

Retail Customer Retention Analysis

A Portfolio Case Study in Business Analysis & ICT

Prepared by: Dhairya Shah

Executive Summary

This case study investigates the dynamics of customer retention within the retail sector, leveraging data-driven analysis and ICT methodologies. The central aim was to understand how customer loyalty impacts revenue, to quantify retention challenges, and to simulate the workflows of a Business Analyst (BA) working in tandem with ICT practices.

Key findings revealed that repeat customers account for over 70% of sales revenue, retention falls drastically within the first three months, and untapped opportunities exist in secondary international markets.

The project demonstrates the ability to transform raw transactional data into actionable insights using structured BA frameworks. The workflow included data cleaning, exploratory segmentation, cohort retention analysis, visualization, and presentation of recommendations through simulated industry-standard tools such as JIRA, Confluence, and Power BI.

Executive Insights

- Repeat customers contribute ~72% of sales revenue.
- Customer retention falls to ~15% after 12 months.
- UK dominates sales, but EIRE and Netherlands offer growth potential.
- Early engagement (first 90 days) is critical to prevent churn.
- Loyalty programs and targeted campaigns are essential for retention.

Introduction & Aim

Customer retention is a cornerstone of profitability. Research consistently shows that acquiring new customers can cost five to seven times more than retaining existing ones. Furthermore, improving retention by as little as 5% can increase profits by 25% to 95%. For retailers competing in saturated markets, these statistics underscore the need for structured retention strategies underpinned by ICT systems and BA workflows.

The aim of this project was twofold. First, to simulate the role of a Business Analyst by applying analytical and ICT skills to a real dataset and producing insights that inform business strategy. Second, to develop a flagship portfolio project that demonstrates professional readiness for BA/ICT roles, with deliverables that mirror real-world consulting or ICT reporting practices. By combining technical rigor with business storytelling, the project provides a recruiter-ready example of my ability to contribute in BA/ICT environments.

Methodology

The methodology followed a structured BA/ICT approach designed to replicate industry practices:

1. Data Cleaning: Processed over 1 million rows. Removed duplicates (~33k), handled ~228k missing IDs, and validated invoice dates.
2. Exploratory Analysis: Conducted loyalty splits, RFM segmentation, and top country contribution analysis to frame business questions.
3. Segmentation & Cohort Analysis: Applied cohort matrices to evaluate retention trends and loyalty splits for churn insights.
4. Visualization: Produced heatmaps, distribution charts, and trend curves to translate technical analysis into stakeholder-friendly visuals.
5. BA Tool Simulation: Structured insights as JIRA epics, Confluence documentation, and Power BI dashboards for stakeholder alignment.

Hypotheses

H1: A small group of repeat customers generate the majority of revenue.

Rationale: Based on the Pareto principle (80/20 rule), it was expected that a minority of loyal customers would contribute disproportionately to sales. Confirming this hypothesis would imply that investments in loyalty programs yield higher ROI than broad acquisition campaigns.

H2: Customer retention declines significantly after the first purchase.

Rationale: Industry evidence suggests customers often disengage early if not nurtured. If validated, this would highlight the importance of structured early engagement strategies in BA recommendations.

H3: The UK dominates sales, but international markets such as EIRE and Netherlands hold growth opportunities.

Rationale: Given the dataset's origin in a UK-based retailer, UK dominance was anticipated. However, validating secondary market potential would provide actionable recommendations for expansion pilots.

H4: Segmentation through RFM and loyalty analysis highlights actionable strategies for retention.

Rationale: Customer segmentation by recency, frequency, and monetary value was expected to identify high-value groups. This would support targeted ICT-enabled workflows (CRM triggers, loyalty rewards).

Analysis & Insights

Step 2: Data Cleaning & Overview

The dataset contained over 1 million rows spanning 2009–2011. After cleaning, the following summary was obtained:

- Raw Rows: 1,067,371
- Rows After Cleaning: 1,007,914
- Removed Duplicates: 33,757
- Date Range: 2009-12-01 → 2011-12-09
- Unique Customers: 5,878
- Unique Invoices: 40,078
- Unique Countries: 43
- Nulls in Customer ID: 228,489

Business Implication: Cleaning ensured data reliability. The large proportion of missing customer IDs (over 228k) suggests CRM process gaps in data capture.

