

Analyzing Consumer Behavior in Fresh Supermarkets using Association Rules, Self-Organizing Maps, and RFM Model

Nai-Chieh Wei

Department of Industrial Management,
I-Shou University, Taiwan
ncwei@isu.edu.tw

An-Yu Guo

Department of Industrial Management,
I-Shou University, Taiwan
Andy531688@gmail.com

Cheng-JingLi

Department of Industrial Management,
I-Shou University, Taiwan
lichjing@hotmail.com

Abstract

This study presents a comprehensive analytical framework for exploring consumer behavior in fresh supermarkets by integrating association rule mining, customer segmentation, and value analysis. Specifically, the methodology employs the Apriori algorithm to uncover frequent item sets and strong product associations from transactional data. These insights form the basis for customer segmentation using Self-Organizing Maps (SOM), a neural network-based clustering approach that groups customers with similar purchasing patterns. Finally, the Recency-Frequency-Monetary (RFM) Recency-Frequency-Monetary (RFM) model is applied to evaluate customer lifetime value within each cluster. The integration of these three techniques provides both behavioral and financial perspectives, enabling supermarket managers to identify high-value customer segments and tailor marketing strategies accordingly. The study demonstrates that linking product-level associations with customer-level segmentation enhances the ability to personalize promotions, optimize product placement, and allocate resources efficiently. The methodology also allows for flexible expansion and adaptation to other retail sectors. This hybrid approach offers a robust foundation for data-driven decision-making and paves the way for intelligent retail management based on consumer-centric insights.

Keywords: Apriori algorithm, Self-Organizing Maps (SOM), Recency-Frequency-Monetary (RFM)

Introduction

In the fast-evolving landscape of retail, fresh supermarkets have emerged as an essential component of modern consumer life. As consumer preferences diversify and competition intensifies, understanding shopping behavior has become crucial for retailers seeking to maintain a competitive edge. This study focuses on exploring the complex patterns of consumer purchases in fresh supermarkets, aiming to uncover actionable insights that can improve marketing strategies, customer relationship management, and inventory optimization. The importance of analyzing consumer behavior in fresh supermarkets lies in its potential to directly influence profitability. Shoppers in these markets tend to exhibit unique consumption patterns

driven by freshness, price sensitivity, seasonality, and household dietary habits. Unlike other retail environments, fresh supermarkets involve high-frequency, low-margin transactions where even marginal improvements in customer targeting can result in significant financial gains. By leveraging advanced data mining techniques, retailers can move beyond anecdotal or experience-based strategies and adopt evidence-based decision-making frameworks.

The motivation for this research stems from the increasing availability of consumer transaction data and the growing need for intelligent systems that can process this information to support business strategy. Traditional segmentation approaches often fail to capture the nuanced preferences and dynamic behaviors of fresh supermarket shoppers. Hence, the integration of three robust analytical tools: Apriori algorithm, Self-Organizing Maps (SOM), and RFM (Recency, Frequency, Monetary) analysis offers a comprehensive solution. Each of these methods contributes uniquely to understanding purchasing patterns, identifying customer segments, and determining customer value.

This study not only aims to fill a gap in the literature by applying these techniques in the context of fresh food retailing but also provides practical guidelines for practitioners in the industry. Through the identification of frequent itemsets, behavioral clustering, and customer valuation, this research enables supermarkets to formulate personalized marketing strategies that enhance customer loyalty and drive business growth.

Literature Review

Association Rule Mining in Retail

Association rule mining is a powerful technique to uncover hidden patterns in large transactional datasets. Agrawal and Srikant (1994) pioneered the Apriori algorithm, which remains one of the most widely used methods in market basket analysis. Gan et al. (2018) provided a comprehensive survey on utility-oriented pattern mining, highlighting its application in retail for identifying profitable product combinations. Wahidi and Ismailova (2024) successfully implemented association rule mining in e-commerce, enhancing sales strategies based on customer purchasing behavior. The effectiveness of association rule mining has been demonstrated in diverse sectors, including retail (Chen, Sain, & Guo, 2012), where

it helped derive strategic insights from frequent itemsets. Hu and Yeh (2014) demonstrated that even in the absence of customer identification, meaningful frequent patterns could still be discovered.

