



**NOVA**

**IMS**

Information  
Management  
School

# MEGI

**Mestrado em Estatística e Gestão de Informação**  
Master Program in Statistics and Information Management

## Market Basket Analysis

Trend Analysis of Association rules in different time periods.

Sohaib Zafar Ansari

M2016027

Dissertation presented as partial requirement for  
obtaining the master's degree in Statistics and  
Information Management

NOVA Information Management School  
Instituto Superior de Estatística e Gestão de Informação  
Universidade Nova de Lisboa

**NOVA Information Management School**  
**Instituto Superior de Estatística e Gestão de Informação**  
Universidade Nova de Lisboa

# **MARKET BASKET ANALYSIS: TREND ANALYSIS OF ASSOCIATION RULES IN DIFFERENT TIME PERIODS**

by

Sohaib Zafar Ansari

M2016027

Dissertation presented as partial requirement for obtaining the master's degree in statistics and Information Management, with a specialization in Market Research and Customer Relationship Management

**Advisor / Co Advisor:** Fernando Bação (Associate Professor)

February 2019

## **ACKNOWLEDGEMENT**

My journey through this dissertation and this degree has required more effort than I was able to deliver on my own. For this reason, I would like to acknowledge all the people who supported me throughout the process.

Firstly, I would like to thank my supportive supervisor Dr. Fernando Bação without whose guidance and knowledge I would not have completed this journey. He has encouraged me to learn new concepts and new tools which will be highly useful in building a successful career in future.

I would like to thank my colleague and friend Jonas Deprez who got me in touch with a retail company who were willing to provide me their transactional data so that I can use it for my thesis.

I would like to thank all the experts that participated in the evaluation of this and guided me with their knowledge and expertise no matter the distance and time.

At last, I would like to thank my family and friends for being supportive and loving at every step and for being there for me to give me a push whenever I needed it.

## **ABSTRACT**

Market basket analysis (i.e. Data mining technique in the field of marketing) is the method to find the associations between the items / item sets and based on those associations we can analyze the consumer behavior. In this research we have presented the variability of time, because with the change in time the habits or behavior of the customer also changes. For example, people wear warm clothes in winter and light clothes in summer. Similarly, customers purchase behavior also changes with the change in time. We study the problem of discovering association rules that display regular cyclic variation over time. This problem will allow us to access the changing trends in the purchase behavior of customers in a retail market, and we will be able to analyze the results which will display the changing trends of the association rules. In this research we will study the interaction between association rules and time. We worked on transactional data of a Belgian retail company and analyzed the results which will help the company to build up time period specific marketing strategies, promotional strategies, etc. to increase the profit of their company.

## **KEYWORDS**

Data mining; Market basket analysis; association rules; apriori algorithm; changing trends of association rules; Comparative Analysis

# Contents

CHAPTER 1: INTRODUCTION .....	1
1.1. DATA MINING .....	2
1.2. MARKET BASKET ANALYSIS .....	2
CHAPTER 2: SCOPE OF THE STUDY .....	5
CHAPTER 3: LITERATURE REVIEW .....	8
CHAPTER 4: METHODOLOGY .....	13
4.1. RESEARCH PURPOSE .....	13
4.2. RESEARCH APPROACH .....	13
4.3. RESEARCH STRATEGY .....	13
CHAPTER 5: STRATEGY FOR MARKET BASKET ANALYSIS .....	15
5.1. KEY TERMS AND CONCEPTS .....	15
5.1.1. Association rules: .....	15
5.1.2. Antecedent and Consequent: .....	16
5.1.3. Causality: .....	16
5.1.4. Frequent Itemset: .....	17
5.1.5. Time Period: .....	17
5.1.6. Transaction: .....	17
5.1.7. Long Tail Effect: .....	18
5.2. DETERMINE SUITABILITY OF Market Basket Analysis .....	18
5.2.1. ADVANTAGES OF USING MARKET BASKET ANALYSIS .....	19
5.2.2. CHECK MBA REQUIREMENTS .....	20
5.2.3. APRIORI ALGORITHM .....	21
5.2.4. OTHER ALGORITHMS USED IN MARKET BASKET ANALYSIS .....	26
CHAPTER 6: PRACTICAL IMPLEMENTATION OF MARKET BASKET ANALYSIS ON THE DATASET .....	30
6.1. DATA COLLECTION .....	30
6.2. DATASET DESCRIPTION .....	30
6.2.1. ATTRIBUTES .....	30
6.3. TOOLS USED FOR ANALYSIS .....	31
6.4. DATA EXPLORING & TRANSFORMATION .....	32
6.5. BINARY REPRESENTATION .....	33
6.6. ANALYSIS & DISCUSSION .....	34
6.6.1. Association rules from dataset .....	34

6.6.2. Time development of association rules .....	35
6.6.3. Top 30 association rules comparison.....	35
CHAPTER 7: CONCLUSION.....	39
7.1. SUMMARY & FUTURE WORK .....	39
7.2. LIMITATIONS .....	40
CHAPTER 8: REFERENCES .....	42
ANNEX .....	46

## Figures

1. Market basket .....	3
2. Transaction example .....	18
3. All possible subsets .....	22
4. If an itemset set id infrequent, we do not consider its super sets .....	22
5. Transactional date to frequent items .....	23
6. Pruning and joining .....	24
7. Pruning and joining again until there are no more frequent items left .....	25
8. Support, confidence and lift calculation .....	25
9. Calculation of confidence .....	26
10. Long tail effect on dataset .....	33
11. Degree of confidence of top 30 rules .....	36
12. Degree of support of top 30 rules .....	37
13. Lift of top 30 association rules .....	38

## Tables

1. Research strategy methods .....	14
2. Attributes description .....	31
3. Data exploring .....	33
4. Category and subcategories distribution .....	33
5. Association rules with different thresholds .....	34
6. 30 association rules with highest degree of confidence .....	36
7. 30 association rules with highest degree of support .....	37
8. 30 association rules with highest lift .....	38

## CHAPTER 1: INTRODUCTION

Retail have evolved since the common corner stores from the 1900s, until the new e-commerce, that have shaken the retail world to its core. This changing process have led to a new era of unlimited possibilities for commerce and the consumers.

Consumers nowadays have a wide range of options, independently in almost every domain. In the past, when the consumer had to buy something, he only could choose a product from the catalogue of the store. However, with the new era of information and globalization, the list of options have increased exponentially. Now consumers can choose between a huge variety of products and their variances. Limitations as geography, season and so on are no longer an issue. Products that were considered as luxury goods are considered as common products now. All of this led the companies to have limitless possibilities nowadays. However, this limitless possibility caused a huge number of new competitors to enter the market. The retail stores seek for marketing strategies in order to attract new customers or keep its current customers. Only new marketing strategies could help this situation by offering efficient promotions and proper product planning. Market basket analysis which have been practiced in other countries have shown remarkable success. As a result, multinational retail stores such as Walmart and Tesco have been using market basket analysis in order to achieve higher profit. But in order to get the insights using market basket analysis we need to have information about our customer purchase regarding what they buy and when they buy it. Hence, comes the importance of the data about the customers purchases which is based on their behavior.

In the last two decades there have been an explosive growth in the data, but not all data is relevant. So, the companies started to use data to discover and extract relevant information. This process of extracting useful information is called Data Mining also known as Knowledge Discovery and Data (KDD) process. Data mining allows a search for valuable information in large volumes of data (Weiss & Indurkha, 1998). It is widely used in several aspects of science such as manufacturing, marketing, CRM, retail trade, psychology, education, etc. There are several data mining techniques which helps to extract meaningful knowledge and find the solution to the organizational problems. Some of them are Neural Networks, Artificial information, Classification, Association, Prediction, Clustering, Regression, Sequence discovery, Visualization.

## 1.1. DATA MINING

According to David Hand, “Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner”. (David Hand, Heikki Mannila, and Padhraic Smyth, 2001)

Application of data mining techniques in various fields have been very effective till now. For example, in the field of healthcare, it can help healthcare insurers detect fraud and abuse, healthcare organizations make customer relationship management decisions, physicians identify effective treatments and best practices, and patients receive better and more affordable healthcare services (Hian Chye Koh and Gerald Tan). In the field of marketing, customer segmentation involves the subdivision of an entire customer base into smaller customer groups, consisting of customers who are similar within each specific segment (Woo, Bae, & Park, 2005). This segmentation technique is useful to identify the customers and cluster them based on their characteristics and features.

One major application domain of data mining is the analysis of transactional data. In a recorded transactional database each transaction is a collection of items. The best technique to analyze and find the relationships and patterns between items is market basket analysis. It is one of the most interesting research areas of the data mining that have received more attention by researchers nowadays.

## 1.2. MARKET BASKET ANALYSIS

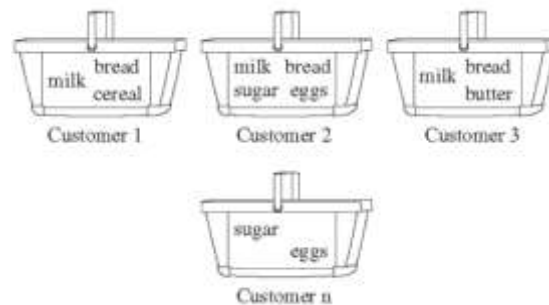
Market basket is defined as an itemset bought together by a customer on a single visit to a store. In our visit to the super market we tend to buy a lot of products from different categories and put them all together in one single basket. Which is considered to be a single transaction. Market basket analysis is the analysis of those baskets all together.

Market basket analysis encompasses a broad set of analytic techniques aimed at uncovering the associations and connections between specific objects, discovering customer behaviors and relation between items. In retail, it is used based in the following idea, if a customer buys a certain group of items, is more (or less) likely to buy another group of items. For example, it is known that when a customer buy beer, in most of cases, buys chips as well. These behaviors produced in purchases is something that the companies selling their products are interested in. The sellers/



supermarkets are interested in analyzing which items are purchased together in order to create new marketing/sales strategies that can be helpful in improving the benefits of the company as well as customer experiences.

Most of the retail markets are more focused on the what their customer's buy. But they ignore the fact about when they buy it. Which is also considered to be a huge factor in their behavior of purchase. This thesis is focused on not just "what" the customer buys but also "when" they buy it. According to Forbes magazine marketers are constantly looking into future, trying to predict next big trend and data driven marketing is the top most trend right now in which time plays a highly significant role. So, data driven marketing with time as a crucial factor will help us predict a better future for the retail company.



