International Journal of Engineering & Technology, 7 (4.42) (2018) 119-123

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Market Basket Analysis of Customer Buying Patterns at Corm Café

N. Isa[1]*, N.A.Kamaruzzaman[2], M.A. Ramlan[3], N.Mohamed[4], M.Puteh[5]

Big Data Analysis and Data Mining Group, Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA, 20800 Kuala Terengganu, Terengganu
**Corresponding author E-mail: norul955@tganu.uitm.edu.my*

Abstract

Market Basket Analysis (MBA) is a technique in data mining used to seek the co-occurrence set of items in a large dataset or database. It is usually used in mining transactions or basket data, especially in retail. This technique has been proven beneficial in understanding customer buying patterns and preferences. It has been widely used in multinational companies. Current business trends have changed dramatically, parallel with the advancement of technology. Changes in customer demand requires an improvement in accuracy of business operations. This paper proposes the implementation of MBA at a Small Medium Enterprise business, a case study at Corm Café. Daily transaction data taken from customer order sheets has been used. A detailed implementation is demonstrated in the paper. The results identify a trend in customer buying patterns, which is useful information for the owner in planning their business operation.

Keywords: *market based analysis; data mining; frequent item set mining*

1. Introduction

Globalization has a huge influence on the business environment nowadays. It provides benefits to both customers and business owners. The business marketspace has become widely available, affecting customer demand and behaviour. In addition, it also opens more business opportunity for business owners to venture into. This is beneficial to those who take this opportunity but may pressure those who remain the same. To succeed in such a challenging environment, businesses need to compete with others and find a way to make their existence significant. Knowing one's customers, especially their preferences, would be one of the best strategies to survive in this challenging environment.

Market Basket Analysis (MBA), also known as Association Rule Mining (ARM), is a data mining technique used to find the co-occurrences of items in the large dataset. This technique can help a business gain better insight into their customer's purchasing behavior [1,2] for example which products are bought together which can help them to make a better decision for their business. Association rule mining has been established a few decades ago by Rakesh Agrawal [3] in 1993. With the advancement of business analytics, MBA has been widely used in many businesses to help them in better decision making [4-7].

MBA has been proven to help businesses in making better decisions, especially in marketing [1]. A literature review shows that this technique has been widely implemented in a big enterprise company. The objective of this paper is to implement market basket analysis technique in a Small Business Enterprise (SME). Corm Café, a contemporary café located in Melaka, has been selected as our case study in this research. The purpose of this MBA implementation is to gain a better understanding of customer preferences based on customers' daily orders. This section contains a discussion on Market Basket Analysis, followed by the Methodology section which will discuss the steps carried out throughout this research. Then results for the experiments will be discussed in the Results and Discussion section, followed by the Conclusion.

2. Market Basket Analysis

Data Mining task can be divided into two categories, first is Descriptive Mining (e.g.: clustering and association pattern discovery) and second is Predictive Mining (e.g.: classification and regression). ARM falls under descriptive mining. ARM tries to find a set of associated items that comes together from the large dataset of transactions. Let say we have item X and item Y. An associated rule $X \rightarrow Y$ indicates that every customer that purchases X will purchase Y too [2]. There are three types of mining in association rule mining [8]. Fig. 1 shows different types of association rule mining. Frequent Itemset mining tries to find a frequent item that appears in the whole transaction dataset while Utility Itemset Mining focuses on finding the set above a utility threshold set by the user. The utility threshold can be any parameter, such as cost, time, and so on [9]. Rare Itemset Mining is converse to Frequent Itemset Mining, as it tries to find a rare set that exists in the larger dataset. Furthermore, [10] introduced sequence MBA, in which rather than finding the co-occurrence itemset, the author was focusing on analyzing the purchasing sequence from large set of dataset.

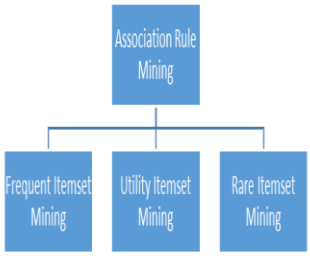

Fig.1: Types of Associatian Rule Mining

In this research, frequent itemset mining is used. To gain more understanding on this area, a review has been performed. Table 1 below depicts research that have been done using frequent itemset mining with the research objectives.

