

Customer Analytics Case Study Report

I. Executive Summary

An online retailer with over 1 million transactions spanning two years sought to understand customer behavior and optimize sales performance. Analysis revealed extreme revenue concentration, with the top 10% of customers generating nearly half of total revenue (\$8.5M of \$17.2M). The business exhibits three distinct seasonal peaks, with the holiday season (August-December) accounting for 53% of annual revenue.

Key findings identified 294 Diamond-tier customers driving 68% of revenue, significant single-item basket opportunities (58.2% of transactions), and high-growth international markets (Netherlands: 15.3% monthly growth, Australia: 267% monthly growth) with different product preferences than the UK core market. Strategic recommendations include implementing a tiered VIP loyalty program (£2.95M investment protecting £11.7M revenue base), pricing strategies to increase basket value by 23-29% (£4-5M annual uplift), and expanding functional product categories for international growth.

II. The Problem

A. Introduction/Background

An online retailer specializing in seasonal and home décor products seeks to improve sales performance and understand its customer base through analysis of over one million transactions spanning December 2009 to December 2011. The dataset contains 1,067,371 transactions across 5,877 customers in multiple countries, primarily UK-based, with international expansion into European and Asia-Pacific markets.

B. Purpose of Study

The primary business objectives are to:

- Increase profitability through customer retention and basket value optimization
- Optimize inventory allocation based on seasonal demand patterns
- Design effective marketing campaigns targeting high-value customer segments
- Identify geographic expansion opportunities
- Develop pricing strategies to maximize revenue per transaction

C. Questions to be Answered

Customer Analytics:

- How can we segment customers based on purchasing behavior?
- Which customers are at risk of churning and how should we prioritize retention efforts?
- Which countries contribute most to sales growth?

Product & Sales Analysis:

- What seasonal purchasing patterns exist and how should inventory be managed?
- How do product preferences differ across geographic markets?

Marketing & Retention:

- Which customers are most likely to respond to promotional campaigns?
- What loyalty program structure would protect high-value customer relationships?

Pricing Strategy:

- How can pricing strategies be adjusted to increase average basket value?

- What bundling and discount opportunities exist?

D. Outline of Remainder of Report

This report details the analytical methodology (Section IV), data preparation processes (Section V), key findings across customer segmentation, churn analysis, promotion targeting, VIP identification, geographic expansion, and pricing optimization (Section VI), followed by strategic recommendations (Section VII).

III. Review of Related Literature

Customer segmentation using RFM (Recency, Frequency, Monetary) analysis is a well-established methodology in retail analytics for identifying valuable customer cohorts. Churn prediction models typically employ recency as a leading indicator of customer disengagement, though seasonal business patterns require threshold adjustments. Market basket analysis techniques identify product bundling opportunities through co-purchase frequency patterns. Geographic expansion strategies balance absolute revenue contribution with growth rate momentum to prioritize resource allocation.

IV. Methodology

A. Identification, Classification and Operationalization of Variables

Dataset Structure:

- Unique Identifiers: Invoice (transaction ID), CustomerID (5,877 unique), StockCode (product ID)
- Numeric Variables: Quantity (units purchased), UnitPrice (£), Revenue (calculated: $\text{Quantity} \times \text{UnitPrice}$)

- Categorical Variables: Description (product name), Country (shipping destination)
- Temporal Variables: InvoiceDate (transaction timestamp)

Derived Variables:

- Recency: Days since last purchase ($\text{analysis_date} - \max(\text{InvoiceDate})$)
- Frequency: Count of distinct invoices per customer
- Monetary: Sum of revenue per customer

B. Statements of Hypotheses/Models

Several analytical models address the business objectives:

Customer Segmentation: K-means clustering ($k=3$) applied to RFM values to identify behavioral customer groups with hypothesis that customers cluster into distinct value tiers.

Churn Risk Analysis: Recency-based classification with seasonal threshold adjustments (180/365/545 days) hypothesizing that inactive periods indicate disengagement risk requiring intervention.

Promotion Response Prediction: Weighted composite scoring model (Recency 50%, Frequency 30%, Monetary 20%) plus segment bonus predicting that customers with optimal recency windows, proven loyalty, and sufficient value respond best to promotional campaigns.

VIP Identification: Percentile-based RFM scoring (Monetary 50%, Frequency 30%, Recency 20%) hypothesizing that top-performing customers warrant tiered premium benefits.

