Product Portfolio Management in F&B using Market Basket Analysis

Peng Ho Oh, Jhun Boon Tan, Jessie Yap

Singapore Management University

Abstract

In the restaurants industry, there are few findings on the actual planning of items offered. As a manager of a restaurant questions should be raised with regards to which products should go together in a set or how does removing a particular item from the menu affect the store. To answer those questions, managers of restaurants and F&B outlets have to understand consumer buying behaviour and popular product purchase combinations. There has been few advancements for analysis-driven decision making in this regard; managers rely heavily on experience and expertise to make estimated decisions on product portfolio management judgements. This paper presents a case of applying an analysis methodology known as Market Basket Analysis in a Japanese F&B chain with a takeaway kiosk concept. Also known as Associations Analysis, it is provides a clear and insightful analytical method for F&B organizations to understand consumer purchase patterns for forming cross-selling and product bundling strategies. Certain considerations based on the nature of the application were made; these includes an analysis of the software used to carry out the analysis and the minimum support used. Frequent Itemset Generation algorithm was used on transactional data from the store's outlets, before generating the association rules. The results show the combination of products for cross-selling in each outlet, and how products can be bundled to increase sales. An analysis of the popularity of products currently being part of a set is also examined. Last but not least, the paper shows how Market Basket Analysis prevents "the profitable-product death spiral".

Keywords: market basket analysis (MBA); association analysis; profitable-product death spiral; F&B; RapidMiner;

1.0 Introduction

In the restaurants industry, there has been an increase in focus on the *science* behind menus – menu planning, menu designs, menu pricing etc. (Ozdemir and Caliskan, 2014). However, few findings have been found on the actual planning of items offered by restaurants. As a manager of a restaurant, besides raising questions on the store's menus should look like or how to price a certain set or item, questions are also raised with regards to which products should go together in a set, how does removing a particular item from the menu affect the store etc. The latter are as important, if not more important, questions to be asked as a store manager as they affect the sales volume and competitiveness of products. To answer those questions, managers of restaurants and F&B outlets have to understand consumer buying behaviour and popular product purchase combinations. This can be done by using an analysis methodology known as Market Basket Analysis. Also known as Associations Analysis, Market Basket Analysis is a method for understanding consumer purchase patterns by analysing transactional data and looking at associations or co-occurrences in each transaction by carrying out association rule discover.

However, little academic attention has been given to product portfolio management or even consumer buying behaviour specifically to a restaurant context. Furthermore, there are no concrete findings of an application of Market Basket Analysis on the assortment of items in a restaurant. Therefore, the purpose of this paper is to demonstrate the application of Market Basket Analysis in product portfolio management in a restaurant setting. This is done by analysing the product offerings of a Japanese F&B chain in Singapore to find popular and unpopular product combinations within existing product sets as well as for products without sets through association rules discovery. This will allow the store managers to identify changes that can be made to the current product portfolios as well as identify products that can be removed from the store's offerings based on product offerings. This study also shows how the application of Market Basket Analysis to products prevents store managers from being susceptible to the "profitable-product death spiral" (Rust, Zeithaml, & Lemon, 2000).

The article is organized as follows. After briefly reviewing the MBA literature, we demonstrate the analysis methodology in the application of MBA on Point-of-Sales transaction data. We first examined the data preparation procedure, followed by the analysis process and the interpretation of the analysis results. Lastly, we then conclude and provide possibilities for future research.

2.0 Literature Review

Market Basket Analysis was first introduced by Agrawal, Imielinski, and Swami in 1993. It aimed to identify when a customer purchases a particular item, a second particular item will be predictably purchased as well. Tan, Steinbach, & Kumar (2006) explains the methodology as follows. Given two items X and Y, a relationship in the form of association rules can be represented as $\{X \rightarrow Y\}$. This suggests that when X is purchased, Y is also likely to be purchased. Support and confidence measures are used as threshold levels in association rules. With reference to the rule $\{X \rightarrow Y\}$, support measures the probability of a transaction containing both X and Y while confidence measures the conditional probability of Y occurring when X occurs.

Historically, a classic example of Market Basket Analysis is the purchase of "beers" and "diapers", two items seemingly unrelated but shown to have high association as they are often bought together. In recent times, common application of Market Basket Analysis can be found in online bookstores like Amazon, where customers are recommended "you may also like" books when they place a particular item in their shopping cart. In various academic literatures, Market Basket Analysis has been used to analyse purchase patterns in a multiple store environment (Chen, Tang, Shen, & Hui, 2005), identify ideal menu items (Ting, Pan, & Chou, 2010), and decide on appropriate product placements in a store (Charlet and Kumar, 2012). Applying Market Basket Analysis in an F&B settings, a restaurant may discover that customers tend to purchase food item X together with food item Y and drink Z. This information help managers to design product bundling strategies and helps floor staff cross-sell and upsell items successfully.

