International Journal of Knowledge and Systems Science
Volume 15 • Issue 1

Market Basket Analysis of a Health Food Store in Thailand:

A Case Study

Kanokwan Singha, Panyapiwat Institute of Management, Thailand

Parthana Parthanadee, Kasetsart University, Thailand*

https://orcid.org/0000-0002-8645-1595

Ajchara Kessuvan, Kasetsart University, Thailand

https://orcid.org/0000-0003-2960-7706

Jirachai Buddhakulsomsiri, Sirindhorn International Institute of Technology, Thammasat University, Thailand

ABSTRACT

This article presents a market basket analysis of a health food store in Thailand. The analysis identifies data attributes that frequently occur together in the dataset. Frequent occurrences of data attributes representing customer purchasing behaviors are extracted as association rules using the frequent pattern growth algorithm. The generated associations are evaluated using standard measures based on occurrence counts and an additional financial measure. Marketing strategies in the form of cross-selling pairs of specific products are then designed based on the data attributes appearing in the significant associations. The cross-selling products are offered at discounted prices and promoted in marketing campaigns. A break-even analysis is performed to estimate the required number of additional sales volumes from each marketing campaign to compensate for the discounted prices. The presented use case demonstrates the effectiveness of extending the market basket analysis to include a financial measure that can lead to practical marketing campaigns.

KEYWORDS

Association rule, Frequent pattern growth, Health food, Market basket analysis, Marketing strategy, Retail

MARKET BASKET ANALYSIS OF A HEALTH FOOD STORE IN THAILAND

A Case Study

Today, businesses must compete within and across industries in a competitive consumer market environment. It has become increasingly crucial for a business to maintain its competitive edge. One key competitive advantage involves insights into customers' purchasing behaviors, which helps the company develop new marketing strategies to increase customers' purchases and revenue. Notably, in

DOI: 10.4018/IJKSS.333617

*Corresponding Author

the retail industry, where direct contact with consumers occurs, having more insights and knowledge of customer purchasing behavior enables the company to strengthen its competitive advantage.

One widely used technique to study retail customer behavior is market basket analysis. This analysis can uncover the patterns of the association behind customers' purchasing behavior (Annie & Kumar, 2012). It can be applied to the retail industry and many other industries, including bioinformatics and education, as discussed in Kaur and Kang (2016). Marketing practitioners have accepted market basket analysis for its usefulness in helping businesses understand their customers' behaviors and thus operate more profitably. Specifically, an analysis output in the form of association rules containing frequent itemsets provides sales and marketing departments with valuable information for designing effective strategies.

The market basket analysis framework has been developed specifically to serve retail and marketing purposes by uncovering sets of items that customers frequently purchase together. The co-occurrence of items or products in a single shopping receipt can be found in transactional data. The analysis outcome is a set of association rules in the form of $A \rightarrow B$. These rules describe the relation between two items: "If a customer purchases item A, then he or she also purchases item B." Upon discovering these relationships, retailers can use this knowledge to offer sales promotions to enhance customer purchases.

Data mining is a technique for discovering traits and patterns to find associations among the items in a sizeable transactional dataset. Regarding mining algorithms, many alternatives exist, including Apriori (Agrawal & Srikant, 1994), frequent pattern growth or FP-growth (Han et al., 2000), Eclat (Han et al., 2001), K-Apriori (Annie & Kumar, 2011), and the two-way cooperative collaborative filtering (CF) approach (Hwang & Lee, 2021). Along with the output, specific measures are used to restrict the number of association rules found. Two common measures include support and confidence. Support is defined as the number of transactions containing both items over the total number of transactions, i.e., the relative frequency of item occurrence. Confidence is the likelihood of purchasing another item when one item is purchased. It is equivalent to the conditional probability of one product purchase given another product purchase (Raorane et al., 2012). Because of its effectiveness, association rule mining is widely used in various applications. Zhang and Guo (2011) developed a novel approach that evaluates the quality of association rules using TOPSIS, a well-known multiple-criteria decision-making technique. Herawan et al. (2012) proposed an SLP-Growth algorithm to identify meaningful association rules for engineering students who suffer from examination anxieties. Tanantong and Ramjan (2021) used the Apriori algorithm to extract valuable insights related to demand and supply patterns from social media data in Thailand. The study aimed to leverage social media's rich and diverse content to gain insights into consumer preferences and market trends to improve social media marketing.

Two main approaches used in the assessment of the market basket analysis are classical and temporal. The classical method does not consider the order in which the products are selected for purchase. It only produces rules based on the principle "If a transaction contains an item A, then it also contains item B." By contrast, the second approach is based on temporal association, meaning that it considers the time aspect. This association type is divided into sequential pattern analysis, where the former is based on a short sequence of transactions and the latter on one long event sequence (Weiß, 2014). Using either type, one can choose based on the desired outcome or computational efficiency. Kamakura (2012) proposed using RFID tags mounted to the shopping cart to study the sequential pattern. The shopping trip's route could then be illustrated using computer software. This analysis allows the researcher to identify the sequence in which products are purchased and determine the sequential patterns accordingly.

