Ecommerce Insights 2019

Turing College
Suja Manandhar
2025/05/26

Table of Content

Executive Summary

This report presents a comprehensive data analysis for the U.S.-based e-commerce company specializing in home and office appliances. The primary objective is to enhance customer experience and satisfaction to drive revenue growth through cohort analysis, Customer Lifetime Value (CLV) analysis, and RFM segmentation.

Key Performance Overview

In 2019, the business recorded 1,468 unique customers generating $4.68 million in revenue, with an Average Order Value of $186.60 and Customer Lifetime Value of $5,596. The company achieved a strong marketing ROI of 169.83% on $2 million invested.

Primary Findings

Customer Segmentation and Revenue Distribution

Female customers represent 63.62% of the customer base and contribute 62.29% of total revenue. However, statistical analysis reveals male customers generate significantly higher average revenue ($3,303 vs $3,120, $p<0.001$). RFM analysis shows that 'Top Customers' and 'Loyal Customers' drive 32% and 27% of revenue respectively, despite representing smaller customer groups.

Critical Retention Challenge

Cohort analysis reveals severe early-stage churn across all customer segments. Retention drops from 100% at acquisition to 6-15% by Month 1, falling below 20% by Month 3. Q1 cohorts (January-March) demonstrate superior long-term retention, with January maintaining the highest 11-month retention at 16%.

Behavioral Loyalty vs. Monetary Value Disconnect

A key finding shows that behaviorally loyal customers don't always yield the highest lifetime value. The Medium CLV segment includes 133 "Loyal Customers" contributing less revenue than smaller High CLV groups. Similarly, the Low CLV segment contains many "Loyal Customers" with significantly lower revenue contributions.

Seasonal Patterns and Product Performance

Revenue peaks occur in January ($460K), April ($442K), and November ($475K). The introduction of Nest products in August created significant impact, quickly becoming the third-largest revenue contributor despite fewer transactions, indicating strong demand for premium home automation products.

Introduction

In a highly competitive e-commerce landscape, understanding customer behavior is crucial to driving sustainable growth. This project was conducted for a U.S.-based online retailer specializing in home and office appliances, with the goal of uncovering actionable insights that improve customer experience and maximize revenue.

Using a combination of data analytics techniques—including cohort analysis, retention tracking, customer lifetime value (CLV) segmentation, and RFM analysis—the project aims to identify high-value customers, evaluate long-term engagement trends, and detect potential gaps in customer value realization.

The findings serve as a foundation for data-driven marketing, retention strategies, and personalized customer targeting.

Project Approach

This project follows the steps of data analysis process below:

1. Ask

This project follows a structured data analysis process, beginning with the Ask phase. Since the company's primary objective is to maximize revenue by improving customer experience and satisfaction, we begin by identifying the key questions that directly support this goal. Each question is designed to uncover insights that guide data-driven business decisions:

a. Who are the company's most valuable customers, and how can we segment them?

Understanding customer value and behavior enables the business to focus its marketing and retention efforts on the segments that drive the most revenue and long-term loyalty.

b. How do specific customer cohorts behave over time in terms of retention?

Analyzing cohort retention reveals patterns in customer engagement and churn, helping to identify when and why customers drop off—critical for designing effective retention strategies.

c. Do the most behaviorally loyal customers also yield the highest lifetime value?

This question tests the alignment between customer engagement and profitability. Addressing any mismatch helps prioritize customers who not only interact frequently but also contribute significantly to revenue.

2. Prepare

For this project we have used the Market Insight for E-commerce dataset from Kaggle. The data was downloaded as CSVs and then exported to BigQuery.

The dataset used includes:

- Customer Data

  This includes demographic data related to the customers which includes:

  - CustomerID: Customer Unique ID
  - Gender: Gender of customer
  - Location: Location of Customer
  - Tenure_Months: Tenure in Months
  - Discount and Coupons Data
  - Discount coupons have been used for different categories in different months.
  - Month: Discount coupon applied in that month
  - Product_Category: Product category
  - Coupon_Code: Coupon Code for given Category and given month
  - Discount_pct: Discount Percentage for given coupon
- Marketing Spend Data

Marketing spend on both offline & online channels daywise.

    - Date: Date
    - Offline_Spend: Marketing spend on offline channels like TV, Radio, NewsPapers, etc while it doesn't specify the particular medium.
    - Online_Spend: Marketing spend on online channels like Google keywords, facebook etc. while it doesn't specify the particular medium.

- Online Sales Data

  This file contains actual orders data (point of Sales data) at transaction level with below variables.

    - CustomerID: Customer unique ID
    - Transaction_ID: Transaction Unique ID
    - Transaction_Date: Date of Transaction
    - Product_SKU: SKU ID – Unique Id for product
    - Product_Description: Product Description
    - Product_Cateogry: Product Category
    - Quantity: Number of items ordered
    - Avg_Price: Price per one quantity
    - Delivery_Charges: Charges for delivery
    - Coupon_Status: Any discount coupon applied

- Tax Data

  This file contains GST Details for given category

    - Product_Category: Product Category
    - GST: Percentage of GST

3. Data Preprocessing

Before proceeding with the analysis, I performed an initial data audit to check for duplicates and null values, ensuring data integrity. I also selected only the relevant columns from each table to streamline the dataset for efficient analysis.

