Proceedings of the 6th Asia Pacific Conference on Industrial Engineering and Operations Management
Bali, Indonesia, August 5-7, 2025

Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/AP06.20250075
Published: August 5, 2025

Visualizing Retail SKU Configurations through Entropy and Market Matrix

Ryuzen Kawabe and Kiyotaka Yoshimura
Graduate School of Engineering and Science
Shibaura Institute of Technology
Saitama, Japan
bq20002@shibaura-it.ac.jp, bq20026@shibaura-it.ac.jp

Yusuke Goto
Department of Engineering and Science
Shibaura Institute of Technology
Saitama, Japan
y-goto@shibaura-it.ac.jp

Abstract

In recent years, small-scale retail stores like convenience stores have expanded their product assortments to meet increasingly diverse consumer preferences, resulting in a growing number of stock-keeping units (SKUs) per product category; however, limited shelf space in small-scale retail stores necessitates strategic SKU optimization. This study proposes a novel analytical framework that extends the traditional BCG matrix (Boston Consulting Group matrix) by incorporating two additional dimensions: purchasing entropy and SKU count. Purchasing entropy quantifies the sales dispersion across SKUs within each category, allowing for evaluating internal structural characteristics beyond market share and growth. Using one year of receipt-based purchase data from four major convenience store chains in Tokyo, Japan, the proposed matrix was applied to visualize category-level sales structures and compare them across chains. The results reveal substantial differences in purchasing behavior and SKU strategies, even within the same product category. The findings demonstrate that the extended matrix offers a multidimensional view of product category performance, facilitating more informed decisions on SKU configuration and sales strategy. This framework provides a practical tool for retail managers to balance product variety with space and performance constraints.

Keywords
Entropy, BCG Matrix, Market Analysis, Product Portfolio Management, Category Management

1. Introduction
In recent years, there has been an increase in small-scale retail stores (such as convenience stores), particularly in urban areas that enable consumers to purchase food items easily. The convenience store industry in Japan has shown long-term growth, with stores exceeding 50,000 by the 2010s. In 2020, the leading chain, 7-Eleven, recorded over 5 trillion JPY in domestic sales, comprising more than 44% of the total market revenue (Lee 2021). These stores must accommodate diverse consumer food preferences; thus, they carry various items such as bread, fried foods, and confectioneries. Additionally, many companies operate small-scale retail stores and develop private-label products and stock offerings from multiple manufacturers to differentiate themselves from competitors. This approach leads to an overall increase in stock-keeping units (SKUs). The trend of SKU growth is observed not only at the store-wide level, but also within individual product categories. For example, it is common for a single store to display ten or more types of bread.

Most convenience stores in Japan are located in space-constrained environments such as commercial buildings and train stations, where physical shelf space is limited; therefore, optimizing the number of SKUs based on sales performance and other data has become a key managerial challenge. Effective SKU optimization is expected to reduce waste due to unsold inventory (Sakoda et al. 2019) and improve profitability using strategic promotional space (Pak et al. 2020).

In this context, optimizing SKU counts to increase sales and profits requires reevaluating the overall number of SKUs at the store level and assessing whether the number of SKUs within each product category is appropriate. To support this evaluation, this study focuses on the sales distribution within product categories—specifically, the degree to which sales are concentrated or dispersed among SKUs. In retail settings, some categories exhibit relatively even sales across multiple products, referred to as dispersed purchasing types; however, in other categories, sales are heavily concentrated in a few SKUs, referred to as concentrated purchasing types. Dispersed purchasing types benefit from a broad assortment, as multiple SKUs contribute stably to overall sales. In contrast, concentrated purchasing types, in which a few SKUs account for most sales, require increased efficiency by promoting high-performing products or eliminating unprofitable ones.

This study conceptualizes variations in purchasing patterns as purchasing entropy which refers to the distribution of sales among SKUs within a given product category, indicating whether sales skew toward specific products or are more evenly distributed. Previous studies analyzed various retail operations; however, relatively few have explicitly focused on purchasing entropy. Moreover, evaluating each product category independently does not provide a relative basis for determining, from a store-wide perspective, which offers room for improvement or should be prioritized.

To address this gap, this study develops a method for evaluating the internal structure of product categories based on purchasing entropy and their relative positioning across categories. This approach aims to support strategic decision-making in SKU optimization for retail management.

1.1 Objectives

This study designs an analytical framework to evaluate purchasing patterns in SKU configurations across product categories in retail stores based on purchasing entropy. We utilize this evaluation to compare the different product categories. Specifically, the study aims to achieve the following two goals:

1. To quantitatively evaluate the variation in sales composition among SKUs within each product category as purchasing entropy.
2. To assess the relative sales structure across product categories based on purchasing entropy.

2. Literature Review

In recent years, customer experiential value within retail environments has gained increasing attention, alongside product price and quality. For example, Verhoef et al. (2009) conceptualized experiential value as a multidimensional construct encompassing emotional, cognitive, and social components. They emphasized the strategic importance of its design and management, identifying product category diversity and SKU abundance as key components. Enhancing these aspects can improve experiential value (Kulkarni 2013), however, shelf space is a limited and critical resource in small-scale retail stores such as convenience stores. The spatial configuration of products within this constrained environment directly impacts profitability and operational efficiency (Gecili and Parikh 2022). Therefore, it is not feasible to expand all categories and their associated SKUs; instead, a strategic promotion and space allocation approach must be adopted at the category level.

