Behavioral Analytics and Product Strategy for a E-Retailer

BIT- 5905

GROUP 4

By Rhea Nair, Samiksha Sarda, Sana Sawant, Vedika Khandelwal

Business Understanding

In today’s rapidly evolving digital commerce environment, how can transactional data be leveraged to understand customer behavior and lifecycle trends, and transform these insights into a sustainable competitive advantage for an online wholesale retailer?

The objective of this project is to build a comprehensive customer intelligence study that not only describes current business performance but also diagnoses the underlying causes, prescribes actionable strategies, and predicts future outcomes. The company specializes in selling unique, all-occasion giftware to wholesale buyers across different countries. Key strategic questions include:

- What products should be sold to which countries, and during which time periods, to maximize sales?
- How do customer behaviors vary by region, season and product type?
- Which products are often purchased together, how can this support effective bundling strategies?
- How can we better understand and serve different customer segments based on their value contribution?

Finally, the insights generated from analysis can support decision-making in several key areas:

Customer Segmentation and Marketing: We used RFM analysis (Recency, Frequency, and Monetary) to identify customer segments such as “Champions” and “At Risk.” This enables the business to run targeted marketing campaigns to retain loyal customers and re-engage inactive ones.

Sales and Product Strategy: Association rule mining (Apriori) uncovers patterns in product combinations that are frequently purchased together. These insights can guide the development of product bundling and cross-selling strategies tailored to specific markets.

Geographical Targeting: Analysis of transaction data by country provides insights about regional purchasing preferences. These insights can guide decisions on which products to promote in which countries and at what times, enabling more effective allocation of sales and marketing resources.

Demand Forecasting: The Holt-Winters method supports forecasting future sales trends by product and region. This aids in planning promotions, managing inventory, and coordinating logistics.

Inventory Optimization: Understanding seasonal demand helps manage inventory more efficiently, reducing storage costs and avoiding product shortages.

Customer Lifecycle Management: Analyzing customer behavior over time reveals changes in engagement levels. This enables timely actions, such as sending personalized offers to maintain relationships.

Strategic Reporting: Interactive Tableau dashboards give stakeholders a clear view of key metrics, customer trends, and business performance, supporting informed and timely decisions.

Data description

The Online Retail II dataset is hosted on the UCI Machine Learning Repository and can be accessed here: https://archive.ics.uci.edu/dataset/502/online+retail+ii. This dataset contains transactional data from a UK-based non-store online retailer, covering the period from December 1, 2009, to December 9, 2011.

The dataset comprises over 1 million rows, each representing a transaction. Key fields include:

- **InvoiceNo**: Unique identifier for each transaction; entries starting with 'C' indicate cancellations.
- **StockCode**: Product code.
- **Description**: Product name.

- **Quantity**: Number of items purchased.
- **InvoiceDate**: Timestamp of the transaction.

  UnitPrice: Price per item.

  CustomerID: Unique identifier for each customer.

  Country: Country of the customer.
- **Sales:** Not in the dataset originally, we created this column using UnitPrice and Quantity

While the dataset is real and valuable for analysis, certain biases and credibility issues should be considered:

- **Missing Values**: Some entries lack CustomerID, which can affect customer-level analyses.
- **Cancellations**: Transactions marked with 'C' in InvoiceNo indicate cancellations and need to be handled appropriately.
- **Temporal Bias**: The data spans only two years, which may not capture longer-term trends.
- **Geographical Bias**: The retailer is UK-based, so the data may not generalize to other regions.

We have addressed licensing, privacy, security, and accessibility in accordance with the nature of the dataset. The Online Retail II dataset is publicly available through the UCI Machine Learning Repository and is intended for research and educational use, which ensures proper licensing. In terms of privacy, all customer data is anonymized, and no personally identifiable information is included, mitigating any privacy concerns. Since the dataset is both static and anonymized, security risks are minimal. Lastly, the dataset is accessible to all users and can be easily downloaded in Excel format from the UCI repository, ensuring broad availability for analysis.

Identified Inconsistencies with the data:

1. **Missing or Placeholder Descriptions**:

2. Some product rows had "?", "blank", or "throw" as descriptions, indicating incomplete or placeholder data. These rows were dropped or flagged depending on context.
3. **Cancelled Transactions**:
   Invoices starting with "C" denote cancelled transactions. These had to be excluded to avoid distorting metrics like total revenue or product frequency.
4. **Bad Debt Records**:
   Transactions with StockCode "B" typically represent bad debt. These had missing CustomerIDs and InvoiceNos starting with "A", and were removed from customer-based analyses.
5. **Postage and Transfer Charges**:
   International Transfers (StockCode: POSTAGE) contain Customer IDs, but Domestic Transfers (StockCode: DOTCOM POSTAGE) often lack them. These inconsistencies affect RFM and churn analysis.
6. **Discounts and Commissions**:
   Discounts (StockCode "D") and CRUK Commission lines (StockCode "CRUK") have negative prices and CustomerIDs present. These were handled carefully to distinguish between promotional activity and actual returns.
7. **Duplicate Records**:
   The dataset contained duplicate rows, which were dropped in the initial cleaning phase to avoid double-counting.
8. **Gift and Damage Stock Codes**:
   Gift items and damaged goods had missing CustomerIDs or ambiguous descriptions. These were excluded from customer-level modeling like RFM and churn analysis.
9. **Temporal Duplication**:
   Some December transactions appeared across both worksheets (likely referring to 2010 and 2011). These needed to be cross-verified and merged accurately.