Step 5: Exploratory Analysis

Objective: Identify customer loyalty contribution, segment revenue, and map regional performance.

Loyalty Split (Repeat vs One-time Buyers)

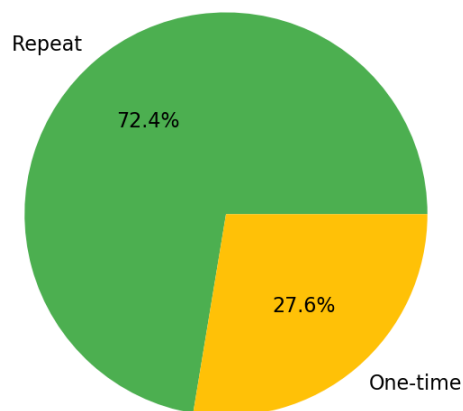


Figure 5.1 – Loyalty Split

Insight: 72% of sales came from repeat customers, validating H1. This emphasizes the disproportionate impact of loyal customers on profitability.

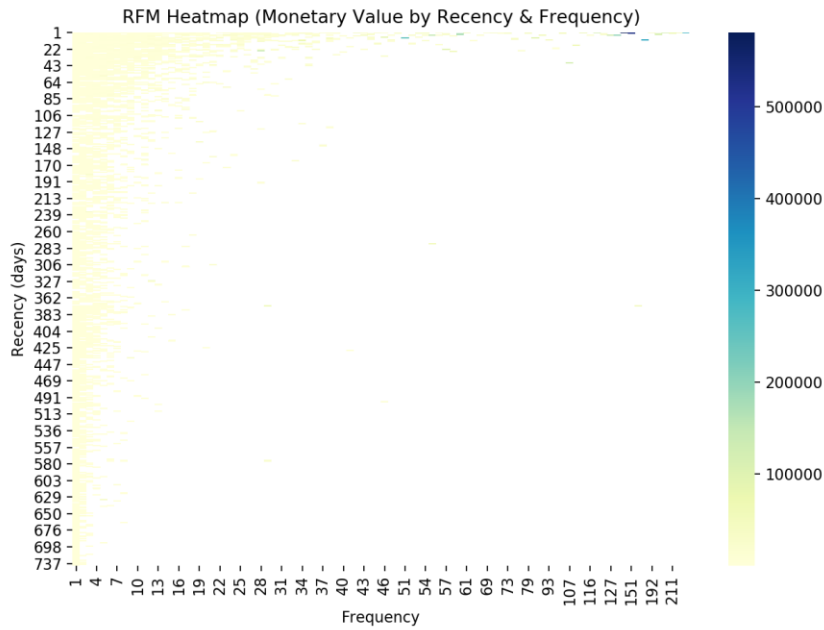


Figure 5.2 – RFM Heatmap

Insight: RFM segmentation highlights customers with high recency and frequency as the most valuable group. These segments warrant targeted ICT workflows for retention.