The set of items which meets the minimum support threshold are also referred to as Frequent Itemsets. There are various methods for generating Frequent Itemsets, the common ones being Apriori and FP-Growth, we will be using the latter in this paper. It is also worth noting that Zaki (2000) also introduced six algorithms for association mining - Eclat (**E**quivalence **CLA**ss **T**ransformation), MaxEclat, Clique, MaxClique, TopDown, and AprClique.

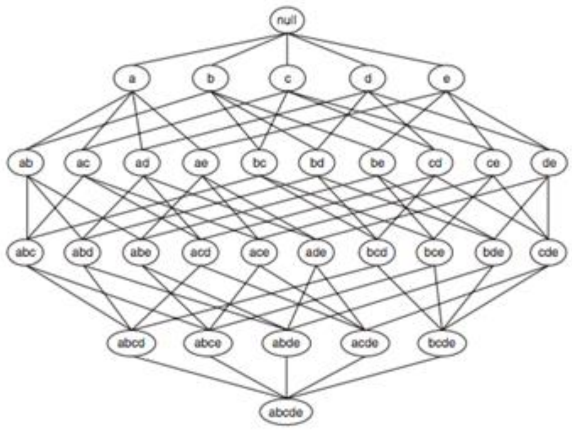

Figure 1. A Frequent Itemset Lattice

Consider the above lattice – each of these are itemsets. Algorithms have to identify the most efficient way to traverse the lattice and identify if a particular itemset is frequent. There are various ways of generating candidates for frequent itemsets and pruning, and this is determined by the algorithm used to carry out association analysis. The way the itemsets are generated and association rules created determine how computationally complex the analysis will be.

Therefore, considerations affecting the computational complexity of an algorithm have to be determined when dealing with mining association rules for large datasets. These include factors such as transaction width, number of products, minimum support level and max itemset size (Tan, Michael, & Kumar, 2005). Since the transaction width and number of products are predetermined, the team has chosen to specifically focus on the latter 2 factors to refine for our analysis - association thresholds and the max itemset size.

An important aspect of association analysis is the generation of frequent itemsets (or the elimination of infrequent itemsets). The minimum support (minsup) and minimum confidence (minconf) is taken into account. These are thresholds used to discover whether the itemsets in A --> B are frequent itemsets and whether A --> B is an acceptable association rule. While the team has explored algorithms to determine the optimal minimal support and minimal confidence levels such as applying Particle Swarm Optimization, the team has examined the data spread to determine appropriate minimum support and confidence levels.

Market Basket Analysis can be used to learn about customer purchase patterns so that customer facing staff can upsell and cross-sell in order to increase sales. Product bundles can also be created to increase attractiveness of products. Ideally, Market Basket Analysis seeks to identify interesting relationships between products in market basket transactions. This means that we seek association rules between products that have not been pre-determinately placed in purchasing relationships, such as items being part of a set or bundle. This is because naturally items within a set has already some form of association between them – an increased likelihood that these products are to be bought together. Therefore it is imperative that in our analysis, we handle items within a set and a la carte items differently. For items within set meals that are currently being offered, we use a subset of Market Basket Analysis to look at the effectiveness of the existing set meals and suggest areas for improvement to increase sales and demand. For a la carte items, we use Market Basket Analysis to identify suitable product combinations and cross-sell or upsell opportunities.

While there has been articles suggesting that association analysis methodology should incorporate item weights and transaction weights to better present analysis findings. In Weighted association rules: Model and algorithm (Ramkumar, Rankar, & Truc, 1998), the following example was suggested: "Caviar is an expensive and hence a low-support item in any supermarket basket. Vodka, on the other hand, is a high to medium-support item. The association: *caviar* => *vodka* is of very high confidence but will never be derived by the existing methods since the itemset {vodka, caviar} is of low support and will not be included." While applying these improved algorithms provides benefits in terms of clarity of results, these advancement in methodology is excluded from the analysis as the variation in support levels are much less.

The "profitable-product death spiral" depicts that organizations or companies constantly attempt to ascertain their products' profitability and because of pressures to cut costs and increase overall profitability, managers within the companies seek to remove the less profitable products. Cannon, Cannon, & Schwaiger (2012) claims that managers ignore "the fact that customers typically want an assortment of products, and that the deletion may weaken assortments that customers want. The resulting loss of sales weakens demand for previously profitable products, subsequently causing them to be dropped. This weakens the assortments even further, and so forth in a downward death spiral. By focusing on customer rather than on product profitability, marketers look at the portfolios of products their customers want rather than disrupting portfolios for the sake of individual product profitability." Therefore, there is definitely worth in analysing the association between products amidst the need to ascertain product profitability.