In response to such constraints, category management has emerged as a managerial approach to enhance promotional efficiency. Category management treats each product category as a strategic business unit within a store, integrating pricing, shelf allocation, and promotional activities to improve the profitability of individual categories (Basuroy et al. 2001). Category-level monitoring and strategy formulation must be employed through category management and related frameworks to optimize SKU configurations within the limited resources and shelf space available in small-scale retail stores.
Nonetheless, several challenges remain within the framework of category management. For instance, Dussart (1998) noted that conventional category management prioritizes efficiency and profitability by reducing unprofitable SKUs and optimizing shelf allocation. This approach often overlooks variations in sales performance and purchasing

behavior among individual items within a category. As a result, uniform strategies may be applied across categories with fundamentally different characteristics and purchasing structures, leading to the risk of implementing strategies misaligned with actual consumer behavior. Given these limitations, category management must incorporate macro-level indicators, such as overall category sales, and a deeper analysis of the internal structure of each category. In particular, effective promotion planning within the constraints of limited shelf space requires evaluating the purchasing structure (that is, how sales are distributed across the available SKUs). Classification techniques such as Activity Based Costing analysis (ABC analysis) and Fast Normal Slow categorization (FNS categorization) are commonly used to understand SKU configurations. Their appropriate application should vary depending on the retail context (Van Kampen et al. 2012); therefore, assessing the degree of purchasing dispersion—whether sales are widely spread across multiple SKUs or concentrated in a few—is critical.

One method for quantitatively evaluating purchasing dispersion is using entropy derived from information theory. Entropy has been widely applied to analyze various aspects of consumer behavior, such as brand preference and basket composition (Guidotti 2015; Mansilla et al. 2022); it is recognized as a robust metric for quantifying behavioral diversity and concentration. For example, Morita et al. (2013) employed entropy based on information theory (Cover and Thomas 2006) to assess the variability in individual purchasing behavior using user purchase history data from a group-buying coupon website.

This present study shifts the analytical focus from individual users to product categories, specifically examining the distribution of purchases across SKUs within each category. Applying entropy in this context allows us to quantitatively measure the skewness in sales composition and, evaluate purchasing dispersion at the product category level. This approach facilitates the development of category-specific sales strategies that align with the intrinsic structural characteristics of each product category.

Product portfolio management (PPM) is a framework that enables such relative evaluation; PPM is designed to support strategic resource allocation across multiple products or business units to manage the product portfolio throughout the product life cycle (Cooper et al. 1999; Jugend and da Silva 2013). Among the tools used in PPM, the BCG matrix (Boston Consulting Group matrix) is one of the most widely adopted. It classifies a company's products or business segments based on two axes—market growth and market share—and guides strategic decision-making appropriate to each segment (Gorb et al. 2022). This classification allows firms to assess their products' market positioning across categories and visualize the direction and prioritization of sales strategies. The traditional BCG matrix effectively evaluates relative positions among product categories but fails to capture internal structural differences such as purchasing dispersion. Consequently, even categories classified within the same matrix segment may exhibit fundamentally different internal sales structures; this disparity can potentially hinder the formulation of appropriate sales strategies and the optimization of SKU configurations.

Recent studies have proposed enhancements to the traditional BCG matrix to address these limitations by integrating additional evaluation dimensions. For example, García-Vidal et al. (2023) introduced the CMQ matrix (Contribution Margin and Quantity matrix), an extended version of the BCG framework that incorporates nonfinancial indicators such as product competitiveness and quality. They demonstrated that expanding the analytical axes makes incorporating new evaluative perspectives previously unaccounted for possible, enabling a more multifaceted portfolio analysis. Similarly, Coronado-Hernández et al. (2020) combined Pareto analysis with the BCG matrix to extract product groups that significantly impact sales and profits, optimizing the product portfolio. Their approach represents an applied example in which elements such as product importance are considered alongside the two conventional axes of the BCG matrix. In line with these developments, our study seeks to construct an analytical framework that extends the BCG matrix by integrating purchasing dispersion as an internal structural evaluation axis. This approach enables the simultaneous assessment of market positioning and internal sales structure offering a more comprehensive tool for category-level management and strategy formulation.

3. Methods

This section presents a method for generating an extended BCG matrix that incorporates purchasing entropy, expanding the traditional two-dimensional framework to a four-dimensional evaluation for product category analysis. We also compared the proposed matrix with the conventional BCG matrix. Figure 1 presents an overview of the analytical framework.