Table 1: Research on Frequent Itemset Mining

Ref	Research Tittle	Description
[4]	Association Rule – Extracting Knowledge Using Market Basket Analysis	• Analyze the large amount of data and obtained consumer behavior pattern based on purchasing record. • Result obtained is used in decision-making for competitive edge over rivals.
[6]	The relationship between (4Ps) & Market Basket Analysis. A Case Study of Grocery Retail Shops in Gweru Zimbabwe	• The authors investigated the relationship between Product, Place, Promotion and Price(4Ps) in Market Basket Analysis. • Then, they established how 4Ps can be applied as tool for competitive advantage
[5]	Application of data mining technique to a selected business organization with special reference to buying behavior	• To understand customer buying behavior with the help of data mining technique and tools. • Secondary data is collected from bills from shopping malls. • Data collected from periodic reports of shopping malls.
[7]	Layout Optimization and Promotional Strategies Design in a Retail Store based on Market Basket Analysis	• To provide prescriptive model for store layout optimization. • To design promotional strategies for up-selling and cross-selling. • Focus on 24 product families with high occurrence in sale ticket.
[1]	Market Basket Analysis: Identify the changing trends of market data using association rule mining	• To provide the information to the retailer to understand the purchase behavior of the buyer, which can help the retailer in correct decision making. • Using daily transactions data for mining.

2.1. Frequent Itemset Mining

There are two steps involved in finding the itemset. The first is finding the frequent item set from a large data set and second is generating associated rules from frequent item set found in the first step. There are several algorithms that have been developed to execute this process. Among them are Apriori algorithm and FP-Growth algorithm.

The Apriori algorithm was developed by [11] in 2005. The basic idea of this algorithm is, first finding the frequent item based on minimum support minimum confidence. Consider a relationship of X→Y. Support is a measurement to measure how many times X and Y appear together in the whole dataset, while confidence measures how many times Y appears in the transaction containing X.

In the FP-Growth algorithm, finding the itemset is done by building a prefix tree or FP-tree [12]. The tree will be pruned to remove the infrequent item. The pruning process will be based on minimum support value such as in the Apriori algorithm. Then associated rules will be generated based on minimum confidence value. In this paper, FP-Growth algorithm was used for frequent item mining.

2.2. Support and Confidence

In frequency mining, generated associated rules are evaluated based on two metrics, which are support and confidence, as shown in equation 1

$$X \rightarrow Y[Support(\%), Confidence(\%)] \quad (1)$$

Here, X is called Left-Hand side (LHS) or antecedent, while Y is Right-Hand side (RHS) or consequent. Equation 2.2 and 2.3 are the formula for support and confidence.

$$Support(X \rightarrow Y) = \frac{\text{Number of item set contains } x \text{ and } y}{\text{Total number of item set}} \quad (2)$$

$$Confidence(X \rightarrow Y) = \frac{Support(X \rightarrow Y)}{Support(X)} \quad (3)$$

3. Methodology

In implementing this project, Cross-Industry Process for Data Mining also known as CRISP-DM has been used. CRISP-DM is a structured approach to plan and execute data mining project. Fig. 2 shows the CRISP-DM framework. It consists of six phases. The next section will discuss all phases in detail.

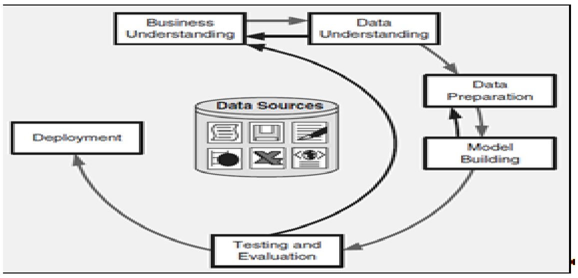

Fig. 2: CRISP-DM [13]

3.1. Business Understanding

In this phase, the problem of the selected case study will be identified. The Corm Café in Jasin, Melaka has been in operation since 2012. Pressure in the current business environment has required the owner to find a new business strategy. The first step is to understand customer preferences. Corm Café offers a contemporary selection of food to their customer ranging from western to local food, as shown in Fig. 4. There are 38 different types of food and drink divided into nine categories. Currently, Corm Café is operated manually, from order processing to record keeping. Every customer order is recorded in single piece of paper as shown in Fig. 3.