3.0 Methodology

This section of the paper is divided into 5 parts: (1) the data involved in the analysis, (2) the selection of the analysis tool used, (3) the breakdown of the analysis methodology as well as (4) the data cleaning and preparation procedure and (5) the analysis measures used.

3.1 Data

Traditionally data related to Market Basket Analysis is three-dimensional: Customers, Orders (i.e. item sets, purchases or baskets), and Items (Beery and Linoff, 2004). A sales order is a most essential and basic piece of information representing a single purchase or transaction made by a customer. Besides main information such as the product bought, the quantity of products bought and total amount of the purchase, the store number, cashier number, type of payment or even the cashier who served is also stored in the order data. The items or rather the contents of the order is most important and founds the basis of identification of association rules. Last but not least, customer information provides a deeper level of analysis by finding associations between certain customer traits and profiles and particular items, allowing the store to carry out market segmentation. (Ting, Pan and Chou, 2010).

A market basket database typically consists of a large number of transaction records. Each record lists all items purchased during a single customer transaction. The objective of this data mining exercise is to identify if certain groups of items are usually purchased together, providing meaningful association rules.

3.2 Analysis Tool Selection

In carrying out Market Basket Analysis certain considerations have to be made. One important factor is the software or tool used to carry out Market Basket Analysis. Based on the client requirements in this project, the tool used must be one that is open-source and easy to use. While the team understands that there are far greater utility in employing paid software such as Clementine (SPSS), Enterprise Miner (SAS), GhostMiner 1.0, Quadstone or XLMiner, this requirement essentially narrows down the tools that the team is able to use (Haughton et. al., 2003). The tools that are open-source are narrowed down into 3 tools: RapidMiner, R and Tanagra.

Table 1

Analysis Tool Selection

<table>
<tr><th>Software and Package</th><th colspan="2">Pros</th><th>Cons</th></tr>
<tr><td>R (arules package)</td><td rowspan="3">- Free to use
- Easy to install
- Association Analysis tutorial document easily available</td><td>- Flexibility and Customizability</td><td>- Difficulty learning curve for using Software
- Difficulty in manipulating input dataset
- Programming Background required</td></tr>
<tr><td>RapidMiner (FP-Growth Operator)</td><td>- Extensive Interestingness Measures
- Gentle learning curve for using Software</td><td>- Set number of operators that can be used
- Set Operator-based processes limits the customizability of processes</td></tr>
<tr><td>Tanagra (Apriori PT Component)</td><td>- Gentle learning curve for using Software</td><td>- Limited interestingness measures
- Lack of customizability in set software processes</td></tr>
</table>

After evaluating the 3 tools, the team realized that though R provided measures and customizability, the learning curve to use R is extremely steep and may not be best for the client based on the non-programming nature of their background. Both RapidMiner and Tanagra is extremely lightweight and easy to use, however the presence of extensive interestingness measures caused the team to choose in favour of RapidMiner.

3.3 Analysis Breakdown

The study will carry out the analysis in the following flow:

1. Products that are already in a set are first analysed. Since there is already an association between items found in a set, the confidence of set components are analysed to identify the most popular side dishes / drinks as well as the most unpopular ones.
2. For products that are not within a set, Market Basket Analysis is carried out to identify association rules between products.
3. Products with low profitability is identified using cost and revenue figures provided by the store; products that may contribute to the “profitable-product death spiral” is prevented from being dropped.

3.4 Data Cleaning and Procedures

Most Point-of-Sales (POS) systems will provide transaction data that consists of information such as transaction id, product id, quantity, price per product, date etc. The data given by the POS system for this example is in the following format:

Table 2

POS Data

Order#	Receipt#	Store#	Date	Time	Cashier	Item	Product ID	Qty	Price
1	3975	1	16/8/15	14:59	John	Katsu Don	M_6	1	$16.00
1	3975	1	16/8/15	14:59	John	Miso Soup	M_21	1	$3.50

The team has formatted and exported the data in a “Comma Separated Values” (CSV) file where the "t_id" is the transaction id, the “p_id” is the product id and the "qty" is the quantity of such products sold in a given transaction.

Table 3

Transaction Data

t_id	p_id	qty
1	m_6	1
1	m_21	1
2	m_14	2
3	o_4	1
4	m_14	1

However, this data requires further transformation to transform from transactional data to market basket data. Ideally, market basket data should be represented in a binary format where each row is a separate transaction id and each column corresponds to a product or item sold. While the quantity is provided and analysed in this example, RapidMiner does not include quantity in the analysis of the result.