Two key preprocessing steps were performed:

1. Invoice Value Calculation

To accurately analyze revenue, I calculated the Invoice Value based on whether a discount code was applied:

a. With Discount Code:

Invoice Value = (Quantity × Avg_Price) × (1 - IFNULL(Discount_pct,0)/100) × (1 + Tax.GST) + Delivery_Charges

b. Without Discount Code:

Invoice Value = (Quantity × Avg_Price) × (1 + Tax.GST) + Delivery_Charges

2. Average Customer Lifetime Value (ACLV)

CLV was calculated using Spotify's approach, defined as:

CLV = AOV × Purchase Frequency × Average Lifetime

Where:

AOV (Average Order Value) = Total Revenue ÷ Number of Customers

Purchase Frequency = Total Orders ÷ Total Customers

Average Lifetime = Average number of active months per user

Based on the calculated values, customers were segmented into:

a. High CLV: Top 25%

b. Medium CLV: Middle 50%

c. Low CLV: Bottom 25%

4. Analyse

- ### Sales, Revenue and Marketing Insights Overview

Over the course of 2019, the business recorded a total of **1,468 unique customers**, generating approximately **$4.68 million in revenue**. This results in an **Average Order Value (AOV)** of **$186.6**, and a calculated **Customer Lifetime Value (CLV)** of **$5,596**.

The company invested **$2 million** in marketing for the year, which yielded an impressive **Return on Investment (ROI) of 169.83%**. This indicates that the marketing efforts effectively contributed to revenue growth, though it's important to note that this ROI calculation is based solely on marketing cost and does not factor in other operational or product-related expenses.

Figure: KPI Overview

- **Monthly Revenue Patterns and Seasonality**

  A clear seasonal trend is visible in the revenue distribution across the year. The months of **January ($460K)** and **November ($475K)** stand out as **peak revenue months**, while **June ($287K)** records the **lowest revenue**.

  From **June to August**, there is a gradual upward trend in both revenue and transaction volume, culminating in **August**, which records the **highest number of transactions (6,150)** for the year. Similarly, from September to December, there is an upward trend in revenue, culminating in November, which records highest revenue.

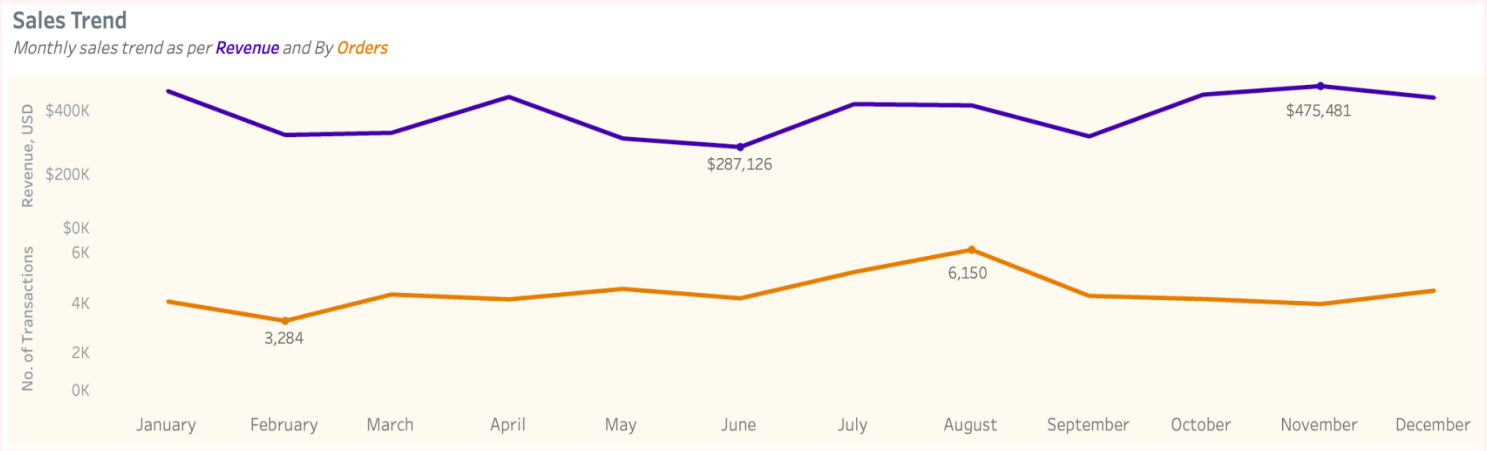


Figure: Sales trend by revenue and order

This surge aligns with a notable increase in sales of high-value product categories, particularly:

- **Nest-USA**
- **Apparel**
- **Nest**, a **new product line introduced in late August**

The **Nest product line**, introduced in **late Q3**, saw a rapid rise in adoption and sales, contributing heavily to the **revenue boost from September through December**. In the final quarter, **Nest products** consistently ranked among the top 3 revenue-generating categories, unlike earlier months where they had little or no contribution.

