



Customer Segmentation & Retention Analysis (RFM Model)

Data Cleaning • Feature Engineering • Segmentation • Insights
• Action Plan

Prepared by: Rimsha Iram
Internship Project — Elevvopath Analytics



Project Goal & Dataset Overview

Goal:

- Segment customers based on Recency, Frequency, and Monetary (RFM) behavior.
- Derive actionable insights and segment-specific strategies.
- Build foundation for future reporting & dashboards.

Dataset:

- Source: online_retail_interim.csv
- Columns: CustomerID, InvoiceDate, InvoiceNo, Quantity, UnitPrice, TotalPrice
- Transformations:
 - Converted InvoiceDate → datetime
 - Created TotalPrice = Quantity × UnitPrice
 - Removed negative/cancelled transactions
 - Dropped missing CustomerIDs
 - Added log transform for Monetary



RFM Modeling Process



Feature Engineering:

- Reference Date: $\text{Max}(\text{InvoiceDate}) + 1 \text{ day}$
- Aggregation:
 - Recency = $(\text{snapshot_date} - \text{last purchase date})$
 - Frequency = number of unique invoices
 - Monetary = total spending
- Scoring (1–5 Quantiles):
 - R: lower recency \rightarrow higher score
 - F, M: higher \rightarrow better
 - RFM_score = concatenation (e.g., 555)

Segment Rules:

- Champions (555)
- Loyal Customers (High F, Medium M/R)
- At Risk (Low R, Medium-High F/M)
- Lost (Low R, F, M)
- Recent Buyers (High R, Low F)
- Others (Remaining combinations)



Segment Summary & Visualization Plan

Segment Overview:

Champions: 938 → 2.62M revenue (Avg 2792)

Loyal Customers: 953 → 0.77M revenue (Avg 807)

At Risk: 260 → 0.26M revenue (Avg 991)

Lost Customers: 1036 → 0.27M revenue (Avg 256)

Recent Buyers: 304 → 0.09M revenue (Avg 311)

Others: 700 → 0.29M revenue (Avg 419)

Core Visuals:

- Bar: Customer Count by Segment
- Bar: Revenue Contribution by Segment
- Optional: Recency vs Frequency scatter, Top Products



Data-Backed Insights



1. Champions: 22.4% customers → 60.9% revenue (avg \$2792)
2. Loyal Customers: 22.7% → 17.9% revenue (avg \$807)
3. At Risk: 6.2% → 6.0% revenue (avg \$991)
4. Lost Customers: 24.7% → 6.2% revenue (avg \$256)
5. Recent Buyers: 7.3% → 2.2% revenue
6. Others: 16.7% → 6.8% revenue

Financial Snapshot:

Total Revenue: ≈ 4.3M | Customers: ≈ 4K

Avg Revenue/Customer: ≈ 1K | Avg Recency: 92 days | Avg Frequency: 4 orders

Behavioral Takeaways:

- Strong inequality → retention programs vital.
- Lost customers show drop-off timing.
- Frequency drives value — build recurring incentives.

Tactical Action Plan

| Segment | Goal | Recommended Actions |
|-----------------|---------------------------|--|
| Champions | Retain & Reward | Early access, referral programs, thank-you notes |
| Loyal Customers | Increase Repeat Purchases | Tiered loyalty, upsell combos |
| At Risk | Reactivate | Limited-time offers, highlight past favorites |
| Lost Customers | Win-Back | Discount emails, "We Miss You" campaigns |
| Recent Buyers | Build Habit | Welcome emails, small incentives |
| Others | Upsell | Bundles, trending items |

Strategic Moves:

1. Launch VIP/Champions Club
2. Create At-Risk Recovery Campaign
3. Track segment transitions monthly
4. Automate RFM scoring in Power BI



Summary & Next Steps

Summary:

- RFM segmentation revealed key tiers & revenue concentration.
 - Retention priority: Champions + Loyal.
 - Reactivation focus: At-Risk + Lost.
 - Future scope: Product-level RFM, CLV modeling.
- 