No	Date	Day	Time	m_1	d_2	d_3	d_4	d_5	d_6	d_7
1	10/21/2015	WED	10:22	0	0	0	0	0	0	0
2	10/21/2015	WED	10:42	0	0	0	0	0	0	0
3	10/21/2015	WED	11:29	0	0	0	0	0	0	0
4	10/21/2015	WED	11:31	0	0	0	0	0	0	0
5	10/21/2015	WED	11:33	0	0	2	0	0	0	0
6	10/21/2015	WED	11:36	0	0	0	0	0	0	0
7	10/21/2015	WED	11:38	0	0	0	0	0	0	0
8	10/21/2015	WED	11:40	0	0	0	0	0	0	0
9	10/21/2015	WED	11:42	0	0	0	1	0	0	0
10	10/21/2015	WED	11:43	0	0	0	0	0	0	0

Figure 2. Transaction data in binary format

Note: our team assumes that the data provided by the POS System is in a most basic relational format as shown above. If the data has been transformed in any particular way that makes certain transformation steps redundant e.g. the data is already in a binomial form, the respective steps can be skipped.

In order to transform the data into a suitable format to apply the identification of frequent itemsets and the generation of the rule, the team carried out the following process (Deshpande, 2012) in RapidMiner:

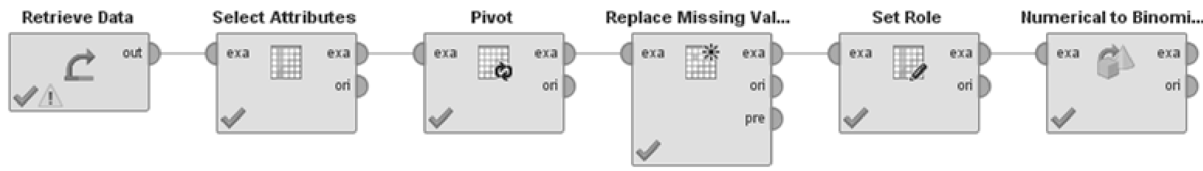

Figure 3. Data preparation in RapidMiner

The team will elaborate on the above stages in the following sections.

***3.4.1 Insert Data*.**

Firstly, we loaded the CSV file into the RapidMiner repository by clicking on "Add Data". RapidMiner should detect the file's format and delimit the rows automatically by ",". Once the file is successfully formatted and stored, the following result should be seen:

Row No.	t_id	p_id	qty	date
1	1	m_10	1	Jan 20, 2015
2	2	m_29	1	Jan 20, 2015
3	3	m_11	1	Jan 20, 2015
4	4	o_16	1	Jan 20, 2015
5	5	m_15	1	Jan 20, 2015
6	6	m_14	1	Jan 20, 2015
7	7	m_11	1	Jan 20, 2015
8	8	m_10	1	Jan 20, 2015
9	9	d_9	1	Jan 20, 2015
10	9	m_27	1	Jan 20, 2015

Figure 4. Data inserted in RapidMiner

Now that the data is in the local repository, we can drag it into the process and it depicts the start of the analysis process. A "Retrieve <name of file>" operator should appear. From the transaction data, we kept only the columns that we require. We connected the operator "Select attributes" to the "out" of our "Retrieve Data" operator and select the filter type "subset" to retain only the relevant columns "t_id", "p_id" and "qty".

Row No.	t_id	p_id	qty
1	1	m_10	1
2	2	m_29	1
3	3	m_11	1
4	4	o_16	1
5	5	m_15	1
6	6	m_14	1
7	7	m_11	1
8	8	m_10	1
9	9	d_9	1
10	9	m_27	1

Figure 5. Data inserted in RapidMiner

3.4.2 Pivot Transformation

Next, we have to group the transactions by the transaction id with products represented in subsequent columns; this operation produces the result that was previously shown in Figure 2. The “Pivot” ‘operator is used and it’s connected to the “out” of our “Select Attributes” operation; under the “group attribute” parameter, we select “t_id” and under “index attribute” we select “p_id”. The output from this set is

Row No.	t_id	qty_m_10	qty_m_29	qty_m_11	qty_o_16
1	1	1	?	?	?
2	2	?	1	?	?
3	3	?	?	1	?
4	4	?	?	?	1
5	5	?	?	?	?
6	6	?	?	?	?
7	7	?	?	1	?
8	8	1	?	?	?
9	9	?	?	?	?
10	10	?	?	?	?

Figure 6. Data after Pivot Transformation

3.4.3 Replacing Missing Values

Since there were some products that had 0 quantity for particular transactions, a “?” or a missing value is retained during our “Pivot” operator. We have to connect a new operator “Replace Missing Values” to the “exa” of the “Pivot” operator to replace these missing values with a “0” to correctly reflect the quantity of the product bought in each transaction. Under the “default” option, “zero” is selected.

