UK RETAIL E-COMMERCE ANALYTICS - PROJECT DOCUMENTATION

Business Context

Analyzed a UK-based online retail company's transactional data spanning December 2010 to December 2011 to uncover customer behavior patterns, product performance insights, and revenue optimization opportunities. This project demonstrates a complete data analytics lifecycle from normalized database design through advanced SQL analysis to interactive Power BI visualization.

Objectives

1. Identify high-value customer segments and churn risks
2. Analyze product performance and category contribution
3. Discover revenue patterns and seasonal variations
4. Segment customers using RFM methodology
5. Perform cohort retention and market basket analysis
6. Build interactive 5-page Power BI dashboard for stakeholder decision-making
7. Provide data-driven recommendations for business growth

Dataset Information

Source

- **Period:** December 2010 - December 2011 (13 months)
- **Format:** 4 normalized CSV files
- **Total Orders/Invoices** | 18,536
- **Total Line Items** | 397,924

Key Metrics

Metric	Value
Total Revenue	£8,911,407
Total Orders	18,536
Total Customers	4,339
Average Order Value	£480
Unique Products	3,665+
Countries Served	38

Data Structure

Normalized 4-Table Schema:

1. **customers** (4,339 rows)
    - customer_id (PK)
    - country
    - first_purchase_date
2. **invoices** (18,536 rows)
    - invoice_no (PK)
    - customer_id (FK)
    - invoice_date
    - country
    - total_amount
3. **products** (3,665+ rows)
    - stock_code (PK)
    - product_name
    - category
4. **invoice_items** (397,924 rows)
    - line_item_id (PK)
    - invoice_no (FK)
    - stock_code (FK)
    - quantity
    - unit_price

Relationships:

```
customers (1) ←→ (N) invoices
invoices (1) ←→ (N) invoice_items
products (1) ←→ (N) invoice_items
```

🗄️ Database Implementation

```
CREATE TABLE customers (
    customer_id INTEGER PRIMARY KEY,
    country VARCHAR(100),
    first_purchase_date TIMESTAMP
);

CREATE TABLE invoices (
    invoice_no VARCHAR(50) PRIMARY KEY,
    customer_id INTEGER REFERENCES customers(customer_id),
    invoice_date TIMESTAMP,
    country VARCHAR(100),
    total_amount DECIMAL(10,2)
);
```

```
CREATE TABLE products (
    stock_code VARCHAR(50) PRIMARY KEY,
    product_name VARCHAR(255),
    category VARCHAR(100)
);

CREATE TABLE invoice_items (
    line_item_id INTEGER PRIMARY KEY,
    invoice_no VARCHAR(50) REFERENCES invoices(invoice_no),
    stock_code VARCHAR(50),
    quantity INTEGER,
    unit_price DECIMAL(10,2)
);
```

🔍 SQL Analysis - 20 Advanced Queries

Query Categories

1. Business Overview (3 queries)

- KPI Scorecard (revenue, orders, customers, AOV)
- Monthly Revenue Trend
- Top 10 Products by Revenue & Quantity

2. Product Analysis (5 queries)

- Category Performance
- Best & Worst Performing Products
- Price Range by Category
- Product Worst Sellers Analysis
- Average Order Composition

3. Customer & Geographic Analysis (3 queries)

- Sales by Country
- Top 10 Customers by Lifetime Spend
- New Customers Acquisition Trend

4. Advanced Segmentation (3 queries)

- Top Category per Country (with ROW_NUMBER ranking)
- Monthly Sales Comparison (highest/lowest with window functions)
- Product Performance by Customer Segment

5. Strategic Analytics (6 queries)

- **RFM Customer Segmentation** (NTILE-based scoring: Champions, Loyal, Regular, At-Risk, Churned, New)
- **Churn Analysis** (90+ day inactivity detection)
- **Churn Summary Statistics** (customer count, revenue by status)

- **Customer Lifetime Value** (CLV calculation with value/recency/frequency scoring)
- **CLV Summary Statistics** (VIP, High Value, Medium, Low segments)
- **Market Basket Analysis** (product co-purchase with confidence & lift metrics)
- **Cohort Retention Analysis** (monthly cohorts × 6-month retention windows)
- **Repeat Purchase Rate** (overall & by country)

SQL Techniques Demonstrated

- Multi-table JOINs (2, 3, and 4-table operations)
- Common Table Expressions (CTEs) for complex business logic
- Window Functions (NTILE, ROW_NUMBER, PARTITION BY)
- Date Functions (EXTRACT, DATE_TRUNC, date arithmetic)
- Subqueries and derived tables
- CASE statements for conditional segmentation
- Aggregate functions (SUM, AVG, COUNT, MIN, MAX)
- HAVING clauses for filtered aggregations
- PERCENTILE functions for statistical analysis