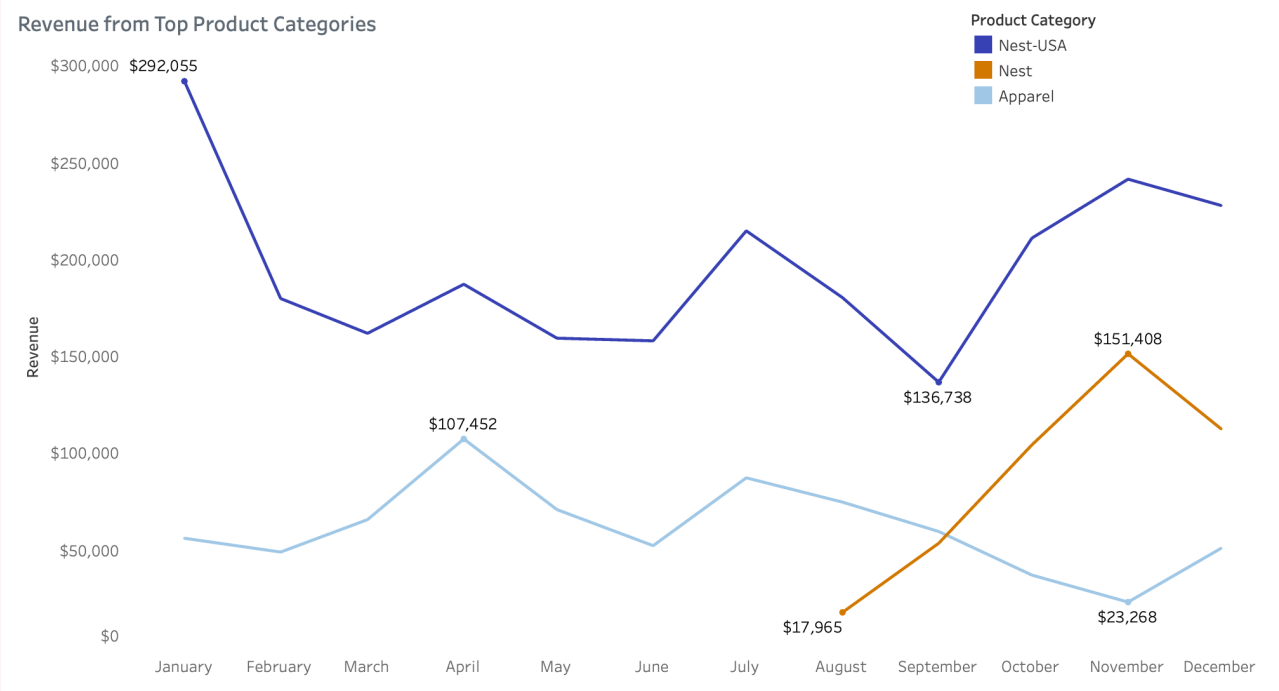


Figure: Revenue by Top 3 Product Category

Additionally, **April's sharp recovery ($442K)** after the post-January decline can be attributed to a spike in **Apparel sales**:

- **Apparel revenue in April peaked at $107K**, significantly higher than the typical $50K–$87K range seen in other months.
- Simultaneously, **Nest-USA** maintained steady sales, further contributing to April's performance.

Interestingly, April—along with January and November—not only shows a peak in revenue but also aligns with **elevated marketing spend**. April's marketing investment is the **highest relative to its preceding month**, which suggests a **targeted marketing push**, potentially aimed at **Apparel promotion**.

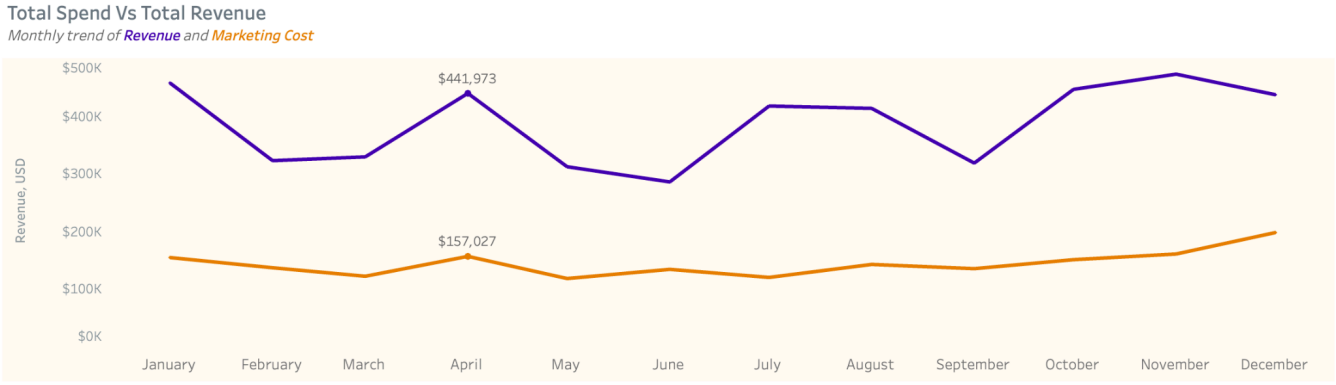


Figure: Total revenue vs total marketing cost

- ## Product Categories

  Based on revenue, the **top three product categories** are **Nest-USA**, **Apparel**, and **Nest**. In contrast, when ranked by the **number of transactions**, the top categories are **Nest-USA**, **Apparel**, and **Office**.