Row No.	t_id	qty_m_10	qty_m_29	qty_m_11	qty_o_16
1	1	1	0	0	0
2	2	0	1	0	0
3	3	0	0	1	0
4	4	0	0	0	1
5	5	0	0	0	0
6	6	0	0	0	0
7	7	0	0	1	0
8	8	1	0	0	0
9	9	0	0	0	0
10	10	0	0	0	0

Figure 7. Data after removing missing values

3.4.4 Set Role: ID

After the above transformation, all the new columns produced are attributes with the same role. However, in order to better represent the data, and also to prepare it for the later parts of the process, we have to give some of the columns a certain “role”. This “role” sets the kind of part that an attribute plays in a data set or a process. In this example, we are specifically changing the role of “t_id” to id, since it is indeed used as a unique identifier for each role. This removes the attribute “t_id” from analysis and leaves it as an identification attribute in the later parts of the process. The “Set Roles” operator is selected and connected to the “exa” of the “Pivot” operator. Under the “attribute name” parameter, “t_id” is selected and the “target role”, “id” is selected.

Row No.	t_id	qty_m_10	qty_m_29	qty_m_11	qty_o_16
1	1	1	0	0	0
2	2	0	1	0	0
3	3	0	0	1	0
4	4	0	0	0	1
5	5	0	0	0	0
6	6	0	0	0	0
7	7	0	0	1	0
8	8	1	0	0	0
9	9	0	0	0	0
10	10	0	0	0	0

Figure 8. Data after setting role "id"

3.4.5 Binomial Representation

Unfortunately, within RapidMiner the quantity of products purchased within a transaction is not relevant but just simply if a product is purchased or not. The algorithm used later requires the transaction data to be represented in "binomial" values – meaning the analysed attribute has only exactly two possible values, "true" if the product is purchased in a transaction and "false" if a product is not purchased in a transaction. In the "Numerical to Binomial" operator in RapidMiner, the attributes are transformed by checking if they are between a minimal and maximal value; if an attribute falls between the these values, it takes on the attribute "false", otherwise "true". By default, the minimal and maximal values are set as "0.0" and "0.0" respectively and hence if an attribute's original value is 0, it will be transformed to "false".

Row No.	t_id	qty_m_10	qty_m_29	qty_m_11	qty_o_16
1	1	true	false	false	false
2	2	false	true	false	false
3	3	false	false	true	false
4	4	false	false	false	true
5	5	false	false	false	false
6	6	false	false	false	false
7	7	false	false	true	false
8	8	true	false	false	false
9	9	false	false	false	false
10	10	false	false	false	false

Figure 9. Data after binomial transformation

Now that the data is prepared, we select the two main operators that carry out the analysis – "FP-Growth" and "Create Association Rules". FP-Growth is one of the algorithms used in generating frequent itemset. "Create Association Rules" is used to find association rules between the frequent itemsets generated. The following is the completed process in Rapidminer:

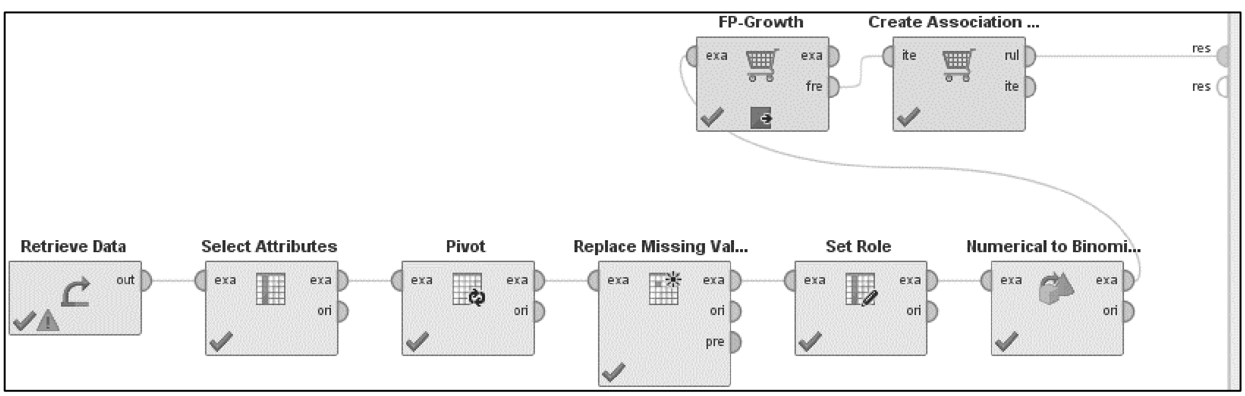

Figure 10. Complete Association Analysis Process in Rapidminer

3.4.6 Frequent Itemset Generation

This particular problem requires us to find sets of items that appear above a percentage threshold of the total number of transactions; as mentioned this is defined as the "minimum support" criteria. Frequent itemset generation is seen as the *prelude* to association rules discovery. In

RapidMiner, the “FP-Growth” operator seeks the common itemsets and the more complex discovery carried out by “Create Association Rules” is derived from the results of “FP-Growth”.