📊 Power BI Dashboard - 5 Pages

Page 1: Executive Overview

Purpose: High-level KPIs for leadership

Key Metrics Displayed:

- Total Revenue: £8,911,407
- Total Orders: 18,536
- Total Customers: 4,339
- Average Order Value: £480

Visuals: Revenue trend, category breakdown, product rankings

Page 2: Customer Intelligence Report

Purpose: Customer segmentation and retention analysis

Customer Status Distribution:

- **Active:** 2,397 customers (55.24% of base, £7.36M revenue)
- **Needs Attention:** 493 customers (11.36%)
- **At Risk:** 589 customers (13.57%)
- **Churned:** 860 customers (19.82% - **CRITICAL**)

CLV Segmentation:

- **VIP:** 117 customers (2.7%) → 37.57% of total revenue (£3.35M)
- **High Value:** 270 customers (6.22%) → 17.63% of revenue
- **Medium Value:** 1,584 customers (36.51%) → 28.76% of revenue
- **Low Value:** 2,368 customers (54.57%) → 16.03% of revenue

RFM Segments: Champions, Loyal Customers, Regular, At-Risk, Churned, New Customers

---

Page 3: Product Performance Analysis

Purpose: Product mix optimization and category insights

Top Products & Categories:

- #1 Product: Paper Craft, Little Birdie (£168,469)
- **Top Category:** General Merchandise (£3,163,097 - 35.5% of revenue)
- **Top 10 Products:** Drive significant revenue concentration

Key Metrics:

- Unique Products: 3,665+
- Average Product Value: £2,431
- Total Quantity Sold: 604,893 units

Page 5: Sales Analytics

Purpose: Geographic and temporal patterns

Geographic Insights:

- **UK Dominance:** £7,308,391 (82% of total revenue)
- **Market Concentration Risk:** Over-reliance on single market
- **International Opportunity:** 38 countries served, secondary markets show growth potential

Temporal Insights:

- **Peak Month:** November 2011
- **Lowest Month:** April 2011
- **Seasonality:** Clear Q4 spike visible

Average Order Composition:

- Avg Items per Order: 162.57 units
- Avg Unique Products: 18.67 per order
- Median Items: 136

Page 4: Advanced Insights (RFM & Retention)

Purpose: Customer segmentation strategy and retention patterns

RFM Analysis Results:

- Champions: High recency, frequency, and monetary value
- Loyal Customers: Strong purchase history
- At-Risk: Low recency despite historical activity
- Churned: Inactive 90+ days

Cohort Retention Analysis:

- **December 2010 Cohort:** 36.6% (M1) → 39.8% (M5)
- **January 2011 Cohort:** 37.9% (M1) → 28.8% (M5)
- Clear retention decline after month 2 for most cohorts

Market Basket Analysis - Top Product Pairs

Product A	Product B	Co-Purchases	Confidence %	Lift
JUMBO BAG RED RETROSPOT	JUMBO BAG PINK POLKADOT	546	34.13	7.26
ROSES REGENCY TEACUP	GREEN REGENCY TEACUP	541	69.09	18.53
ALARM CLOCK BAKELIKE RED	ALARM CLOCK BAKELIKE GREEN	530	60.43	14.19
LUNCH BAG PINK POLKADOT	LUNCH BAG RED RETROSPOT	523	56.18	8.08
LUNCH BAG SUKI DESIGN	LUNCH BAG RED RETROSPOT	519	49.76	7.15

Insight: Top product pair: JUMBO BAG RED RETROSPOT + JUMBO BAG PINK POLKADOT with 546 co-purchases (7.26x lift)

Key Findings & Business Insights

Customer Insights

1. **Severe Churn Crisis**
    - 19.82% of customer base churned (860 customers)
    - Additional 13.57% at risk (589 customers)
    - Average days to churn: 91.5 days
    - **Recommendation:** Deploy 60-day re-engagement campaign immediately
2. **Revenue Concentration Risk**
    - VIP segment (2.7%) generates 37.57% of revenue (£3.35M)
    - Top 20% of customers likely drive 60%+ of revenue
    - **Recommendation:** Implement VIP retention program with white-glove service
3. **Repeat Purchase Success**
    - 65.58% of customers are repeat buyers
    - 7.77% are highly loyal (3+ purchases)
    - **Recommendation:** Leverage this strength; expand loyalty incentives

