

# E-commerce Sales & User Analysis

Yoshihiro Tsunoda

December 11, 2025

---

## Technologies Used

- SQL / BigQuery
- Tableau (Data Visualization)
- GitHub (Repository Management)

---

## Executive Summary

This report analyzes e-commerce sales data to understand sales trends, category structure, customer behavior, purchase patterns, channel efficiency, and revenue concentration.

The aim is to **identify key growth drivers, understand customer value, uncover structural issues, and extract actionable insights for revenue improvement.**

Key findings include:

- **Sales are heavily concentrated during the holiday season**
- **Apparel accounts for 47% of total revenue**, showing significant category dependence
- **Many high-value customers make only a single purchase**, highlighting retention issues
- **A strong bias toward new customers (92%) results in a low repeat rate**, well below industry benchmarks
- Conversely, **revenue is broadly distributed**, with no excessive dependence on a small group of top customers

These insights indicate that improving **repeat purchase rates, optimizing category strategy, developing loyal customers, and strengthening high-efficiency channels** would be the most impactful initiatives.

---

## 1. Introduction

Data-driven decision-making is essential for sustainable growth in e-commerce.

This analysis evaluates key business metrics—sales structure, category performance, customer

distribution, and channel efficiency—to provide a comprehensive understanding of the current business status.

The objectives of this report are:

- Identify the primary drivers of revenue
  - Understand high-value customer characteristics
  - Reveal structural issues contributing to low repeat rates
  - Extract actionable recommendations for CRM, product strategy, and channel optimization
- 

## 2. Data and Methodology

### 2.1 Data

The analysis uses e-commerce transaction logs and behavioral data collected from GA4.

Main fields used:

- Customer information (Customer ID, Region)
- Product information (Category, Product Name, Unit Price)
- Order information (Quantity, Total Price, Order Date)
- Behavioral metrics (active\_days = number of days with purchase activity)
- Traffic channels (organic, direct, referral, other)

**Analysis period: October 2020 – January 2021**

### 2.2 Methods

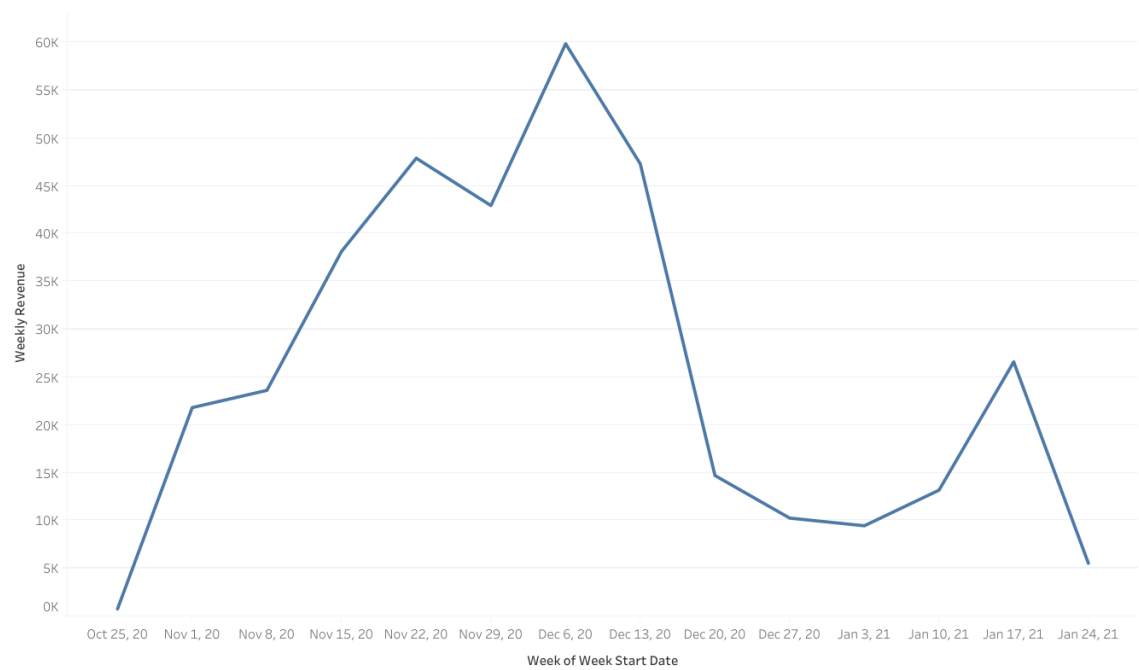
- Aggregation: SQL
- Visualization: Tableau
- Key analytical techniques:
  - AOV (Average Order Value)
  - Purchase frequency analysis
  - AOV × Frequency mapping (Customer Value Map)
  - New vs Repeat classification (via active\_days)
- Channel analysis: CVR calculation
- Revenue concentration: Pareto cumulative sales analysis