The reason why RapidMiner uses FP-Growth is that by building a FP-tree data structure of the data set, a very compressed copy of the data in created. This allows the computation to fit into the main memory even for large data sets. Usually compared to Apriori algorithm, the major advantage of FP-Growth is that it only takes 2 data scan and hence is more suitable for larger data sets. The “FP-Growth” operator only takes in binomial attributes, as we have previously ascertained.

There are two ways of using the “FP-Growth” operator:

a. The first way allows the user to specify a number of products with the highest support to be selected; this is regardless of the minimum support threshold. This mode is used when we do not have a clear idea of a minimum support to set.
b. The second way relies on the minimum support threshold and returns the itemsets with a support greater than the support value provided.

The two modes are determined by the “find min number of itemsets” parameter; when it is set to true the first way is selected, otherwise the second. In this example, after the “FP-Growth” operator is selected, it’s connected to the “exa” of the “Numerical to Binomial” operator. The “find min number of itemsets” is deselected and minimum support threshold of “0.05” is selected. We can select the minimum and maximum size of the itemsets; a “Min. Size” and “Max. Size” of “1” and “3” is selected respectively.

Size	Support ↓	Item 1	Item 2	Item 3
1	0.430	qty_m_11		
1	0.179	qty_s_s_f		
2	0.179	qty_m_11	qty_s_s_f	
1	0.087	qty_s_d_5		
2	0.086	qty_m_11	qty_s_d_5	
2	0.086	qty_s_s_f	qty_s_d_5	
3	0.086	qty_m_11	qty_s_s_f	qty_s_d_5

Figure 11. Frequent Itemsets

3.4.7 Association Rule Discovery

Finally the operator "Create Association Rules" is added and connected to the "fre" output of FP-Growth with the "fre" input of this operator; doing this will deliver both the frequent item sets "fre" and the association rules "rul" to the result ports "res" on the right side. This step is where the analysis of the data to provide association patterns based on the frequent itemsets provided is carried out. The support and confidence criteria is used to identify the most important relationships. In this operator, the antecedents are represented as "Premises" and the consequents are represented as "Conclusions". The confidence criteria is the percentage of a particular premise-conclusion statement appearing in the entire transaction data set. The "min confidence" parameter sets the minimum confidence criteria for a particular statement to be selected. A "min confidence" of "0.8" is selected. Six interestingness measures are provided from this - we'd go into a deeper analysis of these measures in the next part of the report.

No.	Premises	Conclusions	Support	Confidence	LaPlace	Gain	p-s	Lift	Conviction
1799	qty_s_s_f	qty_m_11, qty_s_d_5, qty_s_tp_4	0.019	0.105	0.864	-0.340	0.016	5.578	1.097
1912	qty_s_s_f	qty_m_11, qty_s_d_4, qty_s_tp_1	0.023	0.127	0.867	-0.336	0.019	5.578	1.119
2033	qty_s_s_f	qty_m_11, qty_s_d_5, qty_s_tp_1	0.027	0.152	0.871	-0.331	0.022	5.578	1.147
2036	qty_s_s_f	qty_m_11, qty_s_tp_3, qty_s_d_4	0.028	0.154	0.871	-0.331	0.023	5.578	1.150

Figure 12. Association Rules

3.5 Analysis measures

While association analysis aims to detect relationships between items in a data set, the real value from these analysis is finding connections between items that don't seem to intuitively have a relationship. However, in order to find out if a rule has statistical value and is not purely due to chance, we have to analyse the measures that are used to ascertain the interestingness of an association rule. In RapidMiner, besides support and confidence, there are 4 other common measures that are used, namely – "LaPlace", "Gain", "p-s", "Lift" and "Conviction". The formula of the measures are as follows:

Table 4

Analysis Measures

Measure	Formula	Range
Support	$P(A,B)$	0 … 1
Confidence	$\max(P(B\|A), P(A\|B))$	0 … 1
Lift	$\mathrm{lift}(X \Rightarrow Y) = \frac{\mathrm{supp}(X \cup Y)}{\mathrm{supp}(X) \times \mathrm{supp}(Y)}$	0 … ∞
LaPlace	$\max\left(\frac{NP(A,B)+1}{NP(A)+2}, \frac{NP(A,B)+1}{NP(B)+2}\right)$	0 … 1
Leverage[1]	$P(A,B) - P(A)P(B)$	-1 … 1
Conviction	$\max\left(\frac{P(A)P(\overline{B})}{P(A\overline{B})}, \frac{P(B)P(\overline{A})}{P(B\overline{A})}\right)$	0.5 … 1 … ∞

In order to ascertain which interestingness measures are more meaningful, we analysed the measure based on three key properties (Piatetsky-Shapiro, 1991):

- **Property 1:** $M = 0$ if A and B are statistically independent;
- **Property 2:** M monotonically increases with $P(A, B)$ when $P(A)$ and $P(B)$ remain the same;
- **Property 3:** M monotonically decreased with $P(A)$ (or $P(B)$) when the rest of the parameters ($P(A,B)$ and $P(B)$ or $P(A)$) remain unchanged.