Geographic Growth Analysis: Multi-temporal growth measurement (YoY, QoQ, MoM) identifying markets with sustained expansion momentum.

Pricing Strategy: Market basket analysis hypothesizing that co-purchase patterns reveal bundling opportunities and single-item baskets represent missed cross-sell potential.

C. Sampling Techniques

The analysis used the full dataset of 1,067,371 transactions. Transactions with missing customer IDs were excluded from customer-level segmentation (RFM, churn, promotion targeting, VIP analysis) but retained for basket-level pricing analysis.

D. Data Collection Process

The dataset contains secondary transaction data from an online retailer's point-of-sale system, spanning December 1, 2009 to December 9, 2011 (24 months). Data includes all completed transactions across UK and international markets with order-level details.

E. Modeling Analysis/Techniques Used

- Exploratory Data Analysis (EDA): Revenue trends, seasonal patterns, customer distribution analysis
- K-means Clustering: Applied to standardized RFM values ($k=3$) to segment customers
- Churn Analysis: Recency-based risk categorization validated through sensitivity analysis
- Promotion Propensity Scoring: Weighted RFM model (0-10 base) plus cluster bonus (0-3)
- VIP Scoring: Percentile-ranked RFM components with tier thresholds at 0.70/0.80/0.90/0.95
- Geographic Analysis: Absolute growth, relative growth, and momentum analysis
- Market Basket Analysis: Co-purchase frequency and volume distribution analysis

F. Methodological Assumptions and Limitations

Key assumptions:

- Seasonal business patterns require adjusted churn thresholds (customer inactive 120 days for seasonal products not necessarily churned)
- RFM metrics adequately capture customer value
- K-means clustering with $k=3$ provides actionable segments

Limitations:

- Missing CustomerIDs for subset of transactions
- Dataset spans only 24 months
- No promotional campaign history prevents direct response rate validation
- Case study specificity limits generalizability

V. Data

A. Data Cleaning

- Missing CustomerIDs: Excluded from customer-level analysis
- Duplicate transactions: Removed exact duplicate rows
- Negative quantities: Identified as returns, analyzed separately
- Missing descriptions: Minimal impact, products identifiable via StockCode

B. Data Preprocessing

Transactions with negative or zero monetary values excluded from RFM analysis. Transaction dates converted to multiple time units (Year, Quarter, Month) enabling multi-resolution

temporal analysis. Created derived metrics: Revenue, RFM values per customer, basket-level aggregations, seasonal period classifications.

C. Data Limitations

Primary limitation is absence of customer IDs for all transactions. Dataset lacks promotional campaign history and customer demographic data. Geographic data limited to shipping country without city-level granularity.

VI. Findings (Results)

A. Results Presented in Tables or Charts

TABLE 1: Customer Segmentation Results

Segment	Cluster	Count	Customer_Percent	Avg_Monetary
Wholesale	1	1251	21.3	10468
Casual	2	2359	40.1	393
High-Value	3	2267	38.6	1418

TABLE 2: Churn Risk Distribution

Churn_Risk	Customer_Count	Avg_Frequency	Avg_Monetary
Active	3477	8.8	4290
At Risk	787	3.3	1323

Churn_Risk Customer_Count Avg_Frequency Avg_Monetary

High Risk	1062	2.6	875
Churned	551	1.5	640

TABLE 3: Churn Risk by Customer Segment

Segment Active At Risk High Risk Churned

Wholesale (Cl. 1)	96.9%	1.9%	1.0%	0.2%
High-Value (Cl. 3)	80.2%	12.0%	7.1%	0.8%
Casual (Cl. 2)	18.9%	20.9%	37.7%	22.5%

TABLE 4: Promotion Response Prediction Results

Priority Tier Customers % of Base Avg Score Avg Monetary Total Revenue

Critical VIP	39	0.7%	11.9	\$15,787	\$616K
High Priority	858	14.6%	9.3	\$1,227	\$1.05M
Medium Prior.	1,319	22.4%	8.0	\$253	\$334K
Low Priority	1,695	28.8%	5.6	\$183	\$310K
Do Not Target	1,966	33.5%	3.6	\$41	\$81K

TABLE 5: VIP Tier Distribution

Tier Customers % of Base Avg Spend Avg Frequency Total Revenue Revenue %

Diamond	294	5.0%	\$39,948	53	\$11.7M	68.0%
Platinum	294	5.0%	\$10,657	22	\$3.1M	18.0%
Gold	567	9.6%	\$4,213	9	\$2.4M	14.0%