Data Cleaning

To ensure data integrity, we applied a series of data cleaning and validation steps. Rows were filtered and dropped based on specific stock codes and invoice patterns—for example, entries with codes like B, C, D, and CRUK were excluded to remove bad debt, cancellations, discounts, and commissions that could skew the analysis. Segmented analysis was used where appropriate, such as excluding postage transactions from market basket analysis while retaining them for revenue-related insights. After cleaning, key summary statistics—including total revenue and unique customer counts—were re-checked to validate consistency. These efforts helped preserve meaningful trends in the data and ensured that our analyses, including RFM segmentation, Apriori market basket rules, and Holt-Winters forecasting, were grounded in accurate and reliable inputs.

For data cleaning purposes we have used three tools,

1. **Python (pandas, numpy)**: For data manipulation and cleaning due to its flexibility and wide range of built-in functions for handling missing values, filtering data, and transforming columns.
2. **Matplotlib and Plotly**: For data visualization. Matplotlib helps with static plots, while Plotly is used for interactive visualizations.
3. **OpenPyXL**: Required to read .xlsx files using pandas.read_excel().

After cleaning the data, we have ensured the integrity of the data by verifying that the file was loaded correctly by checking its size and structure. We then used df.info() to examine the data types and identify any anomalies that might affect analysis. Additionally, we checked for missing values in key columns such as Customer ID, Invoice, and StockCode to ensure that critical information required for analysis was present and reliable.

To ensure data quality, we followed several key preprocessing steps. First, we standardized column names for consistency and easier analysis. We removed rows with non-positive quantities or prices, as these likely indicated returns or data entry errors. Product descriptions containing terms like “damaged” or a question mark were filtered out to avoid unclear entries. We then checked for missing values in key columns to assess completeness. Finally, we verified and adjusted data types where needed, for example, converting the description column to string format, to ensure compatibility with downstream analysis.

We verified data's cleanliness and readiness by doing the following steps. First, we ensured that there are no missing values in critical columns such as Customer ID, Invoice, and StockCode. We then confirmed that all numeric fields, such as quantity and price, contain only positive values, as negative or zero entries could indicate errors or returns. We reviewed the Logical consistency, for example, we verified that all dates fall within valid time ranges and that customer IDs are present where required. We validate the number of unique entries, such as customers, invoices, and products, to make sure they align with expected values. Finally, we generate sample visualizations to help detect any outliers or irregular patterns that might indicate issues needing further attention.

Lastly, we have documented the data cleaning steps we performed in our code through comments and organized blocks. Each major cleaning action is explained, and summary statistics are printed to track progress. This ensures that the process is transparent, repeatable, and easy to share with team members or include in project documentation.

Analysis summary

Step 1: Descriptive Analysis

The dataset was first explored through descriptive analysis to understand the underlying structure and emerging patterns across different countries and product categories. The goal was to convert transactional-level data into a form that could support deeper behavioral, associative, and time-based modeling. To this end, product identifiers were standardized, dates were converted into a monthly frequency, and transactions were aggregated by country and stock code. This structure allowed us to observe purchasing dynamics at both customer and product levels across time and geography. Early insights confirmed that the United Kingdom was the primary revenue driver, while countries like Australia and Netherlands, despite having fewer transactions, exhibited clear demand spikes. This base-level understanding validated the quality and richness of the dataset, paving the way for more targeted analyses.

Step 2: Diagnostic Analysis with RFM

We first turned to RFM analysis to diagnose customer engagement and value patterns. RFM, which measures Recency (how recently a customer purchased), Frequency (how often they purchase), and Monetary value (how much they spend), is a diagnostic tool that helps identify which customer segments are driving business performance and which ones may need re-engagement. Customers were scored and segmented per country to reflect geographic behavioral nuances. We chose RFM because it is well-suited for transactional retail data, requires minimal assumptions, and provides intuitive, actionable results for marketing teams. In the UK, we observed a strong cohort of high recency and frequency customers with lower average order value, indicating that UK customers even though high in loyalty and volume, buy less products in each purchase. In contrast, countries like Ireland and Netherlands had frequent buyers who made smaller purchases, suggesting an opportunity for bundling or upselling. Interestingly, Switzerland revealed many recent but low-frequency customers, pointing to a newly engaged segment that could be nurtured. The diagnostic nature of RFM allowed us to pinpoint both strengths (loyal buyers) and opportunities (lapsed or newly acquired customers) at a very granular level, helping inform who to target.

Step 3: Prescriptive Analysis with Apriori Market Basket Analysis

To answer the next strategic question, *what should we sell to each segment?*, we applied Apriori market basket analysis. Apriori is a prescriptive tool that identifies patterns in purchasing behavior by uncovering frequent itemsets and association rules. Its strength lies in recommending product bundles or cross-sell opportunities based on historical co-purchase behavior. Prior to running Apriori, the dataset was reorganized into transaction-level “baskets” grouped by invoice number, with products encoded as sets per transaction. We opted for Apriori due to its interpretability, wide adoption in retail analytics, and ability to surface meaningful product affinities across thousands of SKUs. Rules were generated country-wise using adaptive support and confidence thresholds to account for volume differences, and filtered by lift to ensure statistical significance. In the UK, bundles frequently featured decorative and gift-oriented items such as t-light holders, patterned lunch bags, and novelty storage bags. However, RFM analysis revealed that while UK customers purchase in high volumes, their average order values remain relatively low, suggesting that they tend to buy these items individually rather than in curated sets. In comparison in France and Germany, customers bought practical household goods like night lights and snack boxes and ordered them in sets of 4 or 6 instead of individually. This highlighted the cultural depth of buyer preferences. These prescriptive recommendations were especially valuable when linked back to RFM segments, allowing us to suggest the *right bundle for the right customer*.