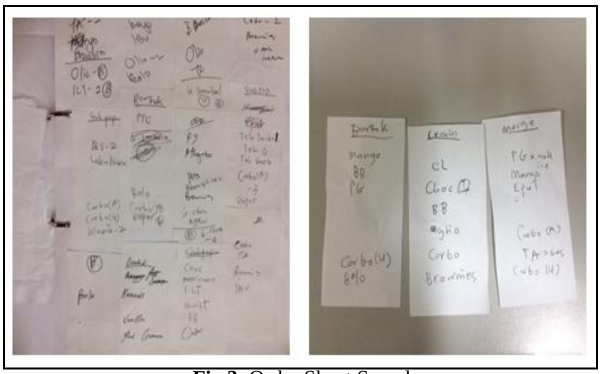

Fig.3: Order Sheet Sample

MINUM		
KOPI		
01.	Espresso	RM5
02.	Latte	RM5
03.	Caramel Latte	RM6
04.	Mocha	RM6
05.	Cappuccino	RM5
06.	Americano	RM5
07.	Affogato	RM6
08.	Conpana	RM6
09.	Coffee tonic	RM5
10.	Flat white	RM5
11.	Kopi lokal	RM2
TEH		
01.	Matcha latte	RM8
02.	Merah klasik	RM2
03.	Lemon	RM3
04.	Earl grey	RM3
05.	Passion fruit	RM3
COKLAT		
01.	Coklat 77000 panas	RM 5
02.	Coklat 77000 sejuk	RM6
KRUSH/JUS		
01.	Epal	RM5
02.	Oren	RM5
03.	Pink guava	RM3
04.	Mango	RM3
ICE BLENDED		
01	Peppermint choc	RM5
02	vanilla	RM4

MAKAN		
KEK		
01.	Keju	RM5
02.	Brownies	RM5
03.	Moist	RM5
04.	Brookies	RM5
TORTILLA		
01.	Daging	RM6
02.	Ayam	RM6
03.	Campur	RM6
PASTA		
01.	Aglio e olio	RM8
02.	Carbonara	RM8
03.	Bolognese daging	RM8
SNEK LOKAL		
01	Cakoi super	RM5
02	Cakoi klasik	RM3
03	Roti kum Sambal/kaya/coklat	RM3
04	Roti kahwin	RM2

Fig.4: Menu Offerred at Corm Cafe

3.2. Data Understanding

In this phase, the dataset used in this research is identified, including daily order sheets from the café. Five months from May until September 2017 have been collected which consist of 1026 order sheets. There are three attributes of the order sheets which are date, name of food and drinks, and quantity for each order. However, this is not enough for mining process. The researcher must combine information from the menu to get the full set of data. as discussed in the next section.

3.3. Data Preparation

In this phase, datasets were prepared for the mining process. There are six attributes involved as shown in Table 2. It is a combination of data from order sheet and food menu. There are four steps involved in data preparation which are i) data consolidation, ii) data cleaning, iii) data transformation iv) data reduction. In data consolidation, all order sheets collected were transformed into single file as shown in Fig. 5. Next was data cleaning, where the values of data sets were preprocessed. Missing, non-existent and incomplete data were identified and treated. In the third step, data were transformed into coded values for algorithm simplification, as shown in Table 3.

Table 2: Attribute Description

Attribute	Description	Data Type
Date	Order date	Date
Order ID	Unique attribute to represent each order sheet.	Numeric
Food and Drink Name	Food and drink ordered by customer	String
Category	Food and drink category	String
Quantity	Food and drink quantity in each order	Numeric
Price	Price for each food and drink	Numeric