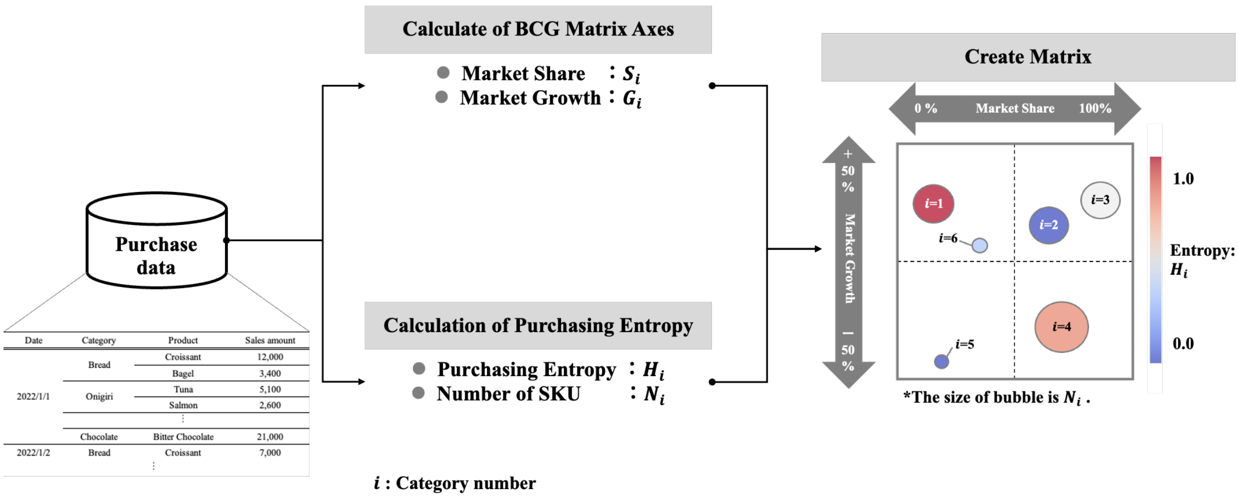

Figure 1. Overview of the methods

Figure 1 shows that this study's proposed method involves computing numerical indicators for each product category based on purchasing data, which are then used to determine the coordinates within the matrix. These indicators are subsequently visualized on the matrix for comparative analysis. Let i denote the index of a product category. The matrix comprises the following four axes.

- Market Share (Horizontal axis of the matrix) : S_i
- Market Growth (Vertical axis of the matrix) : G_i
- Purchasing Entropy (Bubble color) : H_i
- Number of SKU (Bubble size) : N_i

Market share represents the proportion of a company's sales in product category i relative to the total market sales. Market growth indicates the degree to which the market for each product category expands. Based on annual sales data at the category level, this study defines market growth as the year-over-year increase in total sales for the target market segment. Purchasing entropy is a metric that reflects how sales are distributed among SKUs within a product category. A higher value indicates that sales are dispersed across a wide range of SKUs, whereas a lower value suggests that sales are concentrated in a limited number of SKUs. The number of SKUs refers to the number of distinct items handled within each product category. It is calculated by averaging the number of unique items sold per day over the observation period. Since actual in-store SKU data were unavailable for this study, we approximated the SKU count using purchase history records.

The following section explains the calculation methods for each axis. First, Equations (1) and (2) presents the formulas used in the traditional BCG matrix to compute the market share S_i and market growth G_i.

$$S_i = \frac{A_{i,j}}{M_{i,j}} \times 100 \tag{1}$$

($A_{i,j}$: Value of own sales of product category i in year j)

$$G_i = \frac{M_{i,j} - M_{i,j-1}}{M_{i,j-1}} \times 100 \tag{2}$$

($M_{i,j}$:total market sales amount for product category i in year j)

The following section describes the calculation method for purchasing entropy. Purchasing entropy indicates the extent to which sales within a product category are dispersed across different SKUs or concentrated in a limited number of them. This measure is quantitatively evaluated using Shannon entropy, a concept from information theory. The calculation is performed according to the following three steps.

Step-1
The sales ratio of each SKU in product category i on day t, denoted as ${p_{k,t}}^{(i)}$ is calculated.

$$ {p_{k,t}}^{(i)} = \frac{{r_{k,t}}^{(i)}}{{R_t}^{(i)}} \tag{3} $$

$(r_{k,t}$:sales amount of SKU k on day $t)$

$$ {R_t}^{(i)} = \sum_{k-1}^{{N_t}^{(i)}} {r_{k,t}}^{(i)} \tag{4} $$

Step-2
The entropy (purchasing entropy) for each day ($H_{i,t}$) is then calculated. The entropy value inherently depends on the number of SKUs within the category; thus, the raw entropy is normalized by dividing it by $\log_2 {N_t}^{(i)}$is the number of SKUs in category i on day t. Through this normalization, the purchasing entropy is scaled to a range between 0 and 1, allowing for relative comparison across product categories with different SKU numbers.

$$ H_{i,t} = -\frac{\sum_{k=1}^{{N_t}^{(i)}} {p_{k,t}}^{(i)} \log_2 {p_{k,t}}^{(i)}}{\log_2 {N_t}^{(i)}} \tag{5} $$

Step-3
To obtain the overall purchasing entropy (H_i) for product category i, a weighted average of the daily entropy values is calculated using the total daily sales amount of the category ($\omega_{i,t}$) as the weight. Using a weighted average ensures that the purchasing structure on days with higher sales has a greater influence on the overall evaluation. This approach allows the resulting measure to reflect purchasing patterns under more economically significant conditions, supporting more optimal decision-making.

$$ H_i = \frac{\sum_t \omega_{i,t} H_{i,t}}{\sum_t \omega_{i,t}} \tag{6} $$

We then compare and analyze the characteristics of the generated matrix. Figure 2 compares the conventional BCG matrix and the matrix generated using the proposed method.