*figure 1: Market basket*

The market basket analysis is a powerful tool for the implementation of up-selling, cross-selling, inventory management strategies (Chen, Tang, Shen, & Hu, 2005). Market Basket Analysis is also known as association rule mining or affinity analysis, which have been used to understand consumer behavior regarding the types of the purchases they make. It is a Data Mining technique that originated in the field of marketing and was initially used to understand purchase patterns of the customers by extracting associations and co-occurrence from a transactional database (i.e. market basket data). For example, when shopping in a supermarket, consumer rarely buy one product, they far more likely to purchase an entire basket of products, mostly from different product categories. This allows us to uncover nonobvious, usually hidden and counterintuitive associations between items, products, or categories. We are also able to extract products and product categories which are purchased together, and these associations can be represented in the form of association rules. These association rules enable managers to develop marketing strategies like developing interventions, promoting specific product categories, offering promotions, etc.

which eventually leads to get customers spend more money based on two different principles. *Up-selling*, which consists in buying a large quantity of the same product or adding new features and *Cross-selling*, which consists in adding more products from various categories. Market Basket Analysis is also very much useful in *stock management and placement of items*.

## CHAPTER 2: SCOPE OF THE STUDY

Market basket analysis is mostly used in three main domains.

The *first* domain is the creation of personalized recommendations which is a very well known methodology nowadays. During the explosion of the e-commerce, personalized recommendations have appeared as a part of the marketing process. Basically, the idea consists of suggesting items to customer based on his/her preferences. The first way to do it is, by suggesting items like the ones that the customer have purchased in the past which is also called collaborative filtering. The second way is, looking for similar customers and recommending items that had been purchased by others. It is called content-based filtering. There is also a third way to do it which is called hybrid recommendation system. The name of this system explains itself. It is the combination of both collaborative and content-based filtering, which can be very effective in certain cases. These strategies are often used for companies in order to realize cross-selling and upselling strategies.

The *second* domain where market basket analysis is used in the analysis of spatial distribution in chain stores. Due the increasing number of products that nowadays exist, physical space in stores started to be a problem. More and more, stores invest money and time trying to find which distribution of items can lead them to obtain more sells. Due to that, knowing in advance which items are commonly purchased together, the distribution of the store can be changed in order to sell more products. It is also very helpful in inventory or stock management within the store. Also having several stores in different areas of the city or the country helps the chain stores to target more and more customers and develop marketing strategies based on customer demographics. This way the chain stores conduct target marketing which also leads to the variability of price of same product in different store.

The *last* domain is in the creation of marketing strategies which focus on discounts and promotions which is solely based on customer's behavior through which special sales or targeted promotions can be performed. When sales campaigns are prepared, promoted items must be chosen very carefully. The main goal of a campaign is to entice customers to visit the store and buy more than they usually do. Profit margins on promoted items are usually cut; therefore, non-promoted items with the higher profit margin should be sold together with the promoted items. Therefore, the items chosen should make the promotion effective enough to generate higher sales. Customers who buy

a bathroom accessory often also buy several other bathroom accessories. It makes sense that these groups are placed side by side in a retail store so that customers can access them easily. When different additional brands are sold together with the basic brands, the revenue from the basic brands is not decreasing but increasing.

Market basket analysis targets customer baskets in order to monitor buying patterns and improve customer. It is an important component of analytical CRM in retail organizations. By analyzing, recurring patterns in order to offer related goods together, a pattern can be found and therefore the sales can be increased. Sales on different levels of goods classifications and on different customer segments can be tracked easily. For example, if the company knows which items are often purchased together, they can create new offers on those products in order to increase the sale of those items or even they can create combo promotions which increases in the sale of their products.

The main aim of the project is the detection and analysis of purchase behavior of the customers (items purchased together). Based on variability in demographics and difference in the segment of customers, every store will have different results. So, in our research we will be focusing on a single store and its transactions which will be giving a comparative analysis in different time periods. This way we will be able to investigate the fluctuation of the consumer behaviors on purchases with the change in time.

Analyzing transactional database and discovering association rules have helped the managers to increase the sale of the company. Time series analysis allows you to track the changes in the trends and to keep track of the progress or downfall of any market. So, we need to focus on the problem which displays regular variations in the association rules over time. For example, if we compute association rules over monthly sales data, we may observe seasonal variation where certain rules are true at approximately the same month each year. Similarly, changing time variations to hourly, daily, weekly, etc., might display different association rules. This will allow us to find the patterns of purchases by the customers which are different in different time periods. Some changes can be seasonal, they may also vary because of different demographic and sociographic. For example, during summer season we can see that the sale of cold beverages increases a lot whereas during the winter season it drops down. Our research is more specific to the time periods which will allow us to do up-selling and cross-selling in every time lags and help the company to increase its sale and maximize profit.

Similar study was first made by Ramaswamy and Siberschatz (1998) to discover the association rules that repeat in every cycle of a fixed time span. This information about variations in different time periods allowed marketers to better identify the trends in association rules and help in better predictions.

## CHAPTER 3: LITERATURE REVIEW

*S.H. Liao, P.H. Chu and P.Y. Hsiao* reviewed published papers from 2000 to 2011 about data mining techniques and their applications. These techniques have tend to become more expertise-oriented and their application is more problem-centered, leading to development of advanced algorithms and their application in different discipline areas, like, computer science, engineering, medicine, mathematics, earth and planetary sciences, biochemistry, genetics and molecular biology, business, management and accounting, marketing decisions, social science, decision sciences, multidisciplinary, environmental science, energy, agricultural and biological sciences, nursing, material science, neurology, chemical engineering, etc.

Market basket Analysis is rather a practical subject than academical, therefore most of the studies on the matter have been practiced in actual retail stores. MBA as an old field in data mining and is also one of the best examples of mining association rules (Gancheva, Market basket analysis of beauty products, 2013). Rakesh Agrawal and Usama Fayyad as pioneers in data mining, Association Rule Mining (ARM) and Clustering have developed different algorithms to help users achieve their objectives.

Most of the data mining algorithms have existed since decades, but in last decade there have been sudden increase in the data and realization of the importance of data in every field. *S. Linoff, Michael J.A. Berry* in his book suggested for companies in the service sector, data/ information confers competitive advantage. That is why hotel chains records your preference for a non-smoking room, a car rental companies record your preferred type of car. Similarly, credit card companies, airlines, retailers, etc. compete more on services than on price. Many companies find that the information on their customers is not only valuable for themselves but also for others. Like, the information about the customers of a credit card company is also useful for an airline company – they would like to know who is buying a lot of airline tickets. Similarly, google knows what people are looking for on web and it takes advantage of the knowledge by selling this information.

In 2009 *E. Ngai, L.Xiu and D. Chau* presented how data mining in customer relationship management is an emerging trend, which helps in identification, attraction, retention and development of a customer. Customer retention and development are important to maintain a long

term and pleasant relationship with the customers which is very much useful in maximizing the organization's profit. Data mining provides a lot of opportunities in the market sector. For customer identification the methods mostly used are classification, clustering and regression. For customer development the methods usually used are classification, regression, association discovery, pattern discovery, forecasting, etc.

*Aiman Mushtaq* in 2015 highlighted how data mining in marketing helps to increase return on investment or net profit, improve customer relationship management, market analysis, building better marketing strategies, reduces unnecessary expense, etc. With the growing volume of data everyday various techniques are being used to mine the data in the field of marketing and helping in fulfilling the organizational goals.

Market basket analysis is a data mining technique originated in the field of marketing to understand purchase patterns of customers. It has made a lot of advancements since it was first introduced in 1993 by *Rakesh Agrawal*. He proposed the first associative algorithm called apriori algorithm, which have been used as part of technique in association, classification and associative classification algorithms. It was mostly used in stock management and placement of items. Now, market basket analysis is being used to build predictive models and to get interesting insights which are helpful in decision making. Its application is in several fields. *Cascio and Aguinis* in 2008 noted, "there is a serious disconnect between the knowledge that academics are producing and the knowledge the practitioners are consuming." The use of MBA can produce knowledge that is relevant and actionable and market basket analysis can help bridge the science and practical divide.

*S. Kamley, S. Jaloree, R.S. Thakur* in 2014 have developed an association rule mining model for finding the interesting patterns in stock market dataset. This model is helpful in predicting the price of share which will be helpful for stock brokers and investors to invest in the right direction by understanding market conditions. In June 2015, *S.S. Umbarkar and S.S. Nandgaonkar* used association rule mining for prediction of stock market from financial news. Prediction depends on technical trading indicators and closing prices of the stock.

One of the most important use of market basket analysis is product placement in the super market. In 2012 by *A. A. Raorne* market basket analysis was used in understanding the behavior of the customer. The researchers did experimental analysis by employing association rules using MBA,

which improved the strategy in placement of product on the shelf leading to fetch more profit to the seller. This research was effective in fetching more profit but what it was lacking was that changing of consumer behavior. To sustain in a competitive market the organizations must understand consumer behaviors and consumer behavior changes with the change in time.

In 2015, *G. Kapadia* did a study that analyses the pattern of consumer behavior of products of lifestyle store. It gives valuable insights relating to the formation of the basket. This study helped in product assortments, managing the stocks for the likely items sold, making promotions on the likely items sold, give discounts to the loyal customers and cross selling. The limitation of this study was that its scope was limited to one store in a specific region.

Data mining tools are also used in the field of education as well. In April 2015, *Om Prakash Chandrakar* applied association rule mining to analyze the student's performance in their examinations and predicts the outcome of forth coming examination. This prediction allows the students and professors to identify subjects which need more attention even before commencement of the semester.

*Solnet et al.* in 2016 studied the potential of market basket analysis to grow revenue of hotels. The researchers explored and derived the most attractive services and products which could attract and satisfy hotel guests and encourage them to repeat their purchase. This approach can increase revenue without increasing customer counts.

In February 2017, *Roshan Gangurde* did a study on building predictive model using market basket analysis which stated that in a retail business if we are using market basket analysis to come up with product bundles then you are basing past purchase behavior of customers to predict future purchase behavior, which is a predictive model. He also concluded that, with MBA, the leading retailers can attract more customers, increase the value of market basket, drive more profitable advertising and promotion and much more. The study also suggested to design and develop intelligent prediction models to generate the association rules that can be adopted on the recommendation system to make the functionally more operational. Later by the end of 2017, they designed optimized technique for MBA with goal of predicting and analyzing the consumers buying behaviors. In this study, they introduced novel algorithms based on data cleaning, which is one of the most important challenge in every field of data analysis. To overcome this challenge, they combined two data mining algorithms i.e. apriori algorithm and neural networks. They also



highlighted that one of the biggest challenges is that demands of customers are continuously changing with respect to seasons and time. Also output of MBA is totally dependent on time and seasons and so we need to perform it over and over.