	A	B	C	D	E	F
1	date	Order ID	Name of food/drink	Category	Quantity	Price
2	02/05/2017	1	Coklat 77000 panas	Coklat	2	RM5.00
3	02/05/2017	1	Epal	Krush/Jus	2	RM5.00
4	02/05/2017	1	Brownies	Kek	1	RM5.00
5	02/05/2017	1	Tortilla ayam	Tortilla	1	RM6.00
6	02/05/2017	2	Pink Guava	Krush/Jus	1	RM3.00
7	02/05/2017	2	Mango	Krush/Jus	1	RM3.00
8	02/05/2017	2	Epal	Krush/Jus	1	RM5.00
9	02/05/2017	2	Carbonara	Pasta	2	RM8.00
10	02/05/2017	2	Tortilla ayam	Tortilla	1	RM6.00
11	02/05/2017	3	Mango	Krush/Jus	1	RM3.00
12	02/05/2017	3	Keju	Kek	1	RM5.00
13	02/05/2017	3	Pink Guava	Krush/Jus	1	RM3.00
14	02/05/2017	3	Carbonara	Pasta	1	RM8.00
15	02/05/2017	3	Bolognese	Pasta	1	RM8.00
16	02/05/2017	4	Caramel Latte	Kopi	1	RM6.00
17	02/05/2017	4	Coklat 77000 sejuk	Coklat	1	RM6.00
18	02/05/2017	4	Keju	Kek	1	RM5.00
19	02/05/2017	4	Aglio e olio	Pasta	1	RM8.00
20	02/05/2017	4	Carbonara	Pasta	1	RM8.00
21	02/05/2017	4	Brownies	Kek	1	RM5.00
22	02/05/2017	5	Brownies	Kek	1	RM5.00
23	02/05/2017	5	Keju	Kek	2	RM5.00
24	02/05/2017	5	Latte	Kopi	1	RM5.00
25	02/05/2017	5	Mocha	Kopi	1	RM6.00

Sheet1 Sheet2 Sheet3

Edit

Fig. 5. New Data set after Data Consolidation Phase

Table 3: Example of Transformed Data set

Dt	o_no	f_d	ctg	qty	pri
2_5	1	c_pa	c	2	5
2_5	1	kr_e	kr	2	5
2_5	2	kr_e	kr	1	5
2_5	2	p_c	p	2	8
2_5	2	t_a	to	1	6
2_5	3	kr_m	kr	1	3
2_5	3	k_kj	k	1	5

The last step is data reduction. In this step, the dataset was reduced to suit the frequent itemset mining process and requirements. For this research, the data set has been divided into two datasets as shown in Table 4 and Table 5 respectively. Dataset 1 is focusing on order ID (o_no), name of food and drink (f_d) and category (ctg), while in Dataset 2 is date (dt), order ID (o_no) and name food and drink (f_d).

Table 4: Dataset 1

o_no	f_d	ctg
1	c_pa	c
1	kr_e	kr
2	kr_pg	kr
2	kr_m	kr
3	kr_m	kr
3	k_kj	k
5	k_bw	k
5	k_kj	k
5	k_l	kp

Table 5: Dataset 2

dt	o_no	f_d
2_5	1	c_pa
2_5	1	kr_e
2_5	12	kr_e
2_5	12	p_c
2_5	12	kr_m
4_5	13	kr_pg
4_5	13	p_c
4_5	13	p_a

3.4. Model Building

In this research, Rapid Miner has been used for mining process. It covers from data preparation process to generating associated rule. FP-Growth operator and create association rule operator is used for frequent itemset mining process.

3.5. Testing and Analysis

To test the developed model, both datasets have been used. The objective of Dataset 1 is to find an associated rule for the food category based that comes together in each order sheet or transaction or in term of MBA is basket. While the objective for dataset 2 is to find associated rules for food and drink based on order date and order id. To test the model developed, three different minimum support and minimum confidence value were used as shown in Table 6.

Table 6: Minimum Support Confidence

Ref.	Minimum Support	Minimum Confidence
[14]	0.01 (1%)	0.7 (70%)
[15]	0.1 (10%)	0.4 (40%)
[6]	0.25 (25%)	0.6 (60%)

3.6. Deployment

In this phase, the associated rules obtained were analyzed. Redundant rules were eliminated in this phase.

4. Experiment

Three experiments have been conducted using different types of dataset. The researcher also determined several support and confidence levels used by a previous case study. There are two types of datasets, which involve category and food and drink.

4.1. Result for Experiment 1

For the first experiment, minimum support and minimum confidence values were adopted per [6]. Min. support is 0.25 and min confidence is 0.7. Only dataset 1 produced a result as shown in Table 7, whereas no associated rule is found from dataset 2

Table 7: Result from Experiment 1 for Dataset 1

Premises	Conclusion	Support	Confidence
Krush	Pasta	0.302	0.684

4.2. Result for Experiment 2

In second experiment, minimum support and minimum confidence value as suggest by [15] which are min. support is 0.1 and min confidence is 0.4. Tables 8 and 9 show the results obtained from both datasets.