The benefits of the market basket analysis for local and transnational retailers depend on which aspects of shopping transactions are selected for leverage. Discovery of the set of items frequently purchased together through the obtained association rules can lead to a more effective marketing strategy or promotion campaign. For instance, advertising and discounting paired items may be offered (Zekić-Sušac & Has, 2015). Methods for setting original product prices include brand quality-based, value-based (Monroe & Dodds, 1988), and cost-based (Xu, 2009) approaches, similar to discounted pricing. Another alternative entails using the results to launch targeted marketing campaigns for specific customer groups and improve customer loyalty (Moodley et al., 2019).

Marketing campaigns can be tailored not only to specific customer segments but also in terms of seasonality. Regarding sales and marketing, segregating transactional data into seasons can help determine customers' behavior during each season (Surjandari & Seruni, 2010). Performing market basket analysis on such a dataset could produce supporting information for designing effective promotion schemes and cross-selling strategies for each season (Ayu et al., 2018). In addition, the outcome of the analysis could provide information for designing the retail store's layout to guide customers more logically through their shopping route. This guidance can improve customers' shopping activity flow and convenience, potentially leading to loyalty (Ayu et al., 2018).

Notably, boosting sales of a particular item by reducing its price without subsidization from its supplier —whether to create market differentiation or resolve aging stocks—only implies that the business is reducing its margins (Annie & Kumar, 2012). The company could raise the margins of other non-promoted items frequently purchased along with the promoted item to compensate for the promoted item's discounted price. Furthermore, some other applications are worth exploiting. An example is bulk purchase offering, which helps businesses save on the administration costs of transactions by offering campaigns for customers to buy in bulk quantities, consisting of multiple complementary products, in a less frequent manner (Svetina & Zupančič, 2005). Another activity the market basket analysis provides involves educating salespersons to ensure all retail employees are familiar with the item pairs offered together. This activity prepares salespersons to make customer recommendations to improve service satisfaction.

Several researchers have reviewed recent studies on market basket analysis. First, a comprehensive review by Alawadh and Barnawi (2022) discussed various applications of market basket analysis and the use of association rule algorithms to perform the analysis. They reviewed commonly used measures of association rules, such as support, confidence, lift, and leverage. These measures mainly consider the number of occurrences or counts of items frequently purchased together. Samboteng et al. (2022) performed a market basket analysis on a transactional dataset in the food industry. The purpose was to provide users with sales recommendations based on itemsets that are frequently purchased together. They implemented the Apriori algorithm to measure the significance of association using support and confidence. Bohl et al. (2023) introduced a novel approach called the weighted Shapley value index for market basket analysis. This approach includes additional revenue from items that are frequently purchased together. To obtain an accurate estimate of the total revenue contributed by an item, the item's value is weighted based on the frequency with which it is found to be the most expensive purchase in the basket. The authors demonstrate the effectiveness of the proposed approach using a transactional dataset of slow-moving consumable goods.

This article presents a case study using the market basket analysis on a transactional dataset from a health food store in Thailand. The objective is to find important associations between data attributes in the dataset that can lead to practical marketing strategies. However, the associations considered in this article do not consider the sequence of items purchased on the same receipts. This is because the item's sequence on a receipt only represents the order in which the items are scanned at the cashier counter. This sequence does not necessarily represent the order in which the customers pick up the items. Our contributions can be stated as follows:

(1) We extend the widely used market basket analysis to include other data attributes in the associations. Unlike other research studies that only focus on finding associations between itemsets (or products), our study uses additional information, including product category and time of purchase, combined with the product as data attributes in the associations. In other words, we aim to find associations between product categories, between products, and between product and time of purchase.
(2) As mentioned above, most research studies measure the importance of association using parameters based on the number of occurrences, i.e., the number of transactions. In this context, we propose that the significance of association should also be measured in terms of monetary value, i.e., the sales amount. This measure will lead to financially significant associations for retail businesses.
(3) We demonstrate our approach using a case study dataset that leads to market strategies through cross-selling product campaigns with discounted prices that can be used in practice. Finally, we conduct a break-even analysis to estimate the additional sales volume that can compensate for the cost of discounted offerings.

METHODOLOGY

The methodology described in this section adheres to the following notations:

n Total number of transactions (or receipts) in the dataset
A , B Data attributes, including the product category, product, and time of purchase variables
$\left|A\right|$, $\left|B\right|$, $\left|A+B\right|$ Number of transactions containing attribute *A*, attribute *B*, and both attributes *A* and *B*, respectively
$Support_{AB}$ Percentage of transactions containing both *A* and *B* amongst all n transactions
ps Difference between the observed frequency and the expected frequency of an association between attributes *A* and *B*. A positive value of ps indicates that an association is more frequent than expected.