Analyzing the trends is very useful method to understand any businesses performance. *Debaditya and Nimalya* in 2013 attempted to develop a method using association rule mining to find out the most preferable and popular genres which can be represented as movie business's trend. The study can predict the possible movie trend based on the genres. As viewer's taste in movies can change time to time, it was important to get the insights in the changing trends. This study is very useful to the production houses to drive movie business towards profitability.

*Kaur and Kang* (2016) did MBA to identify the changing trends of market data using association rule mining. This study proposed a different approach of periodic mining which will enhance the power of data mining techniques. This study was helpful in finding out interesting patterns from large database, predicting future association rules as well as gives us right methodology to find out outliers. This study shows advancement by not just mining static data but also provides a new way to consider changes happening in data.

*S. Tan and J. Lau* (2014) tried a different approach by summarizing a real-world sales transaction data set into time series format. Rather than applying association rule mining (i.e. often used in market basket analysis), they used time series clustering to discover commonly purchased items that are useful for pricing or formulating cross-selling strategies. This approach uses a data set that is substantially smaller than the data to be used for association analysis which shows that certain market basket analysis can be analyzed more easily using time series clustering instead of association analysis.

*G.N.V.G. Sirisha, M. Shashi & G.V. Padma Raju* in 2013 presented a paper overviewing distinct types of periodic patterns and their applications along with a discussion of the algorithms that are used to mine these patterns. Periodic pattern mining is very much useful in constructing classification/ prediction and recommender systems.

Data keeps on changing with time and interestingness of data differs from person to person. Time to time and task to task. So, attempts are being made to mine the necessary information from a

large amount of transactional data on a seasonal basis. Frequent item sets based on calendric pattern will be mined to generate association rules.

Similar study was first made by *Ramaswamy and Siberschatz* in 1998 to discover the association rules that repeat in every cycle of a fixed time span. This information about variations in different time periods allowed marketers to better identify the trends in association rules and help in better predictions.

Ismail H. Toroslu (2003) extended sequential pattern mining technique by adding repetition support and introduced cyclic pattern mining, which was more efficient than the sequential pattern mining.

Lee and Jiang (2008) relaxed the restrictive crisp periodicity of the periodic association rule to a fuzzy one and developed an algorithm for mining fuzzy periodic association rules. They cover two aspects, finding the pattern and determining the periodicity.

## CHAPTER 4: METHODOLOGY

Methodology are the guidelines or path on how to proceed in validating knowledge on your subject matter. Different areas of science have developed very different bodies of methodology based on which to conduct their research (Little, 2012).

### 4.1. RESEARCH PURPOSE

The ultimate purpose of every business is to find better ways to improve the profit for a long run. But for this research the aim would be to encountering actual case of dependencies among products chosen by customer.

Though several different products could be bought in a single visit to a mega store like, groceries, pillowcase, furniture, an electric toaster, etc. However, we believe that there are no coincidences for these choices. These decisions from several categories results in forming customer's shopping basket. Which with-holds the collection of categories that customer purchased on a specific shopping trip. (Manchanda, Ansari, & Gupta, 1999).

### 4.2. RESEARCH APPROACH

The implementation of Market Basket Analysis results into inventory management, marketing and promotion strategies, etc. Therefore, this a comparative analysis thesis, which displays the relevance of time. This analysis would be done by implementing the idea on a dataset and check the results. Exploratory data mining techniques are used followed by association rules, or pattern recognition, or 'if then' statements. Also, a comparative analysis is done on the data for three months (i.e. April, May and June) all together and individually.

### 4.3. RESEARCH STRATEGY

There are three main types of research strategies that exists namely quantitative, qualitative and mixed (Creswell, 2013). Researches can also be experimental and non-experimental which in some books are falling into another category while Carswell believes they are under three main mentioned categories. In the book research design by John W. Creswell he mentions that these three categories are not discrete, and they actually are at the ends on a continuum (Creswell, 2013). The research tends to be more qualitative than quantitative and vice versa so as a result all other types can be fitted in between this continuum. Therefore, he adds that the mixed methods are in

middle since it contains the elements of both these types. The methods alongside their types are presented in the following tables below explaining the strategies more clearly.

Research Strategy Methods		
Quantitative	Qualitative	Mixed Methods
<ul style="list-style-type: none"> <li>Experimental design</li> <li>Non-experimental designs, such as surveys</li> </ul>	<ul style="list-style-type: none"> <li>Narrative research</li> <li>Phenomenology</li> <li>Ethnographies</li> <li>Grounded theory studies</li> <li>Case Study</li> </ul>	<ul style="list-style-type: none"> <li>Sequential</li> <li>Concurrent</li> <li>Transformative</li> </ul>

Quantitative Methods	→ Mixed Methods ←	Qualitative Methods
<ul style="list-style-type: none"> <li>Pre-determined</li> <li>Instrument based questions</li> <li>Performance data, attitude data, observational data, and census data</li> <li>Statistical analysis</li> <li>Statistical interpretation</li> </ul>	<ul style="list-style-type: none"> <li>Both pre-determined and emerging methods</li> <li>Both open and closed-ended questions</li> <li>Multiple forms of data drawing on all possibilities</li> <li>Statistical text analysis</li> <li>Across databases interpretation</li> </ul>	<ul style="list-style-type: none"> <li>Emerging methods</li> <li>Open-ended questions</li> <li>Interview data, observation data, document data, and audio visual data</li> <li>Text and image analysis</li> <li>Themes, patterns interpretation</li> </ul>

TABLE 1: RESEARCH STRATEGY METHODS

The strategy opted for the research was mixed methods because in our research we are doing data interpretation, pattern interpretation, trying to find insights in the data, doing statistical analysis, data exploring, comparative study. In the coming chapters it will be made clearer.

## CHAPTER 5: STRATEGY FOR MARKET BASKET ANALYSIS

In this section we describe the entire research process. Before getting into the steps of the analysis. First, we clear some of the concepts that we will be coming across in our analysis.

### 5.1. KEY TERMS AND CONCEPTS

#### 5.1.1. Association rules:

Association analysis is also known as affinity analysis or association rule mining, a method commonly used for market basket analysis. ARM is currently the most suitable method for analysis of big market basket data but when there is a large volume of sales transaction with high number of products, the data matrix to be used for association rule mining usually ends up large and sparse, resulting in longer time to process data. Association rules provide information of this type in the form of “IF-THEN” statements. There are three indexes which are commonly used to understand the *presence*, *nature* and *strength* of an association rule. (Berry & Linoff, 2004; Larose, 2005; Zhang & Zhang, 2002)

**Lift** is obtained first because it provides information on whether an association exist or not or if the association is positive or negative. If the value for lift suggests that there is an existence of association rule, then we obtain the value for support.

$$\text{Lift} = \frac{P(A \cap B)}{P(A) * P(B)}$$

**Support** of an item or itemset is the fraction of transactions in our dataset that contain that item or itemset. It is an important measure because a rule that have low support may occur simply by chance. A low support rule may also be uninteresting from a business perspective because it may not be profitable to promote items that are seldom bought together. For these reasons, support is often used to eliminate uninteresting rules.

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

**Confidence** is defined as the conditional probability that shows that the transaction containing the LHS will also contain RHS. Association analysis results should be interpreted with caution. The

inference made by an association rules does not necessarily imply causality. Instead, it suggests a strong co-occurrence relationship between the items in the antecedent and consequent of the rule.

$$\textit{Confidence} = \frac{P(A \cap B)}{P(A)}$$

Confidence and support measure the strength of an association rule. Since the transactional database is quite large, there is a higher risk of getting too many unimportant and rules which may not be of our interest. To avoid these kinds of errors we commonly define a threshold of support and confidence prior to the analysis, so that only useful and interesting rules are generated in our result.

If lift is greater than 1, it suggests that the presence of the items on the LHS has increased the probability that the items on the RHS will occur on this transaction. If the lift is below 1, it suggests that the presence of the items on the LHS make the probability that the items on the RHS will be part of the transaction lower. If the lift is 1, it suggests that the presence of items on the LHS and RHS are independent: knowing that the items on the LHS are present makes **no** difference to the probability that items will occur on the RHS.

While performing market basket analysis, we look for rules with a lift of more than one. It is also preferable to have rules which have high support as this will be applicable to a large number of transactions and rules with higher confidence are ones where the probability of an item appearing on the RHS is high, given the presence of items on the LHS.

#### 5.1.2. Antecedent and Consequent:

In every association rule we have an antecedent and a consequent, also called rule body and rule head accordingly. The generated association rule relates the rule body with the rule head. LHS is the Antecedent and RHS is the Consequent.

#### 5.1.3. Causality:

Ideally, we would like to know that in an association rule the presence of an item/ itemset causes another item/ itemset to be bought. However, "causality" is an elusive concept. Nevertheless, for market-basket data, the following test suggests what causality means. If we lower the price of diapers and raise the price of beer, we can lure diaper buyers, who are more likely to pick up beer

while in the store, thus covering our losses on the diapers. That strategy works because diapers cause beer. However, working it the other way around, running a sale on beer and raising the price of diapers, will not result in beer buyers buying diapers in any great numbers, and we lose money.

#### 5.1.4. Frequent Itemset:

In many (but not all) situations, we only care about association rules or causalities involving sets of items that appear frequently in baskets. For example, we cannot run a good marketing strategy involving items that no one buys anyway. Thus, much data mining starts with the assumption that we only care about sets of items with high support; i.e., they appear together in many baskets. We then find association rules or causalities only involving a high-support set of items. The consequent must appear in at least a certain percent of the baskets, called the support threshold.

#### 5.1.5. Time Period:

Each transaction takes place in a single time period where  $W = \{W_1, W_2, W_3, \dots, W_t\}$  is the set of all the time periods under consideration. Only one transaction occurs per store, per user in a time period.

#### 5.1.6. Transaction:

In general terms transaction is an agreement, contract, understanding, or transfer of cash or property that takes place between two parties and establishes a legal obligation.

In accounting terms events that initiates a change in the asset, liability, or net worth account. Transactions first entry are made in journal and then posted to ledger. From there they move to other accounting books like profit and loss account, balance sheet, etc.

