

Customer Retention & Cohort Analysis Report

1. Introduction

Customer retention is one of the most critical drivers of long-term business profitability. While buying new customers is costly, keeping existing ones helps companies achieve sustainable growth. Cohort analysis is a data-driven approach that groups customers based on their acquisition date or first purchase and tracks their behaviour over time.

This report presents the results of a **Customer Retention & Cohort Analysis** project conducted using an e-commerce dataset. The analysis provides insights into customer retention trends, purchase frequency, monetary contribution, and customer segmentation through RFM (Recency, Frequency, Monetary) analysis.

The key aims were:

- To clean and preprocess transactional data for analysis.
- To build cohort groups and calculate monthly retention rates.
- To visualise customer retention behaviour across cohorts.
- To implement RFM segmentation for actionable customer insights.
- To recommend strategies for improving customer retention.

2. Data Preparation

The dataset consisted of transactional records with the following key fields:

- **InvoiceNo** – unique transaction identifier.
- **InvoiceDate** – date and time of transaction.
- **CustomerID** – unique customer identifier.
- **Quantity** – number of units bought.
- **UnitPrice** – price per unit.
- **Country** – country of the customer.

Data Cleaning Steps:

1. **Managed Missing Values:** Removed rows with missing CustomerID or InvoiceDate.
2. **Type Conversions:** Converted InvoiceDate to a datetime format and ensured CustomerID was an integer type.
3. **Feature Engineering:**
 - TotalPrice = Quantity × UnitPrice to measure the monetary value of each transaction.
 - InvoiceMonth (period-based field) to group transactions by month.
 - CohortMonth to capture each customer's first purchase month.
 - CohortIndex to calculate the number of months since first purchase.

3. Cohort Analysis

3.1 Cohort Creation

Customers were grouped into cohorts based on the month of their first purchase (CohortMonth).

3.2 Retention Calculation

For each cohort, the retention rate was calculated as:

$$\text{Retention Rate} = \frac{\text{Customers active in month } k}{\text{Customers in cohort (month 0)}}$$

3.3 Key Findings

- **Retention Decline:** Retention typically dropped after the first purchase month, with only **20–30%** of customers making repeat purchases in later months.
- **Best Cohorts:** Early cohorts (e.g., July–September 2011) showed stronger retention compared to later months, possibly due to better promotions or customer engagement strategies.

- **Long-Term Customers:** A small proportion of customers remained active beyond 6 months, forming the loyal base.

3.4 Visualization

A **heatmap** was generated to show retention patterns across cohorts. Darker cells showed stronger retention. For example:

- Month 0: 100% (all customers in cohort).
- Month 1: ~20–30% retention.
- Month 3+: <15% retention for most cohorts.

4. RFM Segmentation

To complement cohort analysis, **RFM analysis** was conducted:

- **Recency (R):** Days since last purchase.
- **Frequency (F):** Number of transactions made.
- **Monetary (M):** Total value of purchases.

4.1 Scoring

- Customers were scored 1–4 for each RFM metric (quartile-based).
- Combined score = R_Score + F_Score + M_Score.

4.2 Segmentation

Using the combined score, customers were segmented:

- **Champions** (score ≥ 9): Frequent and recent buyers contributing high revenue.
- **Loyal** (7–8): Repeat buyers with consistent activity.
- **Potential** (5–6): Occasional buyers with growth potential.
- **At Risk** (<5): Inactive customers who need reactivation strategies.

4.3 Key Findings

- Only a small percentage (~10–15%) were **Champions**, but they contributed disproportionately to revenue.
- The majority fell into the **Potential** group, highlighting an opportunity for engagement campaigns.
- A considerable number of customers were **At Risk**, reinforcing the need for retention-focused strategies.

5. Visualisations Produced

1. **Cohort Retention Heatmap** – Monthly retention trends.
2. **Average Retention Curve** – Average percentage of kept customers across all cohorts.
3. **Cohort Sizes Bar Chart** – Number of customers in each cohort (Month 0).
4. **RFM Segment Distribution** – Proportion of customers in Champions, Loyal, Potential, and At Risk.

6. Business Insights

From the analysis, several insights appeared:

- Retention declines sharply after the first month. → onboarding and post-purchase engagement should be strengthened.
- A small but critical group of loyal customers drives long-term revenue → reward them with loyalty programs.
- Many customers fall into “Potential” → personalised campaigns can nurture them into Loyal customers.
- Geographic differences (if analysed by Country) can inform region-specific retention strategies.

7. Recommendations

1. Onboarding Campaigns:

- Send personalised emails to new customers within the first week.
- Offer incentives (discounts, free shipping) for the second purchase.

2. Loyalty Programs:

- Reward repeat buyers (points, VIP tiers).
- Exclusive offers for Champions.

3. Customer Reactivation:

- Target At Risk customers with win-back offers.
- Use reminders and personalised product recommendations.

4. Product & Pricing Strategy:

- Analyse high-retention cohorts' buys to find "sticky" products.
- Bundle or cross-sell these products.

5. Data-Driven Marketing:

- Segment customers by RFM for targeted campaigns.
- Monitor retention metrics monthly to evaluate strategy effectiveness.

8. Conclusion

This analysis highlights the importance of understanding customer behaviour beyond simple sales figures. Cohort analysis revealed clear patterns of churn, while RFM segmentation allowed for actionable customer grouping. By implementing the recommended strategies, businesses can improve retention, increase customer lifetime value, and achieve sustainable growth.

Future extensions may include:

- Predictive modelling for churn.

- Customer lifetime value forecasting.
- A/B testing of retention strategies.