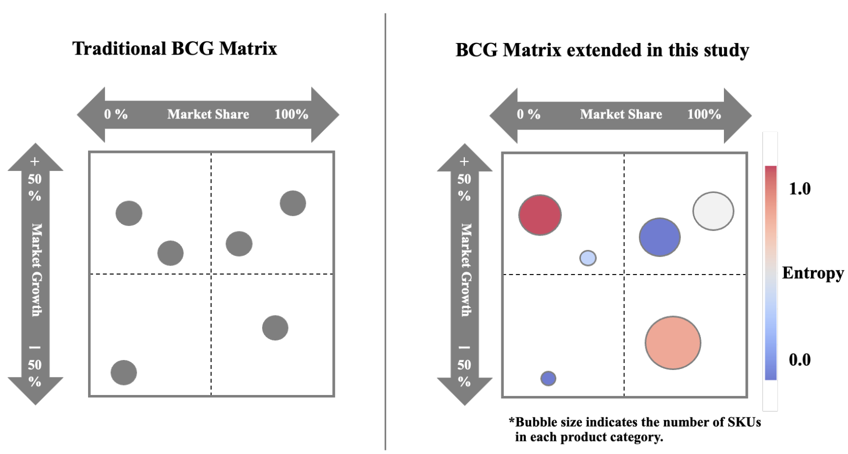

Figure 2. Comparison of the conventional and proposed matrix

Product evaluation in the conventional BCG matrix is based on two axes: market share and market growth. This approach allows for assessing a firm's market position; however, it does not capture internal purchasing structures, such as how sales are distributed across products within a category. In contrast, the matrix proposed in this study enhances the traditional two-axis framework by incorporating two additional dimensions. Specifically, purchasing entropy—representing the degree of sales dispersion—is visualized through the color of each bubble, while bubble size represents the number of SKUs handled in each product category. This approach allows the identification of a category's market position and understanding of its internal sales structure, including sales concentration and product variety.

Moreover, even product categories that fall into the same matrix segment may require different sales strategies depending on their internal structure. For instance, consider two product categories that are classified in the high market growth and high market share segments. Both categories have many SKUs; however, one may exhibit high purchasing entropy (indicating dispersed sales across many SKUs), while the other may show low entropy (indicating sales concentrated in a few items).

For the category with high purchasing entropy, strategic actions may include increasing the inventory volume across a broad range of SKUs or allocating more shelf space to support diversified sales. For the low entropy one, strengthening the promotion of the top-selling SKUs and reducing orders for the low performing ones might be more effective. Thus, the proposed method captures the internal structural differences that the conventional BCG matrix overlooks, enabling the formulation of more tailored and effective sales strategies for each product category.

4. Data Collection

This section describes the dataset to which the proposed method is applied. This study used receipt-based purchase data collected via a mobile application to analyze the sales structure of product categories and the dispersion of purchasing behavior in retail stores. The dataset was collected via CODE, a receipt-reading application provided by Research and Innovation, Inc., as part of the 2024 Data Analysis Competition sponsored by the Joint Association Study Group of Management Science. This dataset allows us to determine when, at which retail chain, and which specific SKUs were purchased. A distinctive feature of this dataset is that it is generated from user input and integrates data from multiple chains, allowing for the quantitative evaluation of market share, one of the axes in our proposed matrix. Table 1 presents the collected raw data.

Table 1 aggregated sales amounts by date, store (defined as a combination of prefecture and chain), and SKU. Subsequently, total sales amounts per SKU were calculated for each store. Table 2 presents the resulting dataset, which serves as the analytical data to which our proposed method is applied.

Table 1. Structure of the raw purchase data based on receipt

Date	**Category**	**SKU**	**Prefecture**	**Chain**	**Sales**
2022-1-1	Bread	Croissant	Tokyo	Chain A	200
2022-1-1	Bread	Croissant	Kyoto	Chain B	180
2022-1-1	Bread	Bagel	Tokyo	Chain A	360

Table 2. Structure of daily purchase data aggregated by SKU

Store	**Date**	**Category**	**SKU**	**Sales amount**
Tokyo_ Chain A	2022-1-1	Bread	Croissant	12,000
Tokyo_ Chain A	2022-1-1	Bread	Bagel	3,400
Tokyo_ Chain A	2022-1-1	Onigiri	Tuna	5,100
Tokyo_ Chain A	2022-1-1	Onigiri	Salmon	2,600

The dataset has two main issues.

1. The purchase data are based on user-entered information; thus, they may not perfectly correspond to actual purchase histories.

2. Only the prefecture and retail chain are available as retailer attributes, so it is impossible to identify the store where the purchase occurred.

As such, certain challenges remain regarding the validity of the dataset. Nonethless, the data are derived from user input; therefore, they are effective for performing quantitative comparisons across chains, which aligns with the objective of the BCG matrix as a tool for relative market evaluation. Considering both the advantages and limitations, this study analyzes a preliminary implementation of the proposed method.

This section outlines the conditions under which the analysis was conducted. The primary analysis uses data from January 1 to December 31, 2022, to construct the proposed matrix and evaluate the sales structure of convenience store product categories. In addition, data from 2021 are used to calculate year-over-year market growth at the category level. The geographical scope of the analysis is limited to the Tokyo Prefecture for two reasons. First, Tokyo recorded the highest number of transactions; second, it is a region where small-scale retail stores, such as convenience stores, are densely concentrated. Due to the dataset's characteristics, purchase data at the individual store level are unavailable; therefore, the analytical unit combines the prefecture and retail chains. The four convenience store chains selected for this analysis vary in scale and store format within Tokyo; Table 3 summarizes their characteristics. The product categories include 28 categories that recorded over 3 million transactions in 2022 based on product category information logged in the mobile application. We established this threshold to ensure a sufficient number of purchase records per category while maintaining the readability of the matrix.