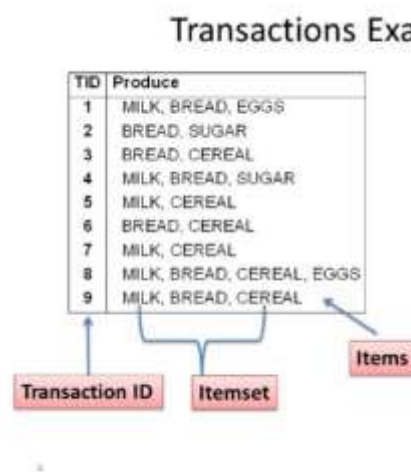
In banking a transaction would be an activity performed by the account holder at his/her request which is affecting a bank account.

In the language of commerce, a transaction will be considered exchange of goods or services between a buyer and a seller. It has 3 components:

1. Transfer of goods or services and money,
2. Transfer of title which may or may not be accompanied by a transfer of possession,
3. Transfer of exchange rights.

In the field of marketing, a typical transaction consists of a set of products purchased by a customer at a retail store or on a website. These transactions contain all the information about each specific transaction which make up the data entered into the database. These can include information on the customer, information of what products were purchased in what quantity, information on time of purchases, information on if the companies marketing strategies are attracting customers or not, etc.

Also, a transaction can take place at one point in time or over time and could involve a day, a quarter, a fiscal year, or even longer periods. Because they are not limited to an event.



*figure 2: Transaction example*

#### 5.1.7. Long Tail Effect:

This term often refers to data products purchase in supermarkets describing their distribution as a long tail in which a small number of products is purchased more frequently whereas a large one is purchased less frequently. This phenomenon creates data sparsity problem and worsens even more their elaboration.

Studying such data with MBA would be practically helpful because its transactional data and best approach to work with transactional data is to do market basket analysis.

## 5.2. DETERMINE SUITABILITY OF Market Basket Analysis

MBA is typically used to examine associations based on data with binary variables. However, it would not be possible to apply MBA when the data set consists of data collected from continuous



or Likert-type scales only. There must be at least some categorical variables among which there is some interest or meaningfulness for deriving categorical or quantitative association rules. Apparently, the data collected by organizations as a part of their business analytics and data collected by scholars for research purposes usually includes both binary as well as categorical variables. Therefore, MBA can be potentially used with many types of data available.

In addition, if the researcher is interested in theory building; examining multilevel, contingency, and dynamic relationships; producing results that can be easily communicated to the practitioner audience; if the data seems too messy or unstable to use, then MBA is likely to be a good methodological alternative. On the other hand, MBA is not appropriate for theory testing purposes due to its exploratory nature.

#### 5.2.1. ADVANTAGES OF USING MARKET BASKET ANALYSIS

- It allows researchers to make use of data from stores which is in abundance to build their theory. This capability of MBA was first highlighted by Locke in 2007. He said using MBA has the potential to lead to important contributions by allowing researchers to implement an inductive approach to theory building, which, despite its advantages is currently underutilized. Indeed, it has led to insights in marketing and other fields. For example, Russell et al. (1999) pointed out multiple-category decision making i.e. theoretical models of purchasing decisions involving products in more than one category. Apart from marketing, other researches like patients with food allergies allowed Kanagawa et al. (2009) to build models regarding which allergens are related to which.
- MBA allows the researchers to use the data which appears to be messy and unusable. Given the affordability and availability of data storage systems, the organizations collect data every day on employees (their performance, training, skills, etc.), customers (frequency of visits, purchases, expenditure, etc.) and many other issues. Most of that data is collected in a very unsystematic and unorganized format without any specific study in mind. MBA is ideally suited to be used inductively with such datasets to uncover association rules that may not be readily apparent (Hafley & Lewis, 1963; Shmueli, Patel, & Bruce, 2010). Messy data often involves dirty data with lots of missing values and outliers. MBA is not immune to the problem of messy data, it just allows the interpretation of missing data as indicating that no option was selected, and association rules are less influenced by outliers

compared to other traditional data analytic approaches. In the context of MBA, outliers result in infrequently occurring associations (He, Xu, Huang, & Deng, 2004)

- MBA can help in building dynamic theories, which states how important is the role of time in theory building. There are mainly two approaches of building dynamic theories via MBA i.e. multiple MBA and sequential MBA. When the available data include transactions as they have occurred overtime, multiple MBA approach is used (Tang et al. 2008). Sequential MBA is used when the available data describes individual events as they have occurred over time. It may uncover the presence of pattern in which event A occurs before event B, which occurs before event C (Han, Kim, & Sohn, 2009).
- MBA can be used to access multilevel relationships. It can be applied across all level of analysis ranging from an individual level to firm level, industry level and country level contexts.

#### 5.2.2. CHECK MBA REQUIREMENTS

Next step would be to check two key requirements for MBA (Marakas, 2003). *Firstly*, it is necessary to have sufficiently huge number of transactions. Fortunately, it is straightforward to meet the sample size requirements because of vast storage of data being collected by firms in response to the analytics movement (Davenport et al., 2010) as well as sharp decrease in cost of data storage technology (Shmueli et al., 2010). In fact, Berry and Linoff (2004) had an argument that firms face the problem of too much data rather than too little.

It is important for the researchers to explain in detail the sources of their data so that readers can be fully aware/ informed regarding the extent to which association rules may generalize to other settings, because it may be the case even if the sample size is large, the data might be collected only in one context.

*Secondly*, it must be verified that the data doesn't have wrong entry or missing values which should be mandatory. It is necessary to check the data for any sort of weirdness. To do so, the researchers can use exploratory data analysis techniques, such as frequency tables, unique values, date of transaction, etc. Since our data will be transactions in retail market, we can choose the important categories that we want to focus in our analysis, rather than looking in to all the categories.

Our illustrative data set meets both requirements. It has sufficiently large number of transactions and data cleaning is the part of the whole process of data exploration before starting to analyze the data.

### 5.2.3. APRIORI ALGORITHM

This part will explain how the algorithm that will be running behind the python libraries for Market Basket Analysis. This will help the companies to understand their clients more and analyze their data more closely and attentively. Rakesh Agrawal proposed the Apriori algorithm which was the first associative algorithm proposed and future developments in association, classification, associative classification algorithms have used it as a part of the technique.

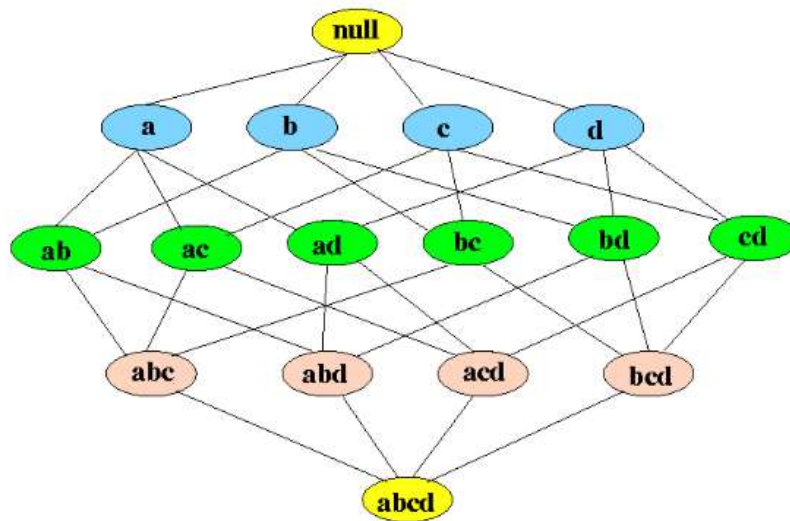
Association rule mining is seen as a two-step approach:

1. **Frequent Itemset Generation:** Find all frequent item-sets with support  $\geq$  pre-determined minimum support count. In frequent mining usually the interesting associations and correlations between item sets in transactional and relational databases are found. In short, Frequent Mining shows which items appear together in a transaction or relation. The discovery of frequent item sets is accomplished in several iterations. Counting new candidate item-sets from existing item sets requires scanning the entire training data. In short it involves only two important steps:
  - a. Pruning
  - b. Joining
2. **Rule Generation:** List all association rules from frequent item-sets. Calculate Support and Confidence for all the rules. Prune rules which fail minimum support and minimum confidence thresholds.

**Frequent Itemset Generation** scan the whole database and find the frequent itemset with a threshold on support. Since it scans the whole database, it is the most computationally expensive step. In the real-world, transaction data for retail, can exceed to Gigabytes and Terabytes of data for which an optimized algorithm is needed to exclude item-sets that will not help in later steps. For this Apriori algorithm is used.

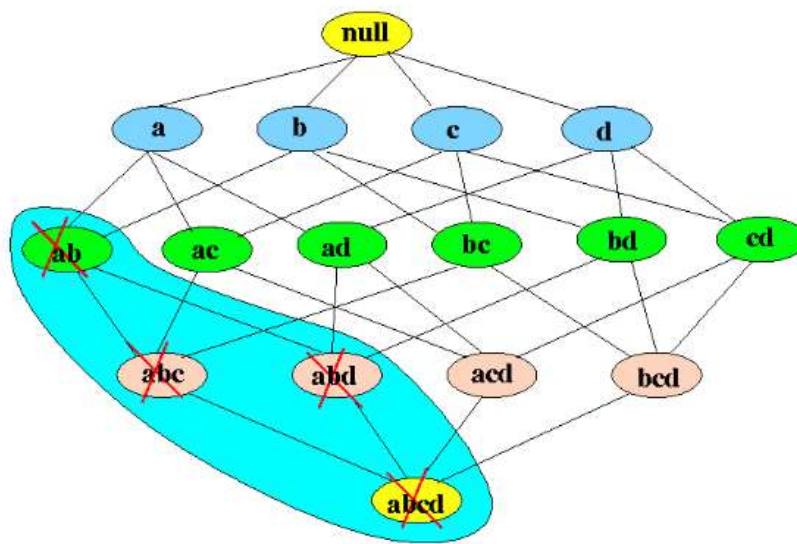
Apriori algorithm states “Any subset of a frequent itemset must also be frequent. In other words, no superset of an infrequent itemset must be generated or tested.”

In the image below, which is a graphical representation of the Apriori algorithm principle. It consists of k-item-set node and relation of subsets of the k-item-set. You can notice in the figure that in the bottom is all the items in the transaction data and then you start moving up creating subsets till it reaches to the null set.



*figure 3: All possible subsets*

This shows that it will be difficult to generate frequent item-set by finding support for each combination. Therefore, in the figure below we can notice that Apriori algorithm helps to reduce the number of sets to be generated.