Step 4: Predictive Analysis with Holt-Winters Exponential Smoothing

To determine *when* these products should be marketed, and to align promotions with demand, we employed Holt-Winters Exponential Smoothing for time series forecasting. This method is inherently predictive, capable of modeling both trends and seasonality in monthly sales data. We selected Holt-Winters because it handles seasonal cycles effectively, is computationally lightweight for thousands of product-country combinations, and does not require external features, making it ideal for SKU-level demand forecasting in an e-retail setting. Prior to modeling, we aggregated sales at a monthly level by product and country, filtered out series with fewer than 24 months of data, and resampled to ensure consistency. The resulting forecasts revealed consistent seasonality in major markets, for instance, November demand spikes in the UK and Germany for gifting and decoration items, and summer dips in Spain. Some smaller markets, like Netherlands and Australia, exhibited surprisingly strong seasonal rhythms, while certain newer products showed growth trajectories in emerging markets, suggesting early-stage adoption curves. This predictive layer helped establish *when to act*, identifying time windows for bundling, advertising, and inventory management.

Step 5: Strategic Mapping Product Bundles to Customer Segments with Seasonal Timing

Bringing these three analytical layers together allowed us to move from raw data to actionable insight. We mapped the bundles that surfaced through Apriori to specific RFM segments, and then used the Winter-Holt forecasts to time those bundles appropriately throughout the year. For example, a holiday-themed bundle of gift was most often purchased by high-frequency, moderate-value UK customers and forecasted to peak in November, making it an ideal candidate for a targeted seasonal campaign. Similarly, summer bundles in Germany aligned with formerly high-value but now lapsed customers, providing a timely opportunity for reactivation. Even smaller countries like Austria and Norway, where bundles had niche but strong associations, became part of the strategic rollout based on their seasonal forecast patterns.

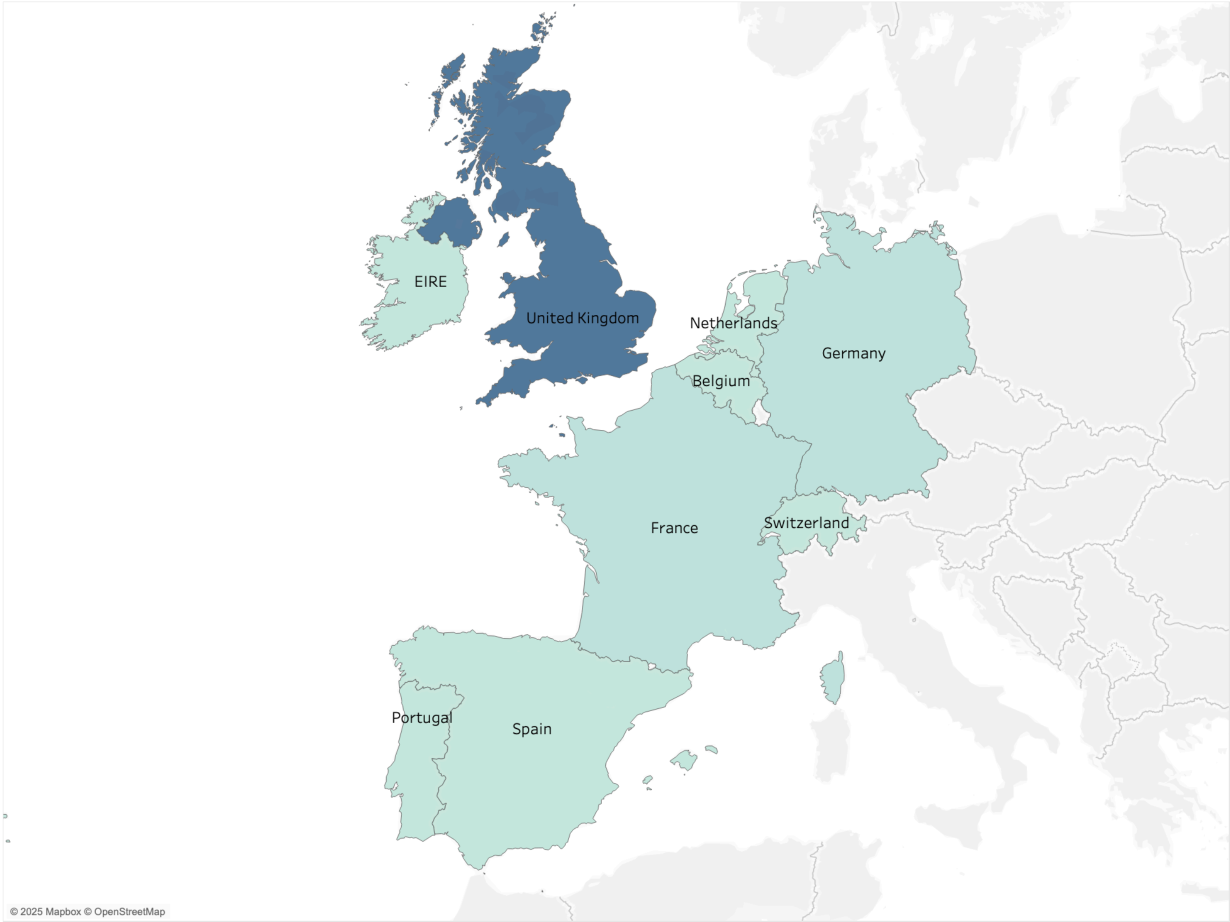
In essence, RFM helped us understand *who* our customers are, Apriori revealed *what* they want to buy together, and Winter-Holt showed us *when* to reach them, forming a complete diagnostic-prescriptive-predictive loop. These methods were chosen not only for their individual strengths but for their synergy, especially in the context of a large-scale e-retailer with varied markets, thousands of products, and evolving customer behaviors.