The above summarizes this study's analysis conditions.

Table 3. Overview of the analyzed convenience store chains

Chain	**Number of stores**	**Characteristics**
Chain A	Large store network	One of Japan's most prominent convenience store chains, offering a wide variety of products
Chain B	Medium-scale store network	A nationwide chain with a medium-scale store network, smaller in scale than Chains A and C
Chain C	Large store network	A major convenience store chain with a comparable store network to Chain A
Chain D	Limited store network	Operated by the same company as Chain C, focusing on health-conscious products

5. Results and Discussion

This section presents the results of the analysis and discusses its implications. Section 5.1 processes the data described in Section 4 using the methodology outlined in Section 3. The resulting numerical values of each matrix axis are used to compare differences across retail chains. Section 5.2 visualizes the matrices for each chain and analyzes the sales structures specific to each case. Based on these findings, Sections 5.3 and 5.4 discuss the effectiveness and implications of the proposed method.

5.1 Numerical Results

First, we present the results of applying the methodology described in Section 3 to each retail chain to compute the values of the matrix axes. Several representative product categories were selected; Table 4 presents, the computed values for each. Table 4 presents two example categories: onigiri (rice balls) and udon (Japanese noodles). The results for each category are discussed below.

Although market growth is negative in the onigiri category, all chains exhibit many SKUs and high purchasing entropy, indicating a dispersed purchasing pattern, where a wide variety of SKUs are offered, and sales are relatively evenly distributed. Onigiri has traditionally been a staple food in Japan. However, recent shifts in dietary habits—such as the diversification of staple foods, including bread—have contributed to a downward trend in overall market sales. Despite this, the category continues to represent a significant share of sales, prompting retailers to maintain a wide assortment of SKUs and design their product offerings to encourage dispersed purchasing across items. This trend is common across all four chains; however, Chains C and D have fewer SKUs than Chains A and B, suggesting that a dispersed purchasing structure is present even with a limited assortment.

In the udon category, Chain A holds a dominant market position with a market share of 76.9%. In comparison, Chains B, C, and D each maintain single-digit market shares, with substantially lower SKU counts and purchasing entropy. For Chains C and D, the number of SKUs is less than two, and the purchasing entropy is around 0.15, indicating that sales are highly concentrated on a minimal number of SKUs. One possible reason for Chain A's strength is its broad assortment of udon products, which may help accommodate a wide range of consumer preferences. Udon varies not only in serving temperature—such as chilled versions in summer and hot ones in winter—but also in flavor and toppings depending on the region. By offering a large number of SKUs, Chain A may be better positioned to reflect this inherent diversity in consumer demand, which could contribute to both a broader purchasing distribution and stronger market share.

Table 4. Matrix axis values by chain for selected product categories

Category	Chain	Share: S_i	Growth: G_i	Entropy: H_i	SKU: N_i
Onigiri	A	47.96	−4.44	0.89	467.30
Onigiri	B	1.04	−4.44	0.87	27.73
Onigiri	C	17.52	−4.44	0.87	200.31
Onigiri	D	1.01	−4.44	0.82	12.79
Udon	A	76.89	−1.57	0.82	58.46
Udon	B	0.31	−1.57	0.09	1.70
Udon	C	6.73	−1.57	0.63	9.40
Udon	D	0.50	−1.57	0.23	1.40

These findings illustrate that the sales structure can vary significantly across chains, even within the same product category. The matrix indicators visualize differences in the SKU strategy and purchasing trends. In particular, analyzing the purchasing entropy and the number of SKUs reveals whether a category follows a high-variety, dispersed sales structure (such as udon in Chain A) or a low-variety, concentrated sales structure (such as udon in Chains C and D).

As observed above, both categories exhibit distinct sales patterns across retail chains. To statistically verify whether these differences are significant, a non-parametric statistical test was conducted. While the entropy values shown in Table 4 represent the weighted average of daily entropy per chain, the present analysis focuses on the distribution of daily entropy values to examine inter-chain differences. First, the normality of the daily entropy distributions for each chain was assessed using the Shapiro–Wilk test, and the results confirmed that none of the distributions followed a normal distribution. Based on this and considering that the data consist of more than three independent groups, the Kruskal–Wallis test, a non-parametric alternative to ANOVA, was employed. The analysis yielded extremely small p-values: $p = 1.13 \times 10^{-11}$ for the onigiri category and $p = 1.04 \times 10^{-46}$ for the udon category. These results statistically substantiate that the sales structures significantly differ across chains in both product categories.

5.2 Graphical Results

Next, we describe the matrices generated by the proposed method. Figure 3 presents the matrices for each of the four retail chains.

Section 5.1 compared the chains based on the numerical indicators. This section examines the sales structure of each chain through the visualized matrices.