  Despite being launched later in the year (August), **Nest products** quickly rose to become the **third-highest revenue contributor**, even though they had **fewer transactions** than other categories. This indicates a **higher average order value** for Nest products and strong market interest immediately post-launch.

  Overall, this pattern suggests that **customers are especially drawn to high-value home automation systems**, with **Nest and Nest-USA products clearly leading the demand**.

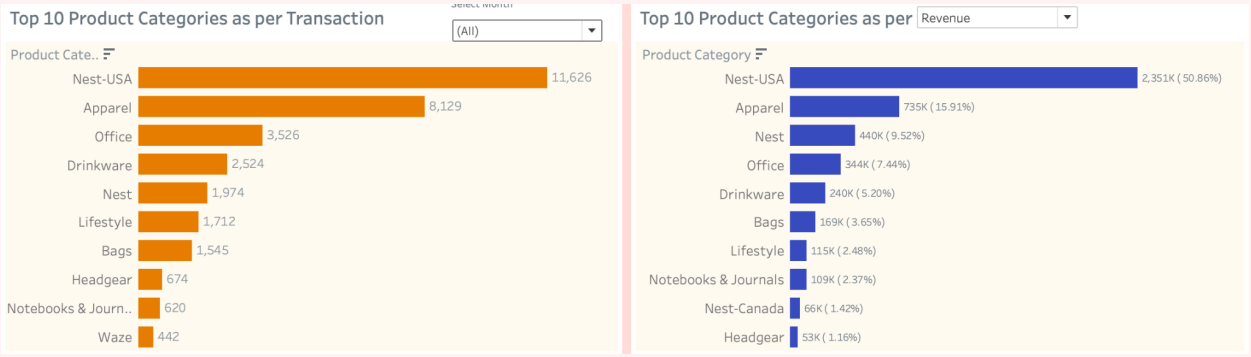


Figure: Top 10 products by categories and transaction volume

- **Revenue Distribution by Gender**

  Around 64% of users are female and 36% are male. The female customers contribute to around 62% of the total revenue.

Gender	# Customers	% Customer	% Revenue	Revenue
Female	934	63.62%	62.29%	$2,913,833
Male	534	36.38%	37.71%	$1,763,725

  A/B Test: Is the gap in revenue significant?

  To understand if the Gender affects Revenue, a T-test is conducted.

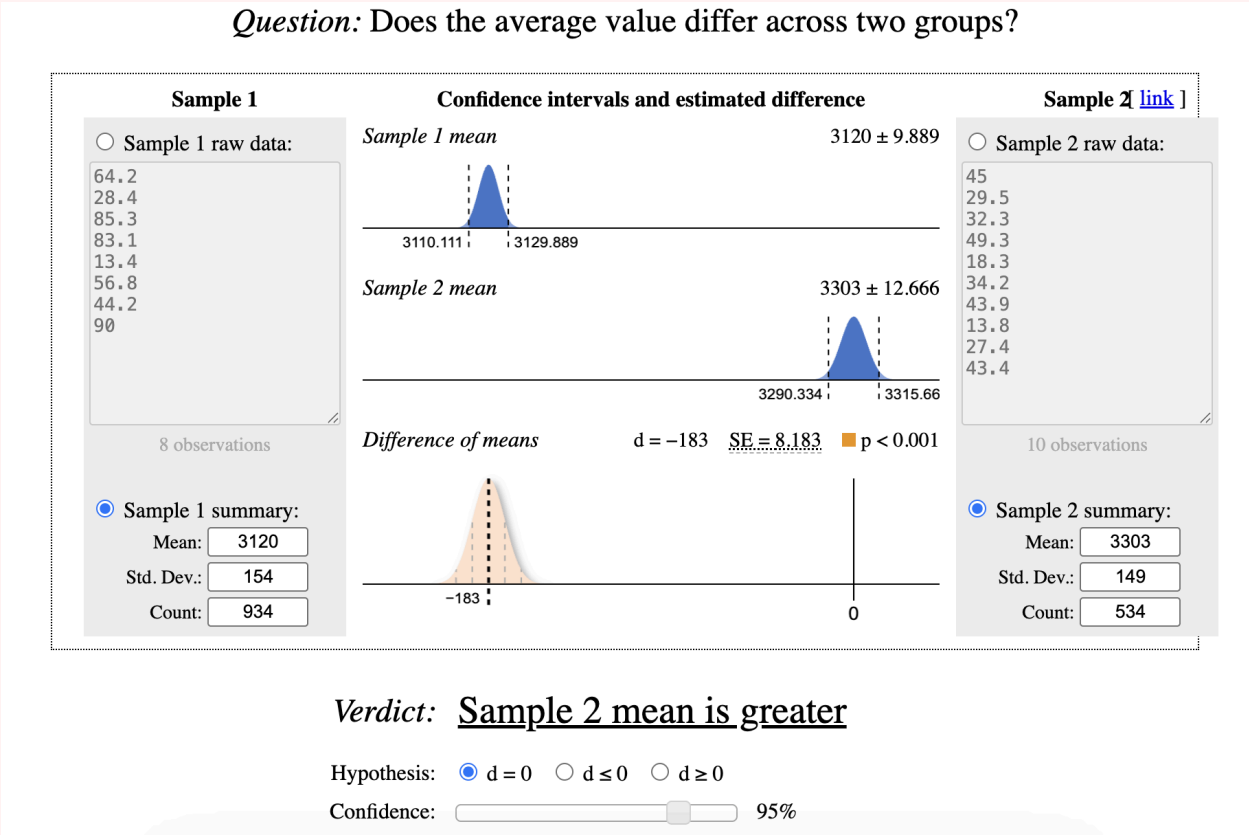
  Hypothesis Formulation:

  - Null (H0): There is no significant difference in average revenue between female and male customers.
  - Alternate Hypothesis (H1): There is a significant difference in average revenue between male and female customers.

  Test Inputs:

Metric	Female	Male
Mean Revenue	$3120	$3303
Std. Deviation	154	149
No. of Customers	934	534

Test Result:



Conclusion

Since p<0.001, we reject the null hypothesis.

Therefore, male customers generate significantly more revenue than female customers, on average.

Business Implication

1. Segment Prioritization

Male customers are contributing more revenue per head. You may want to prioritize high-value male segments for:

- Upselling/cross-selling
- Loyalty programs
- Premium products

2. Tailored Marketing Strategies

Consider designing gender-specific campaigns:

- If male customers respond better to specific products or offers, optimize future campaigns accordingly.
- For females, explore new messaging or bundles to increase their average order value (AOV).

3. Further analysis on Product & Experience Optimization
   - Dive deeper: Why are male customers spending more?
   - Is it product type? Discount sensitivity? Purchase frequency?

- **Customer Retention Analysis**

  These two heatmaps provide a month-by-month view of customer retention across the year, shown both in **absolute numbers** (Figure 1) and **percentages** (Figure 2), relative to each cohort's initial size (Month 0).

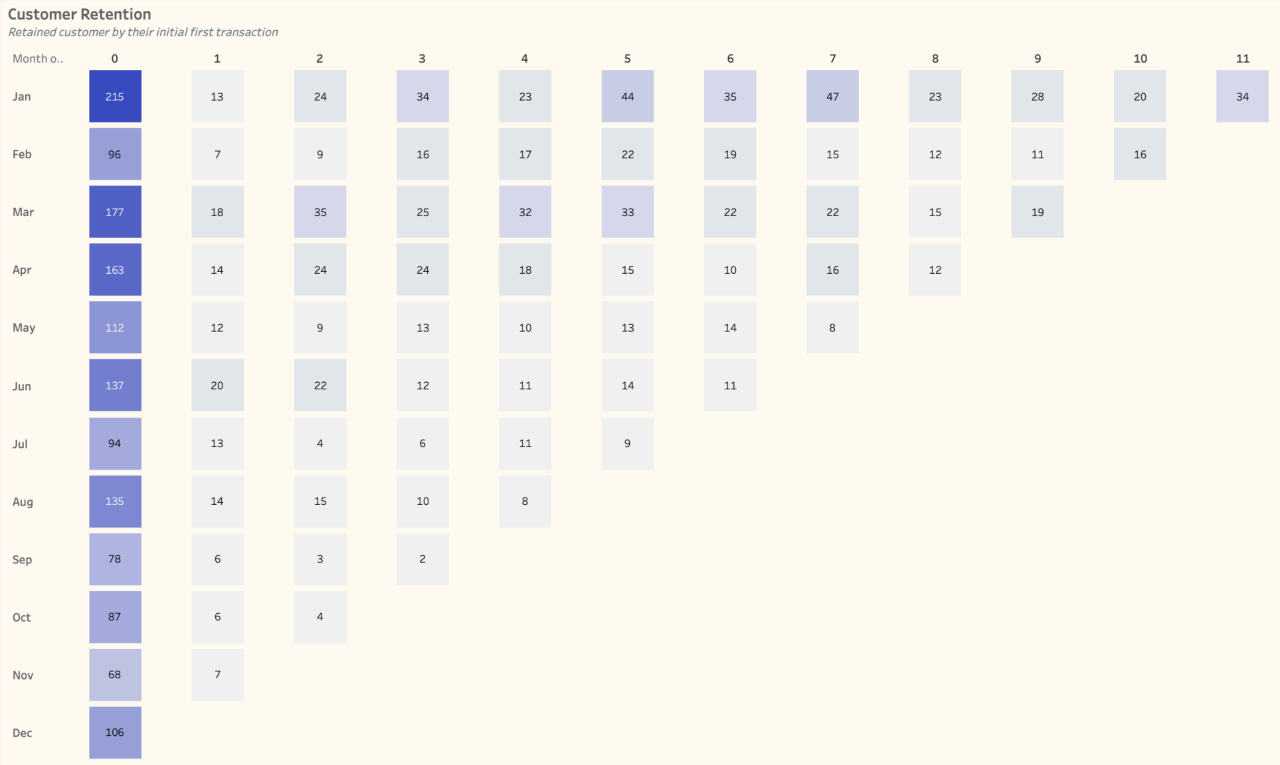


Figure 1: Customer retention by number

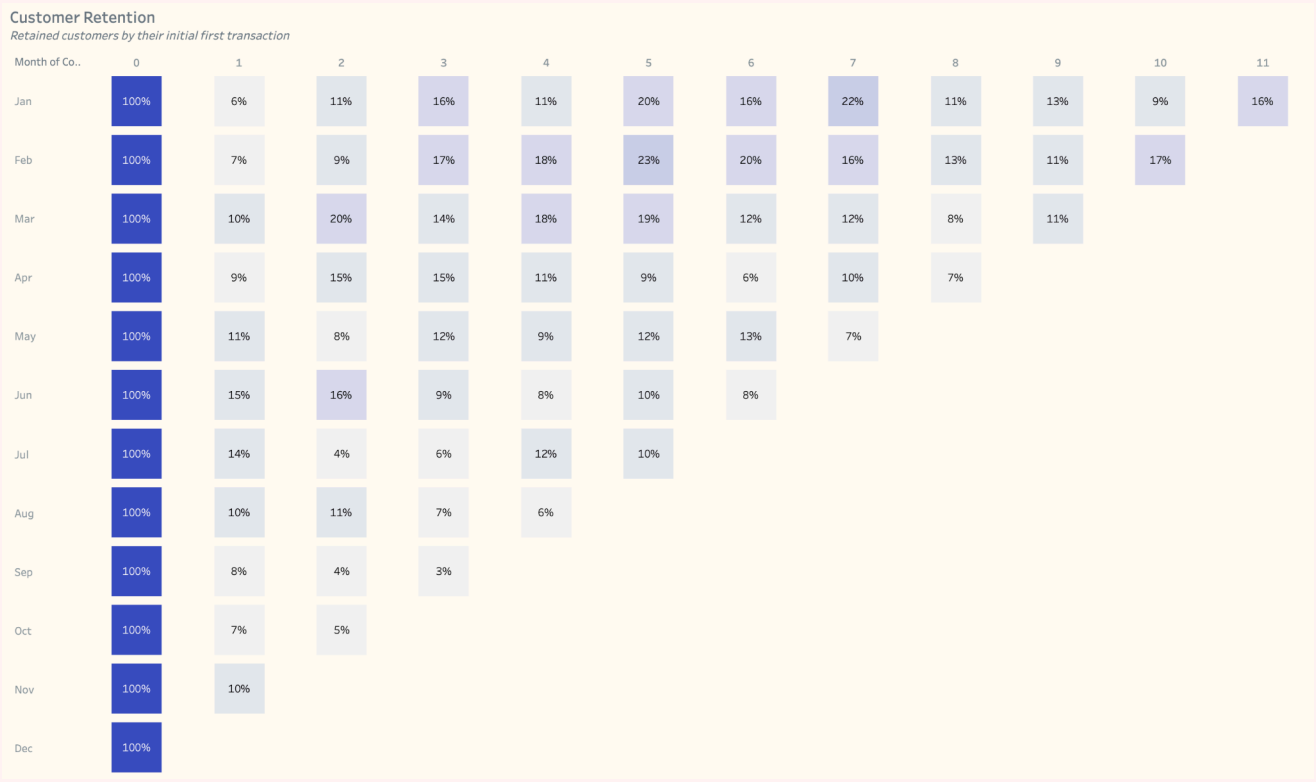


Figure 2: Customer retention by rate