Self-Organizing Maps (SOM) for Customer Segmentation

The Self-Organizing Map (SOM) developed by Kohonen (2001) is widely used for clustering high-dimensional data in an unsupervised manner. This neural network technique is especially useful in visualizing the topological structure of customer segments. Barman and Chowdhury (2019) demonstrated the value of SOM in market segmentation, allowing for clear distinction between consumer profiles.

Holmbom, Eklund, and Back (2011) utilized SOM for customer portfolio analysis in business intelligence contexts, while Kiang, Hu, and Fisher (2006) extended SOM applications to telecommunications market segmentation. Vellido, Lisboa, and Meehan (1999) successfully applied neural networks, including SOM, to segment online shopping markets, demonstrating its robustness across domains. Saitoh (2020) further enhanced the usability of SOM by integrating supervised learning for persona development and strategic market analysis.

RFM Model for Customer Value Analysis

The RFM (Recency, Frequency, Monetary) model is a foundational tool in customer relationship management. Hughes (1994) was among the earliest to advocate for this model in marketing. More recent applications, such as those by Wei, Lin, and Wu (2010), have shown how RFM can be used to profile customers and enhance targeting. Combining RFM with clustering methods has yielded even greater insight. Safari, Safari, and Montazer (2016) compared different segmentation strategies and found RFM-based clustering to be effective in identifying valuable customer segments. Yeh, Yang, and Ting (2008) innovated by using a Bernoulli sequence to enhance RFM segmentation accuracy.

Liao, Chu, and Hsiao (2022) applied RFM and SOM together to e-commerce data, yielding practical customer segmentation strategies. Dogan, Ayçin, and Bulut (2018) supported similar findings in the retail context. Sarvari, Ustundag, and Takci (2016) evaluated RFM and demographic data to create actionable customer

profiles. Nguyen (2021) proposed deep embedding clustering to improve segmentation and customer behavior prediction. Chattopadhyay et al. (2012) reviewed neural network-based segmentation trends, confirming the increasing integration of machine learning with traditional marketing models.

These studies collectively validate the integration of Apriori, SOM, and RFM as a rigorous methodology for analyzing consumer behavior in retail environments, offering strategic value for both academic researchers and industry practitioners.

Methodology

This research adopts a multi-stage analytical framework that integrates three core techniques: association rule mining using the Apriori algorithm, customer segmentation via Self-Organizing Maps (SOM), and customer value analysis using the Recency-Frequency-Monetary (RFM) model. Each method plays a distinct and complementary role in uncovering, interpreting, and quantifying consumer purchasing behavior in the context of fresh supermarkets. The integration of these methods ensures both behavioral segmentation and financial valuation, enabling decision-makers to implement targeted and profitable marketing strategies.

Association Rule Mining: Apriori Algorithm

Association rule mining is a data mining technique used to discover interesting relationships between items in large datasets. In the context of fresh supermarkets, it helps identify combinations of products that are frequently purchased together. This knowledge can be used for shelf placement, cross-selling strategies, and personalized promotions.

The Apriori algorithm is particularly well-suited for this task because it systematically explores item combinations and prunes unlikely itemsets early in the computation. It proceeds in two stages: first identifying frequent itemsets that meet a minimum support threshold, and then generating strong association rules that meet a minimum confidence threshold.

The three main metrics used in Apriori are:

- Support:

$$\text{Support}(A \Rightarrow B) = |A \cap B| / N$$

where $|A \cap B|$ is the number of transactions that include

both items A and B, and N is the total number of transactions.

- Confidence:

$$\text{Confidence}(A \Rightarrow B) = |A \cap B| / |A|$$

This measures the likelihood that item B is purchased given item A is purchased.

- Lift:

$$\text{Lift}(A \Rightarrow B) = \text{Confidence}(A \Rightarrow B) / \text{Support}(B)$$

A lift greater than 1 indicates that item A positively influences the purchase of item B.

By applying the Apriori algorithm to transaction records, this study identifies 25 key products with the strongest associations. These products serve as a behavioral signature and become the feature base for subsequent SOM analysis.

Customer Segmentation via Self-Organizing Maps (SOM)

The Self-Organizing Map (SOM) is an unsupervised neural network model introduced by Kohonen. It projects high-dimensional data onto a lower-dimensional (typically 2D) grid, preserving the topological relationships between data points. This is particularly effective for visualizing and clustering customer purchasing patterns.