Below is an analysis of the 5 measures based on the above 3 properties:

Table 5

Analysis of Analysis Measures

Measure	Property 1	Property 2	Property 3
Support	No	Yes	No
Confidence	No	Yes	No
Lift	Yes[2]	Yes	Yes
LaPlace	No	Yes	No
Leverage	Yes	Yes	Yes
Conviction	No	Yes	No

[1] Leverage, Piatetsky-Shapiro Measure (p-s)

[2] Yes if measure is normalized

Based on the analysis of the 3 properties, the team has decided upon Lift and Leverage as the analysis measures used[3].

3.5.1 Lift, Interest

Lift measures how many more times are two products likely to be bought together as compared to the products being bought together if they were independent. A lift of 1 represents independence, and the results approaches infinity if a product is completely dependent. Any number greater than 1 indicates dependency between the two products. Consider the following example:

Table 6

Lift Example

Product X	Product Y	P(X n Y)	Lift	Leverage
Bread	Coffee	0.005	6.141	0.00417

The probability of Bread and Coffee being bought together is 6 times more likely than them being bought if they were independent of each other. Assume that there were 100,000 transactions, 500 of them contains the Bread and Coffee. Their individual probabilities are 0.0235 (Bread) and 0.0351 (Coffee), and if there were totally independent, the probability of them occurring together would be 0.0235 x 0.0351 = 0.000826. The number of transactions that Bread and Coffee were to appear if they were totally independent would be approximately 83 times. The actual number of transactions is 6 times more than the expected value if the two products were independent. This means that when either product is bought, it is 6 times more likely that the other product is bought than by the other product's individual probability.

3.5.2 Leverage, Piatetsky-Shapiro Measure (PS)

Similar to lift, leverage measures the additional probability of products X and Y being bought together over the probability of products X and products Y being bought independently. If

[3] Note that these measures are used for item sets that are not within a set. An analysis measure is discussed further in the next part of the paper.

the additional probability is 0 or lower, it shows that the purchase of these two products are independent. If the result is near to 1 then it is an indication of association between the two products.

Consider the above example, the leverage is 0.00417. This means that the actual increase in probability would be 0.004 or the actual increase in number of transactions is 417 (out of 100,000 transactions).

While both lift and leverage seems to provide a similar implication, leverage provides a clearer business implication in the actual increase in probability based on the popularity of selected products.

3.6 Analysis measures for Sets

With regards to the assortments within sets, the team found that the two selected measures lift and leverage is not applicable. This is simply because the products are statistically dependent i.e. side dishes have to be bought together with a main dish. As such, the team considered the **conditional probability** of main dishes being bought together with sides and analysed the popularity of these combinations:

4.0 Analysis Results

The analysis results are three-fold based on the breakdown of the analysis methodology: (1) analysis results of products that are already in a set in order to identify the most popular and unpopular components of the sets; (2) analysis results of products that are not within a set; (3) analysis results of product profitability.

4.1 Analysis results of products in a set.

The team found that certain drinks and side dishes are more likely to be bought together with a certain main dish than others. Recommendations based on the analysis results can be made as to whether the products should be retained in the product offering. Consider the following:

Table 7

Set Analysis Results

Main Dish	Drink	Support
Main Dish X	Hot Green Tea	0.473
Main Dish X	Cold Green Tea	0.385
Main Dish X	Ayataka	0.124

Recommendations can be made on which drinks are more popular with Main Dish X in a set – that "Ayataka" can be possibly removed from the set as well as by recommending highly associated drinks such as "Hot Green Tea" or "Cold Green Tea" will increase the satisfaction that customers derive. This is purely made on the assumption that the notion that "safety in numbers" is true. Stan (2016) suggests this is a belief based on the fact that a large number of consumers can't all be wrong about the quality of value of the product combinations but "it's entirely possible for a large number of people to be wrong, especially if few consumers research their options before making a purchase. As a result, a popularity appeal by itself might fail to convince savvy consumers." It is then suggested that in order to make a quality recommendation is to back the popularity appeal with facts – in this case, it is that there is greater pleasure between "Main Dish X" and "Hot Green Tea" for example.

4.2 Analysis results of products without a set

With an acceptable minimum support threshold of 0.005 (since the average support for products is 0.02), the team discovered new association rules between products that not only provide a high association between the two products in the itemset, but also holds a substantial amount of support in the transaction data set.