### 3. Results

---

#### 3.1 Sales Trend

- Sales increased sharply from late October to early December
- The peak occurred during the week of **December 6 (~\$60,000)**
- Sales dropped significantly after Christmas and hit the lowest levels in early January
- A typical “**Holiday Peak → Year-end decline**” pattern



**Figure 1. Weekly Sales Trend**

## 3.2 Category Revenue Structure

Key insights:

- **Apparel accounts for 47.4% of total revenue**, nearly half of the site's sales
- The top 5 categories generate **72%** of total revenue
- The site exhibits a **strong reliance on Apparel** with a long-tail distribution for the other categories

Implications:

- Category portfolio optimization is needed
- Low-volume categories hold potential for improvement

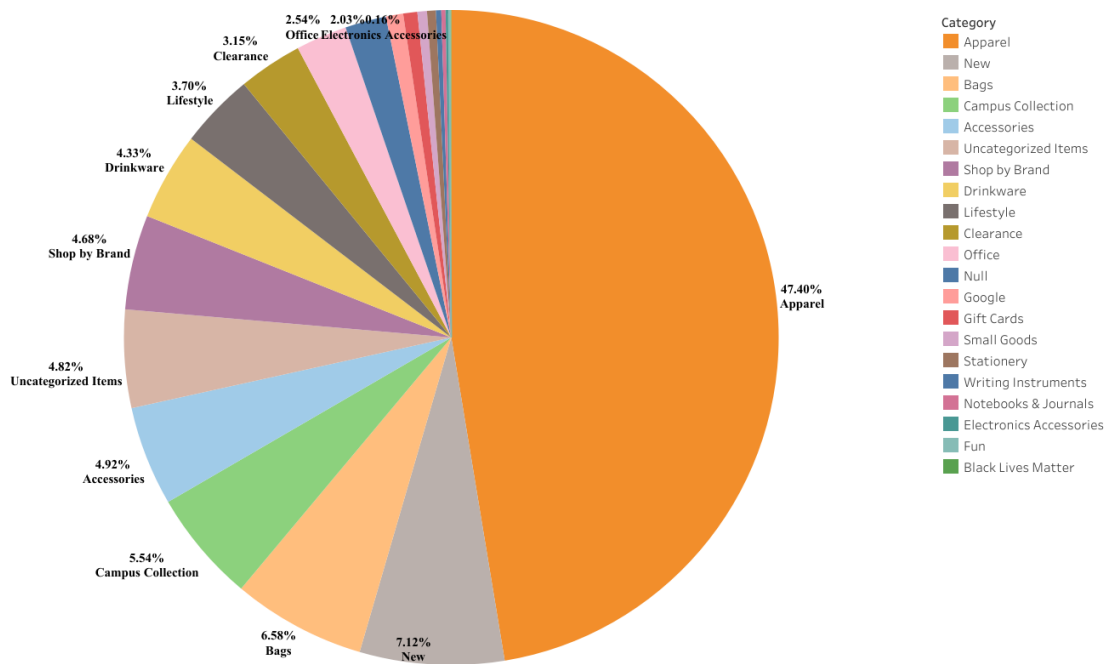
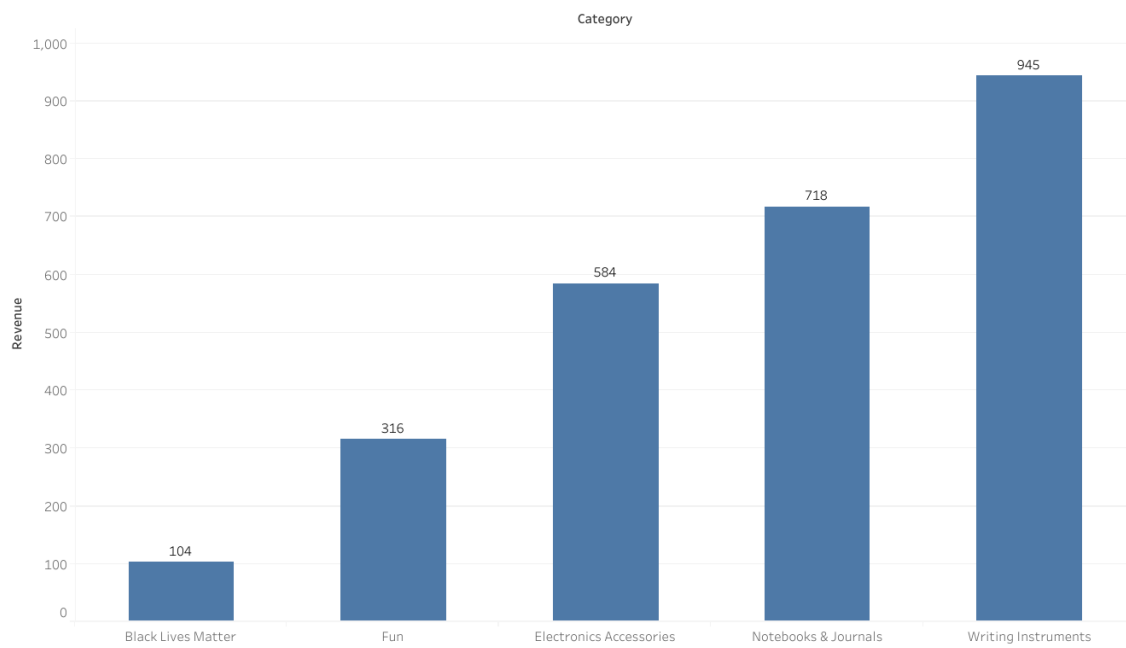


