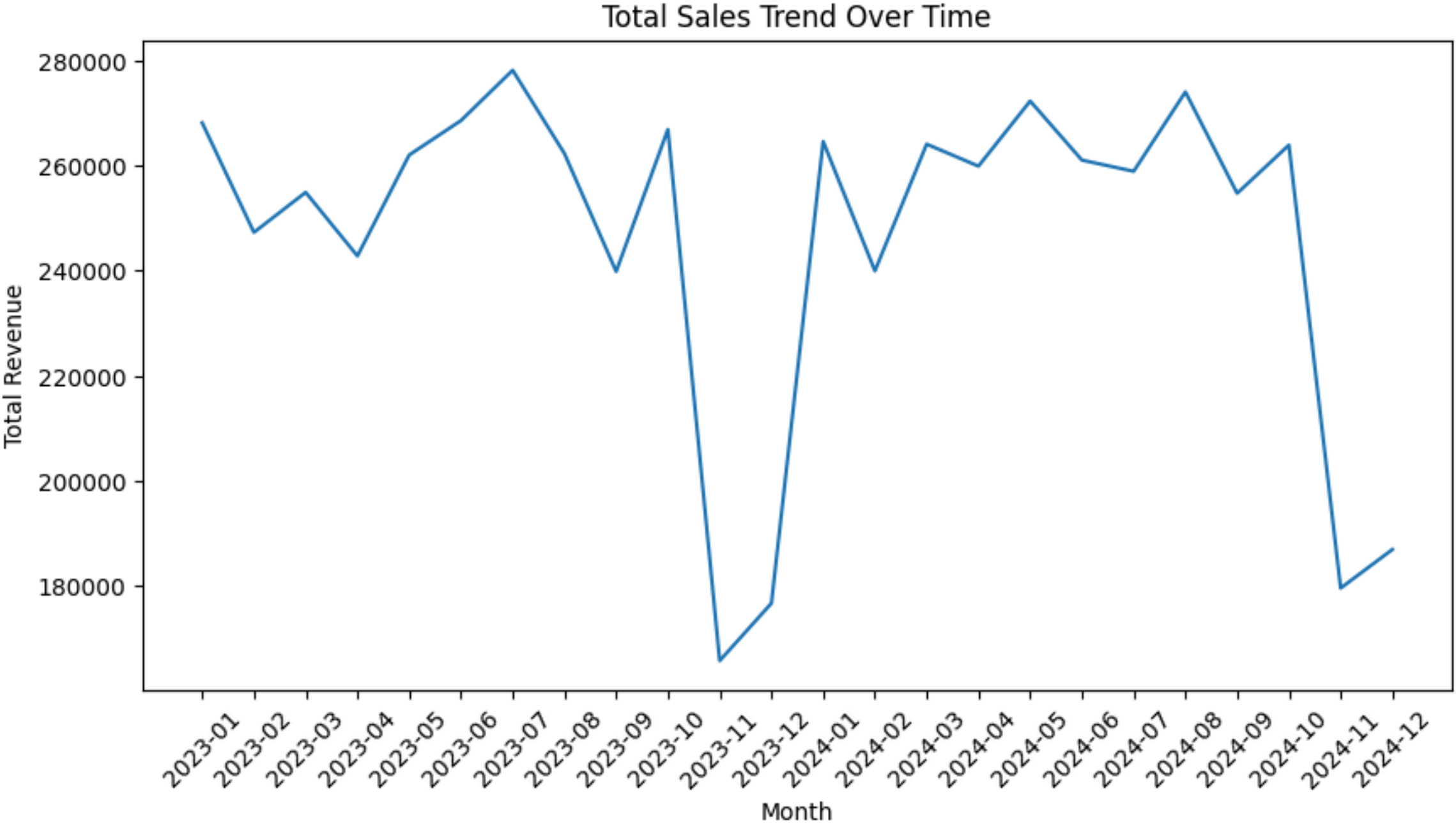- Bottom ~40% in spend → Lower-value customers

	customer_segment	number_of_customers	avg_spend	avg_orders
0	High-Value / Loyal	828	4369.326957	17.632850
1	Inactive	894	399.844474	1.640940
2	Occasional	783	281.420294	1.717752
3	Regular	1103	1031.470526	4.116047