Chain A has a consistently high market share across product categories, indicating a strong market presence. Regarding purchasing entropy, most categories displayed a dispersed pattern, suggesting that sales were distributed across various SKUs. In contrast, specific categories—such as karaage (fried chicken)—are positioned in the lower left of the matrix with relatively low entropy values (around 0.6). These categories also exhibit small bubble sizes, implying a smaller number of SKUs compared to others.

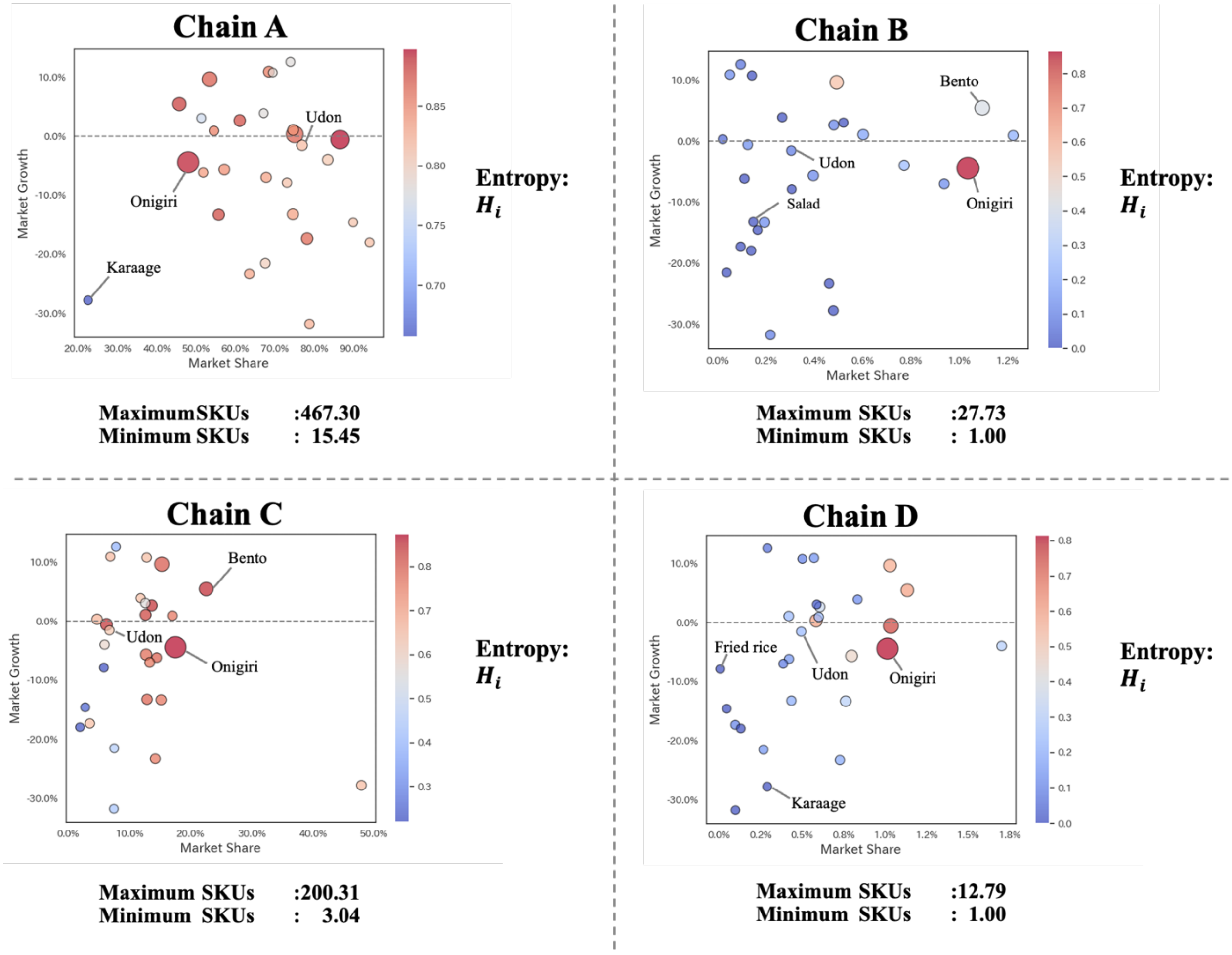

Figure 3. Visualization of matrices for each chain

Chain B generally exhibits a low market share, indicating a weaker competitive position in the market. The purchasing entropy varies considerably across categories. For instance, while the onigiri category shows many SKUs and a dispersed purchasing structure, other categories (such as bento and salads) demonstrate a more concentrated structure with fewer SKUs. Chain B maintains a low-variety, concentrated sales structure relative to the other chains.

Chain C exhibits a high market share in several categories, reflecting strong market positioning within the matrix. It also maintains high purchasing entropy across many categories, including onigiri and bento, which indicates a trend toward dispersed purchasing. These characteristics, in conjunction with many SKUs, suggest that Chain C adopts a high-variety, dispersed sales structure that leverages a broad product lineup to meet diverse consumer needs.

Chain D exhibits low market share across all categories, placing many in marginal positions within the BCG matrix. The chain also handles a few SKUs, and its purchasing entropy is comparatively low; in many categories, sales are concentrated in a small number of SKUs. Notably, categories like fried rice and karaage exhibit minimal SKU counts and near-zero entropy, indicating a mono-product dominant structure. In summary, Chain D follows a concentrated sales strategy focused on a narrow product range.