Tier	Customers	% of Base	Avg Spend	Avg Frequency	Total Revenue	Revenue %
Silver	577	9.8%	\$3,408	7	\$2.0M	11.6%
Regular	4,145	70.5%	\$732	6	\$3.4M	19.7%

TABLE 6: Geographic Sales Growth - Top Markets

Country	Total Revenue	Growth Amount	Growth Rate	Avg Monthly Growth	Latest Revenue
UK	\$17.1M (85.3%)	+\$7.3M	517%	3.5%	\$571K/mo
Netherlands	\$550K (2.7%)	+\$260K	816%	15.3%	\$12K/mo
EIRE	\$659K (3.3%)	+\$253K	847%	20.4%	\$7K/mo
Germany	\$386K (1.9%)	+\$186K	998%	—	\$7K/mo
France	\$326K (1.6%)	+\$180K	1,115%	17.2%	\$7K/mo
Australia	\$169K (0.8%)	—	5,932%	267%	—

TABLE 7: Basket Value Distribution

Basket Segment	Baskets	% of Baskets	Revenue	% of Revenue
Over £1,000	2,621	7.1%	\$6.6M	38.2%
£250–£500	12,963	35.2%	\$4.6M	26.5%
£100–£250	10,905	29.6%	\$1.8M	10.7%
£50–£100	2,097	5.7%	\$160K	0.9%
Under £50	2,375	6.4%	\$54K	0.3%

Current Metrics: Avg Basket = £467.68 (mean), £302.46 (median) | 58.2% single-item baskets

B. Results Reported with Respect to Hypotheses/Models

CUSTOMER ANALYTICS

Customer Segmentation: K-means clustering identified three distinct segments validating the hypothesis of tiered customer value. Wholesale segment (21.3% of customers) generates 76% of revenue with average \$10,468 lifetime value and 18.9 purchase frequency, indicating B2B wholesale/retail accounts. High-Value segment (38.6%) contributes 18.6% revenue with moderate engagement. Casual segment (40.1%) shows minimal engagement contributing only 5.4% revenue. This concentration exceeds traditional 80/20 Pareto principle.

Churn Analysis: Recency-based classification successfully identified current at-risk customers. Churn concentrates in low-value Casual segment (81% at-risk/churned) contributing only 5.4% revenue, while high-value segments demonstrate strong retention (Wholesale: 97% active, High-Value: 80% active). Key finding: 572 previously valuable customers (\$1.84M historical revenue) currently at-risk require immediate intervention. Sensitivity analysis confirmed threshold robustness.

Geographic Analysis: UK remains core market contributing 85.3% revenue and 84.2% of growth. Netherlands emerges as international success with highest sustained monthly growth (15.3%). Australia shows explosive percentage growth (5,932%) but from smaller base. Product analysis reveals high-growth international markets prioritize functional items unlike UK's decorative/seasonal focus.

PRODUCT & SALES ANALYSIS

Seasonal Patterns: Sales concentrate in three distinct peaks. February-March (13% revenue) driven by Valentine's and Easter preparation. April-May (13%) dominated by outdoor/garden

season. August-December (53%) heavily Christmas-focused with extreme product seasonality—some items showing 95-100% sales concentration in this period.

Top Products: Product preferences vary significantly by market. UK top products: decorative/seasonal. Netherlands: storage/organization dominate. Australia: functional items lead. Universal products sell across markets but at vastly different scales.

MARKETING & RETENTION

Promotion Prediction: Weighted scoring model identified 897 customers (15.3%) as high-response targets representing £1.67M potential recovery, while flagging 1,966 customers (33.5%) as unlikely responders. Critical VIP tier (39 customers) shows exceptional characteristics requiring immediate personal intervention. Model validation: high scorers average 20+ purchases while low scorers average 1.6 purchases.

VIP Loyalty Programs: Top 5% (Diamond tier, 294 customers) generate 68% of total revenue. Program investment (£2.95M annually) functions as business continuity insurance protecting £11.7M concentrated revenue base. Expected returns yield 12% direct ROI, but primary value is risk mitigation.

PRICING STRATEGY

Basket Value: Current performance shows average £467.68 but median £302.46 indicating right-skewed distribution. Critical finding: 58.2% of baskets contain only one item. Basket distribution reveals 7.1% of baskets (over £1,000) generate 38.2% revenue.