This analysis reveals how well newly acquired customers are retained over time:

a. **Strong Initial Acquisition in Q1 (Jan–Mar)**
   i. **January** had the highest acquisition volume with **215 new customers**, followed by **March (177)** and **April (163)**.
   ii. These cohorts also showed **relatively stronger long-term retention**, with January maintaining **16% retention in Month 11**, the highest observed in any cohort.

b. **Sharp Drop After First Month**
   i. Across all cohorts, there is a **steep decline** from Month 0 to Month 1:
   ii. Retention typically drops to **6%–15%** in Month 1.
   iii. This early drop indicates potential issues in onboarding, product satisfaction, or post-purchase engagement.

c. **Moderate Mid-Term Retention for Jan–Mar Cohorts**

i. **January to March** show consistent engagement, with retention stabilizing around **10–20%** from Months 3 to 7.
ii. For example, March has **14% retention in Month 3**, and still maintains **11% in Month 9**, signaling a healthier retention curve.

d. **Lower Retention in Later Cohorts (Sep–Dec)**
   i. Cohorts from **September onwards** show lower initial volume and weaker retention:
   ii. September starts with only 78 customers and drops to **3% by Month 3**.
   iii. October and November show even weaker follow-through with limited data beyond Month 2.

e. **Retention Peaks (Spikes)**
   i. There are slight **spikes or plateaus** in Month 5 for some cohorts (e.g., January: 44 users retained).
   ii. This could reflect periodic reactivation campaigns or seasonality (e.g., mid-year promotions).

