

# Ecommerce Customer Churn Report

## 1. Project Objective

The goal of this analysis is to identify **why** customers are churning, segment the customer base by risk, and provide data-driven recommendations to improve retention. The project follows a standard ETL (Extract, Transform, Load) pipeline using **Python** for data cleaning and **SQL** for strategic analysis.

---

## 2. Phase I: Data Cleaning & Preparation (Python)

**Tool Used:** Python (Pandas) **Goal:** To ensure data integrity before loading it into the database.

- **Standardization:** Column names were converted to `snake_case` to ensure SQL compatibility.
  - **Data Integrity:**
    - **Duplicates & Nulls:** Checked for duplicates (0 found) and dropped null values in critical columns like `Customer ID`.
    - **Logic Checks:** Removed rows with negative prices/quantities and ensured `tenure_days` was always positive (filtered out errors where `last_login < signup_date`).
  - **Feature Engineering:** Calculated `tenure_days` (Last Login - Signup Date) to enable time-based churn analysis.
  - **Loading:** Cleaned data (486k+ records) was pushed to PostgreSQL using SQLAlchemy.
- 

## 3. Phase II: Strategic Analysis (SQL)

**Tool Used:** PostgreSQL **Goal:** To answer specific business questions regarding customer behavior and retention.

### A. Acquisition & Marketing Strategy

#### 1. Acquisition Quality (Churn by Marketing Channel)

- **The Analysis:** Calculated the churn rate for each **Marketing\_Channel** to compare acquisition sources.
- **Outcome & Strategic Insight:**
  - **Insight:** High volume does not equal high quality. If "Paid Ads" brings in 10,000 users but has a 60% churn rate, while "Referrals" brings 1,000 users with only 10% churn, the "Referral" channel is actually more valuable.
  - **Action:** Shift marketing budget from high-churn channels to high-retention channels to improve Return on Ad Spend (ROAS).

## 2. Acquisition Seasonality (The "Holiday Shopper" Effect)

- **The Analysis:** Grouped customers by **Signup\_Month** to calculate churn rates for different times of the year.
- **Outcome & Strategic Insight:**
  - **Insight:** A "January Hangover" effect often appears where customers acquired in December (holiday shoppers) churn at much higher rates than those acquired in July (deliberate shoppers).
  - **Action:** Adjust Customer Lifetime Value (CLV) models. Do not spend the same ad budget acquiring a "December user" as a "July user" if the December user is statistically worth half as much.

## B. Pricing & Operational Health

### 3. Price Sensitivity (Impact of Discounts)

- **The Analysis:** Segmented customers into **Discount\_User** vs. **Full\_Price\_User** and compared their retention rates.
- **Outcome & Strategic Insight:**
  - **Insight:** If only customers who received discounts are staying, the business has attracted "bargain hunters" rather than loyalists.
  - **Action:** If churn is high among discount users, the pricing strategy is unsustainable. The brand needs to focus on product value rather than heavy discounting.

### 4. Operational Friction (Delivery Time Impact)

- **The Analysis:** Created "bins" for delivery speed (e.g., 0-3 days, 7+ days) to correlate logistics performance with churn.
- **Outcome & Strategic Insight:**
  - **Insight:** Churn isn't always a product issue. If customers receiving items in 7+ days churn at double the rate of those receiving them in 2 days, the problem is logistics.
  - **Action:** This provides a business case to invest in faster shipping or new warehouse locations to reduce delivery times.

## C. Customer Segmentation & Risk

### 5. High-Value Customer Loss (VIP Churn)

- **The Analysis:** Calculated the revenue impact of churned users specifically within high-value segments (e.g., VIPs).
- **Outcome & Strategic Insight:**
  - **Insight:** Not all churn is created equal. Losing 1 VIP customer might cost \$1,000, whereas losing 10 new customers costs \$100.
  - **Action:** If VIP churn is spiking, immediate "white glove" retention campaigns (personal outreach) are required, rather than generic automated emails.

### 6. Predicting "At-Risk" Customers (RF Analysis)

- **The Analysis:** Used Recency (days since last login) and Frequency (total purchases) to classify *active* customers into risk zones.
- **Outcome & Strategic Insight:**
  - **Insight:**
    - **High Risk VIP:** Buy often (>5 times) but haven't visited in 60+ days.
    - **Drifting:** Casual shoppers who haven't visited recently.
  - **Action:** This is a *proactive* tool. Trigger a "Win-Back" email (e.g., "We miss you, here is 10% off") for High Risk VIPs *before* they officially churn.

## D. Behavioral Insights

### 7. Behavioral Cooling (Basket Size Decline)

- **The Analysis:** Compared the quantity of items in a user's *first* order vs. their *last* order for churned customers.
- **Outcome & Strategic Insight:**
  - **Insight:** Customers rarely just vanish; they "cool down" first. A negative % change (e.g., -30% basket size) confirms gradual disinterest.
  - **Action:** If data shows basket sizes shrinking, trigger a "Volume Discount" (e.g., "Buy 5 items, get 20% off") the moment a VIP's average basket size drops by 10%.

### 8. Share of Wallet (Cross-Category Analysis)

- **The Analysis:** Compared the revenue share of different product categories for Active vs. Churned users.
- **Outcome & Strategic Insight:**
  - **Insight:** Users might not stop buying *everything*. If churned users spent disproportionately on "Home Decor" before leaving, it suggests that specific category has quality issues.

- **Action:** Investigate quality control or pricing for the specific categories that are driving customers away.

## 9. The "Critical Window" (Tenure Analysis)

- **The Analysis:** Grouped churned customers by their tenure (lifespan) to find the "drop-off" point.
- **Outcome & Strategic Insight:**
  - **Insight:** If the majority churn after the first week, it is an **onboarding issue**. If they churn after 6 months, it is an **engagement issue**.
  - **Action:** Set up automated emails or offers to trigger *just before* that critical drop-off moment to extend the customer lifecycle.

## 10. Cohort Analysis (The Retention Curve)

- **The Analysis:** Tracked the retention rate of different monthly signup cohorts over time.
- **Outcome & Strategic Insight:**
  - **Insight:** This validates product updates. If the "March Cohort" has higher retention than the "January Cohort," it proves that features released in March improved product stickiness.
  - **Action:** Use this to measure the long-term impact of product changes rather than just looking at immediate sales.

---

## 4. Conclusion

This project successfully transformed raw transactional data into actionable business intelligence. By leveraging Python for robust data cleaning and SQL for targeted segmentation, we identified that **logistics (delivery time)** and **seasonality** are key churn drivers. The analysis provided the marketing team with a prioritized list of **"At-Risk VIPs"** for immediate intervention, moving the strategy from reactive to proactive retention.

---

## Report Prepared by

Anshu Raj

ar.anshu2611@gmail.com