Pricing Initiatives: Market basket analysis identified multiple opportunities. Co-purchase analysis revealed natural product pairs bought together 700-1,200+ times enabling bundling strategies. Volume analysis identified products predominantly sold as single units presenting tiered volume discount opportunities. Price point analysis shows £1-£5 items drive 66.3% of revenue. Conservative implementation projects 23-29% revenue growth (£4-5M annually).

VII. Conclusions and Recommendations

A. Discuss Alternative Methodologies

The RFM-based segmentation was selected for interpretability and actionability, though alternatives exist. Hierarchical clustering could validate optimal cluster count through dendrogram analysis. Predictive churn modeling using logistic regression could forecast future churn probability rather than identifying current at-risk status. Market basket analysis using association rules (Apriori algorithm) could identify complex multi-item patterns beyond pairwise co-purchases.

B. Recommendations

CUSTOMER SEGMENTATION & RETENTION

Critical VIPs: Immediately implement personal intervention for 39 at-risk Wholesale customers. Each represents significant business relationship (average £15,787 value) requiring CEO/executive phone calls within 30 days. These 39 customers represent £616K at risk.

High-Value Customers: Launch targeted automated promotional campaigns for 431 at-risk High-Value customers within 60 days. Utilize personalized email sequences with 15-20% discount offers. Expected 20-25% response rate would recover 86-108 customers worth £122K-£153K annually.

Casual Customers: Avoid significant investment in winning back 1,912 at-risk Casual customers. With average £393 lifetime value, ROI on intensive retention efforts likely negative.

Loyalty Program: Implement tiered VIP loyalty program with Diamond tier receiving dedicated account managers, 2% quarterly cashback, and priority services. Program investment (£2.95M annually) protects £11.7M revenue base. Phase 1 (Months 1-2): Diamond tier launch. Phase 2 (Months 3-4): Platinum/Gold rollout. Phase 3 (Months 5-6): Silver tier.

GEOGRAPHIC EXPANSION

UK Investment: Maintain 85% resource allocation to UK market. UK contributes 84.2% of growth with £571K monthly run rate demonstrating continued expansion potential.

Netherlands Blueprint: Conduct deep-dive analysis of Netherlands success factors within 90 days. Netherlands combines meaningful absolute contribution (\$260K growth) and highest sustained monthly growth (15.3%). Replicate learnings to Belgium, Luxembourg, Scandinavia.

Product Strategy for International Markets: Expand functional product categories (storage/organization, children's items, night lights) demonstrating success in high-growth markets. Phase 1: Netherlands optimization (£200K investment, £110K expected return). Phase 2: Australia expansion (£175K investment, £130K expected return). Phase 3: UK functional category test.

PRICING AND SALES OPTIMIZATION

Free Shipping Threshold: Implement £75 UK threshold immediately (Month 1). Expected 30% conversion of £50-£75 baskets generates incremental revenue. International threshold: £150. Expected impact: £500K annual.

Product Bundling: Create pre-made bundles for top co-purchased items (Months 1-2). Implement dynamic "Frequently Bought Together" recommendations. Expected impact: Convert 20% of single-item baskets = £722K incremental. Total: £2.1M annual.

Volume Discounts: Implement tiered volume pricing for high single-unit products (Months 3-6). Example: Building Block Words—tiered pricing structure. Expected impact: £800K incremental.

Segment-Specific Strategies: Develop differentiated approaches by customer segment (Months 3-6). Wholesale: Minimum order value £300. Casual: Free shipping progress bar. High-Value: Bundle recommendations. Expected impact: £4M incremental.

Add-On Optimization: Create checkout "Impulse Zone" featuring sub-£5 items (Months 6-12). Expected impact: £750K incremental.

Implementation Priority: Phase 1 Quick Wins (Months 1-2): Free shipping + Bundling = £2.6M potential. Phase 2 Medium Term (Months 3-6): Volume discounts + Segment strategies = £4.8M potential. Phase 3 Advanced (Months 6-12): Add-on optimization = £750K potential. Total conservative projection: £4-5M annual (23-29% growth).

Success Metrics: Track weekly—average basket value (target: £302 → £400), single-item basket percentage (target: 58.2% → 40%). Track monthly—Diamond retention (target: 98%+), bundle attachment rate (target: 25%). Track quarterly—revenue per segment, cumulative revenue impact.