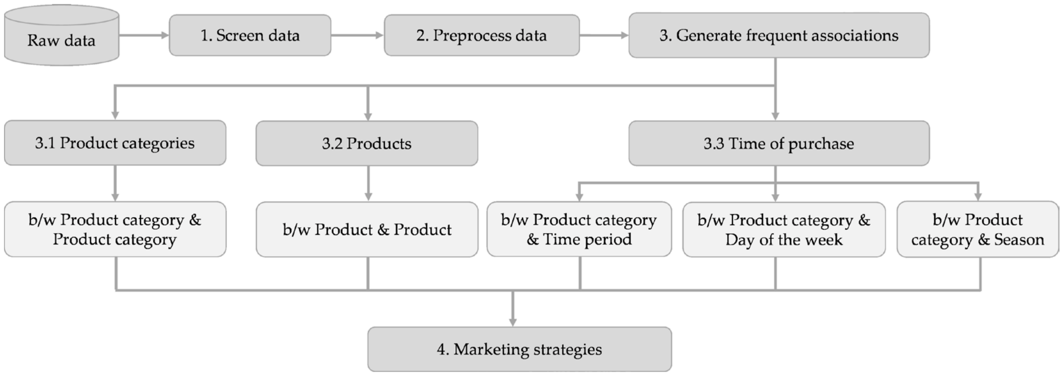
Data Collection

The dataset in this study contains transactional data from a health food store belonging to a retail chain in Thailand. The selected products include fresh vegetables, fresh fruits, and organic products. The transactions span 12 months, with a total of 112,991 transactions and 333,853 items sold. Amongst these, there are 1,518 unique items. The total value of all transactions is 14,975,283 THB, with an average of approximately 1.25 million THB per month.

Market Basket Analysis Framework

The market basket analysis framework is shown in Figure 1. First, we screened the transactional data to contain only 238 unique items, each with an annual sales amount of at least 20,000 THB. We then pre-processed the data to retain only the data attributes used in the analysis. These attributes include Transaction ID, receipt date, day of the week, time of purchase, product category code and name, product code and name, unit price (in Thai Baht, THB), product quantity, and total sale amounts. We then analyzed the dataset using the FP-growth algorithm to generate frequent associations in three aspects: (1) finding frequent associations between two product categories, (2) finding frequent associations between two unique products, and (3) finding frequent associations between product category and time of purchase attributes. The time of purchase attribute was divided into time of day, day of the week, and season. The time of day was defined as morning (8:00–11:00), late morning (11:01–13:00), afternoon (13:01–15:00), and

Figure 1. Analysis framework



late afternoon to evening (15:01 to 19:00). All days of the week were included, from Monday to Sunday. Seasons in Thailand include the rainy, winter, and summer seasons. We then combined the resulting frequent associations to derive market strategies that can increase the sales of the selected products in the marketing campaigns.

Market Basket Analysis Using FP-Growth Algorithm

By pivoting the data concerning its transaction ID, we generated frequent associations to identify frequent data attributes and their respective measures, namely support, *ps*, and the association's total sale amount, which is the proposed measure in this article. The association generation process involves applying the FP-growth algorithm to extract frequent pairs of data attributes: product category, unique product, and time of purchase attributes.

Regarding the algorithm, the FP-growth implements the frequent-pattern tree approach, also known as the FP-tree method. The algorithm scans the dataset twice to generate the constructed tree. The first scan identifies the frequent attributes that satisfy the minimum support criterion specified by the user. The second scan then determines the next attribute that follows from the first to create a complete tree structure (Liu & Guan, 2008).

While other methods, such as Apriori and K-Apriori, can efficiently extract associations from a large dataset, this article uses the FP-growth algorithm. This is because, in other methods, the generated association rules take the form $A \rightarrow B$. Upon reviewing the generated rules, it is evident that to find frequent data attributes, the pair $A \rightarrow B$ is identical to that of $B \rightarrow A$. There is no identifier for the sequence in which data attributes were scanned. These contrariwise pairs can be identified from two duplicate rules' reversed order of data attributes. These rules also have equivalent support values, calculated from the number of transactions that include both data attributes A and B out of the total number of transactions. Therefore, without considering the order of data attribute appearance in the association, one of the analogous associations is omitted through a cleansing process, which searches for associations with swapped orders of precedent – data attribute A – and subsequent – data attribute B – and only retains the contrariwise pair.

We computed the support and *ps* measures values using the FP-growth algorithm. Table 1 is an excerpted summary of one pair taken from the transactional database of the retail store that provided the data for this study to demonstrate the definition of the measures used in the data analysis. In this example, the data attributes are simply items (or products).

Support refers to the fraction of transactions containing both A and B. Hence, for the pair of iceberg lettuce (pack) 300g and cabbage (pack) 500g, the support can be determined from Equation 1.