Visualizations and Key findings

To effectively represent our customer intelligence analysis, we used a variety of visualization techniques, each selected to match the nature of the insights we wanted to uncover and communicate clearly to stakeholders:

1. **Choropleth Maps**: These were used to display both average and total sales by country, enabling a quick geographical comparison of sales performance. This technique is ideal for identifying high-value markets and regional sales disparities at a glance.

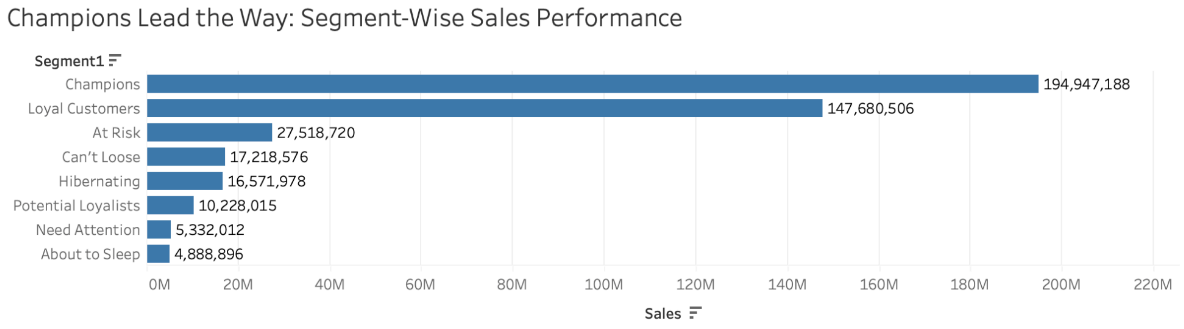
Identifying High-Grossing Countries by Sales Volume



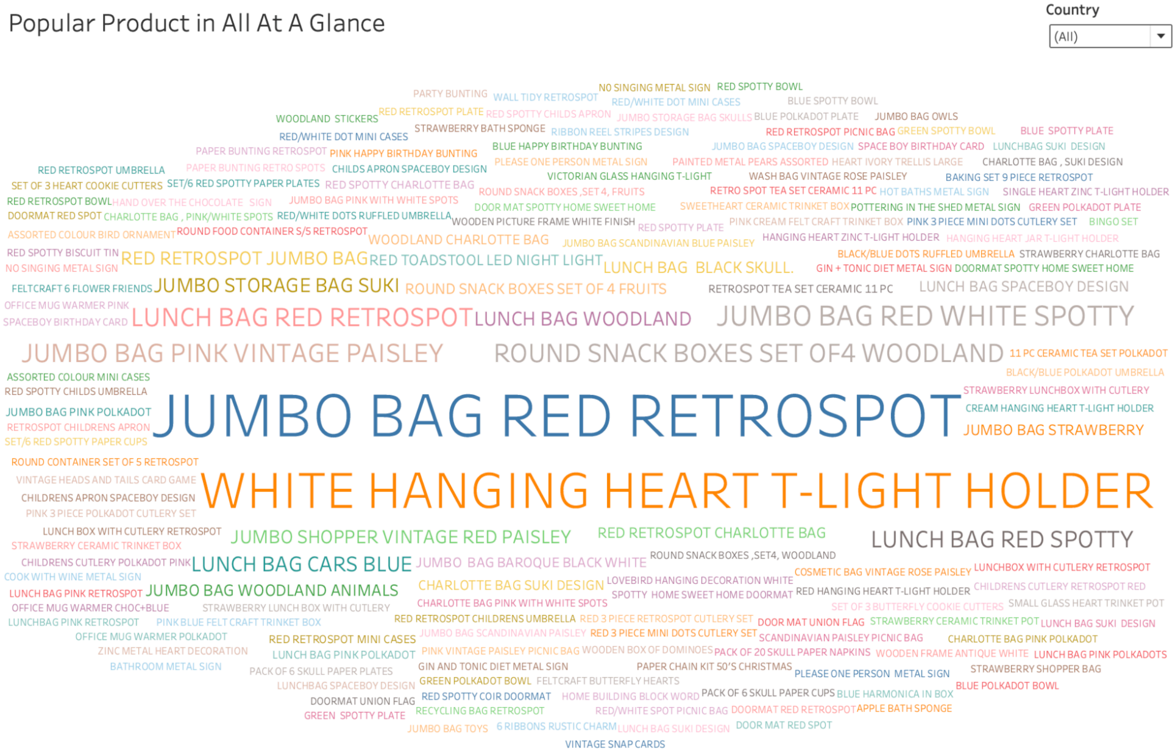
© 2025 Mapbox © OpenStreetMap

2. **Bar Charts**: Horizontal bar charts helped visualize total sales by country and by RFM-based customer segments (e.g., "Champions," "At Risk"). These are excellent

for comparing absolute values across categories and highlighting the contribution of each segment or region to overall revenue.


