**Customer Retention Cohort Analysis Report**

**Project Title**: Building a Modern Marketing Analytics Solution for Customer Retention

**Date**: July 2025

**Tools Used**: Fivetran, GCP Cloud SQL, Databricks (Delta Lake, SQL, Dashboards)

horizontal line

# 

### 1. Project Objective

The goal of this project was to build an end-to-end modern data analytics solution to understand and improve **customer retention** in an e-commerce environment. We analyzed cohort behavior to answer:

* How long it takes customers to make a second purchase
* How repeat purchase rates evolve over time
* How retention varies across monthly cohorts
* How new customer acquisition trends have changed over the past 6 months

### 2. Data Pipeline and Architecture

We used a **modern data stack** to ingest, transform, and visualize data:

* **Data Source**: E-commerce sales data stored in **GCP Cloud SQL**
* **Ingestion**: **Fivetran** was used to automatically replicate tables and schema into **Databricks Delta Lake**
* **Transformation**: Raw data was transformed using **SQL (CTEs, subqueries)** in Databricks to:  
  + Identify each customer's **first** and **second purchase date**
  + Group users into **monthly cohorts**
* **Visualization**: Interactive **Databricks Dashboards** were built to monitor retention, repeat purchase behavior, and cohort sizes

### 3. Key Visualizations & Insights

#### **3.1 Retention Rate by Cohort (Time to Second Purchase)**

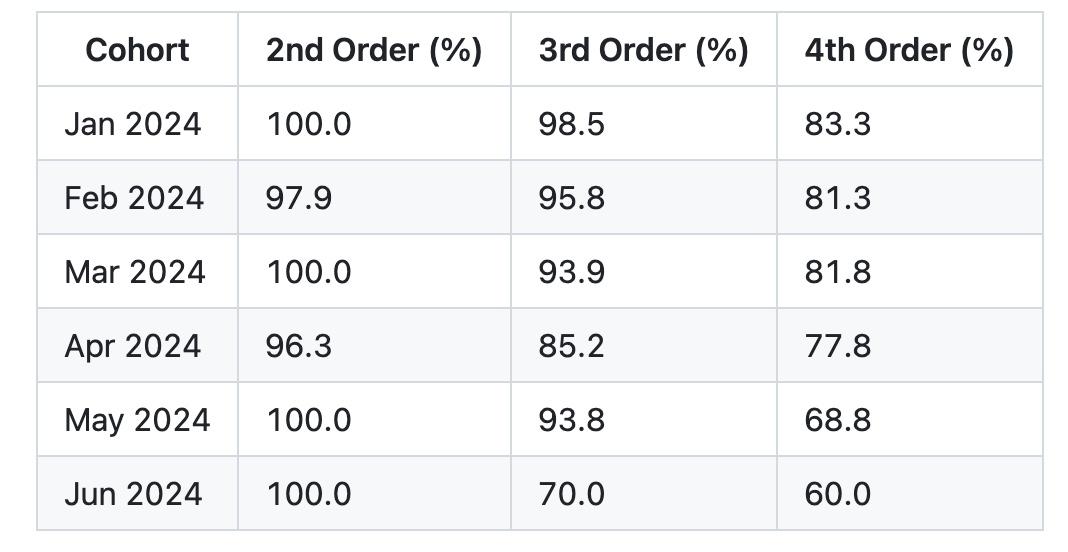
#### 

#### 📌 **Trend**:

* Early cohorts (Jan–Mar) had **lower 1-month retention** (~30–38%)
* Later cohorts (Apr–May) improved to **48–50%**

#### 

#### **3.2 Repeat Purchase Rate by Cohort**



#### 📌 **Observations**:

* **2nd Purchase** rates are consistently **high (96–100%)**
* Drop-off begins at **3rd and 4th orders**, especially for **newer cohorts**
* **Older cohorts** show stronger long-term loyalty

📌 **Interpretation**:

* May reflect **shorter observation window** for recent users
* Could also indicate **weaker engagement or seasonal effects**

#### **3.3 Cohort Size Trends**

#### 

#### 📌 **Observation**:

* Steady **decline in new customer acquisition**
* June’s cohort is **85% smaller** than January’s

### 

### 

### 

### 

### 

### 4. Key Insights Summary

* 📈 **Short-term retention improved** in newer cohorts (48–50%)
* 📉 **Long-term retention is dropping**, especially for 3rd and 4th orders
* 🔽 **Customer acquisition is slowing down**, suggesting reduced reach or changes in marketing strategy
* 📊 High 2nd purchase rate suggests **onboarding or first experience is strong**

### 5. Recommendations

* 🧪 **Deep dive into 3rd/4th purchase drop-off** – investigate loyalty, satisfaction, and touchpoints
* 🎯 **Optimize acquisition campaigns** – revisit ad channels, landing pages, and offers
* 📦 **Enhance post-purchase experience** – upsell, loyalty programs, follow-up emails
* 📅 **Continue monitoring** newer cohorts over time as they mature
* 🔄 **Repeat cohort analysis quarterly** to track changes and improvements

### 

### 

### 

### 6. Conclusion

This project showed how tools like Fivetran and Databricks can work together to build a **scalable, automated cohort analysis solution**. It provided key insights into retention patterns and helped identify opportunities for **marketing and product strategy improvements**.

By understanding customer behavior over time, the company can take **data-driven actions** to improve retention, engagement, and long-term value.