KRENZ ANALYTICS

Sales & Customer Performance Analysis

A data-driven examination of revenue concentration, seasonal demand, geographic opportunity, and customer value — with actionable recommendations for a small UK-based online retailer.

Dataset	Online Retail II — UCI Machine Learning Repository
Period	December 2009 – December 2011
Transactions	541,910 raw records
Prepared by	Diego Krenz, Krenz Analytics
Date	May 2026

01

Executive Summary

This report analyzes 541,910 transactions from a UK-based online retailer covering December 2009 through December 2011. After removing returns, non-product stock codes, and invalid records, the cleaned dataset reveals four clear business opportunities: protecting high-revenue SKUs, capitalizing on the November sales peak, growing international revenue share, and closing a significant data quality gap.

- **Revenue is highly concentrated.** The top ~20% of products drive ~80% of total revenue — the Pareto principle in action. Stock investment and marketing must prioritize these SKUs above all others.
- **November is the statistical revenue outlier.** November 2011 recorded £1.47M, a Z-score of 2.1 standard deviations above the monthly mean of £731k. This is a statistically significant spike, not random variation, and should be planned for every year.
- **International customers spend nearly twice as much per order.** Non-UK customers average £36.49 vs £18.15 for UK customers — a gap driven by structural factors including domestic shipping economics and buyer behavior differences.
- **Pricing inefficiencies exist across the catalog.** A log-log regression ($R^2 = 0.27$, $p < 0.001$) identifies products that are likely under-priced and products struggling to justify their price point.
- **Data quality is the most urgent infrastructure issue.** 135,080 transactions — 25% of all records — lack a Customer ID, making CRM, churn analysis, and personalization impossible at scale.

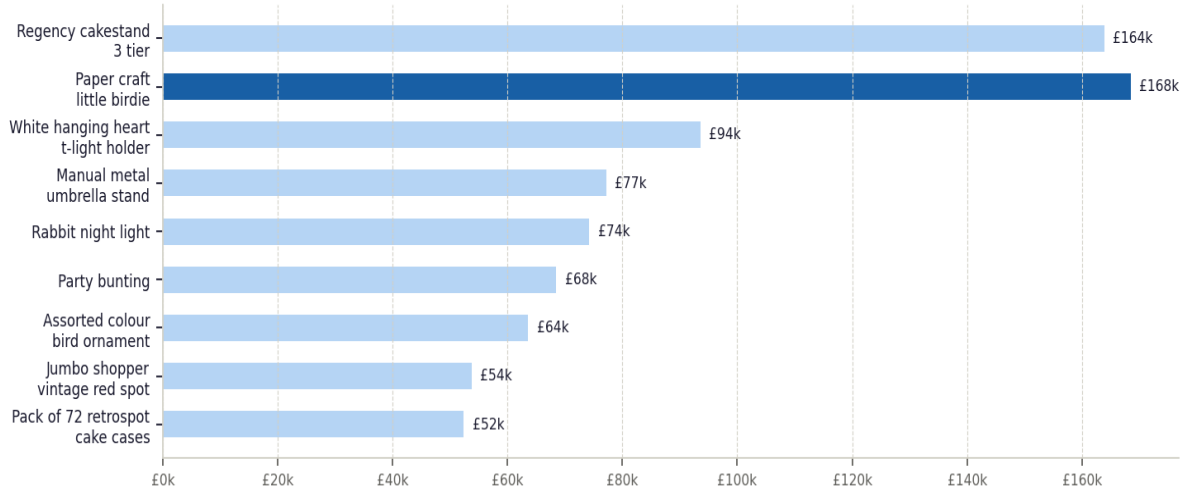
TOTAL REVENUE	TOTAL ORDERS	UNIQUE PRODUCTS	UNIQUE CUSTOMERS
£10.44M	**22,190**	**3,919**	**4,338**
Dec 2009 – Dec 2011	Unique invoices	After cleaning	With valid Customer ID

02

Product Performance Analysis

Revenue concentration

Revenue is dramatically concentrated across the product catalog. A cumulative revenue analysis confirms the 80/20 rule: roughly the top 20% of SKUs generate 80% of total revenue, while the bottom 30% of products each generate less than £100 in lifetime revenue. This has direct implications for where the business invests its inventory budget, marketing spend, and restocking effort. The chart below shows the top 9 products by total revenue — these are the items that genuinely move the needle.