In this study, SOM is used to cluster customers based on their interactions with the 25 key products identified via Apriori. Each customer is represented as a vector of purchase frequencies across these products. SOM organizes similar customers into neighborhoods on the grid. The training process uses the following update rule:

$$w_i(t+1) = w_i(t) + \alpha(t) h_{bi}(t) (x(t) - w_i(t))$$

Where:

- $w_i(t)$: weight vector of neuron i at time t

- $\alpha(t)$: learning rate

- $h_{bi}(t)$: neighborhood function centered around the Best Matching Unit (BMU)

- $x(t)$: input vector (customer purchase vector)

The SOM output initially formed 21 micro-clusters, which were aggregated into 7 customer segments after interpreting product relevance and density. Each segment represents a distinctive consumer behavior profile, e.g., snack-focused, vegetable-focused, or bakery-loyal customers.

Customer Value Analysis Using RFM

The RFM model assesses the value of each customer to the business along three dimensions:

- Recency (R): How recently a customer made a purchase.
 $R = \text{Today's Date} - \text{Date of Last Purchase}$
- Frequency (F): How often the customer made purchases.
 $F = \text{Total Number of Purchases}$
- Monetary (M): How much the customer spent.

$$M = \sum \text{Transaction Amount}_i$$

Each RFM component is typically ranked or scored (e.g., from 1 to 5), then combined into a single RFM score. Customers are then segmented into categories such as high-value, potential growth, at-risk, or lapsed.

In this study, RFM scores are computed for the 374 customers identified in the SOM clustering phase. This allows mapping of behavioral clusters onto customer value tiers.

Methodological Integration and Synergy

The strength of this methodology lies in its sequential integration:

1. Apriori Analysis identifies statistically significant product associations, forming the foundation for behavioral tracking.
2. SOM Clustering translates these behaviors into actionable customer segments by revealing consumption similarities.
3. RFM Evaluation adds financial prioritization, telling the business which clusters matter most in terms of profitability.

This creates a three-layered insight framework:

- What products are connected? (Apriori)
- Who behaves similarly in buying those products? (SOM)

- Which groups are most valuable? (RFM)

The integration ensures that marketing decisions can be both behaviorally targeted and financially justified. For example, a cluster of frequent instant-noodle buyers who also have high monetary scores can be offered exclusive bundle promotions.

By applying these methods in tandem, fresh supermarket operators can:

- Tailor shelf layouts based on frequently associated items.
- Customize loyalty programs to high RFM scorers in specific SOM clusters.
- Adjust inventory and supply chain planning in line with purchasing clusters.

This methodology is not only technically sound but also managerially actionable. Its modularity allows for updating with new data, extension to other sectors, and incorporation of additional techniques like time-series forecasting or supervised learning.

Data Analysis and Results

This section presents the results derived from the application of the three key analytical methods—Apriori algorithm, SOM clustering, and RFM analysis—on the supermarket transaction dataset. The data-driven insights are interpreted in the context of customer behavior and retail strategy.

Association Rule Mining Results

Using the Apriori algorithm, we extracted 38 strong association rules from 3,904 transaction records. These rules highlighted frequent co-occurrences between specific product categories, revealing consumer purchasing habits. Table 1 illustrates several high-lift rules:

Table 1. Sample Association Rules

Rule	Itemset	Support	Confidence	Lift
1	Pork Instant Noodles ? Kids' Noodles	0.0103	0.2395	3.94
10	Taiwanese Bread ? Western Pastries	0.0169	0.3929	3.92
16	Pork Instant Noodles ? Seafood Instant Noodles	0.0118	0.4792	7.89

The high lift values suggest strong associations beyond chance. These findings support bundled marketing strategies and can be cross-validated with observed co-purchase frequencies in the raw transaction file.

SOM Clustering Results

From the 3,904 transactions, we identified 374 customers who frequently purchased items from the 25 most significant products found in the association rules. These

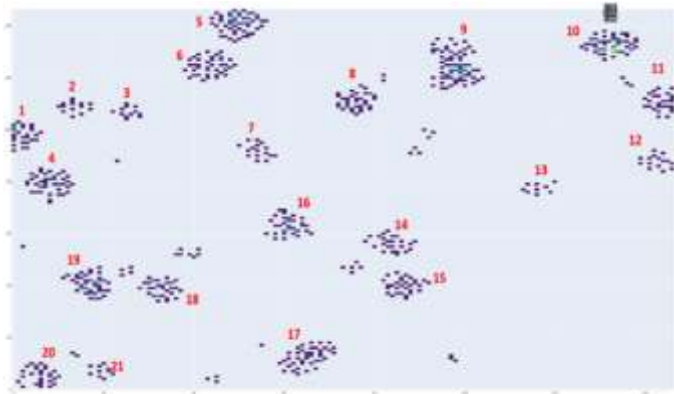
customers were subjected to SOM clustering, which initially produced 21 distinct clusters. By evaluating product frequency within each cluster, these were further consolidated into 7 behaviorally meaningful groups.