Champions Lead the Way: Segment-Wise Sales Performance

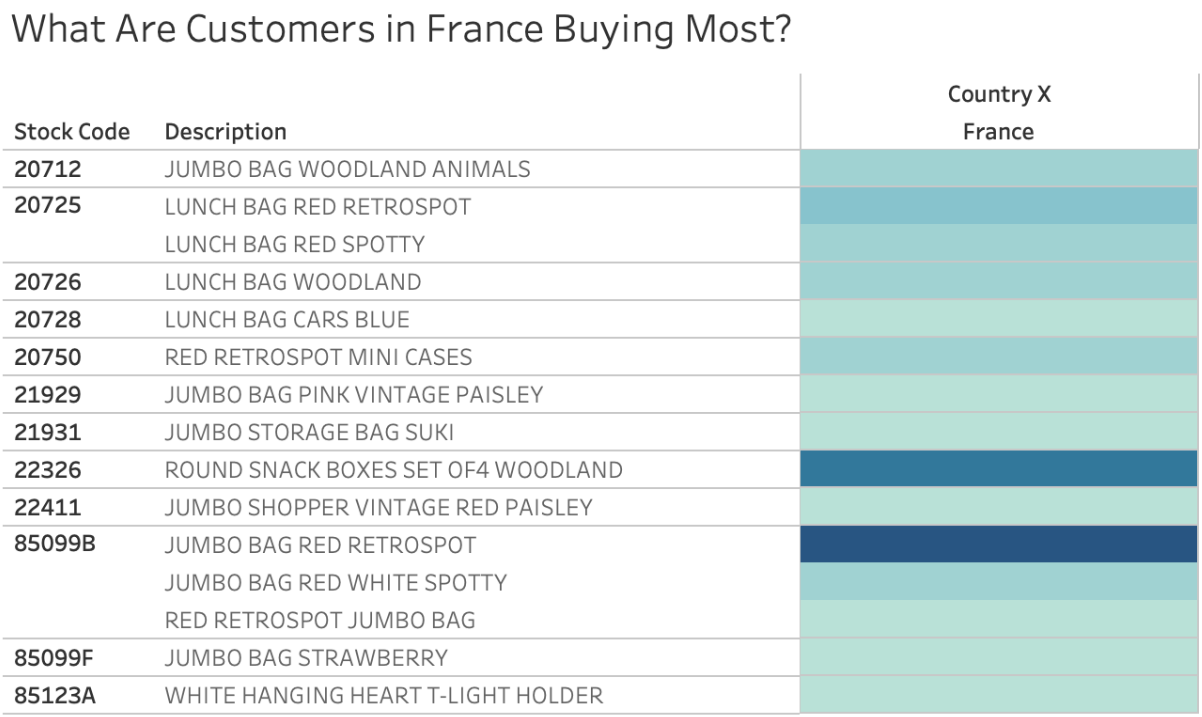
3. **Word Cloud**: To identify popular products, we used a word cloud based on product descriptions. This visual technique quickly surfaces high-frequency items, which is particularly useful for exploratory analysis and identifying trending products.


Popular Product in All At A Glance

Country: (All)

4. **Heatmaps**: Heatmaps were employed to show sales performance by product and country, as well as product associations. The color intensity effectively conveys

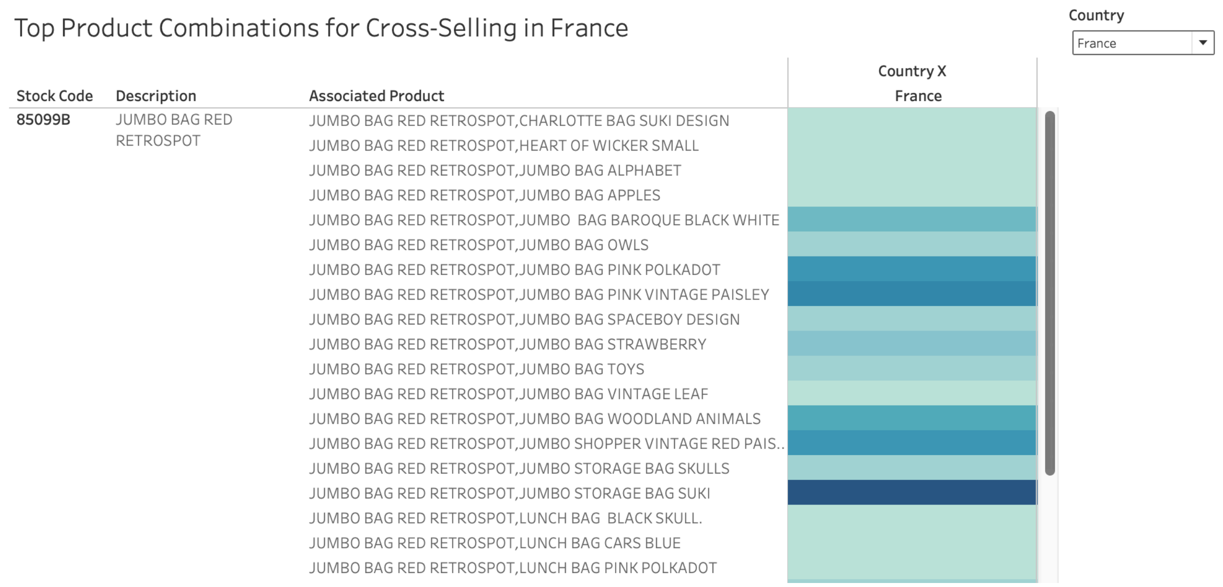
variations in sales volume, helping detect strong performers and bundling opportunities.



What Are Customers in France Buying Most?