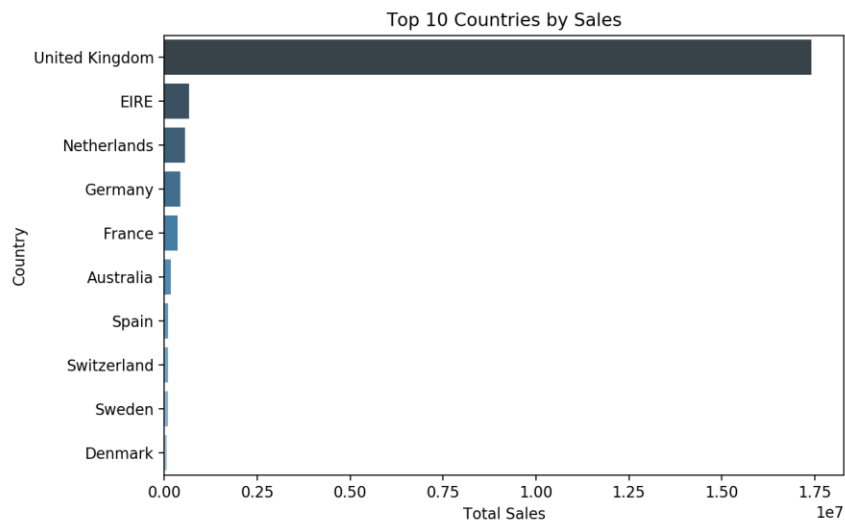


Figure 5.3 – Top Countries by Sales

Insight: The UK dominates sales, but markets such as EIRE and Netherlands provide room for targeted growth initiatives.

Step 6: Cohort Analysis

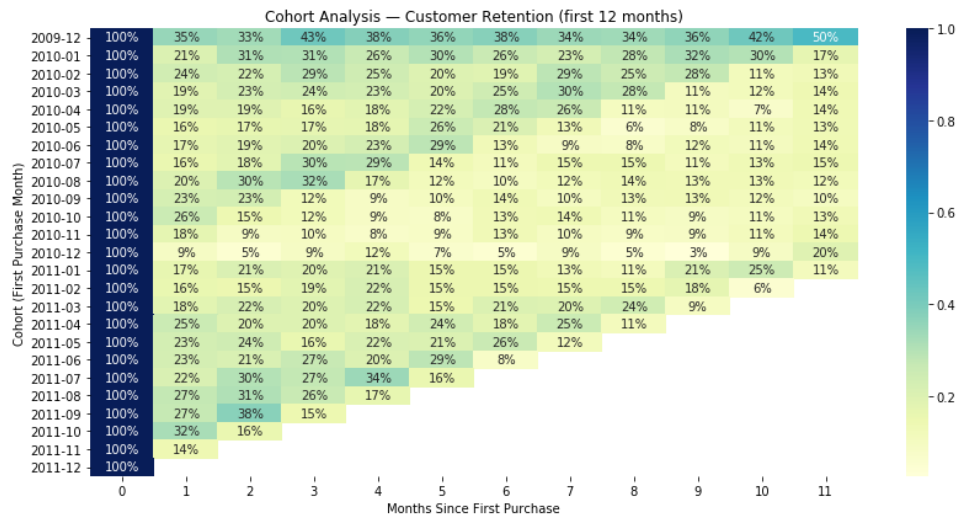


Figure 6.1 – Cohort Retention Heatmap

Insight: Retention drops from 100% in month 0 to ~35% by month 3 and ~15% by month 12. This validates H2 and highlights churn risk.

