Beyond the First Click: A Power BI Case Study on Repeat Behavior and Churn Risk in E-Commerce

TL;DR

This case study explores repeat behavior within a fast-growing e-commerce business that struggled with stagnant customer retention despite strong acquisition performance. Using SQL, Excel, and Power BI, I analyzed a dataset of over 4,500 customers to segment loyalty levels, identify churn risk, and overlay fraud insights. Key findings revealed that only 8% of customers were highly loyal, while over 70% of revenue originated from high churn risk segments. The interactive dashboard I built translated raw data into strategic recommendations for improving retention, targeting fraud-prone cohorts, and tailoring loyalty incentives by category preference.

Project Summary

A fast-scaling e-commerce company saw strong customer acquisition but stagnant repeat purchases. Despite rising marketing spend, loyalty metrics showed no meaningful growth. I stepped in as a data analyst to answer a pivotal question:

“What drives repeat purchases, and where are customers slipping away?”

The Challenge

Although the dataset covered 4,500+ customers across 10 countries, the company lacked behavioral segmentation. Key gaps included:

- No clear definition of customer loyalty or churn
- No cohort-based analysis of revenue drivers
- Limited visibility into how fraud skews behavioral metrics

Tools Used:

- **SQL**- Engineered loyalty and churn logic, segmented customer behavior
- **Microsoft Excel**- Used for preprocessing and intermediate cleaning
- **Power BI**- Built an interactive dashboard to explore patterns across loyalty, fraud, and revenue

Segmentation Strategy:

- **Loyalty Tiers**- Grouped customers based on the total number of orders
- **Churn Risk**- Identified based on revenue volatility and order recency
- **Fraud Overlay**- Mapped fraud presence against loyalty/churn behavior

Key Business Insights

Loyalty Distribution is Highly Concentrated: Only 8% of customers qualified as Super Loyal (15+ orders). Meanwhile:

- 87% fell into the Loyal (10-14 orders) and Moderately Loyal (5-9 orders) brackets
- The rest remained Low Engaged or Inactive despite high acquisition costs

Opportunity: The Moderately Loyal group is the most scalable revenue target, prime for conversion through loyalty perks and personalization.

High Churn Segments Drive Majority Revenue: Although only 43% of customers were high risk for churn, they drive 70% of total revenue ($3.49M).

- Medium Risk customers: $1.1M
- Low Risk customers: $340K

Action: Strategic retention in this segment can protect revenue at scale.

Fraud is entangled with Churn Risk: Within the High Churn Risk cohort:

- 2.5% were identified as fraudulent
- In contrast, Medium/Low churn segments had <1% fraud

Recommendation: Integrate fraud detection more tightly into churn prevention initiatives.

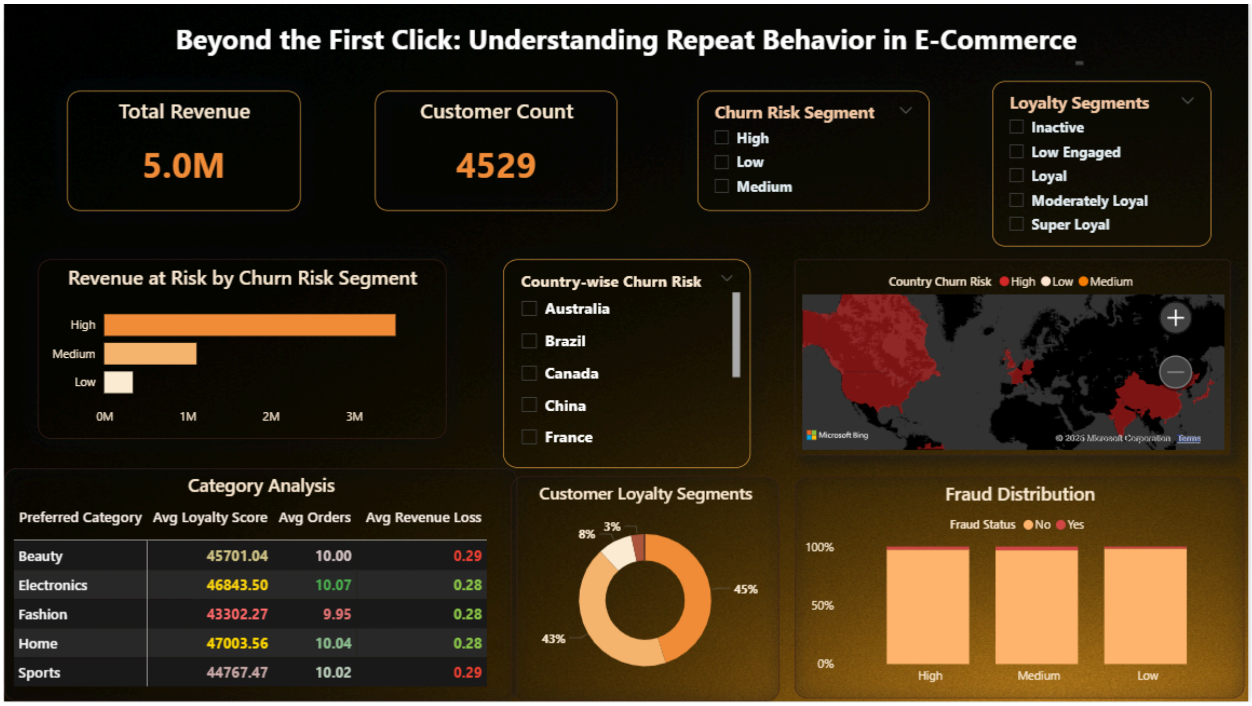
Category Loyalty Varies Widely:

- Home and Sports buyers were deeply loyal (average orders ~10; high loyalty scores)
- Fashion buyers spent more but churned faster

Tactic: Launch retention campaigns for fashion buyers; introduce VIP tiers for Home/Sports segments.

Dashboard Preview:

(Power BI screenshot here)



SQL Logic Preview:

(Screenshot of segmentation logic used)

```sql
WITH CRM AS (
  SELECT Preferred_Category,
  ROUND(AVG(Loyalty_Score), 2) AS Avg_Loyalty_Score,
  ROUND(AVG(Total_Orders), 2) AS Avg_Orders,
  ROUND(AVG(Churn_Risk), 2) AS Avg_Revenue_Loss
  FROM ecommerce_table
  GROUP BY Preferred_Category
  )
  SELECT
  e.Customer_Id,
  e.Country,
  e.Total_Orders,
  CASE
   WHEN e.Total_Orders BETWEEN 15 AND 19 THEN 'Super Loyal'
   WHEN e.Total_Orders BETWEEN 10 AND 14 THEN 'Loyal'
   WHEN e.Total_Orders BETWEEN 5 AND 9 THEN 'Moderately Loyal'
   WHEN e.Total_Orders BETWEEN 1 AND 4 THEN 'Low Engaged'
   ELSE 'Inactive'
 END AS Loyalty_Segment,
```

```sql
END AS Loyalty_Segment,
ROUND(e.Avg_Order_Value * e.Total_Orders, 2) AS Total_Revenue,
 CASE
 WHEN (Avg_Order_Value * Total_Orders) >= 1000 THEN 'High'
 WHEN (Avg_Order_Value * Total_Orders) >= 500 THEN 'Medium'
 ELSE 'Low'
 END AS Churn_Risk_Segment,
 CASE
 WHEN e.Is_Fraudulent = 1 THEN 'Yes'
 ELSE 'No'
 END AS Fraud_Status,
 e.Preferred_Category,
 c.Avg_Loyalty_Score,
 c.Avg_Orders,
 c.Avg_Revenue_Loss
FROM ecommerce_table e
LEFT JOIN CRM c
 ON e.Preferred_Category = c.Preferred_Category
 WHERE e.Country IS NOT NULL;
```

Why This Case Study Matters

Many firms focus on top-line growth but ignore retention drivers. This case study demonstrates:

- Full cycle analysis from SQL logic to interactive BI
- Stakeholder-ready dashboard built with real business context
- Strategic thinking, turning data into decisions.

What I'd Explore Next

- **Customer Lifetime Value (CLV):** Building CLV models would help prioritize retention efforts and segment high-potential customers more effectively.

- **Cohort Analysis by Acquisition Channel:** Understanding whether acquisition source (e., paid ads, email, referrals) impacts long-term engagement or churn would help optimize marketing spend.

- **Time-Based Purchase Behavior:** Incorporating a temporal lens (e.g., seasonality, time since last purchase) could reveal patterns around re-engagement windows and buying cycles.

- **A/B Testing Simulations for Loyalty Programs:** Simulating different retention strategies on segments like Moderately Loyal customers would help estimate potential ROI before rollout.

- **Customer Sentiment (if available):** Combining transaction data with survey or review sentiment would add a qualitative layer to loyalty understanding.

Limitations

This analysis was based on a synthetic dataset. While all insights are logically structured, accrual user behavior may vary. The absence of time-series data and marketing attribution limits some advanced behavioral modeling.

Explore the Project

- GitHub Repo
- Portfolio