*figure 4: if an item set is infrequent, we do not consider its super sets*


If an item-set {a, b} is infrequent then we do not need to consider all its super sets.

We can also look at it in the form of a transactional data-set. In the following example you can notice why Apriori algorithm is much more effective and it generates stronger association rules step by step.

### Step: 1

- Create a table containing support count of each item present in the dataset.
- Compare support count with minimum support count (in this case we have minimum support count = 2 and if support count is less than the minimum support count then remove those items), this gives us a new set of items.

TID	items
T1	I1, I2 , I5
T2	I2,I4
T3	I2,I3
T4	I1,I2,I4
T5	I1,I3
T6	I2,I3
T7	I1,I3
T8	I1,I2,I3,I5
T9	I1,I2,I3



Itemset	sup_count
I1	6
I2	7
I3	6
I4	2
I5	2

*figure 5: transactional data to frequent items*

### Step: 2

- This step is known as join step. We generate another set by cross joining each item with one another.
- Check if the subsets of an itemset are frequent or not and if not remove that itemset. For example, in the case below we can see that the subset of {I1, I2} are {I1}, {I2} and are frequent. We must check for the each itemset in the same way.
- Now find the support count of these item-sets by searching in the dataset.

- Since we have already specified a threshold of minimum support count of 2. We compare the minimum support count and if the support count is less than the minimum support count, then remove those items. Gives us another itemset as we can see below.

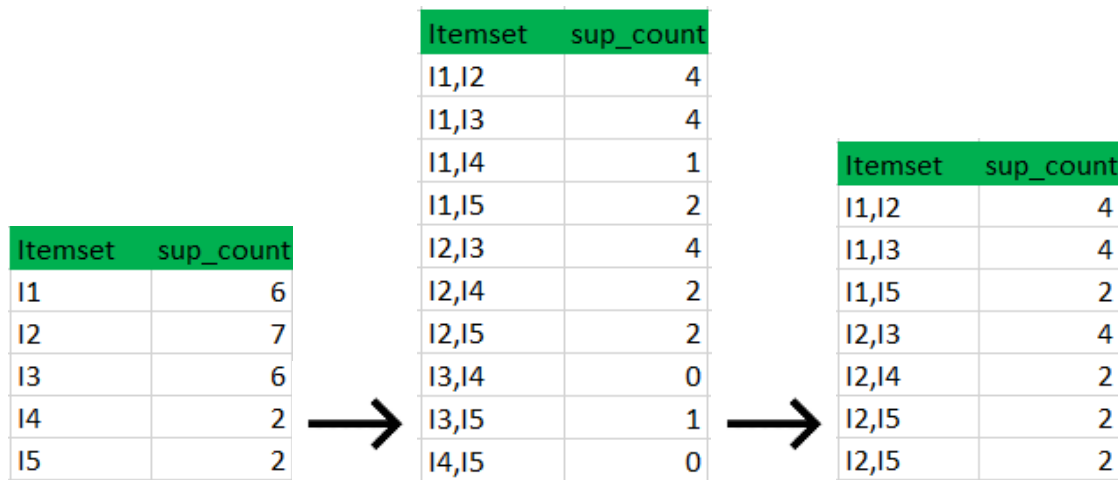



figure 6: pruning and joining

### Step: 3

- After getting another dataset we follow the same step (I.e. join step). We cross join each itemset with each other. So, the itemset generated after this step will be:
  - {I1, I2, I3}
  - {I1, I2, I4}
  - {I1, I2, I5}
  - {I1, I3, I5}
  - {I2, I3, I4}
  - {I2, I4, I5}
  - {I2, I3, I5}
- Check all subsets of these item sets are frequent or not, if not then remove that item sets. For example, in this case the subset of {I1, I2, I3} are {I1, I2}, {I1, I3}, {I2, I3} which are frequent. But for {I2, I3, I4} one of the subsets is {I3, I4}, which is not frequent. So, we remove this. We do the same for every itemset.
- Once we have removed all the non-frequent item sets, find support count of the remaining itemset by searching in the dataset.

- Compare the minimum support count and if the support count is less than the minimum support count, then remove those items. It gives us another itemset as we can see below.

Itemset	sup_count
I1,I2	4
I1,I3	4
I1,I5	2
I2,I3	4
I2,I4	2
I2,I5	2
I2,I5	2



Itemset	sup_count
I1,I2,I3	2
I1,I2,I5	2

*figure 7: pruning and joining again until there are no more frequent items left.*

#### Step: 4

- We follow the same procedure again. First, we do the join step and we cross join each itemset with one another. In our example the first two elements of the item set should match.
- After, check all subsets of these item sets are frequent or not. In our example the itemset formed after join step is {I1, I2, I3, I5}. So, one of the subsets of this itemset is {I1, I3, I5} which is not frequent. Therefore, there is no itemset left anymore.
- We stop here because no more frequent itemset are found anymore.

This was the first step of association rule mining.

The next step will be to list all frequent item-sets and generate how strong are the association rules. For that we calculate the confidence of each rule. To calculate confidence, we use the following formula:

$$\begin{array}{l}
 \text{Rule: } X \Rightarrow Y \\
 \begin{array}{l}
 \nearrow \text{Support} = \frac{\text{freq}(X, Y)}{N} \\
 \rightarrow \text{Confidence} = \frac{\text{freq}(X, Y)}{\text{freq}(X)} \\
 \searrow \text{Lift} = \frac{\text{Support}}{\text{Supp}(X) \times \text{Supp}(Y)}
 \end{array}
 \end{array}$$

*figure 8: support, confidence and lift calculation*

By taking an example of any frequent Item (we took {I1, I2, I3}) we will show how rule generation is done:

Rules	Formula	Confidence
$\{I1, I2\} \Rightarrow \{I3\}$	$2 / 4 * 100$	50,00%
$\{I1, I3\} \Rightarrow \{I2\}$	$2 / 4 * 100$	50,00%
$\{I2, I3\} \Rightarrow \{I1\}$	$2 / 4 * 100$	50,00%
$\{I1\} \Rightarrow \{I2, I3\}$	$2 / 6 * 100$	33,33%
$\{I2\} \Rightarrow \{I1, I3\}$	$2 / 7 * 100$	28,57%
$\{I3\} \Rightarrow \{I1, I2\}$	$2 / 6 * 100$	33,33%

*figure 9: calculation of confidence*

So, in this case if the minimum confidence is 50% then the first 3 rules can be considered strong association rules. For example, take  $\{I1, I2\} \Rightarrow \{I3\}$  having confidence equal to 50% tells that 50% of people who bought **I1** and **I2** also bought **I3**.

#### 5.2.4. OTHER ALGORITHMS USED IN MARKET BASKET ANALYSIS

Even though Apriori algorithm is the mostly used algorithm for market basket analysis. There have been different approaches taken by the researchers to build different algorithms since accuracy and efficiency are crucial factors in data mining.

The AIS algorithm was the first algorithm proposed for mining association rule. In this algorithm only one item consequent association rules are generated, which means that the consequent of those rules only contains one item, for example we only generate rules like  $X \cap Y \rightarrow Z$  but not those rules as  $X \rightarrow Y \cap Z$ . The main drawback of the AIS algorithm is too many candidate itemset that finally turned out to be small are generated, which requires more space and wastes much effort that turned out to be useless. At the same time this algorithm requires too many passes over the whole database. Apriori is more efficient during the candidate generation process. Apriori uses pruning techniques to avoid measuring certain item sets, while guaranteeing completeness. These are the item sets that the algorithm can prove will not turn out to be large. However, there are two bottlenecks of the Apriori algorithm. One is the complex candidate generation process that uses most of the time, space and memory. Another bottleneck is the multiple scan of the database. Based



on Apriori algorithm, many new algorithms were designed with some modifications or improvements.

As we know that apriori algorithm was an algorithm proposed by Rakesh Agrawal in 1993. Since then there have been future developments in association, classification, associative classification algorithms, which have used apriori as part of the technique.

In 1998, Liu et al. Proposed CBA the first associative classification algorithm. CBA implements the famous Apriori algorithm in order to discover the frequent rule items. CBA selects high confidence rules to represent the classifier. Finally, to predict a test case, CBA applies the highest confidence rule whose body matches the test case. Experimental results designated that CBA derives higher quality classifiers with regards to accuracy than rule induction and decision tree classification approaches.

Han et al., in 2000 presented a new association rule mining approach that does not use candidate rule generation called FP-growth that generates a highly condensed frequent pattern tree (FP-tree) representation of the transactional database. Each database transaction is represented in the tree by at most one path. FP-tree is smaller in size than the original database the construction of it requires two database scans:

1. Frequent item sets along with their support in each transaction are produced
2. FP-tree is constructed.

The mining process is performed by concatenating the pattern with the ones produced from the conditional FP-tree. One constraint of FP-growth method is that memory may not fit FP-tree especially in dimensionally large database.

In 2001, Li et al., recommended Classification based on Multiple Association Rules (CMAR). This method is an extension of FP-growth, constructs a class distribution-associated FP-tree, and mines large database efficiently. Moreover, it applies a CR-tree structure to store and retrieve mined association rules efficiently, and prunes rules effectively based on confidence, correlation and database coverage. The classification is performed based on a weighted  $\chi^2$  analysis using multiple strong association rules. Extensive experiments on 26 datasets from UCI machine learning database repository show that CMAR is consistent, highly effective at classification of various kinds of databases and has better average classification accuracy in comparison with CBA.

In 2003, Yin et al., suggested Classification based on Predictive Association Rules (CPAR). CPAR integrates the features of associative classification in predictive rule analysis. It has the following advantages:

1. CPAR generates a much smaller set of high-quality predictive rules directly from the dataset.
2. Avoids redundant rules, CPAR generates each rule by considering the set of “already generated” rules.
3. Predicting the class label of an example, CPAR uses the best  $k$  rules the example satisfies.

Moreover, CPAR employs the following features to further improve its accuracy and efficiency:

1. CPAR uses dynamic programming to avoid repeated calculation in rule generation.
2. Rule generation considers all the close-to-the-best literals are selected so that important rules will not be missed.

CPAR generates a smaller set of rules, with higher quality and lower redundancy in comparison with associative classification. As a result, CPAR is much more time efficient in both rule generation and prediction but achieves as high accuracy as associative classification.