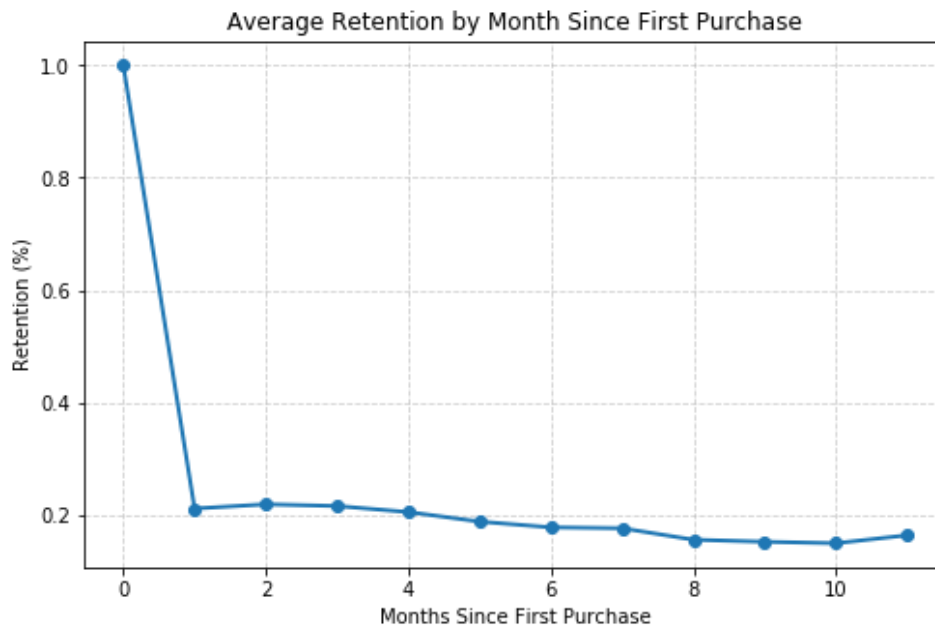


Figure 6.2 – Retention Curve

Business Implication: Without structured onboarding, customers disengage quickly. Early engagement workflows (discounts, onboarding campaigns) are critical.

Step 7: Segmentation & Loyalty Analysis

Loyalty Split (Repeat vs One-time Buyers)

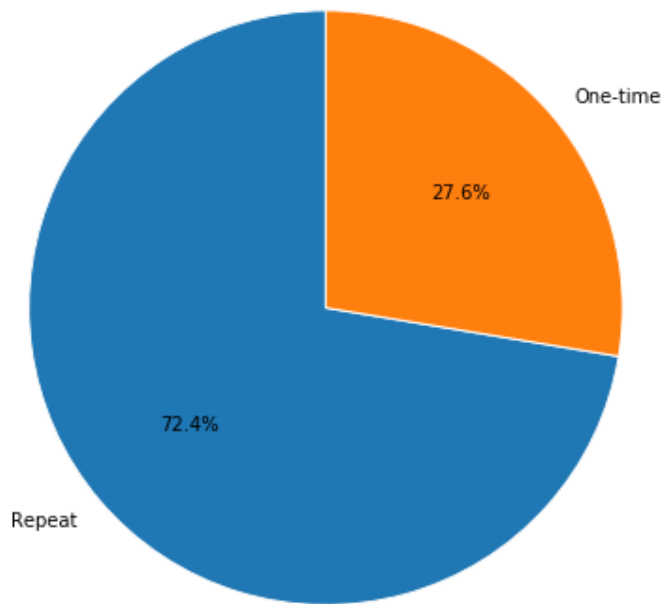


Figure 7.1 – Loyalty Split Detail

Insight: Further detail confirms repeat buyers generate higher lifetime value. Acquisition alone cannot sustain profitability.

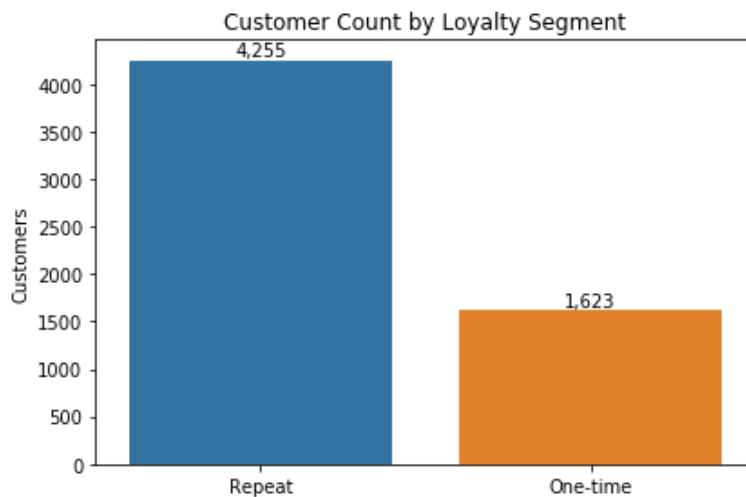


Figure 7.2 – Loyalty Contribution by Group

Implication: Targeting repeat customers with loyalty incentives directly drives revenue growth.