Stock Code	Description	Country X France
20712	JUMBO BAG WOODLAND ANIMALS	▒
20725	LUNCH BAG RED RETROSPOT	▓
	LUNCH BAG RED SPOTTY	▒
20726	LUNCH BAG WOODLAND	▒
20728	LUNCH BAG CARS BLUE	░
20750	RED RETROSPOT MINI CASES	▒
21929	JUMBO BAG PINK VINTAGE PAISLEY	░
21931	JUMBO STORAGE BAG SUKI	░
22326	ROUND SNACK BOXES SET OF4 WOODLAND	█
22411	JUMBO SHOPPER VINTAGE RED PAISLEY	░
85099B	JUMBO BAG RED RETROSPOT	█
	JUMBO BAG RED WHITE SPOTTY	▒
	RED RETROSPOT JUMBO BAG	░
85099F	JUMBO BAG STRAWBERRY	░
85123A	WHITE HANGING HEART T-LIGHT HOLDER	░

5. **Association Product Tables**: These tables linked frequently bought-together products (from Apriori analysis) and revealed valuable cross-sell and bundling insights. Presenting them alongside sales volumes added further context for strategic recommendations.

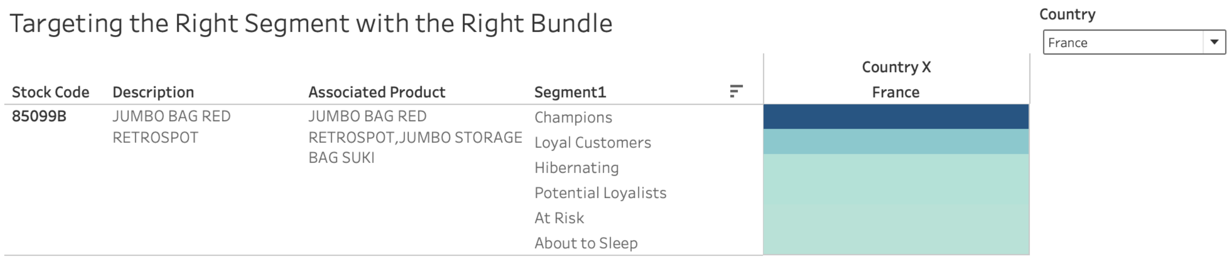


Top Product Combinations for Cross-Selling in France

Country: France

Stock Code	Description	Associated Product	Country X France
85099B	JUMBO BAG RED RETROSPOT	JUMBO BAG RED RETROSPOT,CHARLOTTE BAG SUKI DESIGN	
		JUMBO BAG RED RETROSPOT,HEART OF WICKER SMALL	
		JUMBO BAG RED RETROSPOT,JUMBO BAG ALPHABET	
		JUMBO BAG RED RETROSPOT,JUMBO BAG APPLES	
		JUMBO BAG RED RETROSPOT,JUMBO BAG BAROQUE BLACK WHITE	
		JUMBO BAG RED RETROSPOT,JUMBO BAG OWLS	
		JUMBO BAG RED RETROSPOT,JUMBO BAG PINK POLKADOT	
		JUMBO BAG RED RETROSPOT,JUMBO BAG PINK VINTAGE PAISLEY	
		JUMBO BAG RED RETROSPOT,JUMBO BAG SPACEBOY DESIGN	
		JUMBO BAG RED RETROSPOT,JUMBO BAG STRAWBERRY	
		JUMBO BAG RED RETROSPOT,JUMBO BAG TOYS	
		JUMBO BAG RED RETROSPOT,JUMBO BAG VINTAGE LEAF	
		JUMBO BAG RED RETROSPOT,JUMBO BAG WOODLAND ANIMALS	
		JUMBO BAG RED RETROSPOT,JUMBO SHOPPER VINTAGE RED PAIS..	
		JUMBO BAG RED RETROSPOT,JUMBO STORAGE BAG SKULLS	
		JUMBO BAG RED RETROSPOT,JUMBO STORAGE BAG SUKI	
		JUMBO BAG RED RETROSPOT,LUNCH BAG BLACK SKULL.	
		JUMBO BAG RED RETROSPOT,LUNCH BAG CARS BLUE	
		JUMBO BAG RED RETROSPOT,LUNCH BAG PINK POLKADOT	

6. **Segment-Country-Product Analysis Tables**: These combined views of product, associated product, and customer segment across countries are powerful for personalized marketing, targeting strategies, and understanding behavioral clusters.