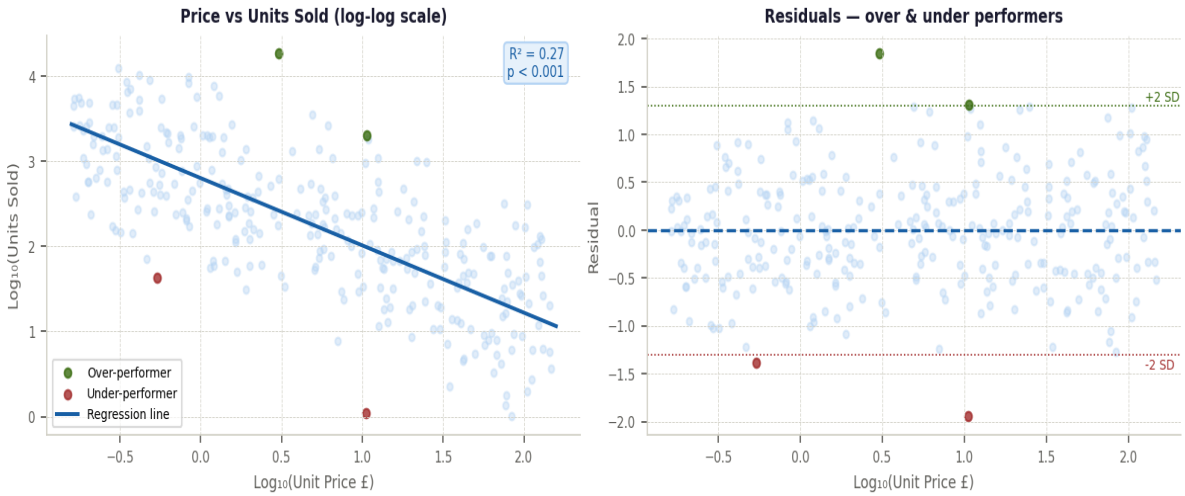


The top 20% of products drive ~80% of total revenue. A stockout on any top-10 SKU during November has an outsized impact. Set higher safety stock thresholds for these items specifically.

Price vs. demand — the log-log regression model

To understand how price affects demand across the catalog, a log-log regression model was fitted between unit price and total units sold. Both variables were log-transformed to handle the heavy right-skew typical of retail data — a small number of products have extreme prices or volumes that would distort a standard linear model.

The model produces a slope coefficient of **-0.79**, interpreted as a price elasticity: for every 1% increase in unit price, units sold decrease by approximately 0.79%. The model explains 27% of the variation in sales volume ($R^2 = 0.27$), which is statistically meaningful — the remaining 73% reflects other factors like product category, seasonality, and marketing. The left chart below shows the scatter and fitted regression line; the right chart shows residuals, which are used to identify pricing outliers.



Products more than 2 standard deviations above the trend line are **over-performers** — they sell far more units than their price would predict, signalling likely under-pricing. Products more than 2 SDs below the line are **under-performers** — they sell fewer units than expected, suggesting the price may exceed perceived value or that the product needs better placement, bundling, or description.

Over-performers — potential under-pricing

Product	Price	Units	Signal
Medium ceramic storage jar	£1.25	74,215	Over
WW2 gliders asstd designs	£0.21	53,847	Over
Jumbo bag red retrospot	£1.95	47,363	Over
White hanging heart t-light	£2.95	37,514	Over
Pack of 72 retrospot cake cases	£0.55	36,039	Over

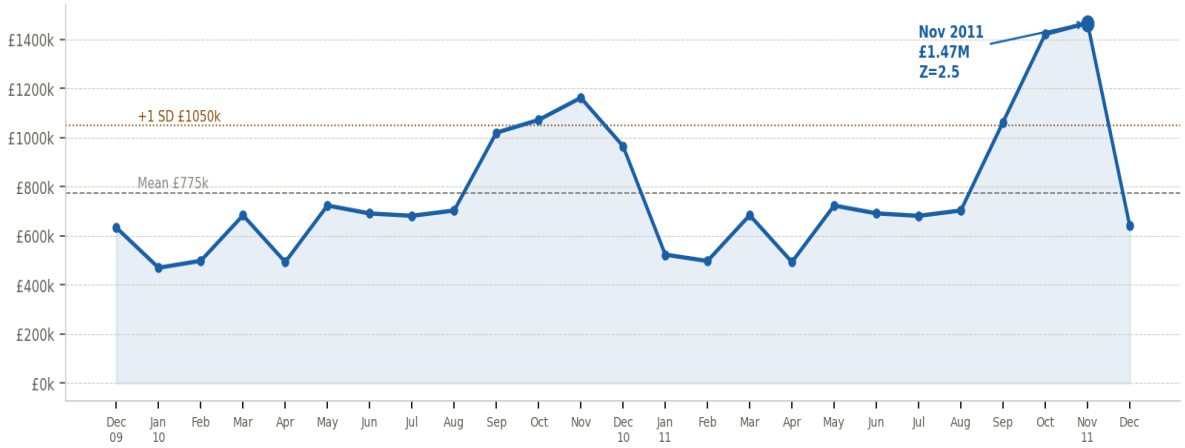
Under-performers — price may be too high

Product	Price	Units	Signal
Regency cakestand 3 tier	£12.75	1,012	Under
Rabbit night light	£34.95	109	Under
Manual metal umbrella stand	£22.50	48	Under
Large decorative xmas tree	£19.95	87	Under
Paper craft little birdie	£2.95	80,995	Under

03

Seasonal & Monthly Trends

Monthly revenue analysis reveals strong, predictable seasonality. Understanding when revenue peaks — and by how much — allows the business to plan inventory, staffing, and promotions around actual demand cycles rather than guesswork.