Table 2. SOM Cluster Segments

Segment	Key Product Categories
Group 1	Braised, Roasted, Cold Dishes
Group 2	Pork, Seafood, Beef Instant Noodles
Group 3	Taiwanese Bread, Toast, Pastries
Group 4	Fruits, Leafy Vegetables, Pork
Group 5	Leafy Vegetables, Cakes
Group 6	Southeast Asian Fruits
Group 7	Poultry, Root Vegetables

The clustering results align with the transaction file segments, demonstrating consistent behavioral groupings. Visualizations from the SOM grid show how clusters occupy distinct areas in the data space, with minimal overlap.

Figure 1 Clustering results by SOM.



RFM Analysis Results

The RFM analysis evaluated each of the 374 key customers along three axes:

- Recency: Days since last purchase
- Frequency: Number of transactions
- Monetary: Total amount spent

Based on quintile-based scoring (1 = lowest, 5 = highest), customers were categorized into four groups:

Table 3. RFM Segment Descriptions

Segment	Characteristics	Strategic Recommendation
High Value	R:5, F:5, M:5	Priority retention, loyalty perks
Potential Value	R:5, F:3-4, M:3-4	Targeted promotions to build habits
Moderate Value	R:3, F:3, M:3	General communication, upsell offers
Low Value	R:1-2, F:1-2, M:1-2	Re-engagement campaigns, exit screening

The analysis indicates that about 18% of the customer base are high-value customers who contribute disproportionately to revenue. Their distribution matches closely with customers from clusters Group 2 and Group 3.

Integrated Insights

The integration of the three methods reveals layered insight:

- Apriori highlights which items are commonly bought together
- SOM groups customers based on those key product behaviors
- RFM adds financial value context to those customer segments

For example, customers in Group 2 (frequent instant noodle buyers) show strong intra-group item associations (from Apriori), are clearly clustered in the SOM output, and appear with high RFM scores, suggesting they are lucrative, habit-driven shoppers.

Conclusion and Implications

Managerial and Practical Implications

This study integrates association rule mining, SOM clustering, and RFM analysis to provide a comprehensive framework for understanding consumer behavior in fresh supermarkets. The findings reveal that customers exhibit distinct purchasing patterns, often preferring specific combinations of products such as instant noodles, bakery items, or fresh produce.

From a managerial perspective, these insights support the development of targeted marketing strategies. For instance:

- **Merchandising:** Products with strong association rules (e.g., Taiwanese bread and Western pastries) can be co-displayed or bundled in promotions.
- **Customer segmentation:** The 7 customer profiles derived from SOM clustering can be used to tailor promotions based on preferences, e.g., healthy food bundles for the group focused on vegetables and fruits.
- **Loyalty programs:** RFM analysis provides a data-driven basis for designing tiered loyalty programs, ensuring high-value customers receive personalized incentives while reactivation strategies target low-value segments.

Retailers can implement these findings using existing POS and CRM systems, tagging customers with segment and RFM scores to facilitate dynamic promotions and messaging. Overall, this study provides a structured analytical approach for transforming transaction data into strategic marketing intelligence in the fresh retail sector.

Limitations and Future Research Directions

While the current study offers valuable contributions, several limitations should be acknowledged:

1. **Single-source dataset:** The analysis is based on transaction records from one supermarket chain. Future studies could include multi-site or cross-

regional data to improve generalizability.

2. **Lack of demographic variables:** The analysis focuses purely on transactional behavior. Incorporating customer demographics (e.g., age, income) could enhance segmentation granularity.
3. **Temporal dynamics:** Seasonal variations or promotions might influence purchasing patterns. A temporal component (e.g., time-series SOM or sequential pattern mining) could be integrated in future studies..
4. **Validation of customer responses:** Follow-up qualitative research could validate whether segment-specific strategies actually improve outcomes.