Figure 2. Category Revenue Share



**Figure 3. Lowest Revenue Categories**

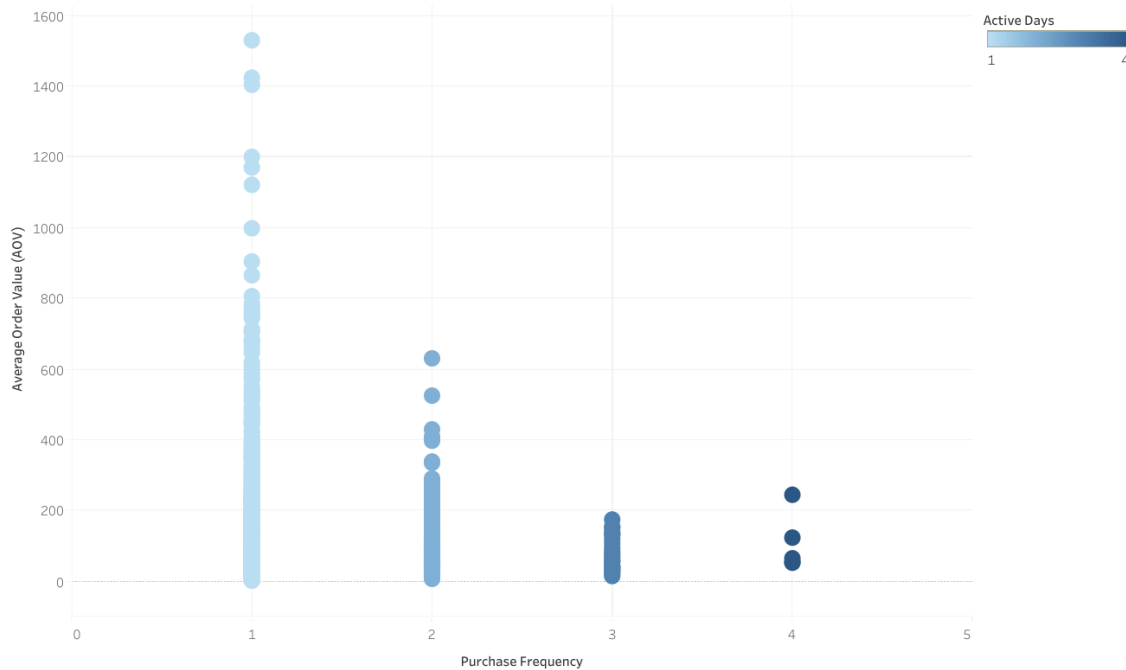
### 3.3 Customer Value Map (AOV × Frequency)