The MCAR algorithm introduced by Fadi et al., in 2005 uses an intersection technique for discovering frequent rule items. MCAR takes advantage of vertical format representation and uses an efficient technique for discovering frequent items based on recursively intersecting the frequent items of size  $n$  to find potential frequent items of size  $n+1$ . MCAR consists of two main phases:

1. Rule generation and
2. A classifier builder.

In the first phase, the training data set is scanned once to discover frequent one rule items, and then MCAR recursively combines rule items generated to produce potential frequent rule items (candidate rule items) involving more attributes. The supports and confidences for candidate rule items are calculated simultaneously, where any rule item with support and confidence larger than minimum support and minimum confidence, respectively, is created as a potential rule.

In the second phase, the classifier builder ensures that each training instance is covered by at most one rule, which has the highest precedence among all rules applicable to it. Furthermore, there are

no useless rules in the MCAR classifier since every rule correctly covers at least one training instance. This approach is similar to the CBA classifier builder as each rule in CBA also covers at least one training instance. However, the way MCAR builds the classifier by locating training instances is more efficient than that of CBA due to abounding going through the training data set multiple times.

## CHAPTER 6: PRACTICAL IMPLEMENTATION OF MARKET BASKET ANALYSIS ON THE DATASET

### 6.1. DATA COLLECTION

MBA was originally designed for the use with large datasets that are usually collected by others (i.e., not the scholars or the research team). Thus, the researchers can seek out partnerships with organizations that will provide data in exchange for conducting analysis and presenting results to those who provided the data. It is also possible for the research team to collect the data but given the time involved, effort and issues to access the data sources, most MBA studies are done on the data collected by other parties. In the chapter of literature review we were able to see that researchers have been able to create data-sharing partnerships in many different fields.

In addition to using data collected by organizations, researchers can implement a third-party data collection strategy consisting of reliance on publicly available information. For example, O’Boyle and Aguinis (2012) investigated issues regarding individual performance by using data on academics, entertainers, politicians, and amateur and professional athletes.

The data for this research was collected from a Belgian known super market chain and are related to customer transactions for the period from April 2018 to June 2018. Data is extracted from the information system of the company and come exclusively from customer transactions through loyalty cards. Every transaction is a record of the purchase carried out by the customer whoever used the loyalty card at the time of the purchase.

### 6.2. DATASET DESCRIPTION

Number of Attributes: 11

Time period of the data: 3 months (April, May, June)

Number of transactions: 78839

Number of customers: 42468

#### 6.2.1. ATTRIBUTES

S. No	Attribute	Description
-------	-----------	-------------

1	random_cust_no	This is a customer ID which was replaced by random number by the company itself.
2	orig_invoice_id	This is the most important attribute of the data. It can be called invoice ID or Transaction ID. The whole analysis is based on transactions. It is explained in detail under define the transactions.
3	date_of_day	This is the day when the transaction took place.
4	minute_desc	For more accuracy we do not just keep track of day of the transactions but also time of transaction. Since in case of a super market there are hundreds of transactions happening every day.
5	mikg_art_no	It is the code of the article / item.
6	art_name	Name of the article / item
7	catman_buy_domain_desc	The categories are in hierarchy level. This is the top level of the category under which the article falls. Category_1
8	pcg_main_cat_desc	This is the second level of category under which the article comes. Category_2
9	pcg_cat_desc	This is the third level of category under which the article comes. This is the important level of category since we used this category in our analysis. The association rules are based on this level of category. Category_3
10	pcg_sub_cat_desc	This is the last and final level of category under which the article comes. Category_4
11	quantity	This is the quantity of product bought by the customer. We used this attribute to change the transactional data to binary representation.

TABLE 2: ATTRIBUTES DESCRIPTION

### 6.3. TOOLS USED FOR ANALYSIS

- Jupyter notebook

- Python Libraries
  - Pandas
  - Datetime
  - Matplotlib
  - Mlxtend
  - Random
  - xlswriter
- MS Excel

#### 6.4. DATA EXPLORING & TRANSFORMATION

Before doing anything with the data as soon as we read in the data in jupyter notebook, we start exploring it. Data exploring is the most important step for any analysis. While, exploring the data the issues that needed fixing were as follows:

1. The attribute *date\_of\_day* was read as string variable but in this research, time is a very crucial factor, so its data type was changed to date type.
2. For comparison of data of each month we needed to separate the data based on each month. Since *date\_of\_day* attribute was changed to date type. It was very easy to get the month of each transaction.
3. As explained earlier in the long tail effect the data sparsity problem tackled up to some extent by considering only the transactions in which total items purchased were more than 8 items. While exploring the data it's important to look at the data distribution. For market basket analysis it has always been observed to have the long tail effect. The long tail effect has assumed its present connotation and dynamics by Chris Anderson [21]. This term often refers to data products purchase in supermarkets describing their distribution as a long tail in which a small number of products is purchased more frequently whereas a large one is purchased less frequently. This phenomenon creates data sparsity problem and worsens even more their elaboration. For this research from the total number of transactions there were selected, those that included at least the purchase of 8 products. That is how, to some extent, it was tried to resolve the data sparsity problem and prevent the removal of any market basket product, which may present a risk of information loss. Figure below shows the curve of long tail effect.

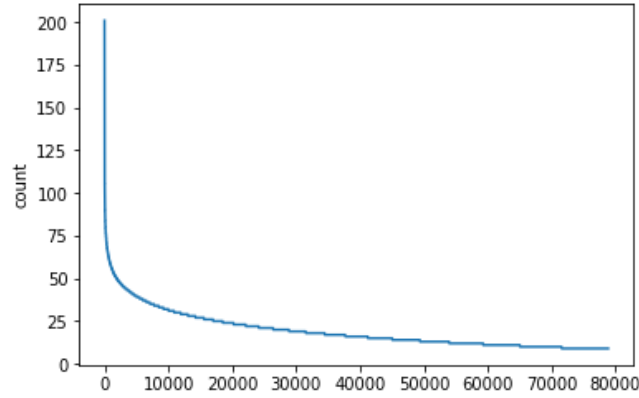


figure 10: long tail effect in our dataset

Some general exploring of data was done after all the fixes. In the table we can see the number of unique transactions, number of unique customers, number of unique articles sold, and number of articles sold in the period of 3 months as well as in each month (i.e. April, May, June).

Time	3 months	April	May	June
<b>Number of Transactions</b>	78839	24803	27710	26326
<b>Number of Customers</b>	42468	18743	20436	19347
<b>Number of Unique Articles Sold</b>	42694	31286	31035	31462
<b>Number of Total Articles Sold</b>	1541130	482776	544023	514331

TABLE 3: DATA EXPLORING

Since there are several categories and sub categories in the super market. We can easily identify the hierarchy of the categories based on the number of unique categories in the table below.

Category/Sub category	Total
catman_buy_domain_desc (Category_1)	87
pcg_main_cat_desc (Category_2)	348
pcg_cat_desc (Category_3)	1334
pcg_sub_cat_desc (Category_4)	3783

TABLE 4: CATEGORY AND SUBCATEGORIES DISTRIBUTION

## 6.5. BINARY REPRESENTATION

Binary data are used in information systems because data are categorized for attributes value by 0 and 1. It represents either the attribute category is present or absent in the data.

For super markets, the databases are very large and usually represented in binary format. The binary format is a matrix in which Transactions forms the rows and the items/ attributes forms the

columns. For a specific transaction, if an item is purchased then the matrix position is made as 1. If the item is not purchased in the transaction, then the matrix position will be made 0. Binary data sets are interesting and useful due to its computational efficiency and minimal storage capacity.

## 6.6. ANALYSIS & DISCUSSION

### 6.6.1. Association rules from dataset

Since we have the data for three months, we divide the data into subset of each month to compare the analysis output of each month.

Firstly, we look at the comparison of number of rules when the minimum support count is less than 0.01 and minimum lift is 1.

Secondly, for all the data and for each of the subsets we have applied market basket analysis and recorded the results of the three indicators, i.e. confidence, support and lift. For every application of a market basket analysis process, the minimum level of confidence, support and lift established in the program, is 0.5, 0.01 and 1 respectively. The results of the number of rules that came out, are registered in table below.

<b>Time</b>	<b>Minimum support = 0.01 Lift &gt;= 1</b>	<b>Minimum support = 0.01 Confidence &gt;= 0.5 Lift &gt;= 1</b>
3 months	17282	1534
April	16564	1371
May	19380	1760
June	19272	1900

TABLE 5: ASSOCIATION RULES WITH DIFFERENT THRESHOLDS

In the first section we looked at the number of association rules with minimum support 0.01 and lift greater than 1. This criterion was applied to overall data as well as to monthly data. Our results were:

**May > June > 3 months > April**

We also calculated the number of association rules by adding degree of confidence greater than 0.5, the result changed to:

**June > May > 3 months > April**



### 6.6.2. Time development of association rules

The 128 rules, at the level of 1334 subcategories of products, resulting from the extraction process of overall 3 months data were used as a base for additional rules analysis at monthly level. For each of the 128 rules, there have been recorded the values for the three indicators, lift, support and confidence during each month (i.e. April, May, June). Thus, the table was created with the monthly development of association rules measurement indicators. The table can be found at the end of the document in ANNEX section in the end of the document.

In the end of the ANNEX we can notice that there are 2 association rules **{'COLA KZH'} => {'GROENTEN'}** and **{'GROENTEN'} => {'COLA KZH'}** which do not exist in the month of May, because they didn't pass the threshold of minimum support of 0.01 and lift greater than equal to 1. But these two rules only do not exist in the month of May They exist in the month of April and June.

Similarly, there are 2 association rules **{'GROENTEN'} => {'KOEKJES'}** and **{'KOEKJES'} => {'GROENTEN'}** which do not exist in the month of April and June. But they do exist in the month of May.

Whereas, the 128 association rules were chosen based on the threshold of minimum support as 0.05 and lift greater than and is equal to 1 from overall dataset. Whereas, to match with these 128 association rules the threshold set for monthly data was minimum support 0.01 and lift  $\geq 1$

Through these examples we can also say that there is a difference in the strength of the association rules in different time periods Its not just minor fluctuations in their strengths but sometimes it can be greater.

### 6.6.3. Top 30 association rules comparison

To show the fluctuation of degree of confidence of top 30 rule in each month. We have selected the rules with highest degree of confidence. In the table below, we can see the top 30 rules and their degree of confidence in 3 months all together as well as individually.

We can notice some major differences in rule number 4, 9, 18 and 19 in which we can see that those rules had lower degree of confidence in the month of April compared to the month of May and June. Also, in rule number 30 we can see that the degree of confidence for this rule is increasing with time.