Basic Sales Analysis

Sales Trend Over Time

What the chart shows

- Monthly total revenue from January 2023 to December 2024

- Overall revenue remains relatively stable
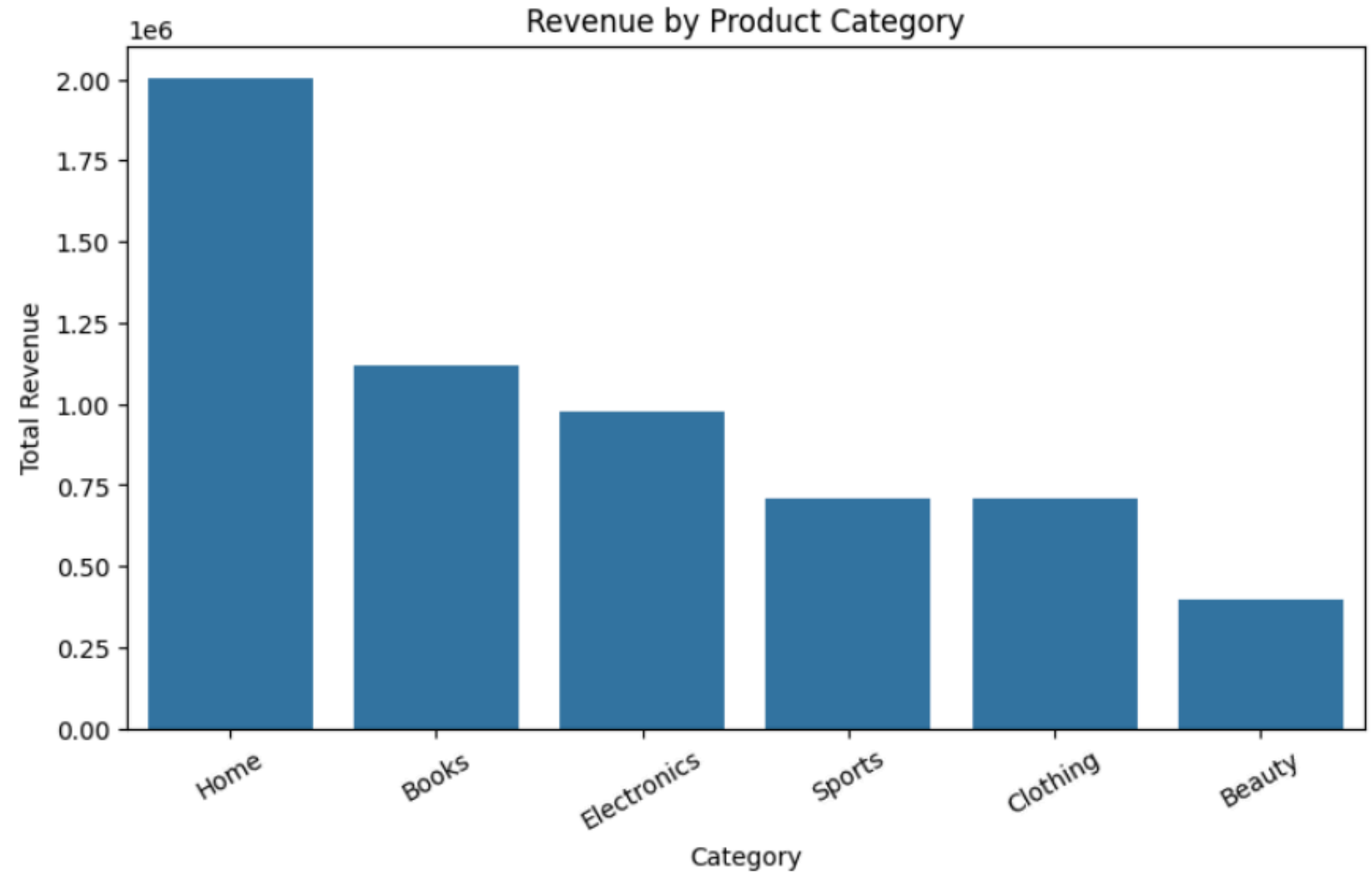
- Some clear peaks and dips across months

Basic Sales Analysis

Revenue by Product Category

What this chart shows

- Total revenue generated by each product category

- Comparison across categories:

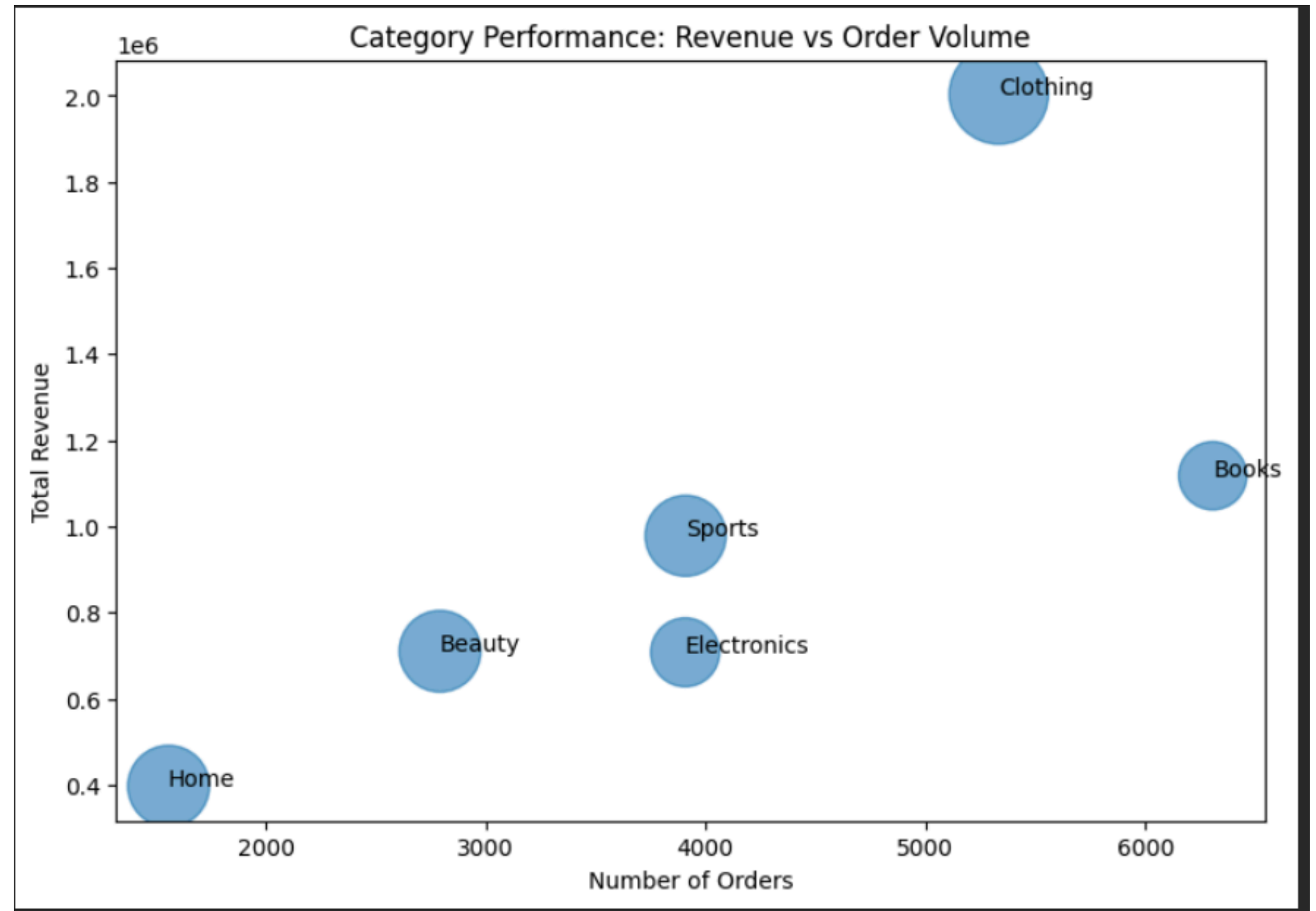 Home, Books, Electronics, Sports, Clothing, Beauty

Basic Sales Analysis

Revenue vs Order Volume by Category

What this chart shows

- X-axis: Number of orders
- Y-axis: Total revenue
- Bubble size: Average order value
- Each bubble represents a product category

Key Findings & Recommendations

KEY FINDINGS

- Sales are generally stable over time, with noticeable drops around October–November each year.
- Revenue is concentrated in a few categories, mainly Home, Books, and Electronics.
- Order volume does not equal revenue:
- Books has the most orders but lower value per order.
- Clothing generates high revenue with fewer orders due to higher average order value.
- Customer segmentation shows clear differences between high-value, regular, occasional, and inactive customers.

RECOMMENDATIONS

Prioritize high-revenue categories

  Focus marketing and premium positioning on Clothing, which shows strong revenue with fewer orders.

  Protect and maintain performance in Home, as it remains a top revenue contributor.

Increase order value in high-volume categories

  Use bundles, cross-selling, or targeted offers in Books and Beauty to improve average order value