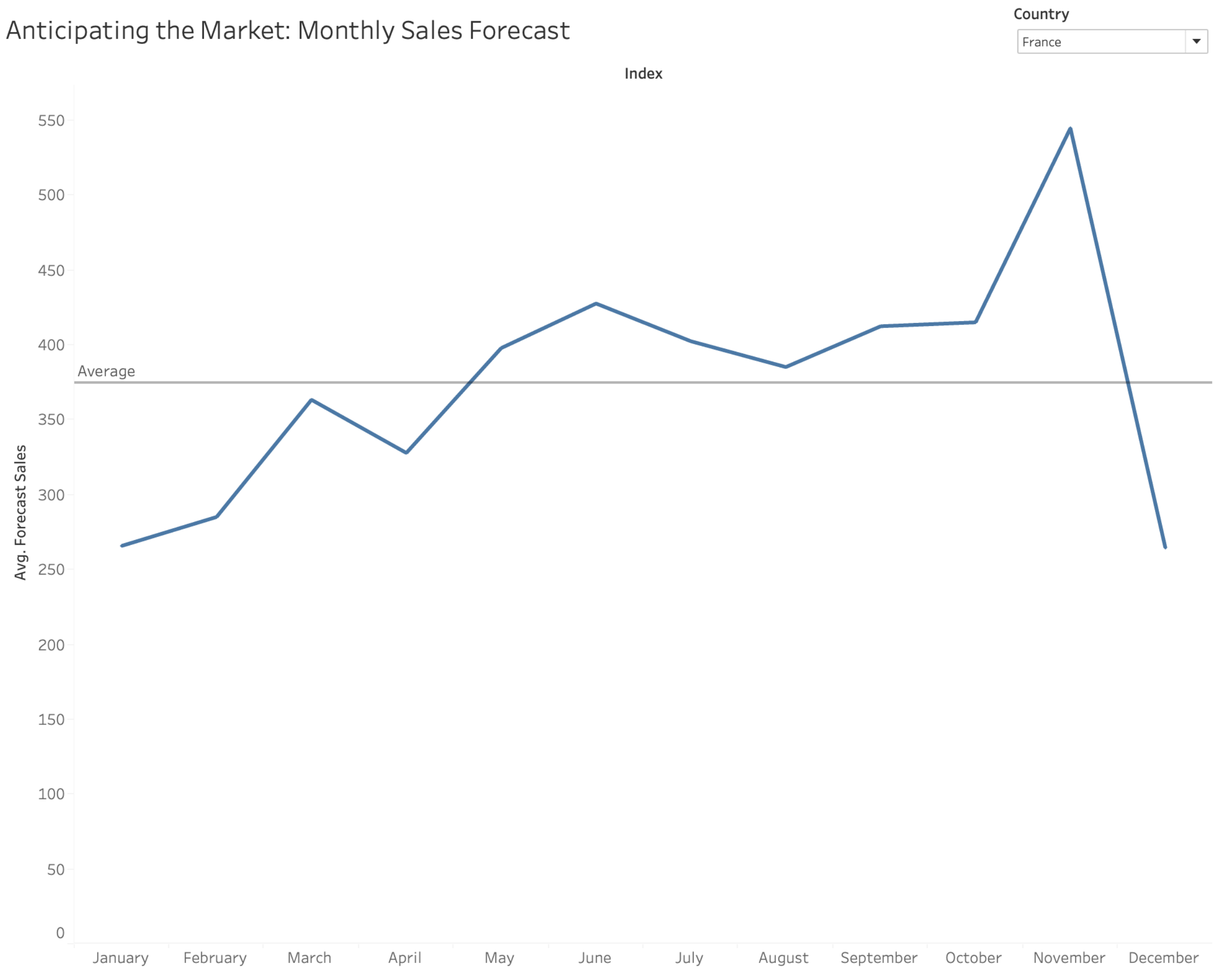


Targeting the Right Segment with the Right Bundle

Country: France

Stock Code	Description	Associated Product	Segment1	Country X France
85099B	JUMBO BAG RED RETROSPOT	JUMBO BAG RED RETROSPOT,JUMBO STORAGE BAG SUKI	Champions	
			Loyal Customers	
			Hibernating	
			Potential Loyalists	
			At Risk	
			About to Sleep	

7. **Line Charts**: A time series line chart was used to show average monthly sales trends for selected customer segments. This helps monitor engagement over time and detect lifecycle shifts such as churn risk or reactivation potential.



To address the main question of how transactional data can be used to understand customer behavior, identify lifecycle trends, and build a sustainable competitive advantage for an online wholesale retailer, we selected a range of visualization techniques that aligned directly with our strategic objectives. Choropleth maps were used to visualize total and average sales by country, helping us identify regional demand patterns and guide geographical targeting decisions. Bar charts were effective in showing sales contributions by customer segment and country, providing clarity on who our most valuable customers are and where they are located. Word clouds allowed us to quickly surface popular product descriptions, revealing customer preferences and potential bundling opportunities. Heatmaps provided a clear view of product performance within specific markets, while also highlighting frequently purchased product combinations uncovered through association rule mining using the Apriori algorithm. Line charts were used to track average monthly sales trends for specific customer segments, offering insights into engagement shifts and lifecycle stages over time. Additionally, we used combined tables to explore relationships between products, associated products, customer segments, and countries, supporting highly targeted marketing and inventory decisions. Together, these visualizations created an integrated Tableau dashboard that enabled us to not only describe current business performance but also diagnose issues, prescribe data-driven strategies, and forecast future outcomes. Ultimately, the visualizations played a vital role in transforming raw transactional data into actionable insights that support long-term business growth and competitive differentiation.

Data story

Imagine you’re managing a wholesale online giftware store with thousands of transactions coming in from buyers across Europe, each one holding hidden clues about customer preferences, timing, and behavior. But making sense of this data isn’t easy. Sales vary dramatically across countries, some loyal customers go quiet, and while certain products sell well together, others don’t move at all. Marketing feels generic, and inventory planning is reactive. As we dug deeper, the core issue became clear: without understanding who the customers are, what they buy together, and when they’re most likely to purchase, the business risks losing sales opportunities and wasting resources. Through RFM segmentation, we uncovered key groups like “Champions” and “At Risk” across countries, revealing where to focus retention and reactivation efforts. Market basket analysis using Apriori surfaced high-affinity product bundles, like themed storage and lunch bags, that are frequently bought together, differing by country. Then, Holt-Winters forecasting allowed us to pinpoint when demand for specific products peaks, such as the UK’s November holiday spike. With targeted visualizations like maps, heatmaps, bar charts, and time-series trends, we connected these insights to regions and customer segments. The result? A clear, data-driven strategy: offer the right products to the right customers at the right time.