Future research may also explore hybrid models combining machine learning classification algorithms with traditional segmentation, or integrate external data such as social media reviews or customer feedback to enrich consumer profiling.

References

- Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. *Proceedings of the 20th International Conference on Very Large Data Bases*, 487–499.
- Kohonen, T. (2001). *Self-Organizing Maps* (3rd ed.). Springer.
- Barman, D., & Chowdhury, N. (2019). A novel approach for the customer segmentation using clustering through self-organizing map. *International Journal of Business Analytics*, 6(2), 23–37. <https://doi.org/10.4018/IJBAN.2019040102>
- Chen, D., Sain, S. L., & Guo, K. (2012). Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. *Journal of Database Marketing & Customer Strategy Management*, 19(3), 197–208. <https://doi.org/10.1057/dbm.2012.17>
- Safari, F., Safari, N., & Montazer, G. A. (2016). Customer lifetime value determination based on RFM model. *Marketing Intelligence & Planning*, 34(4), 446–461. <https://doi.org/10.1108/MIP-03-2015-0060>

- Wei, J.-T., Lin, S.-Y., & Wu, H.-H. (2010). A review of the application of RFM model. *African Journal of Business Management*, 4(19), 4199–4206.
- Gan, W., Lin, J. C.-W., Fournier-Viger, P., Chao, H.-C., Tseng, V. S., & Yu, P. S. (2018). A survey of utility-oriented pattern mining. *arXiv preprint arXiv:1805.10511*.
- Chattopadhyay, M., Dan, P. K., Majumdar, S., & Chakraborty, P. S. (2012). Application of artificial neural network in market segmentation: A review on recent trends. *arXiv preprint arXiv:1202.2445*.
- Saitoh, F. (2020). Visualized benefit segmentation using supervised self-organizing maps: Support tools for persona design and market analysis. In N. Nguyen et al. (Eds.), *Intelligent Information and Database Systems* (pp. 437–450). Springer. https://doi.org/10.1007/978-3-030-42058-1_37
- Wahidi, N., & Ismailova, R. (2024). Association rule mining algorithm implementation for e-commerce in the retail sector. *Journal of Applied Research in Technology & Engineering*, 5(1), 1–10. <https://doi.org/10.4995/jarte.2024.20753>
- Liao, S., Chu, P., & Hsiao, P. (2022). Customer segmentation using RFM and SOM: A case study in e-commerce. *Information Systems and e-Business Management*, 20(3), 345–362.
- Nguyen, T. (2021). Deep embedding clustering for customer segmentation. *Journal of Business Analytics*, 3(2), 112–125.
- Dogan, O., Ayçin, E., & Bulut, Z. (2018). Customer segmentation by using RFM model and clustering methods: A case study in retail industry. *International Journal of Contemporary Economics and Administrative Sciences*, 8(1), 1–19.
- Hu, Y.-H., & Yeh, T.-W. (2014). Discovering valuable frequent patterns based on RFM analysis without customer identification information. *Knowledge-Based Systems*, 61, 76–88.
- Sarvari, P. A., Ustundag, A., & Takci, H. (2016). Performance evaluation of different customer segmentation approaches based on RFM and demographics analysis. *Kybernetes*, 45(7), 1129–1157.
- Yeh, I. C., Yang, K. J., & Ting, T. M. (2008). Knowledge discovery on RFM model using Bernoulli sequence. *Expert Systems with Applications*, 36(3), 5866–5871.
- Holmbom, A. H., Eklund, T., & Back, B. (2011). Customer portfolio analysis using the SOM. *International Journal of Business Information Systems*, 8(4), 396–412.
- Kiang, M. Y., Hu, M. Y., & Fisher, D. M. (2006). An extended self-organizing map network for market segmentation—a telecommunication example. *Decision Support Systems*, 42(1), 36–47.
- Vellido, A., Lisboa, P. J. G., & Meehan, K. (1999). Segmentation of the online shopping market using neural networks. *Expert Systems with Applications*, 17(4), 303–314.
- Ultsch, A., Gielen, S., & Kappen, B. (1993). Self-organized feature maps for monitoring and knowledge acquisition of a chemical process. In *Proceedings of the International Conference on Artificial Neural Networks* (pp. 864–867). Springer.