---

Product Insights

1. **Category Concentration**
    - General Merchandise dominates: £3.16M (35.5% of revenue)
    - Significant drop to secondary categories
    - **Recommendation:** Diversify product mix; reduce SKU dependency
2. **Seasonality Effect**
    - November 2011 peak vs April 2011 trough shows 2.6x variance
    - Clear Q4 holiday shopping pattern
    - **Recommendation:** Optimize inventory and marketing spend for Q4
3. **Market Basket Opportunities**
    - JUMBO BAGS: 546 co-purchases (7.26x lift)
    - TEACUPS: 541 co-purchases (18.53x lift)
    - **Recommendation:** Create themed product bundles and gift sets

---

Geographic Insights

1. **UK Market Over-Reliance**
    - 82% of revenue from UK (£7.31M)
    - 18% from 37 other countries
    - **Recommendation:** Develop international expansion strategy
2. **Order Composition**
    - Average order contains 162.57 items (HIGH - suggests bulk/B2B)
    - 18.67 unique products per order
    - **Recommendation:** Leverage bulk ordering; create tiered pricing

---

Business Recommendations

1. **Churn Prevention Campaign**
    - Target 860 churned + 589 at-risk customers
    - Estimated revenue at risk: £556K+
    - Deploy personalized win-back emails at 60, 75, 90 days inactive
2. **VIP Retention Program**
    - Create exclusive tier for 117 VIP customers
    - Expected ROI: 15-20% retention improvement
3. **Market Basket Implementation**
    - Bundle JUMBO BAGS + complementary products
    - Add "Frequently Bought Together" recommendations

- Target: 10% AOV increase

4. **Inventory Optimization**
    - Stock top 50 products heavily for Q4
    - Reduce slow-moving SKUs (bottom 30%)
5. **Cohort-Based Retention**
    - Analyze why Month 2 retention drops sharply
    - Implement email nurture sequence at day 30
6. **Geographic Expansion Pilot**
    - Localize offerings for top 5 secondary markets
    - Target: 5% international revenue growth
7. **Predictive Analytics**
    - Build churn prediction model using CLV + RFM signals
    - Implement automated intervention workflows

Technologies & Tools

Component	Technology
Database	PostgreSQL (SQL)
BI Tool	Power BI Desktop
Visualizations	5 pages, 40+ visuals
Dashboard Features	Interactive slicers, cross-filtering, drill-through

Data Pipeline:

```
Raw CSV Files → PostgreSQL Database → 20 SQL Queries → Power BI Import → 5-Page Dashboard
```

Project Impact

Analytical Achievements

- 20 advanced SQL queries across 5 analytical categories
- 5 comprehensive Power BI pages with 40+ visuals
- RFM segmentation of 4,339 customers
- Cohort retention analysis (13 cohorts × 6 months)
- Market basket analysis (546+ co-purchase pairs)
- Churn identification (1,449 at-risk customers)

🎓 Skills Demonstrated

Technical Proficiency

- Database Design (normalized 3NF schema)
- Advanced SQL (CTEs, window functions, multi-table joins)
- Statistical Analysis (RFM segmentation, cohort analysis)
- Business Intelligence (Power BI dashboard design)
- Data Visualization (40+ interactive visuals)

Business Acumen

- Customer Segmentation & Lifecycle Analysis
- Churn Risk Assessment & Mitigation
- Product Performance & Category Analysis
- Geographic & Temporal Pattern Recognition
- Strategic Business Recommendations with ROI projections

📁 Repository Structure

```
uk-retail-analytics/
├── data/
│   ├── customers.csv (4,339 rows)
│   ├── products.csv (3,665+ rows)
│   ├── invoices.csv (18,536 rows)
│   └── invoice_items.csv (397,924 rows)
├── sql/
│   └── uk_retail_queries.sql (20 queries)
├── dashboards/
│   └── uk_retail.pbix (5 pages)
│   └── uk_retail.pdf
├── documentation/
│   └── project_documentation.pdf
└── readme.md
```

Project Highlights

> **"Transformed 397K+ retail transactions from a 4-table normalized database into enterprise-grade business intelligence through 20 advanced SQL queries and a 5-page interactive Power BI dashboard. Identified £3.35M revenue concentration in VIP segment, uncovered 19.82% churn crisis affecting £556K, and discovered 7.26x product affinity in JUMBO BAG category enabling strategic bundling opportunities."**

Analysis Performed: 13 months of UK retail data | **Customers Analyzed:** 4,339 | **Transactions:** 18,536 | **Total Revenue Tracked:** £8.91M | **Strategic Recommendations:** 8 actionable initiatives

---

About This Project

- **Author:** S. M. Sharifuzzaman
- **Email:** smsharifuzzaman113@gmail.com
- **LinkedIn:** https://www.linkedin.com/in/smsharif/
- **GitHub:** https://github.com/sm-sharif

- **Project Date:** November 2025
- **Tools Used:** PostgreSQL, Power BI, Python