Key findings:

1. **Single-purchase customers dominate the user base**
2. **Higher purchase frequency correlates with lower order value**
3. **High-value but low-frequency customers represent the largest growth opportunity**
4. **Customers with more active days are strong candidates for loyalty programs**

Implications:

- Personalized offers and coupons may convert high-value one-time buyers
- Subscription or membership programs suit repeat-purchase segments



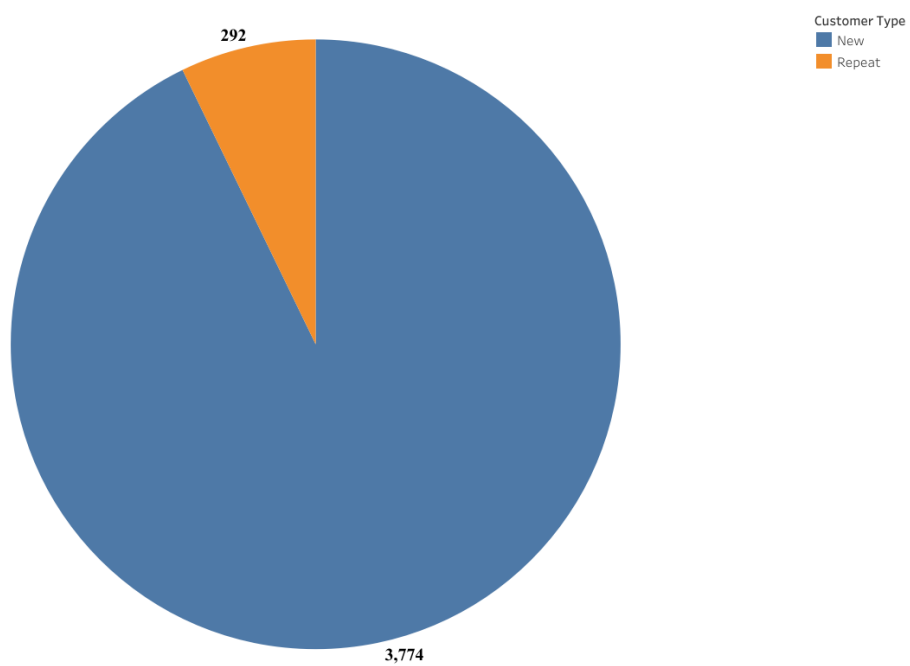
**Figure 4. Customer Value Map (AOV × Frequency)**

### 3.4 New vs Repeat Customers

- **New customers: 92.8% (3,774 users)**
- **Repeat customers: 7.2% (292 users)**  
→ Significantly below the industry average (20–40%)
- Repeat customers show **2–5× higher average value**

Implication:

**Increasing second-purchase conversion offers the highest ROI**



**Figure 5. New vs Repeat Customer Ratio**

### 3.5 Top 50 Revenue-Generating Users

Findings:

- A smooth long-tail revenue curve
- Top 3 customers exceed **\$1,400** each
- A strong middle tier (between \$800–\$1,200)
- Many high-value customers have **active\_days = 1**

Implication:

- Revenue depends more on **high-value one-time buyers** than on frequent shoppers
- Loyalty development presents a substantial opportunity