Business Implications:

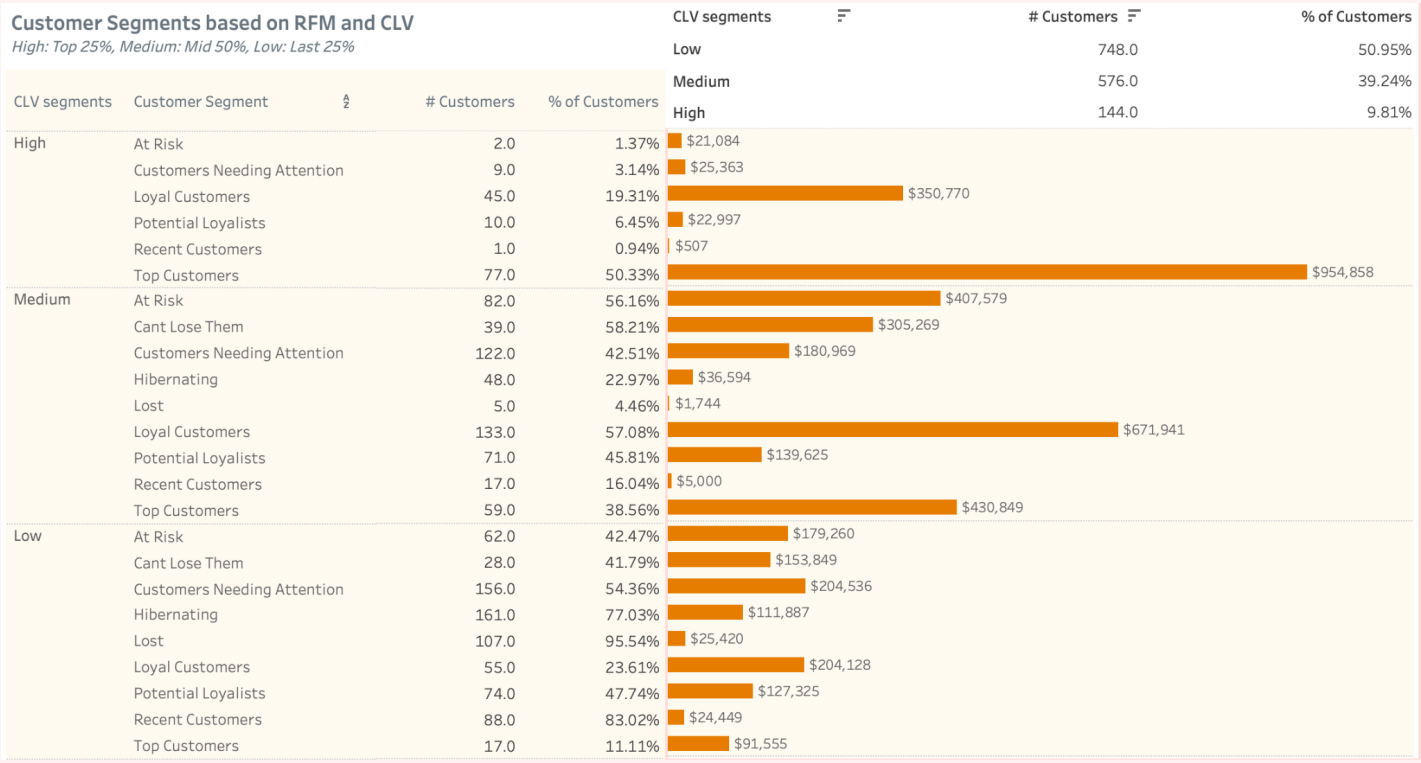
- **Early churn** is the most significant challenge: retention plummets after Month 0. This suggests a need to strengthen **onboarding experiences, post-purchase follow-up**, or **incentives for 2nd purchases**.
- **Focus on replicating Q1 strategies**: Cohorts from January–March retain customers better in the long term, indicating that acquisition or engagement strategies in Q1 are more effective.
- **Consider seasonality**: Poorer retention in later months (Sep–Dec) could be tied to **holiday churn**, **end-of-year fatigue**, or **low acquisition intent** — review marketing calendar alignment.
- **Explore win-back campaigns**: Spikes in Month 5–7 retention hint that reactivation efforts may be effective and should be expanded.

Recommendations:

- Implement **targeted re-engagement campaigns** after Month 1.
- Analyze **differences in acquisition sources or products** for Q1 vs Q3–Q4 cohorts.
- Invest in **automated onboarding journeys** to convert more new users beyond their first month.
- Segment retention by customer type or geography to personalize strategies.

- **Customer Segmentation based on CLV and RFM analysis**

  This dashboard presents a combined segmentation of customers using RFM behavioral profiles (e.g., Loyal, At Risk, Lost) and Customer Lifetime Value tiers (High, Medium, Low). The goal of this analysis is to identify customer quality across behavioral and monetary dimensions, and prioritize efforts to drive retention, growth, and revenue uplift.