Table 8: Result from Experiment 2 for Dataset 1

Premises	Conclusion	Support	Confidence
Kek, Snek Lokal	Krush	0.140	1
Kek, Coklat	Kopi	0.140	0.857
Krush, Tortilla	Pasta	0.116	0.833
Coklat	Kopi	0.186	0.727
Kek, Tortilla	Krush	0.116	0.714
Kopi, Tortilla	Pasta	0.116	0.714
Tortilla	Kopi	0.163	0.700
Tortilla	Pasta	0.163	0.700
Krush	Pasta	0.302	0.684
Kopi, Krush	Pasta	0.140	0.667
Snek Lokal	Krush	0.186	0.667
Snek Lokal	Pasta	0.186	0.667
Coklat	Kek	0.163	0.636
Pasta, Kek	Krush	0.116	0.625
Pasta, Snek Lokal	Kopi	0.116	0.625

Table 9: Result from Experiment 2 for Datase 2

Premises	Conclusion	Support	Confidence
Pasta aglio e olio	Pasta carbonara	0.155	0.432

4.3. Result for Experiment 3

Table 10: Result from Experiment 3 for Dataset 1

Premises	Conclusion	Support	Confidence
Kopi, Krush, Kek, Tortilla, Snek Lokal, Coklat	Pasta	0.023	1
Pasta, Krush, Kek, Coklat	Kopi	0.047	1
Tortilla, Snek Lokal, Coklat	Kek	0.023	1
Krush, Tortilla, Snek Lokal	Kek	0.070	1
Kek, Ice blended, Coklat	Kopi	0.047	1
Tortilla, Snek Lokal	Krush	0.070	1
Snek Lokal, Coklat	Kopi	0.047	1
Ice blended, Coklat	Kopi	0.093	1
Snek Lokal, Tortilla	Pasta	0.070	1
Pasta, Coklat	Krush	0.093	1
Snek Lokal, Coklat	Pasta	0.047	1
Kek, Tortilla	Pasta	0.070	0.750
Kopi, Kek	Coklat	0.140	0.750
Krush, Snek Lokal	Kek	0.140	0.750
Krush, Tortilla	Pasta	0.116	0.833

For the third experiment, minimum support and minimum confidence values were set per [16] with min. support of 0.01 and min confidence of 0.7. Tables IX and X depict the results obtained from both dataset. Both datasets generated more than ten rules. From the results obtained, support and confidence values play an important role in this experiment. In addition, as we can see, dataset 1 generates more rules because the rules are generated based on food and drink category, for which occurrence in the data sets is more frequent compared to dataset 2. Dataset 2 contains food and drink names, for which the number of occurrences in the whole dataset for each item would be lesser.

Table 11: Result from Experiment 3 for Dataset 2

Premises	Conclusion	Support	Confidence
Teh Lemon, Mocha	Pasta carbonara	0.019	0.731
Pasta aglio e olio, Teh Lemon, Cakoi super	Pasta carbonara	0.011	0.733
Pasta aglio e olio, Cakoi super, Tortilla campur	Pasta carbonara	0.0011	0.733
Pasta aglio e olio, Pasta Bolognese, Cakoi super	Pasta carbonara	0.016	0.762
Pasta carbonara, Teh Lemon, Cakoi Klasik	Pasta aglio e olio	0.011	0.786z

5. Result and Discussion

Tables 12 and 13 show the associated rules obtained from datasets 1 and 2 for all experiments. Only the top five selected rules are chosen.

Table 4: Associated Rules from Dataset 1

Support & Confidence	Associated Rules
support: 0.25 confidence: 0.7	{Krush} → {Pasta}
support: 0.1 confidence: 0.4	{Kek, Snek Lokal} → Krush {Kek, Coklat} → {Kopi} {Krush, Tortilla} → {Pasta} {Coklat} → {Kopi} {Kek, Tortilla} → {Krush}
support: 0.01 Min confidence: 0.7	{Ice belended, Coklat} → {Kopi} {Pasta, Coklat} → Krush {Tortilla, Snek Lokal} → {Krush} {Krush, Tortilla, Snek Lokal} → {Kek} {Snek Lokal, Coklat} → {Pasta}