Table 1. Excerpt summary of one itemset from the transactional data of the case study retail store

Item *A*	Item *B*	Number of transactions		
		Containing *A*	Containing *B*	Containing A and *B*
Iceberg lettuce (pack) 300g	Cabbage (pack) 500g	$\lvert A \rvert$	$\lvert B \rvert$	$\lvert A+B \rvert$

$$Support_{AB} = \frac{\lvert A+B \rvert}{n} \tag{1}$$

Another selected measure is *ps*, which is computed as

$$ps = \lvert A+B \rvert - \frac{\lvert A \rvert \cdot \lvert B \rvert}{n} \tag{2}$$

where *ps* is the difference between the *observed* number of transactions, $\lvert A+B \rvert$, and its expected value, $\frac{\lvert A \rvert \cdot \lvert B \rvert}{n}$. Essentially, *ps* indicates how much more frequent a pair of itemsets in an association is compared to a scenario in which there is no association between them. In addition, *ps* is similar to another measure of performance, *lift*, which is the ratio of the "observed" frequency to the "expected" frequency of itemsets. In this application, *ps* is relatively simple for users to understand as it is expressed in terms of the number of transactions.

In addition to the measures of performance expressed in terms of the number of transactions, namely support and *ps*, the monetary value of an association is used to evaluate its importance. Specifically, it is computed by multiplying the number of transactions containing both itemsets, $\lvert A+B \rvert$, and each item's total price(s) in the association. With this measure, an association is assessed in terms of its frequency and its total value. Finally, marketing campaigns that involve frequent and high-value itemsets can be designed to increase the sale of the items in the association.

RESULTS AND DISCUSSION

Results of Market Basket Analysis Using FP-Growth Algorithm

The FP-growth algorithm used to generate association results for the case study dataset is subject to the minimum support specified by the user. Appropriately setting the minimum support results in a number of associations that is neither too low nor too high. In this article, the minimum support is set at 0.1% to ensure frequent itemsets are not excluded from the result.

Associations Between Two Product Categories

The analysis results between product categories found 55 associations from a total of 101,580 transactions (i.e., receipts). First, we preprocessed the data to include only receipts in which customers purchased items from two or more product groups, reducing the number of receipts to 36,485. Among the 55 associations, the significant associations are those with a positive *ps*, which indicates that the associations occur more frequently than expected (see Table 2). We ranked the associations by the total sales amount. For example, the first association is between the fresh vegetables and fresh fruits

Table 2. Significant associations between two product categories

No.	Category A	Category B	Support (%)	*ps*	Sales (THB)	Sales (%)
1	Fresh vegetables	Fresh fruits	8.46	319.69	563,095	3.76
2	Juice & herbal drinks	Jam & honey	3.26	112.83	498,042	3.33
3	Juice & herbal drinks	Tea, herbal tea, & coffee	2.21	98.42	443,013	2.96
4	Juice & herbal drinks	Nut products	1.22	34.59	322,438	2.15
5	Jam & honey	Tea, herbal tea, & coffee	1.18	232.52	282,851	1.89
6	Juice & herbal drinks	Cereal drinks	1.70	210.93	237,889	1.59
7	Fresh vegetables	Sauce & salad dressing	3.72	119.70	180,890	1.21
8	Processed vegetables	Sauce & salad dressing	1.31	165.68	38,507	0.26

categories, with a support of 8.46%. This implies that, on average, 8 in every 100 receipts contain items in these categories together. For this association, the number of transactions is 319 times greater than expected, with a total sales amount of 563,095 THB compared to the total annual sales of all product categories (14,975,283 THB) or 3.76%.

Associations Between Two Products

The analysis results between products found 155 associations out of 101,580 receipts. As above, we preprocessed the total number of receipts to include only those in which customers purchased two or more items per receipt. This reduced the number of receipts to 59,935. Out of 155 associations, 42 associations had a positive *ps*. Table 3 lists examples of five associations from various product items. For example, the first association between iceberg lettuce and cabbage has a support of 3.27%, which occurs 518 times higher than expected. This association has total sales of 114,000 THB or 0.76% of the total annual sales.

Product Category and Time-of-Purchase Associations

An analysis between product category and each time-of-purchase variable found the following results: 15 associations with time of day, 21 associations with day of the week, and 19 associations with season. Table 4 contains the most significant association between one product category and the three time-of-purchase variables. For example, the association between fresh vegetables and the late

Table 3. Examples of associations between two items

Item A	Item B	Support (%)	ps	Sales (THB)	Sales (%)
Iceberg lettuce	Cabbage	3.27	518.30	114,000	0.76
Fried pumpkin crackers	Fried carrot crackers	2.25	1222.05	87,925	0.59
Chinese cabbage	Spinach	1.80	537.01	72,450	0.48
Sweet milk tablets	Chocolate milk tablets	1.55	871.03	50,970	0.34
Japanese cucumber	Cos salad	1.15	337.99	42,400	0.28

Table 4. Examples of associations between product category and time of purchase variables

Category	Time of purchase	Support (%)	*ps*	Sales (THB)	Sale (%)
Fresh vegetables	Late morning	15.11	2197.52	1,271,857	8.49
Fresh vegetables	Saturday	11.05	1831.50	949,130	6.34
Fresh vegetables	Rainy season	23.32	1533.29	2,001,985	13.37

morning period has a support of 15.11%, which is 2,197 times greater than its expected value. This association has total sales of 1,271,857 THB, or 8.49% of the total annual sales.