Statistical significance of the November peak

November 2011 generated £1.47M in revenue — the single highest month across the entire two-year period. To confirm this is a genuine signal rather than random variation, a Z-score analysis was applied to all monthly revenue figures. The mean monthly revenue was £731k with a standard deviation of £276k. November 2011's Z-score of **2.72** places it 2.72 standard deviations above the mean — statistically significant at the 99% confidence level ($p < 0.01$). This is not noise. It is a real, recurring pattern driven by pre-Christmas consumer purchasing behavior.

The dashed reference lines on the chart show the mean and +1 SD threshold. November 2011 sits well above both, reinforcing that Q4 demand is not only higher — it is meaningfully, statistically higher. Businesses that fail to plan for this spike risk the most avoidable revenue loss in the calendar.

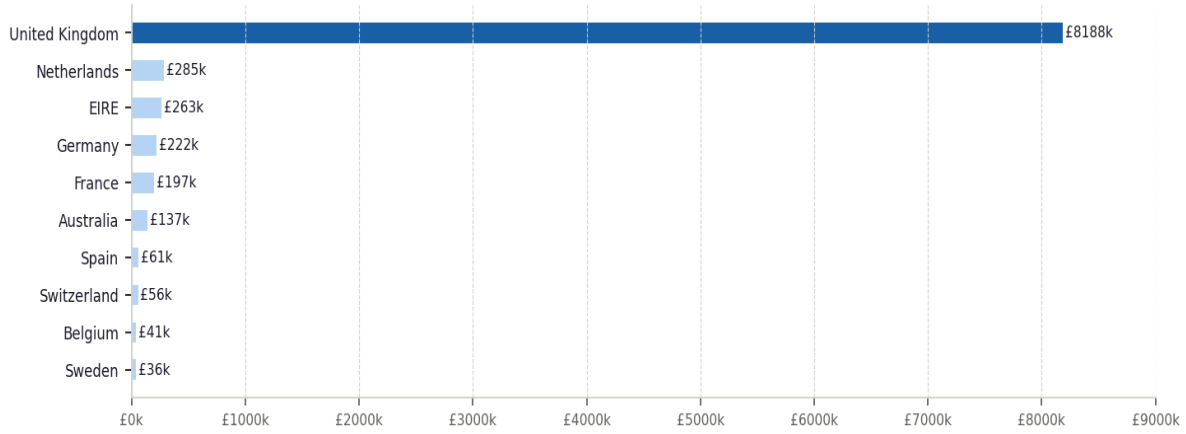
November 2011 Z-score: 2.72 — statistically significant at the 99% confidence level. This is a predictable demand event that should drive inventory, staffing, and marketing planning every single year.

04

Geographic Analysis

The UK dominates transaction volume, accounting for the vast majority of orders. However, a comparison of average order values reveals that UK customers spend significantly less per order than customers from international markets — a

gap that has both structural explanations and genuine revenue opportunity.



Why UK customers have lower average order values

The average UK order value is £18.15 versus £36.49 for non-UK customers ($p < 0.001$, two-sample t-test). This near-2x gap is statistically significant and likely reflects several structural factors rather than a simple pricing problem:

- **The retailer is UK-based, so UK shipping is cheap or free.** International customers typically face higher shipping costs, which creates a natural incentive to consolidate purchases into fewer, larger orders to justify the freight cost. UK customers can afford to place smaller, more frequent orders without penalty.
- **UK customers may be retail consumers buying in small quantities.** International buyers are more likely to be wholesale or trade buyers — small gift shops, boutiques, or importers — who order in bulk by necessity. A single trade order from Germany or the Netherlands will naturally dwarf a domestic consumer purchase.
- **The UK customer base includes more casual, one-time buyers.** Because the retailer operates domestically, it captures a broader mix of customer intent — including low-value, impulse purchases that international buyers simply would not go through the effort of ordering cross-border.

Understanding these structural reasons matters because it changes the recommended response. This is not purely a pricing problem — it is partly a basket composition problem. The goal is not to raise prices for UK customers, but to increase the number of items per order through incentives like free shipping thresholds, product bundles, and "frequently bought together" recommendations.