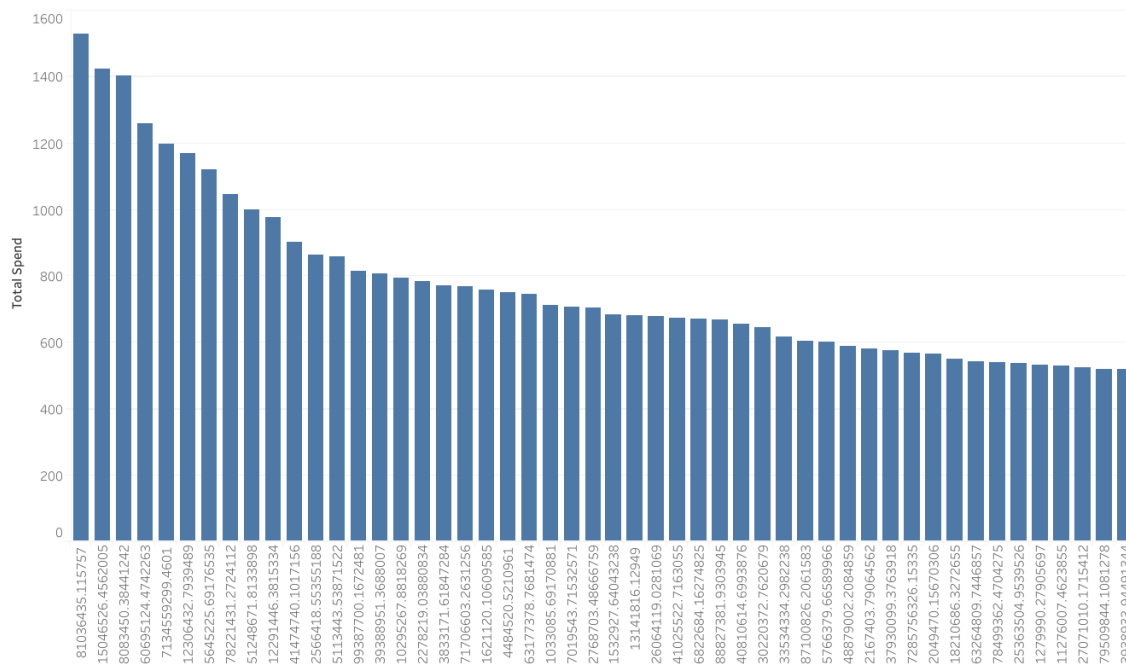


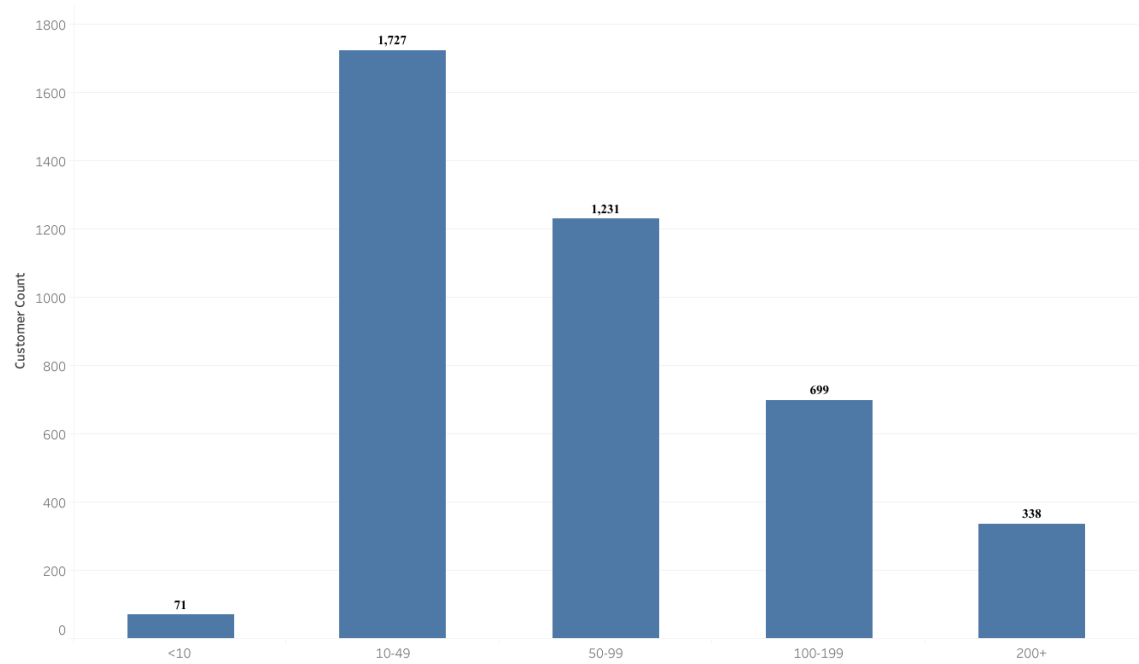
Figure 6. Revenue Top 50 Users



### 3.6 Sales Distribution by Spend Range

- **\$10–49** is the largest group (42%)
- **\$50–99** follows at 30%
- **\$100+** represents only 25% of customers but contributes disproportionately to revenue

Implication:  
Upsell strategies targeting mid-tier users are effective.