These observations illustrate each chain's distinct sales structure as visualized through the matrix, offering insights into their SKU strategies and purchasing trends.

5.3 Proposed Improvements

This section discusses the current limitations of the proposed method and outlines the directions for future development. This study's analytical framework was designed to visualize the internal structure of product categories using a matrix-based approach, thereby supporting strategic decision-making in sales planning; however, we identified the following issues in the current implementation due to methodological and data-related constraints. The first limitation is that the method does not consider the temporal variations in purchasing entropy**.** Daily purchasing data are used to compute entropy values for each category; however, these values are aggregated into a single annual figure via a weighted average based on sales volume. As a result, temporal fluctuations in purchasing behavior, such as seasonal trends or sudden changes, are not captured. Time-series analysis of entropy would detect shifts in consumer behavior, enabling retailers to determine more precisely when to implement strategic interventions such as promotional campaigns or assortment changes; however, the method does not currently incorporate this form of temporal analysis. The second limitation is the inability to perform detailed comparisons between individual stores or across regions. The fundamental constraint of the dataset used in this study is that it does not include identifiers for individual store locations; thus, the unit of analysis was limited to combinations of the prefecture and retail chain. Furthermore, due to significant imbalances in the amount of data available across prefectures, conducting a comparative analysis of purchasing structures between different regions was not feasible. As a result, this study could not investigate variations in store-level strategies or regional differences in consumer behavior within the same retail chain.

Given these two issues, we propose the following two directions for future research to enhance our framework's capacity to support strategic planning in retail settings based on category-specific purchasing structures. The first direction is incorporating the time-series analysis of purchasing entropy. As mentioned, analyzing the temporal dynamics of entropy can reveal the timing of behavioral changes, which is crucial for developing timely and effective interventions. Future studies should treat entropy as a time-series variable rather than aggregating daily entropy values via weighted averages, enabling the detection of anomalies or abrupt changes in consumer behavior. The second direction is applying the method to alternative datasets and regional comparisons. Although the limitations of the dataset constrained the current analysis, the proposed framework for matrix construction and visualization was shown to function effectively. Future research should apply the method to datasets with higher granularity, such as those allowing store-level analysis or containing balanced data across regions. This approach would allow researchers to conduct broader and more generalizable analyses.

5.4 Validation

This section examines the validity of the proposed methodology and its results. The matrix developed in this study visualizes purchasing entropy and SKU configurations within product categories, aiming to capture retail chain's sales structure and support product strategy development. This study's findings are considered consistent with those of prior research.

The convenience store chains analyzed in this study can broadly be classified into two types: large-scale chains, such as Chain A and Chain C, and medium- to small-scale chains, such as Chain B and Chain D. Sections 5.1 and 5.2 showed that the large-scale chains are characterized by high SKU counts and high purchasing entropy across multiple product categories, indicating a dispersed purchasing structure. This observation aligns with the findings of Marshall (2016), who reported that major convenience store chains such as Seven-Eleven Japan manage a large number of SKUs by continuously introducing new products and deploying a wide variety of private-label items.

In contrast, medium- to small-scale chains tend to exhibit a concentrated purchasing structure; they have low SKU counts and low purchasing entropy, where sales are concentrated in a limited number of SKUs. This pattern is similar to the "scaling law" in sales quantity described by Fukunaga et al. (2016). Based on POS data analysis, their study demonstrated a proportional relationship—known as Taylor's law—between sales quantities' mean and standard deviation (SD). Specifically, categories with fewer total sales tend to exhibit Poisson-like scaling, where the SD was proportional to the square root of the mean. This finding implies that in categories with lower sales volumes, sales are more likely to be concentrated in specific SKUs—a tendency observed in the results for Chain B and Chain D in this study.

Overall, this study's insights are consistent with the existing literature that explores similar perspectives, suggesting that the proposed method offers a valid framework for understanding purchasing structures at the product category level. Future research could use store-level purchasing data rather than user-submitted data to enhance the framework's reliability and practical applicability through empirical validation.

6. Conclusion

This study proposed a novel analytical framework to support strategic decision-making in retail environments, such as convenience stores by visualizing the purchasing structure of product categories. Building on the traditional BCG matrix, the proposed method incorporates the number of SKUs and purchasing entropy as additional evaluation axes. This approach enables the visualization of intra-category purchasing bias and a deeper understanding of internal structural characteristics not captured by conventional BCG analysis. We applied the proposed method to real-world purchase data collected in 2022 from four major convenience store chains operating in Tokyo. Through this application, the study successfully visualized the differences in sales structures across product categories and retail chains. These results highlight how the same product category can exhibit divergent sales patterns depending on the chain, illustrating the proposed matrix's analytical power. The primary contribution of this study lies in developing a multidimensional analytical tool that incorporates purchasing dispersion into portfolio evaluation. This framework offers practical value for optimizing SKU configurations and planning sales strategies based on empirical purchasing patterns. Future work should focus on enhancing the robustness and applicability of the proposed method by applying it to high-resolution transaction data from actual stores and conducting temporal and regional analyses to investigate the dynamic and spatial dimensions of purchasing behavior.

Acknowledgements

We utilized the purchase data collected by CODE, a receipt-reading application provided by Research and Innovation, Inc. through the 2024 Data Analysis Competition sponsored by the Joint Association Study Group of Management Science.