Marketing Strategies

Marketing Campaigns From Top-Selling Associations

We used the associations involving top-selling products from the previous sections to design promotional marketing campaigns. Specifically, cross-selling product campaigns were derived from these associations, as seen in Table 5.

Each row of the table presents a pair of cross-selling products chosen from their respective categories. For example, from the association between fresh vegetables and fresh fruits categories, beetroot and passion fruit were selected for one of the cross-selling product campaigns. We chose this pair to boost the sales of a regular item (beetroot) by taking advantage of a popular item (passion fruits). Table 6 shows that passion fruit has a higher purchase volume than beetroot. Specifically, the ratios are 3:1, 2.3:1, and 5.1:1 for the number of receipts, annual sales, and sales amount (THB), respectively. The table also lists this pair's support, *ps*, sales amount, and sales percentage. Note that the sales percentage is expressed as the total sale of the cross-selling products within the total sale of the product categories to which the two products belong, not the total sales of all products combined. That is, passion fruit and beetroot have a total sales amount of 12,400 THB, which is 2.20% of the total sales amount for fresh fruits and fresh vegetables of 563,095 THB.

In addition, Table 7 contains a list of cross-selling product campaigns selected from the significant associations between product category and the time of purchase variables (time of day, day of the week, and season). Three examples for each purchase variable are as follows: (1) For the association

Table 5. Cross-selling product campaigns from significant associations between two product categories

No.	Cross-selling products	Support (%)	ps	Sales (THB)	Category Sale (%)*
1	Passion fruit & beetroot	0.16	43.90	12,400	2.20
2	Plum juice & Chitralada honey	0.23	9.54	10,016	2.01
3	Pomegranate juice & green tea	0.41	20.86	32,000	7.27
4	Mixed vegetable juice & macadamia	0.21	6.84	14,980	4.65
5	Doi Kham honey & green tea	2.43	9.02	29,710	10.50
6	Pomegranate juice & soy milk	0.44	18.49	23,906	10.05
7	Iceberg lettuce & jam jar salad dressing	0.91	84.79	19,505	10.78
8	Vegetable salad & salad dressing	6.06	32.61	5,448	14.15

Note: *Category sale (%) was computed only within the two product categories of the cross-selling products.

Table 6. The ratio of the number of receipts, quantity, and sales of passion fruit to beetroot

Product	Number of transactions	Quantity (pack/year)	Sale (THB)
Passion fruit	2613	4,086	228,100
Beetroot	864	1,768	42,720
Ratio	3:1	2.3:1	5.1:1

Table 7. Cross-selling product campaigns from significant associations between product category and time of purchase

Category & time of purchase	Cross-selling products	Support (%)	*ps*	Sales (THB)	Category Sale (%)*
Fresh vegetables & morning	Cabbage & Chinese cabbage	4.80	208.98	32,030	2.86
Fresh vegetables & late-morning	Cabbage & Chinese cabbage	3.99	240.32	34,495	2.71
Snack & afternoon	Fried pumpkin crackers & fried carrot crackers	7.70	186.27	21,150	6.62
Snack & late afternoon to evening	Fried pumpkin crackers & fried carrot crackers	7.13	231.61	27,325	6.51
Juice and herbal drinks & Monday	Pomegranate juice & mangosteen Juice	1.07	26.7	10,000	3.11
Snacks & Tuesday	Fried pumpkin crackers & fried carrot crackers	8.53	107.56	14,250	8.61
Snacks & Thursday	Fried pumpkin crackers & fried carrot crackers	7.66	86.88	10,625	6.3
Fresh vegetables & Friday	Cabbage & Chinese cabbage	4.35	144.98	18,935	3.14
Fresh vegetables & Saturday	Iceberg lettuce & Cos salad	4.23	143.13	30,110	3.17
Fresh vegetables & Sunday	Iceberg lettuce & Cos salad	3.55	136.38	23,975	2.94
Fresh vegetables & rainy season	Iceberg lettuce & Cos salad	3.93	338.69	68,755	3.43
Fresh vegetables & winter	Cabbage & Chinese cabbage	4.98	397.1	52,550	3.19
Snacks & summer	Fried pumpkin crackers & fried carrot crackers	7.50	168.69	20,025	6.58

Note: *Category sale (%) was computed only within the two product categories of the cross-selling products.

between fresh vegetables and the morning period, the chosen pair of cross-selling products includes cabbage and Chinese cabbage; (2) for the association between fresh vegetables and Sunday, the chosen pair comprises iceberg lettuce and Cos salad; and (3) for the association between snacks and summer, fried pumpkin crackers and fried carrot crackers were chosen. Similar to Table 5, the sales percentage for each pair is based on the total annual sales of the product categories within the associated time of purchase category. These indicate that they represent strong associations. Noticeably, Table 7 includes four pairs of cross-selling products. For instance, the iceberg lettuce and Cos salad pair is a frequent itemset on Saturday and Sunday (weekends) and during the rainy season.