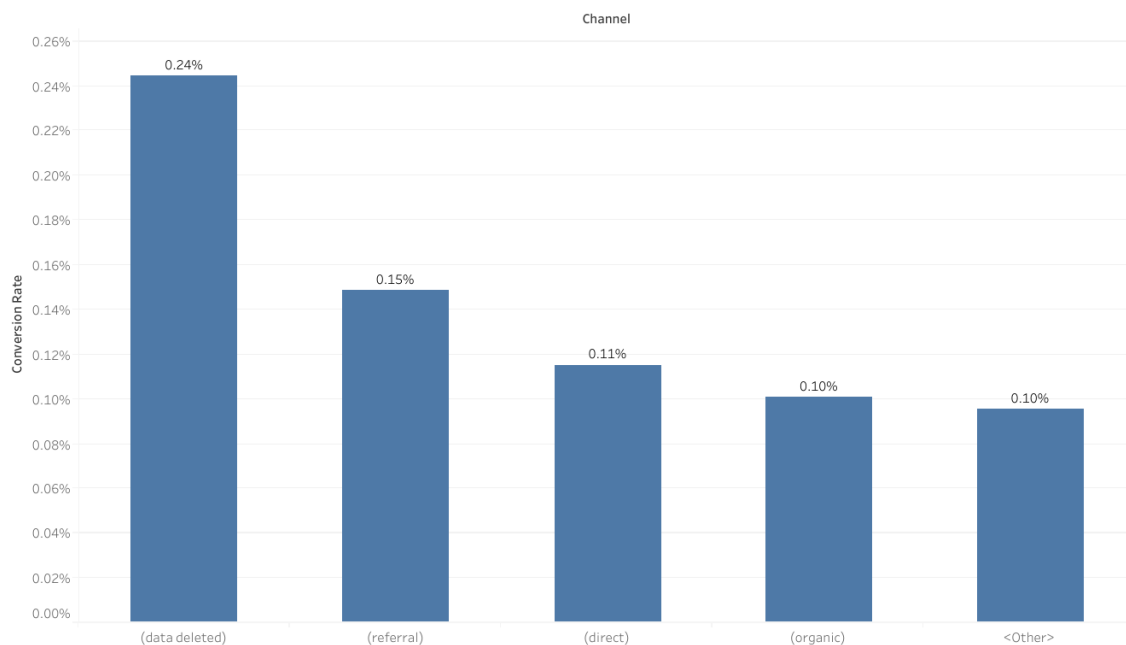
**Figure 7. Sales Distribution by Spend Range**

### 3.7 Traffic Channel Conversion Rates

- Highest CVR: **a specific high-efficiency channel (0.24%)**
- **Referral** performs strongly in both traffic quality and CVR
- **Organic** generates large traffic but low conversion
- **Other** shows the lowest efficiency

Implication:

Investment should prioritize high-CVR channels, while low-efficiency channels require optimization or reduction.