Step 10: Power BI Mock Dashboard

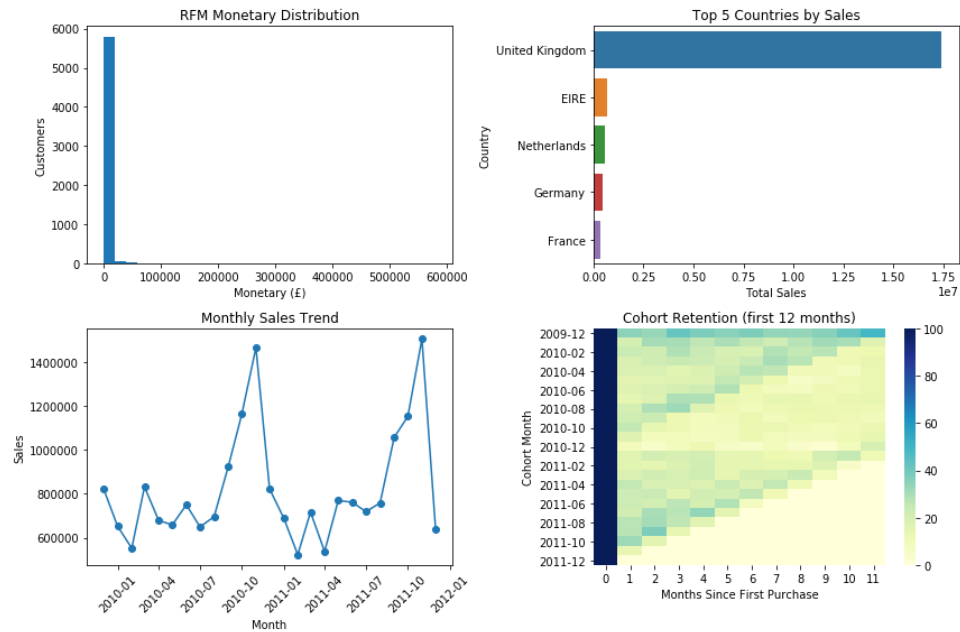


Figure 10.1 – Power BI Dashboard Overview

The simulated dashboard consolidates KPIs such as retention, sales, and loyalty. This demonstrates ability to present technical results in executive-friendly formats.

Key Findings

1. Repeat customers drive ~72% of sales revenue, proving loyalty is the backbone of profitability.
2. Customer retention declines rapidly, with <15% retained beyond 12 months.
3. UK dominance is clear, but EIRE and Netherlands represent growth opportunities.
4. RFM segmentation shows high recency/frequency buyers as the most valuable customers.
5. BA/ICT integration is required: data gaps (228k missing IDs) must be addressed to strengthen CRM accuracy.

Recommendations

- Loyalty Programs: Introduce structured digital rewards systems to convert one-time buyers into repeat customers.
- Early Engagement Funnels: Prioritize the first 90 days with automated onboarding emails, discounts, and personalized campaigns.
- International Expansion: Run pilot campaigns in secondary markets like EIRE and Netherlands.
- ICT Enhancements: Ensure CRM captures complete customer IDs to avoid missing critical retention data.
- BA/ICT Integration: Use JIRA epics to manage retention initiatives, Confluence for documentation, and Power BI dashboards for ongoing executive reporting.

Learnings

This project strengthened my ability to combine technical analysis with BA/ICT practices:

- Developed technical skills in Python, Pandas, and visualization.
- Applied BA frameworks such as hypothesis-driven analysis and stakeholder storytelling.
- Simulated ICT workflows through mock JIRA epics, Confluence-style documentation, and Power BI dashboards.
- Gained experience in translating raw data into business recommendations that align with strategic goals.

Conclusion

The project validated its hypotheses: repeat customers are central to revenue, retention challenges emerge early, and secondary markets represent strategic opportunities. More importantly, it demonstrates my ability to work as a BA/ICT professional by not only analyzing data but also framing findings as business recommendations supported by industry-standard workflows. This positions me strongly for BA and ICT roles requiring both technical and business acumen.