Marketing Campaign Using Cross-Selling Products With a Discounted Price

To promote the cross-selling products, we suggest the retail store launch a promotional campaign that offers a discounted price for the pairs of cross-selling products. Figure 2 presents sample advertisements

Figure 2. Examples of promotional campaigns for cross-selling products

displaying the special discounted price of cross-selling products. The regular prices of the individual items are listed along with the discounted prices when purchased together.

Cross-selling two products at discounted prices would give customers a purchase incentive that can increase sales of items in pairs. As for the pairing of beetroot and passion fruit, for example, we found from the original data that, on average, only one in every 28 customers who purchased passion fruit also purchased beetroot (i.e., 3.52%). Suppose the retail store runs this cross-selling promotion with a discounted price. Also, suppose that the promotion effectively persuades customers who usually purchase only passion fruit to purchase beetroot as well. Success rates of 1%, 5%, and 10% would represent an increase of 26, 130, and 260 receipts including both items, respectively. This is equivalent to sales increases for beetroot of 530, 2,650, and 5,300 THB, respectively, even after deducting a 5% discount. This example demonstrates the effectiveness of pairing a popular item (passion fruit) with another less popular item (beetroot) at this retail store. Note that we chose these two items since they are common ingredients in a healthy fruit-and-vegetable mixed juice or smoothie, which makes customers interested in purchasing them together.

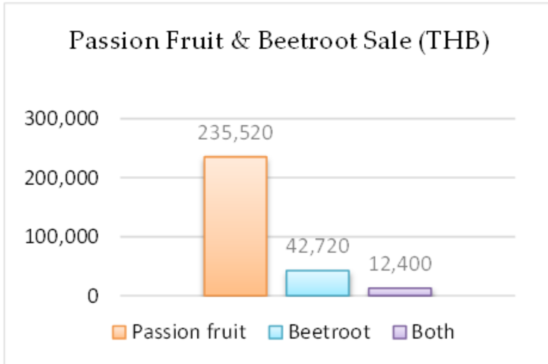
Figure 3 illustrates the potential sales boost of the items in two examples of cross-selling campaigns. In Figure 3(a), the total annual sales of the popular item in the campaign, passion fruit, is 235,520 THB, while it is only 42,720 THB for beetroot, the regular item in the campaign. Without the cross-selling campaign, although the association between these items outperforms its expected number of transactions, the total sales amount from these transactions is relatively small, 12,400 THB, only a fraction of the total sales amount of both items combined. A significant difference between the sales amount for individual items and the sales amount for the pair at original prices indicates that pairing these items together while offering a discounted price has a significant potential to boost sales. In other words, many customers purchase only passion fruit or beetroot. If the campaign can successfully boost some proportion of these customers, the sale of the other item in the pair would naturally increase. For this particular pair, the popular passion fruit can increase the sale of beetroot.

Similarly, Figure 3(b) displays a cross-selling campaign of cabbage and Chinese cabbage, explicitly designed to entice customers who purchase either item during the morning and late morning periods. The graph shows the reasonable sales increase that can be gained from the discounted prices of the bundle.

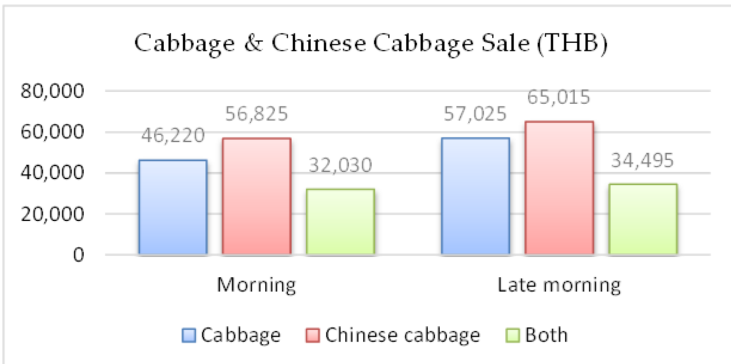
Break-Even Analysis

We performed a break-even analysis to estimate the incremental sales of cross-selling products required to compensate for the cost of offering 5% discounted prices. First, we computed the equivalent numbers of cross-selling pairs based on the items' regular and discounted prices. Then, we used the

Figure 3. Potential sales boost from the pleasant match campaign (a) passion fruit and beetroot cross-selling campaign, (b) cabbage and Chinese cabbage campaign during busy periods



(a)



(b)

difference between the two numbers of pairs to obtain the estimated number of pairs required to run the promotional discounted prices, i.e., the break-even point.