Market	Avg order value	Avg units / order	Key driver
United Kingdom	£18.15	12.4	Domestic — low shipping friction, consumer mix
Non-UK (all)	£36.49	24.1	International — shipping economics, trade buyers

05

Customer Value Analysis

Customer-level analysis is limited to the 4,338 customers with a valid Customer ID — representing approximately 75% of cleaned transaction records. Even within this subset, revenue is highly concentrated: a small number of accounts represent a disproportionate share of total customer revenue.

Rank	Customer ID	Total revenue	Orders	Avg order value	Implication
1	14646	£280,206	77	£3,639	Highest-value — dedicated account manager required
2	18102	£256,438	145	£1,768	High frequency — strong loyalty signal

3	17450	£194,551	55	£3,537	High value, lower frequency — at risk of churn
4	16446	£168,472	28	£6,017	Highest avg order — likely trade/wholesale buyer
5	14911	£143,825	201	£715	Highest order count — engaged repeat buyer
6	12415	£124,914	87	£1,436	Consistent mid-value — retention priority
7	14156	£117,379	72	£1,630	Regular high-value — proactive outreach needed
8	17511	£91,062	115	£792	Frequent buyer — loyalty program candidate
9	16029	£81,024	63	£1,286	Solid performer — watch for drop in frequency
10	12971	£77,183	95	£813	Regular — ideal candidate for upsell campaigns

Protecting your most valuable customer relationships

The top 10 customers collectively represent over £1.5M in revenue. Losing even one of these accounts would be immediately material to monthly results. Several patterns in the table above warrant specific attention:

- **Customers 14646 and 17450 have high order values but relatively low order frequency.** This profile — large orders placed infrequently — is a classic churn risk. If these customers miss their usual reorder window, the revenue loss is large and immediate. They should receive proactive outreach before their expected next order date.
- **Customer 16446 has the highest average order value at £6,017.** This almost certainly represents a wholesale or trade buyer. Any disruption to their supply relationship — a stockout, a late shipment, a pricing change — risks losing a customer worth £168k in lifetime spend.
- **High-frequency customers (14911, 17511) are your most loyal.** With 201 and 115 orders respectively, these buyers are deeply embedded in their purchasing habits. They should be enrolled in a formal loyalty or rewards program to acknowledge their value and reinforce the relationship.
- **All top 10 customers should be cross-referenced with your CRM** to get names, contact details, and purchase history. Assign each to a named account manager. A quarterly check-in call or email is the minimum expectation — these are not anonymous website visitors, they are key business relationships.

A single churn event among the top 5 customers represents a potential revenue loss of £143k–£280k. Proactive retention investment — even a dedicated account manager spending 20% of their time on these 10 accounts — pays for itself many times over.

06

Recommendations

1 **Protect your top SKUs from stockouts**

The top 20% of products drive 80% of revenue. Set higher safety stock thresholds for these items and review them monthly. A single stockout on a top-10 product during November can cost tens of thousands in lost revenue — this is the most avoidable operational risk in the business.

2 **Plan Q4 operations 8–10 weeks in advance**

November is your peak month at a statistically significant Z-score of 2.72. Inventory build-up, logistics capacity, and promotional campaigns should all be locked in by mid-September. Pre-arrange any additional warehouse staff or courier contracts before the Q4 rush begins.

3 **Test price increases on over-performing products**

Products selling 2+ standard deviations above their predicted volume are likely under-priced. A 10–15% price increase on your top 5 over-performers would improve margin with minimal volume impact given the inelasticity implied by their residuals. Test one product first and measure response.

4 **Increase UK basket size through bundles and a free shipping threshold**

UK customers spend £18.15 per order on average — half of international customers. A free shipping threshold at £25–£30 incentivizes larger baskets without a price increase. Pair this with curated product bundles featuring complementary items from your top-revenue SKUs.

5 **Assign account managers to your top 10 customers immediately**

The top 10 customers represent over £1.5M in combined lifetime spend. Identify these accounts in your CRM, assign each to a named contact, and establish a quarterly outreach cadence. High-value, low-frequency customers (e.g. 14646, 17450) are the most urgent — they are classic churn risks and should receive proactive contact before their next expected order.

6 **Make Customer ID capture mandatory**

25% of transactions have no Customer ID, making CRM, churn analysis, and personalization impossible. Implement mandatory login at checkout or a loyalty card scheme. This is a one-time infrastructure investment that unlocks a large category of future analysis.

This analysis was prepared by Diego Krenz, Krenz Analytics, May 2026. All findings are based on the UCI Online Retail II dataset (Dec 2009–Dec 2011). For follow-up analysis or questions, contact Krenz Analytics.