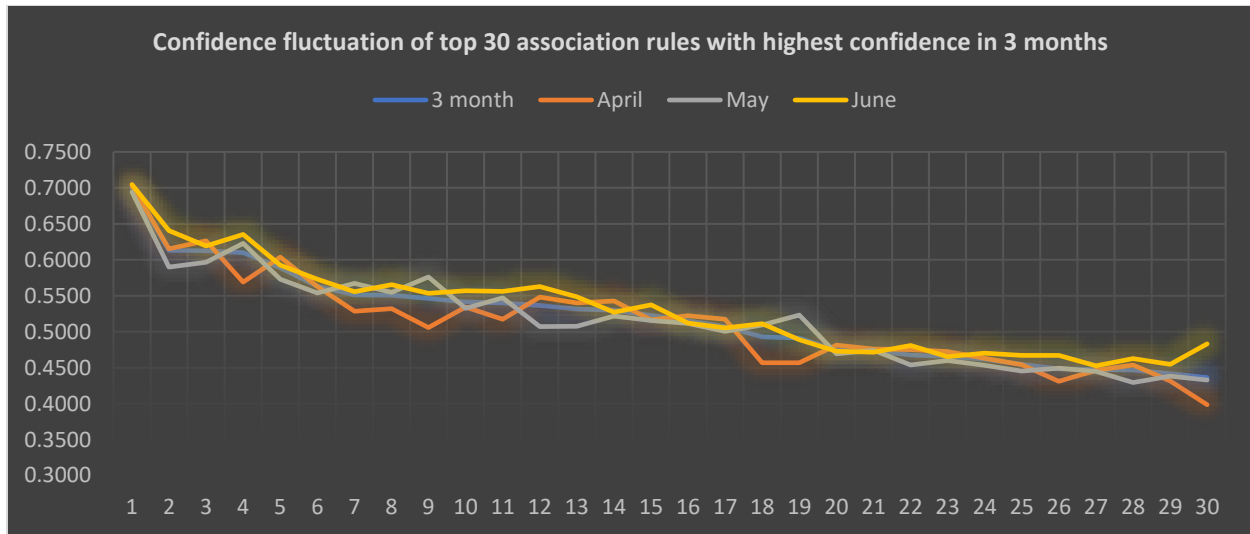


figure 11: Degree of confidence of top 30 rules

1	{'GROENTEN', 'CUCURBITACEAE'} => {'TOMATEN'}	16	{'GEBAK'} => {'BROOD'}
2	{'CUCURBITACEAE'} => {'TOMATEN'}	17	{'TOMATEN', 'BROOD'} => {'GROENTEN'}
3	{'CUCURBITACEAE', 'TOMATEN'} => {'GROENTEN'}	18	{'KRUIDEN + BLOEMEN'} => {'TOMATEN'}
4	{'PAPRIKA'} => {'TOMATEN'}	19	{'GROENTEN', 'TOMATEN'} => {'CUCURBITACEAE'}
5	{'SALADES', 'TOMATEN'} => {'GROENTEN'}	20	{'SCHARRELKIP'} => {'GROENTEN'}
6	{'SALADES', 'GROENTEN'} => {'TOMATEN'}	21	{'BANANEN'} => {'BROOD'}
7	{'BOLLEN'} => {'TOMATEN'}	22	{'STENGELGROENTEN'} => {'GROENTEN'}
8	{'KRUIDEN + BLOEMEN'} => {'GROENTEN'}	23	{'KIP'} => {'GROENTEN'}
9	{'PAPRIKA'} => {'CUCURBITACEAE'}	24	{'SALADES'} => {'GROENTEN'}
10	{'PAPRIKA'} => {'GROENTEN'}	25	{'EXOTISCH/TROPISCH'} => {'GROENTEN'}
11	{'WORTELGROENTE'} => {'TOMATEN'}	26	{'EXOTISCH/TROPISCH'} => {'TOMATEN'}
12	{'CUCURBITACEAE'} => {'GROENTEN'}	27	{'SALADES'} => {'BROOD'}
13	{'BOLLEN'} => {'GROENTEN'}	28	{'EUROPA'} => {'GROENTEN'}
14	{'TOMATEN'} => {'GROENTEN'}	29	{'SALADES'} => {'TOMATEN'}
15	{'WORTELGROENTE'} => {'GROENTEN'}	30	{'EXOTISCH/TROPISCH'} => {'BESSEN'}

TABLE 6: 30 ASSOCIATION RULES WITH HIGHEST DEGREE OF CONFIDENCE

Degree of support of top 30 rule in each month. In figure 12 and table 7 we can see the top 30 associations rules with highest support in 3 months all together as well as individually.

When we look at the figure, we can notice that there is a difference in the support level of most of the association rules. But for most of them the support is increasing with the increase in time. Most of the association rules had lower support in the month of April but it eventually increased in the month of June.

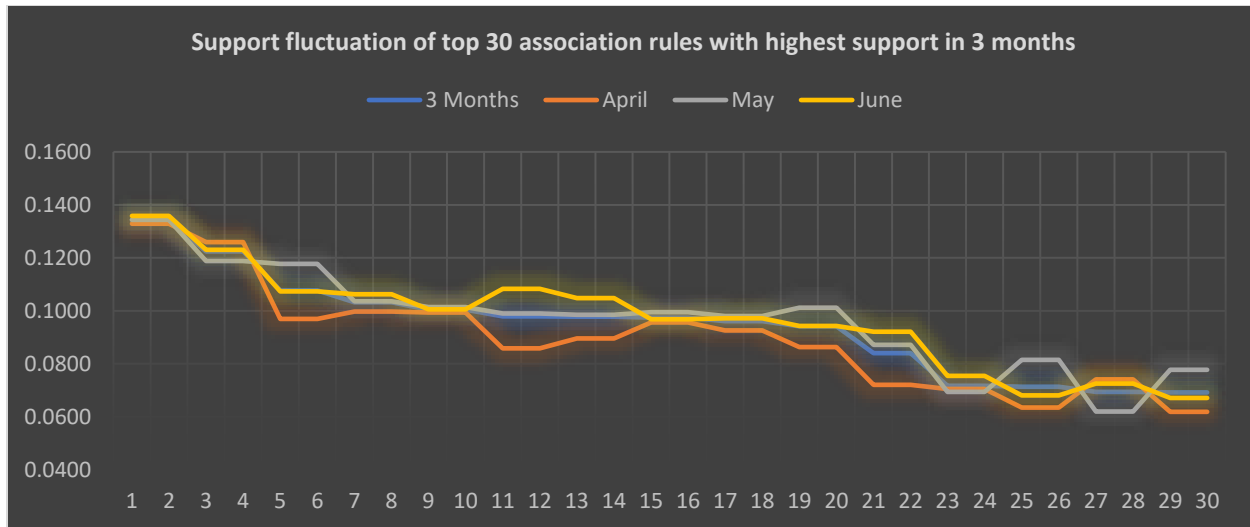


figure 12: Degree of support of top 30 rules

1	{'GROENTEN'} => {'TOMATEN'}	16	{'SALADES'} => {'BROOD'}
2	{'TOMATEN'} => {'GROENTEN'}	17	{'SALADES'} => {'TOMATEN'}
3	{'GROENTEN'} => {'BROOD'}	18	{'TOMATEN'} => {'SALADES'}
4	{'BROOD'} => {'GROENTEN'}	19	{'GROENTEN'} => {'CUCURBITACEAE'}
5	{'CUCURBITACEAE'} => {'TOMATEN'}	20	{'CUCURBITACEAE'} => {'GROENTEN'}
6	{'TOMATEN'} => {'CUCURBITACEAE'}	21	{'TOMATEN'} => {'BESSEN'}
7	{'TOMATEN'} => {'BROOD'}	22	{'BESSEN'} => {'TOMATEN'}
8	{'BROOD'} => {'TOMATEN'}	23	{'KIP'} => {'GROENTEN'}
9	{'SALADES'} => {'GROENTEN'}	24	{'GROENTEN'} => {'KIP'}
10	{'GROENTEN'} => {'SALADES'}	25	{'CUCURBITACEAE'} => {'BROOD'}
11	{'BESSEN'} => {'BROOD'}	26	{'BROOD'} => {'CUCURBITACEAE'}
12	{'BROOD'} => {'BESSEN'}	27	{'GEBAK'} => {'BROOD'}
13	{'BESSEN'} => {'GROENTEN'}	28	{'BROOD'} => {'GEBAK'}
14	{'GROENTEN'} => {'BESSEN'}	29	{'CUCURBITACEAE'} => {'SALADES'}
15	{'BROOD'} => {'SALADES'}	30	{'SALADES'} => {'CUCURBITACEAE'}

TABLE 7: 30 ASSOCIATION RULES WITH HIGHEST DEGREE OF SUPPORT

In the table 8 we can see what the top 30 association rules with the highest Lift are and in figure 13 we can see what the lift values of those 30 association rules in 3 months are all together and each month periodically.

In the comparison of lift we can notice that there is a clear difference in the lift of association rules number 1 and 2 in all the 3 months. But as the lift gets smaller and smaller the difference in different time periods decreases. Rule 5 and 6 also had higher lift in the month of April but in the month of May and June it dropped.