For example, for the cross-selling of cabbage and Chinese cabbage, the regular and discounted prices are 30 and 28.50 THB, respectively. The total annual sales of the two items cross-selling at regular prices in the morning and late morning combined is 66,525 THB. Assuming a 20% profit margin at regular prices leads to a profit of 20% × 66,525 = 13,305 THB at regular prices. To maintain the same profit, the store must sell 13,305 / (30 × (20% – 5%)) = 2,956.67 pairs at the discounted prices. Also note that at regular prices, the number of pairs sold is 66,525 / 30 = 2,217.50. In other words, an increase of (2,956.67 – 2,217.50) = 739.17 pairs in annual sale volumes is required to run this promotion at break-even. If the increase in sale volume exceeds this amount, the promotion can be considered a successful marketing campaign.

Managerial Insights

This section summarizes the managerial insights that can be drawn from the market basket analysis using the FP-growth algorithm presented in this article. From a theoretical standpoint, the proposed market basket analysis introduces a measure of association rule that reflects the monetary value of the association, in addition to traditional measures based on the count or frequency of occurrence of itemsets. This measure can help users evaluate the generated associations in terms of their contribution to a company's total revenue. From a practical perspective, effective marketing strategies can then be designed, and the required budget to run or launch such a marketing campaign can be determined.

Practitioners can use the generalized analysis framework to implement the proposed market basket analysis for other transactional datasets from various industries using the following steps:

1. Acquire transactional data.
2. Pre-process the data, which may include removing irrelevant/redundant data attributes, managing missing values, identifying and handling outliers, and converting data into an analysis-ready format.
3. Select additional data attributes of interest to be included in the associations. For example, a user may be interested in finding strong, high-value associations between frequent itemsets or item categories and customer geographical data, demographical data, purchase timing, etc.
4. Specify minimum support and implement the FP-growth algorithm that generates an appropriate number of associations between frequent itemsets and selected data attributes.
5. Evaluate the monetary value of each association from the FP-growth algorithm results.
6. Design marketing campaigns from frequent and high-value itemsets in the associations.

CONCLUSION

This article presents a market basket analysis that applied the FP-growth algorithm to case-study transactional data of fresh vegetables, fruits, and organic products at one of Thailand's major health food retail stores. The study found frequent associations in customer purchasing behaviors to design appropriate marketing campaigns, i.e., cross-selling products. In addition to commonly used performance measures of the generated associations, our study demonstrates the effectiveness of measuring the financial strength of the associations in terms of the total sales amount. Another significant contribution of our research is that knowledge and findings from significant associations can lead to practical marketing campaigns. The future direction of this study is to perform a market basket analysis on transactional data from other stores in the retail chain to develop effective marketing campaigns for all stores in the chain as well as specific stores.

REFERENCES

Agrawal, R., & Srikant, R. (1994, September). Fast algorithms for mining association rules. In *Proceedings of the 20th International Conference on Very Large Data Bases, VLDB* (Vol. 1215, pp. 487–499). Michigan State University.

Alawadh, M. M., & Barnawi, A. M. (2022). A survey on methods and applications of intelligent market basket analysis based on association rule. *Journal of Big Data*, *4*(1).

Annie, L. C., & Kumar, A. D. (2011). Frequent item set mining for market basket data using K-Apriori algorithm. *International Journal of Computational Intelligence and Informatics*, *1*(1), 14–18.

Annie, L. C., & Kumar, A. D. (2012). Market basket analysis for a supermarket based on frequent itemset mining. [IJCSI]. *International Journal of Computer Science Issues*, *9*(5), 257.

Ayu, S. K., Surjandari, I., & Zulkarnain, Z. (2018, July). Mining association rules in seasonal transaction data. In *Proceedings of the 2018 5th International Conference on Information Science and Control Engineering (ICISCE)* (pp. 321–325). IEEE. doi:10.1109/ICISCE.2018.00074

Bohl, H., Craig, J., & Shellshear, E. (2023). A Shapley-value index for market basket analysis: Weighting Shapley's value. *Journal of Business Analytics*, 1–10.

Han, J., Kamber, M., & Pei, J. (2001). *Data mining: Concepts and techniques*. Morgan Kaufmann Publishers.