Table 5: Associated Rules from Dataset 2

Support & confidence	Associated Rules
support: 0.25 confidence: 0.7	No rules found.
support: 0.1 confidence: 0.4	{Pasta Aglio e Olio} → {Pasta Carbonara}
support: 0.01 confidence: 0.7	{Teh Lemon, Mocha} → {Pasta Carbonara} {Pasta aglio e olio, Teh Lemon, Cakoi super} → {Pasta Carbonara} {Pasta aglio e olio, Cakoi super, Tortilla campur} → {Pasta Carbonara} {Pasta aglio e olio, Pasta Bolognese, Cakoi super} → {Pasta Carbonara} {Pasta carbonara, Teh Lemon, Cakoi Klasik} → {Pasta Aglio e Olio}

From the results obtained, support and confidence values play an important role in this experiment. In addition, as we can see, dataset 1 generates more rules because the rules are generated based on food and drink category, for which occurrence in the data sets is more frequent compared to dataset 2. Dataset 2 contains food and drink names, for which the number of occurrences in the whole dataset for each item would be lesser.

6. Conclusion

This research proposes an implementation of MBA technique to a Corm Café. The objective of this research was to understand a customer buying pattern from the Corm Café customer's daily order. The data was obtained from daily order sheets recorded manually by Corm Café employees during the ordering process. Three different experiments with two datasets have been conducted. As a result, we can see the associated menu ordered by customer. For example, 70% of customers who order krush as a drink will order pasta for their meal. This is based on 25% of the whole data. This information can be used for future planning for Corm Café in terms of inventory and marketing strategy.
In this research, only five months data have been used. For future research, more data should be used. This would increase the accuracy of the results. For future research, more data should be used. This would increase the accuracy of the results.

Acknowledgement

Special thanks to Universiti Teknologi MARA and Research Management Centre for the continuous commitments given towards the completion of this research. Special gratitude also goes to the Ministry of Education Malaysia for the sponsorship granted to the main author.

References

[1] Kaur M and Kang S. Market basket analysis: identify the changing trends of market data using association rule mining. Procedia Comput. Sci., 2016, 85: 78–85.
[2] Chen Y L, Tang K, Shen R J, and Hu Y H. Market basket analysis in a multiple store environment, Decis. Support Syst. 2005, 40(2): 339–354.
[3] Agrawal R, Imielinski T, and Swami A. Mining association in large databases, Proc. 1993 ACM SIGMOD Int. Conf. Manag. Data 1993, 207–216.
[4] Raorane A A, Kulkarni R V, and Jitkar B D. Association rule – extracting knowledge using market basket analysis, Res. J. Recent Sci. Feb. Res.J.Recent Sci, 2012, 1(2), 19–27.
[5] Hilage T and Kulkarni R. Application of data mining techniques to a selected business organization with special reference to buying behavior, Int. J. Database Manag. Syst., 2011, 3(4), 169–181.
[6] Musungwini S, Zhou T G, Gumbo R, and Mzikamwi T, The relationship between (4ps) & market basket analysis: A case study of grocery retail shops in Gweru Zimbabwe. International Journal of Scientific and Technology, 3(10), 2014.
[7] Bermudez I J, Apolinario I K, and Abad A G. Layout optimization and promotional strategies design in a retail store based on a market basket analysis, International Multi-Conference for Engineering, Education, and Technology, 2016, pp. 20–22.
[8] Pillai J. Overview of itemset utility mining and its applications, Int. J. Computer Appl., 2010, 5(11): 9–13,
[9] Pillai J. User centric approach to itemset utility mining in market basket analysis, Int. J. Comput. Sci. Eng., 2011, 3, 393–400.
[10] Kamakura W A. Sequential market basket analysis. Mark. Lett., 2012, 23(3): 505–516.
[11] Agarwal B R, Srikant R, and Ahmad M A. Fast algorithms for mining association rules, Proceedings of the 20th VLDB Conference Santiago, Chile, 1994, 487-498
[12] Borgelt C, An implementation of the fp-growth algorithm, Discovery, 2005.
[13] Turban E, Sharda R, and Delen D, Business Intelligence and Analytics: Systems for Decision Support, Tenth. Harlow, United Kingdom: Pearson Education Limited, 2014.
[14] Kouzis-loukas M. Analysing customer baskets, 2014.
[15] Velislava G. Market basket analysis of beauty products, Erasmus University Rotterdam, 2013.