Table 8

Non-Set Analysis Results

Product X	Product Y	Support	Confidence	Leverage	Lift
Fried Dish A	Fried Dish B	0.005	0.140	0.004	5.881
Onigiri A	Onigiri B	0.006	0.140	0.005	4.070
Main Dish A	Drink A	0.005	0.076	0.002	1.776

The following recommendations can be made: (1) Providing discounts for either of Fried Dish A and Fried Dish B as well as Onigiri A and Onigiri B will see an increase in sales of the other. An alternative will be to provide sets that allows customers to choose these two as an option. Thirdly, the stores should consider continuing placing the products close together as this will increase the association between the products further. Ultimately, this recommendations will increase sales volume of the selected products. (2) A new set meal containing Main Dish A and Drink A can be recommended since there is an association between both products. Since it is already a popular choice to purchase the drink with this meal, placing them together in a set at a reduced price will increase sales volume of these two products further.

4.3 Analysis results of product profitability

The profitability of products are ascertained using Cost of Goods Sold (COGS) and the pricing of the products. Under the assumption that products with lower profitability should be removed, the team has identified various products that have low profitability. One example is Onigiri B with a profitability of less than 26% (average for Onigiris are about 40%). In examining Onigiris to be removed from the product offering to save cost for the stores, managers should intuitively remove Onigiri B. However, with association analysis as we observed above, that there is a high association between Onigiri A and Onigiri B and the team uncovered that Onigiri A is in fact extremely profitable (profitability of 46%). By removing Onigiri B, a decrease in sales volume will be experienced by Onigiri A and hence a lowered overall store performance will be experienced.

5.0 Conclusion and Future Work

We have presented a new model of applying Market Basket Analysis in aiding a store's product portfolio management by looking associations between products. This allows a store manager to have more targeted bundling options. Furthermore, the removal of the products based on the products' profitability can now be analysed with consideration to the implications it has on other products.

Further research can be carried out on the actionable recommendations of Market Basket Analysis to F&B stores; examples could be menu management or the analysis of products being placed in a set. For stores with greater fluctuations in product support levels, an examination of a weighted model of Market Basket Analysis can be applied. However, in this regard it is imperative that the Delphi method is to be implemented to ascertain the weight of the product association; in other words, expert opinions from the sales staff or managers have to be harnessed or observed. Another likely advancement of the product portfolio management is to represent each product in a "social network", where the size of the node is a factor of the product's profitability and its importance based on the strength of association with other nodes and their respective importance. As such, a more accurate understanding of a product's importance can by studied to prevent the "profitable-product death spiral".

6.0 Acknowledgements

We would like to thank Professor Kam Tin Seong for useful discussions and wise guidance during the formative stages of the paper.

References

Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *ACM SIGMOD Record*, *22*(2), 207–216. doi:10.1145/170036.170072

Cannon, J. N., Cannon, H. M., & Schwaiger, M. (2012). Modeling the "Profitable-Product Death Spiral" Accounting for Strategic Product-Mix Interactions in Marketing Simulation Games. *Simulation & Gaming*, 43(6), 761-777.

Chen, Y.-L., Tang, K., Shen, R.-J., & Hu, Y.-H. (2005). Market basket analysis in a multiple store environment. *Decision Support Systems*, *40*(2), 339–354. doi:10.1016/j.dss.2004.04.009

Deshpande, B. (2012, November 13). *Market basket analysis: generating recommenders using RapidMiner*. Retrieved from http://www.simafore.com

Loraine Charlet Annie M.C., & D Ashok Kumar. (2012). Market Basket Analysis for a Supermarket based on Frequent Itemset Mining. *International Journal of Computer Science Issues, 9*(5), 257-264.

Mack, S. *What Is Popularity Appeal in Advertisement?* Retrieved from http://smallbusiness.chron.com/popularity-appeal-advertisement-66060.html on 24, March, 2016

Ozdemir, B., & Caliskan, O. (2014). A review of literature on restaurant menus: Specifying the managerial issues. *International Journal of Gastronomy and Food Science*, *2*(1), 3-13.

Ramkumar, G. D., Ranka, S., & Tsur, S. (1998, August). Weighted association rules: Model and algorithm. In Proc. ACM SIGKDD.

Rust, R. T., Zeithaml, V. A., & Lemon, K. N. (2000). *Driving customer equity: How customer lifetime value is reshaping corporate strategy*. New York, NY: The Free Press.

Tan, P.-N., Steinbach, M., & Kumar, V. (2006). Chapter 6: Association Analysis: Basic Concepts and Algorithms. In *Introduction to data mining* (pp. 327–414). United States: Addison Wesley Publi.

Ting, P.-H., Pan, S., & Chou, S.-S. (2010). Finding ideal menu items assortments: An empirical application of market basket analysis. *Cornell Hospitality Quarterly*, *51*(4), 492–501. doi:10.1177/1938965510378254

Zaki, M. J. (2000). Scalable algorithms for association mining. IEEE Transactions on Knowledge and Data Engineering, 12(3), 372–390. doi:10.1109/69.846291