Han, J., Pei, J., & Yin, Y. (2000). Mining frequent patterns without candidate generation. *SIGMOD Record*, *29*(2), 1–12. doi:10.1145/335191.335372

Herawan, T., Vitasari, P., & Abdullah, Z. (2012). Mining interesting association rules of students suffering study anxieties using SLP-growth algorithm. [IJKSS]. *International Journal of Knowledge and Systems Science*, *3*(2), 24–41. doi:10.4018/jkss.2012040102

Hwang, W. Y., & Lee, J. S. (2021). Further improvement on two-way cooperative collaborative filtering approaches for the binary market basket data. *Applied Sciences (Basel, Switzerland)*, *11*(19), 8977. doi:10.3390/app11198977

Kamakura, W. A. (2012). Sequential market basket analysis. *Marketing Letters*, *23*(3), 505–516. doi:10.1007/s11002-012-9181-6

Kaur, M., & Kang, S. (2016). Market basket analysis: Identify the changing trends of market data using association rule mining. *Procedia Computer Science*, *85*, 78–85. doi:10.1016/j.procs.2016.05.180

Liu, Y., & Guan, Y. (2008, November). Fp-growth algorithm for application in research of market basket analysis. In *Proceedings of the 2008 IEEE International Conference on Computational Cybernetics* (pp. 269–272). IEEE. doi:10.1109/ICCCYB.2008.4721419

Monroe, K. B., & Dodds, W. B. (1988). A research program for establishing the validity of the price-quality relationship. *Journal of the Academy of Marketing Science*, *16*(1), 151–168. doi:10.1007/BF02723333

Moodley, R., Chiclana, F., Caraffini, F., & Carter, J. (2019). Application of uninorms to market basket analysis. *International Journal of Intelligent Systems*, *34*(1), 39–49. doi:10.1002/int.22039

Raorane, A. A., Kulkarni, R. V., & Jitkar, B. D. (2012). Association rule–extracting knowledge using market basket analysis. *Research Journal of Recent Sciences ISSN*, *2277*, 2502.

Samboteng, L., Rulinawaty, R., Kasmad, M. R., Basit, M., & Rahim, R. (2022). Market basket analysis of administrative patterns data of consumer purchases using data mining technology. *Journal of Applied Engineering Science*, *20*(2), 339–345. doi:10.5937/jaes0-32019

Surjandari, I., & Seruni, A. (2010). Design of product placement layout in retail shop using market basket analysis. *Makara Journal of Technology*, *9*(2), 43–47.

Svetina, M., & Zupančič, J. (2005). How to increase sales in retail with market basket analysis. *Systems Integration*, 418–428.

Tanantong, T., & Ramjan, S. (2021). An association rule mining approach to discover demand and supply patterns based on Thai social media data. [IJKSS]. *International Journal of Knowledge and Systems Science*, *12*(2), 1–16. doi:10.4018/IJKSS.2021040101

Weiß, C. H. (2014). Association rule mining. In *Wiley StatsRef* (pp. 1–6). Statistics Reference Online.

Xu, X. (2009). Optimal price and product quality decisions in a distribution channel. *Management Science*, *55*(8), 1347–1352. doi:10.1287/mnsc.1090.1023

Zekić-Sušac, M., & Has, A. (2015). Discovering market basket patterns using hierarchical association rules. *Croatian Operational Research Review*, *6*(2), 475–487. doi:10.17535/crorr.2015.0036

Zhang, Z., & Guo, C. (2011). Association rules evaluation by a hybrid multiple criteria decision method. [IJKSS]. *International Journal of Knowledge and Systems Science*, *2*(3), 14–25. doi:10.4018/jkss.2011070102

Kanokwan Singha is a Lecturer at the Faculty of Logistics and Transportation Management, Panyapiwat Institute of Management, Nonthaburi, Thailand. She received her PhD in Industrial Engineering from the School of Manufacturing Systems and Mechanical Engineering, Sirindhorn International Institute of Technology, Thammasat University, Thailand, in 2018. Her research interests focus on logistics and supply chain management, statistical data analysis, inventory control systems, discrete event system simulation, and production planning and control.

Parthana Parthanadee is an Associate Professor in the Department of Agro-Industrial Technology at Kasetsart University, Bangkok, Thailand. She received her PhD in 2004 from the Department of Industrial and Manufacturing Engineering, Oregon State University, USA. Her research interests include applied operations research in agro-industry, vehicle routing, production planning and control, and discrete-event system simulation and analysis.

Ajchara Kessuvan is an Assistant Professor in the Department of Agro-Industrial Technology, Faculty of Agro-Industry, Kasetsart University (Tel. +66818433029, Email. Ajchara.ke@ku.ac.th). She received her DBA in Marketing from Chulalongkorn University, Thailand, her MBA in Management from the University of Wisconsin (Milwaukee), USA, and her BA in Economics from Thammasat University, Thailand. Her expertise is in marketing management and strategy, marketing research and market modeling, consumer research and customer journey design, and international marketing strategy.

Jirachai Buddhakulsomsiri is an Associate Professor at the School of Manufacturing Systems and Mechanical Engineering, Sirindhorn International Institute of Technology, Thammasat University, Thailand. He received his PhD in Industrial Engineering from the Department of Industrial and Manufacturing Engineering, Oregon State University, USA, in 2003. His research interests include applied operations research, data mining, statistical data analysis, discrete event system simulation, production planning and control, and logistics and supply chain management.