References

Basuroy, S., Mantrala, M. K. and Walters, R. G., The impact of category management on retailer prices and performance: theory and evidence, *Journal of Marketing*, vol. 65, no. 4, pp. 16-32, 2001.

Cooper, R. G., Edgett, S. J. and Kleinschmidt, E. J., New problems, new solutions: making portfolio management more effective, *Research-Technology Management*, vol. 43, no. 2, pp. 18-33, 2000.

Coronado-Hernández, J. R., Ospina-Mateus, H., Baron-Villamizar, J., Vidal-Pacheco, L., Piñeres-Castillo, A. P., Rojas-Millán, R. H. and Silva, J., A method to rationalize the product portfolio in retail stores, *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*, pp.509-515 , vol.643, Springer, Singapore, 2020.

Cover, T. M. and Thomas, J. A., *Elements of Information Theory*, 2nd Edition, Wiley-Interscience, New York, 2006.

Dussart, C., Category management: strengths, limits and developments, *European Management Journal*, vol. 16, no. 1, pp. 50-62, 1998.

Fukunaga, G., Takayasu, H. and Takayasu, M., Property of fluctuations of sales quantities by product category in convenience stores, *PLoS ONE*, vol. 11, no. 6, e0157653, 2016.

García-Vidal, G., Sánchez-Rodríguez, A., Pérez-Campdesuñer, R. and Martínez-Vivar, R., Contribution margin and quantity matrix to analyze the product portfolio in the context of SMEs: criticism of the BCG matrix and its alternatives, *Cogent Business & Management*, vol. 10, no. 3, pp. 1-16, 2023.

Gecili, H. and Parikh, P. J., Joint shelf design and shelf space allocation problem for retailers, *Omega*, vol. 111, 102634, 2022.

Gorb, O., Dorohan-Pysarenko, L., Yehorova, O., Yasnolob, I. and Doroshenko, A., Boston Consulting Group matrix: opportunities for use in economic analysis, *Scientific Horizons*, vol. 25, no. 7, pp. 20-30, 2022.

Guidotti, R., Coscia, M., Pedreschi, D. and Pennacchioli, D. Behavioral entropy and profitability in retail, *Proceedings of 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, Paris, France, pp. 1-10, October 19-21, 2015.

Jugend, D. and Da Silva, S. L., Product-portfolio management: a framework based on methods, organization, and strategy, *Concurrent Engineering: Research and Applications*, vol. 22, no. 1, pp. 17-28, 2014.

Kulkarni, V., A study of impact of merchandise variety and assortment on shopping experience of customer in convenience stores in organized retail in India, *International Journal of Management*, vol. 4, no. 1, pp. 85-94, 2013.

Lee, K. K., A study on the competitive advantage strategy of convenience store (CVS) in Japan: Focusing on 7-Eleven, *East Asian Journal of Business Economics*, vol. 9, no. 1, pp. 1-16, 2021.

Mansilla, R., Smith, G., Smith, A. and Goulding, J., Bundle entropy as an optimized measure of consumers' systematic product choice combinations in mass transactional data, *Proceedings of the 2022 IEEE International Conference on Big Data,* pp. 1044-1053, Osaka, Japan, December 17-20, 2022.

Marshall, D., Convenience stores and discretionary food consumption among young Tokyo consumers, *International Journal of Retail & Distribution Management,* vol. 44, no. 10, pp. 1013-1029, 2016.

Morita, H., Shirai, Y. and Nakamoto, M., Analysis method for flash marketing incorporating diversification on purchasing, *Proceedings of the 2013 Fall National Conference of the Japan Society for Management Information*, pp. 329–332, Kobe, Japan, October 26-27, 2013.

Pak, O., Ferguson, M. E., Perdikaki, O. and Wu, S.-M., Optimizing stock-keeping unit selection for promotional display space at grocery retailers, *Journal of Operations Management*, vol. 66, no. 5, pp. 501-533, 2020.

Sakoda, G., Takayasu, H. and Takayasu, M., Data science solutions for retail strategy to reduce waste keeping high profit*, Sustainability*, vol.11, no.13, pp.1-30, 2019.

Van Kampen, T. J., Akkerman, R. and van Donk, D., SKU classification: a literature review and conceptual framework, *International Journal of Operations & Production Management*, vol. 32, no. 7, pp. 850-876, 2012.

Verhoef, P. C., Lemon, K. N., Parasuraman, A., Roggeveen, A., Tsiros, M. and Schlesinger, L. A., Customer experience creation: determinants, dynamics and management strategies, *Journal of Retailing*, vol. 85, no. 1, pp. 31-41, 2009.

Biographies

Ryuzen Kawabe is a graduate student in the College of Systems Engineering and Science at the Shibaura Institute of Technology, Japan. His research interests include social simulation, narratives communication, and data science.

Kiyotaka Yoshimura is a graduate student in the College of Systems Engineering and Science at the Shibaura Institute of Technology, Japan. His research interests include the foundations of quantum mechanics, and quantum information science from physical, mathematical, informational, and philosophical view points.

Yusuke Goto is a Professor in the College of Systems Engineering and Science at the Shibaura Institute of Technology, Japan. His research interests include systems engineering, risk analysis, social simulation, and social data science. He received a Ph.D. in Engineering from Waseda University in 2010.