Customer Segments based on RFM and CLV

High: Top 25%, Medium: Mid 50%, Low: Last 25%

CLV segments	Customer Segment	# Customers	% of Customers
High	At Risk	2.0	1.37%
	Customers Needing Attention	9.0	3.14%
	Loyal Customers	45.0	19.31%
	Potential Loyalists	10.0	6.45%
	Recent Customers	1.0	0.94%
	Top Customers	77.0	50.33%
Medium	At Risk	82.0	56.16%
	Cant Lose Them	39.0	58.21%
	Customers Needing Attention	122.0	42.51%
	Hibernating	48.0	22.97%
	Lost	5.0	4.46%
	Loyal Customers	133.0	57.08%
	Potential Loyalists	71.0	45.81%
	Recent Customers	17.0	16.04%
	Top Customers	59.0	38.56%
Low	At Risk	62.0	42.47%
	Cant Lose Them	28.0	41.79%
	Customers Needing Attention	156.0	54.36%
	Hibernating	161.0	77.03%
	Lost	107.0	95.54%
	Loyal Customers	55.0	23.61%
	Potential Loyalists	74.0	47.74%
	Recent Customers	88.0	83.02%
	Top Customers	17.0	11.11%

Figure: Customer Segmentation based on CLV and RFM

Some of the Key Observations by CLV Segment:

- **High CLV Segment (Top 25%)**: Total customers: 144 (9.81%)
  Despite being the smallest segment by volume, this group generates the largest share of revenue, led primarily by:
    - Top Customers: 77 users (50.33%) contributing $954,858
    - Loyal Customers: 45 users contributing $350,770

  Notably, even small segments like:
    - Potential Loyalists (10 customers) generated nearly $23K
    - While Recent Customers and At Risk users in this group show minimal revenue, these likely represent high-value new or disengaging customers worth re-engaging.

  Implication:
  This group should be the top priority for retention and expansion, especially the Top and Loyal segments. Personalized offers, exclusive loyalty perks, and predictive retention strategies should be applied to preserve and grow this cohort.

- **Medium CLV Segment (Middle 50%)**:Total customers: 576 (39.24%)
  This group is large and generates moderate value — a mix of opportunity and risk:
    - Loyal Customers (133 customers) still bring $671,941, showing potential for conversion into High CLV
    - Top Customers (59 users) also contribute $430,849, comparable to the High CLV segment — indicating misalignment in CLV tiering or recent drops in frequency/monetary activity
    - Customers Needing Attention (122) and At Risk (82) still generate over $400K combined, but are vulnerable to churn

  Implication:

- The Medium CLV–Loyal and Top customers are ripe for upsell and nurture—they behave like high-value customers but haven't yet reached full potential.
- The At Risk and Needing Attention groups signal where to intervene early with reactivation and win-back campaigns.

- **Low CLV Segment (Bottom 25%)**
  - Total customers: 748 (50.95%) — the largest segment by volume
  - Generates the least revenue:
    - Loyal Customers (55 users) bring $204,128, showing loyalty doesn't always translate to value
    - Lost (107 users) and Hibernating (161 users) are still significant in size, contributing marginal revenue
  - Even the Top Customers in this CLV tier (17 users) contributed just $91,555, far below the Medium/High equivalents

  Implication:

  This segment reflects the classic volume-value imbalance — many users who are frequent but low-spending or have churned. Focus on:
  - Moving "Loyal" and "Potential Loyalists" up the value ladder through bundling, upselling, or targeted promotions
  - Reducing acquisition costs on customers likely to fall into this tier
  - De-prioritizing reactivation for consistently Low CLV-Lost customers

Key Recommendations

Below summarized are the key recommendation from the analysis:

- **Address Early Retention Crisis**

  The most critical finding—retention dropping to single digits within Month 1—requires immediate intervention. Implement comprehensive onboarding sequences and early engagement campaigns. Analyze and replicate Q1

acquisition strategies that demonstrated superior retention patterns across all seasons.

- **Implement Value-Based Customer Segmentation**
  The disconnect between behavioral loyalty and revenue contribution necessitates a shift from treating all "loyal" customers equally. Prioritize High CLV customers for retention efforts while targeting Medium CLV loyal customers for revenue expansion through upselling and premium positioning.

- **Optimize Product Portfolio Strategy**
  The rapid success of Nest products demonstrates market appetite for premium home automation solutions. Accelerate expansion in high-value categories while leveraging identified seasonal patterns (January, April, November) for major launches and campaigns.

- **Gender-Specific Revenue Strategy**
  Prioritize high-value male segments ($3,303 avg revenue vs $3,120 female) for premium products and loyalty programs. Develop tailored campaigns to increase AOV among the larger female customer base (63.62%).

- **Seasonal Product Optimization**
  Leverage peak revenue months (January, November) for major launches. Accelerate expansion in premium categories like Nest products, which quickly became the third-largest revenue contributor despite late introduction.

Conclusion

This analysis reveals both significant opportunities and critical challenges for sustainable growth. While the company achieved strong performance with 169.83% marketing ROI and $4.68 million revenue, key inefficiencies in customer retention and value optimization present substantial improvement opportunities.