**Figure 8. Traffic Channel Conversion Rate Comparison**

---

### 3.8 Pareto Analysis: Revenue Concentration

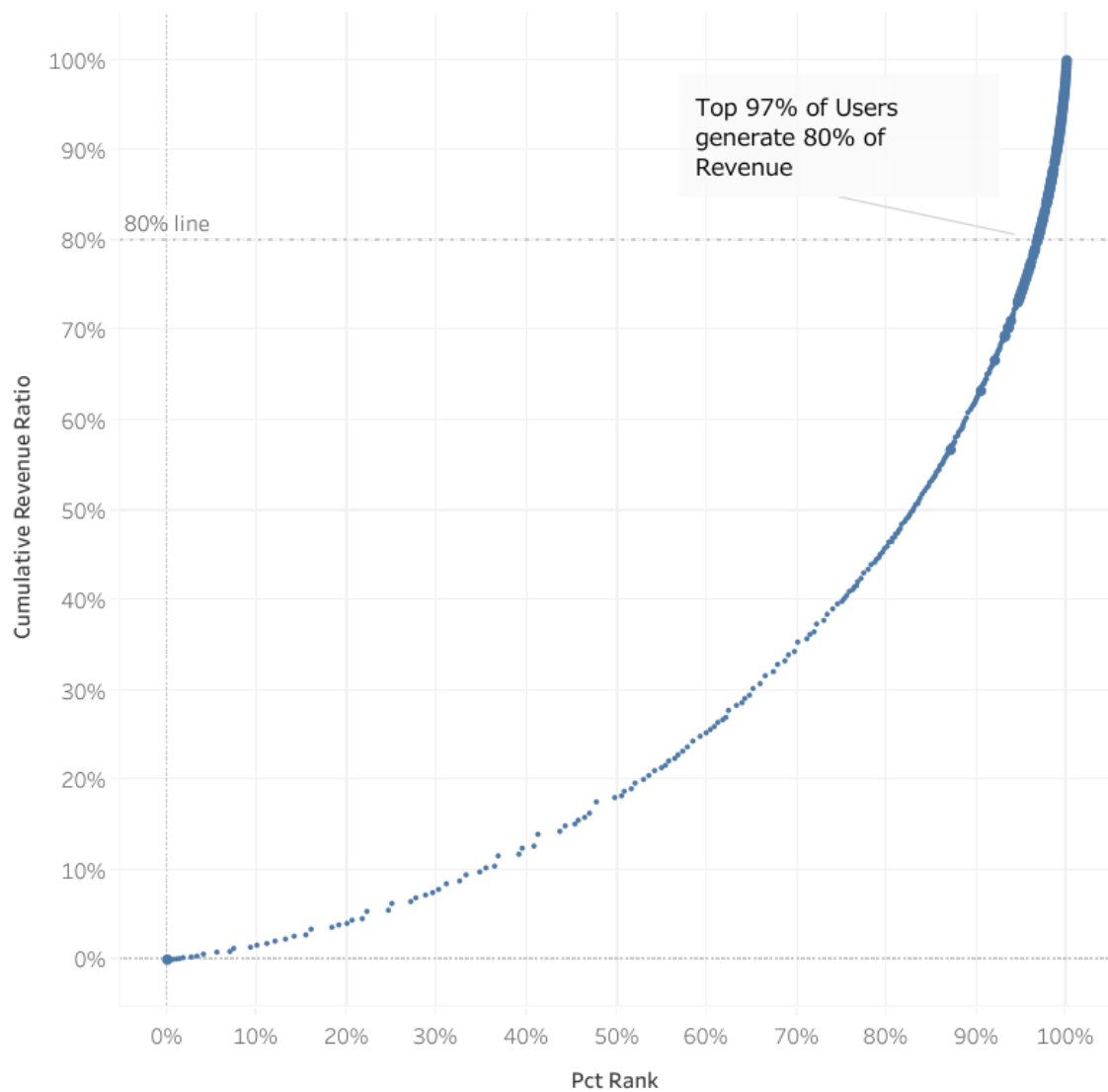
This analysis ranked customers by revenue contribution and calculated cumulative revenue share.

Results:

- **80% of revenue is generated by the top 97% of customers**
- This differs from the typical **80/20 Pareto pattern**, where a small group dominates revenue

#### Implications:

- The business does **not** depend on a small set of top customers
- Revenue is **widely distributed**, indicating a stable long-tail structure
- The most effective investments are:
  - Increasing **overall purchase frequency**
  - Improving **first-to-second purchase conversion**



**Figure 9. Pareto Analysis — Revenue Concentration by User**

## 4. Discussion

From the analysis, several structural issues and growth opportunities emerge.

---

### 1) Extremely low repeat rate due to high dependence on new customers

- LTV is low, reducing business stability
  - **Priority: Strengthen the second-purchase pathway (highest ROI)**
- 

### 2) The largest growth opportunity lies in high-value, low-frequency customers

- Many high-value customers churn after a single purchase
  - **Loyalty development could significantly increase revenue**
- 

### 3) Heavy reliance on Apparel and strong seasonality

- Nearly half of revenue is driven by Apparel
  - Sales are highly seasonal
  - **Category diversification is needed**
- 

### 4) Revenue is broadly distributed, not concentrated

- Dependence on top customers is low
- A wide user base supports revenue stability
- **Focus should be on improving overall purchase frequency and developing mid-tier customers**

## 5. Conclusion

To drive sustainable e-commerce growth, the following initiatives are most effective:

1. **Increase repeat rates among first-time buyers**
2. **Develop high-value, low-frequency customers into loyal users**
3. **Reduce dependence on Apparel and cultivate additional categories**
4. **Upsell to mid-tier customers & personalize offers for valuable segments**
5. **Scale high-CVR channels and optimize low-efficiency ones**

Implementing these strategies will improve revenue stability, raise LTV, and mitigate seasonal fluctuations.

---

## 6. References / Data Sources

- Google Analytics 4 (GA4)
  - Traffic channel data
  - Event logs
  - Purchase behavior data