Our original question asked how transactional data could be used to understand customer behavior, track lifecycle trends, and ultimately create a sustainable competitive advantage for a niche online wholesale retailer. Our findings directly address this by revealing *who* the customers are through RFM segmentation, *what* they buy together using Apriori market basket analysis, and *when* to engage them via Holt-Winters time series forecasting. Visualizations such as segment-wise sales trends, association heatmaps, and country-level demand forecasts brought clarity to otherwise complex datasets. These insights provided targeted strategies for product bundling, customer retention, and seasonal promotion planning, effectively closing the loop from raw data to business impact, precisely as the original question intended.

Our Audience

Our primary audience includes marketing strategists, sales managers, and business decision-makers within the wholesale retail organization, professionals who rely on clear, actionable insights to plan campaigns, manage inventory, and retain customers. The best way to communicate with them is through a combination of visual storytelling and focused dashboards. Interactive Tableau visualizations allow stakeholders to explore key metrics such as segment value, country-specific trends, and product performance in a digestible way, while a clear narrative, delivered via a concise executive summary or presentation, connects the data to strategic business goals. This balance of visuals and story ensures clarity, engagement, and decision-making confidence.

Impact of Data Visualisations

Data visualization played a critical role in sharing our findings. It translated complex analyses, like RFM segmentation, market basket associations, and sales forecasts into intuitive visuals that made patterns easy to understand at a glance. Choropleth maps helped us highlight country-level differences in sales performance, bar charts made customer segment contributions clear, and line charts revealed seasonal trends. These visuals allowed us to tell a cohesive story, turning raw transactional data into actionable insights that are easy for stakeholders to grasp and act on.

Our presentation is accessible to the intended audience, which includes marketing, sales, and strategy teams. We structured the dashboard and findings in a way that balances technical depth with business relevance. By using clear labels, simple chart types, consistent color schemes, and concise narrative explanations, we ensured that even non-technical stakeholders can understand the insights. Interactive elements within Tableau dashboards also allow users to explore specific regions, customer segments, or product categories based on their interest, making the experience both informative and user-friendly.

Recommendations

1. Champions & Loyal Customers (High Recency, High Frequency, High Monetary Value)

Profile: These are your most engaged and profitable customers. They frequently purchase and respond well to new products or exclusive offerings.

Recommendation:

- **Product Strategy**: Offer **early access to newly bundled gift sets**, particularly high-performing decorative items identified in Apriori analysis (e.g., *T-light holders, patterned bags, themed storage boxes* in the UK).
- **Engagement Tactic**: Position this as “VIP exclusive bundles” or “preferred buyer collections.”
- **Timing**: Launch just ahead of peak demand periods (e.g., **late October** for holiday products in the UK and Germany, **early summer** for household bundles in Germany).
- **Channel**: Email campaigns, private landing pages, or custom offers for repeat wholesale buyers.

2. Potential Loyalists & Promising Customers (Moderate Frequency, Good Recency, Growing Monetary)

Profile: These are customers who are close to becoming loyal buyers. A small push can convert them into high-value segments.

Recommendation:

- **Product Strategy**: Introduce **mid-tier practical bundles**, such as *kitchen sets or snack boxes* seen in France and Germany. Offer a discount for bulk purchase or repeat orders.
- **Engagement Tactic**: Create an incentive-based loyalty tier (e.g., “Buy 3 bundles, get next 10% off”) and upsell during peak cycles.

- **Timing**: Align offers with medium-volume demand periods (e.g., **May–June or September** in continental Europe).
- **Channel**: Newsletter inserts, re-targeting ads, or sales rep outreach for high-potential accounts.

3. At Risk & About to Sleep Customers (Low Recency, High Past Frequency/Monetary)

Profile: These were once valuable customers but haven't engaged recently. They need personalized reactivation strategies.

Recommendation:

- **Product Strategy**: Offer **limited-time bundles** that include frequently bought products from their past purchases (e.g., themed bags, gift items) based on Apriori results. Consider bundling with a small, free add-on to encourage re-trial.
- **Engagement Tactic**: Use urgency tactics like "We miss you!" or "Last chance to get your favorites before the season ends." Include reorder links to previously purchased products.
- **Timing**: Launch **2–3 weeks before forecasted seasonal peaks** (e.g., early November in the UK, August in Germany) to re-engage before your highest-traffic periods.
- **Channel**: Personalized email, SMS, and direct sales calls for B2B customers.

Conclusion

Our overall analysis directly answers the central question: how can transactional data be leveraged to understand customer behavior, track lifecycle trends, and build a sustainable competitive advantage for a niche online wholesale retailer? By combining descriptive, diagnostic, prescriptive, and predictive techniques, we uncovered clear patterns in who buys, what they buy together, and when they buy.

Moving forward, these insights can guide the marketing team in launching segment-specific campaigns (e.g., reactivating “At Risk” customers before seasonal peaks), and help sales and operations teams better align product bundles and inventory with country-level demand patterns.

Next steps would include operationalizing these findings into campaign calendars, bundle rollouts, and forecast-based stock planning. We also see potential to expand this work by incorporating marketing campaign performance data or customer demographics to sharpen personalization and test bundle effectiveness across regions.

In short, this analysis offers a scalable framework for data-driven decisions and with additional inputs, can evolve into a predictive system for long-term customer and product strategy.

Appendix:

Code:

https://colab.research.google.com/drive/1CX1ePRNl7A-9AyjdIPyTD0gVa-Lf_o7C?usp=sharing