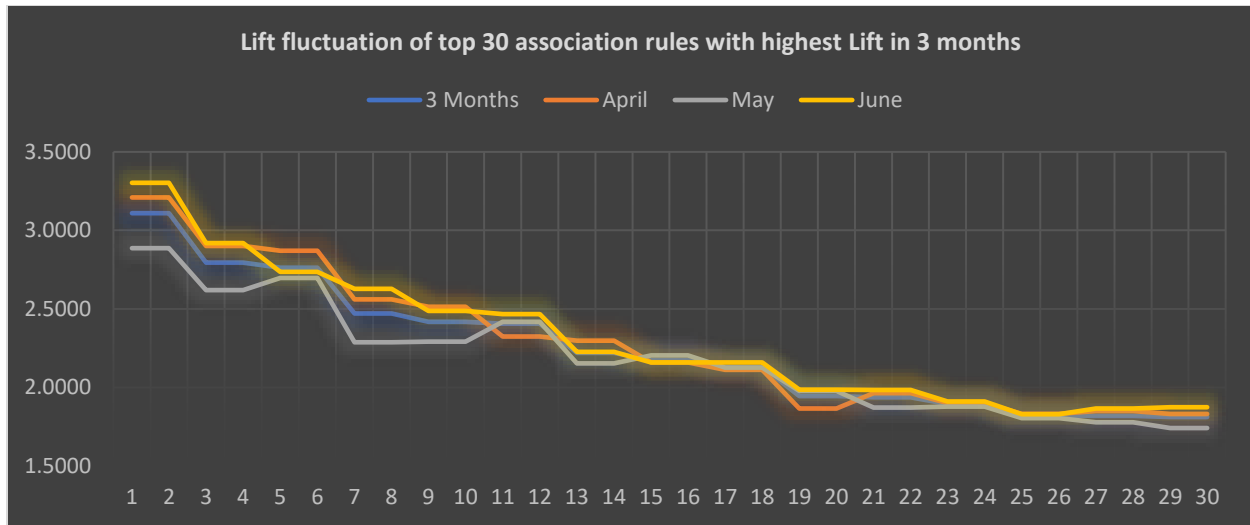


figure 13: Lift of top 30 association rules

1	{‘CUCURBITACEAE’} => {‘PAPRIKA’}	16	{‘TOMATEN’} => {‘BOLLEN’}
2	{‘PAPRIKA’} => {‘CUCURBITACEAE’}	17	{‘TOMATEN’} => {‘WORTEL GROENTE’}
3	{‘GROENTEN’, ‘TOMATEN’} => {‘CUCURBITACEAE’}	18	{‘WORTEL GROENTE’} => {‘TOMATEN’}
4	{‘CUCURBITACEAE’} => {‘GROENTEN’, ‘TOMATEN’}	19	{‘TOMATEN’} => {‘KRUIDEN + BLOEMEN’}
5	{‘TOMATEN’} => {‘GROENTEN’, ‘CUCURBITACEAE’}	20	{‘KRUIDEN + BLOEMEN’} => {‘TOMATEN’}
6	{‘GROENTEN’, ‘CUCURBITACEAE’} => {‘TOMATEN’}	21	{‘GROENTEN’, ‘TOMATEN’} => {‘SALADES’}
7	{‘CUCURBITACEAE’} => {‘WORTEL GROENTE’}	22	{‘SALADES’} => {‘GROENTEN’, ‘TOMATEN’}
8	{‘WORTEL GROENTE’} => {‘CUCURBITACEAE’}	23	{‘GROENTEN’} => {‘CUCURBITACEAE’, ‘TOMATEN’}
9	{‘CUCURBITACEAE’} => {‘TOMATEN’}	24	{‘CUCURBITACEAE’, ‘TOMATEN’} => {‘GROENTEN’}
10	{‘TOMATEN’} => {‘CUCURBITACEAE’}	25	{‘SALADES’, ‘TOMATEN’} => {‘GROENTEN’}
11	{‘PAPRIKA’} => {‘TOMATEN’}	26	{‘GROENTEN’} => {‘SALADES’, ‘TOMATEN’}
12	{‘TOMATEN’} => {‘PAPRIKA’}	27	{‘EXOTISCH/TROPISCH’} => {‘BESSEN’}
13	{‘TOMATEN’} => {‘SALADES’, ‘GROENTEN’}	28	{‘BESSEN’} => {‘EXOTISCH/TROPISCH’}
14	{‘SALADES’, ‘GROENTEN’} => {‘TOMATEN’}	29	{‘CUCURBITACEAE’} => {‘SALADES’}
15	{‘BOLLEN’} => {‘TOMATEN’}	30	{‘SALADES’} => {‘CUCURBITACEAE’}

TABLE 8: 30 ASSOCIATION RULES WITH HIGHEST LIFT

## CHAPTER 7: CONCLUSION

### 7.1. SUMMARY & FUTURE WORK

This research shows that time is a very important factor which should be kept in mind while performing any market basket analysis. It can help reveal very interesting information / insights about the customers which contribute in profit maximization. For example, the discovery of complementary or supplementary products can lead to cross-selling or promotional opportunities.

This research is a superior and cheaper approach than the traditional customer surveys, which are usually very costly and are time consuming. They also include errors at every step of the survey research.

In this research we have eliminated less frequent items from the dataset by setting minimum threshold. The results which we got after detecting the purchase behaviour of the customer can be used in cross selling recommendation and to enhance their marketing strategies while deciding for promotions.

This research also shows that how market basket analysis also has its applicability in many domains and applying market basket analysis considering time as an important factor will be able to solve several problems in much effective and efficient way.

We used graphs to see the pattern of the association rules in 3 months periods, which is a good way to visualize top association rules and compare them at different time periods. With more data, we will be able to easily identify which association rules are more active during certain time periods. This insight will be very useful for the super markets to preplan their marketing strategies for that period.

We not only did comparison of association rules but also found top association rules during the 3 months of period which can help the company is inventory management as well as which categories of products should be kept close to each other to upscale the purchase. Mining into their data provides managers with a unique over view of what is happening with their business so that they can implement strategies efficiently and can move faster than their competitors.

Market basket analysis can also focus just on one or more than one Category/ categories which will give us insight within some important categories based on the requirements of the marketing team.

In our research we used the monthly time period. But MBA can give more in-depth insights if the analysis was done on a daily or weekly period. The association rules that we will be getting will display more rapid fluctuation in confidence and lift. It will be interesting to get more insights at a shorter period basis.

In future research it will be very interesting to do in-depth understanding of the association rules by evaluating the changes in the lift and confidence values, which can be made possible by calculating the standard deviation. This way we will be able to witness the evolution of association rules.

In future we can also find association rules using time series clustering method. Since it was called a superior alternative for market basket analysis in SIM University, School of Business in Singapore.

Future work would also be to design and develop an intelligent prediction model to generate the association rules that can be adopted on a recommendation system to make the functionality more operational.

## 7.2. LIMITATIONS

- Data was only for 3 months. If the data set was very large (i.e. for one whole year). It would have been very difficult for a normal computer to analyse the data faster. Hence, we had to stick to a smaller data set.
- Thresholds for support and confidence doesn't have any specific criteria to decide.
- Due to the long tail effect we only considered transactions which had more than 8 items. We might have missed some relevant information in that discarded data.
- Archival data collected by firms, professional associations and other organizations are usually different form data collected for academic research. Third party dataset might have errors in labelling of fields and recording of information.
- Finally learning a new methodological tool or statistical technique requires hands-on experience. Though we provided a description of how to use MBA through an illustrative

dataset, there are several complications when you are practically analysing the data. It requires several iterations to get to a better result.

## CHAPTER 8: REFERENCES

1. Liao, S. H., Chu, P. H., & Hsiao, P. Y. (2012). Data mining techniques and applications - A decade review from 2000 to 2011. *Expert Systems with Applications*. <https://doi.org/10.1016/j.eswa.2012.02.063>
2. [Book] *Data mining techniques: for marketing, sales and CRM. (Third Edition)* by Gordon S. Linoff, Michael J.A. Berry
3. Mushtaq, A. (2015). Data Mining For Marketing. *Journal of Marketing and Consumer Research*, 9, 17–29.
4. Ngai, E. W. T., Xiu, L., & Chau, D. C. K. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36(2 PART 2), 2592–2602. <https://doi.org/10.1016/j.eswa.2008.02.021>
5. Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining Association in Large Databases. *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data - SIGMOD '93*, 207–216. <https://doi.org/10.1145/170036.170072>
6. Gangurde, R., Kumar, B., & Gore, D. (2017). Building Prediction Model using Market Basket Analysis. *International Journal of Innovative Research in Computer and Communication Engineering*, 5(2). <https://doi.org/10.15680/IJIRCCE.2017.0502136>
7. Hibino, A., & Niwa, Y. (2008). Graphical representation of nuclear incidents/accidents by associating network in nuclear technical communication. *Journal of Nuclear Science and Technology*, 45(5), 369–377. <https://doi.org/10.1080/18811248.2008.9711446>
8. Gangurde, R., Kumar, B., & Gore, D. (November 2017). Optimized Predictive Model using Artificial Neural Network for Market Basket Analysis. *International Journal of Computer Science & Communication*, 9(1). <https://doi.org/10.090592/IJCSC.2017.014>
9. Liao, S. H., Chu, P. H., & Hsiao, P. Y. (2012). Data mining techniques and applications - A decade review from 2000 to 2011. *Expert Systems with Applications*. <https://doi.org/10.1016/j.eswa.2012.02.063>



10. Solnet, D., Boztug, Y., & Dolnicar, S. (2016). An untapped gold mine? Exploring the potential of market basket analysis to grow hotel revenue. *International Journal of Hospitality Management*, 56, 119–125. <https://doi.org/10.1016/j.ijhm.2016.04.013>
11. Kamley, S., Jaloree, S., & Thakur, R. S. (2014). An Association Rule Mining Model for Finding the Interesting Patterns in Stock Market Dataset. *International Journal of Computer Applications*, 93(9), 975–8887. Retrieved from <http://research.ijcaonline.org/volume93/number9/pxc3895792.pdf>
12. Umbarkar, S. S., & Nandgaonkar, S. (2013). Using Association Rule Mining: Stock Market Events Prediction from Financial News. *International Journal of Science and Research (IJSR) ISSN (Online Index Copernicus Value Impact Factor*, 14(6), 2319–7064. Retrieved from <http://www.ijsr.net/archive/v4i6/SUB155622.pdf>
13. Raorane AA, Kulkarni RV, & Jitkar BD. (2012). Association Rule – Extracting Knowledge Using Market Basket Analysis. *Research Journal of Recent Sciences Feb. Res. J. Recent Sci*, 1(2), 19–27. Retrieved from <https://paulallen.ca/documents/2015/01/raorane-a-et-al-association-rule-extracting-knowledge-using-market-basket-analysis-2012.pdf>
14. Kapadia, G., & Kalyandurgmath, K. (2015). Market Basket Analysis of Consumer Buying Behaviour of a Lifestyle Store. *International Conference on Technology and Business Management*, 406–412.
15. Chandrakar, O. (2015). Predicting Examination Results using Association Rule Mining, 116(1), 7–10.
16. Debaditya Barman, Nirmalya Chowdhury (July 2013). Movie Business Trend Prediction using Market Basket Analysis, *International Journal of Computer Applications*, Volume74(9).
17. Kaur, M., & Kang, S. (2016). Market Basket Analysis: Identify the changing trends of market data using association rule mining. *Procedia - Procedia Computer Science*, 85(Cms), 78–85. <https://doi.org/10.1016/j.procs.2016.05.180>
18. Tan, S. C., & Lau, J. P. S. (2014). Time series clustering: A superior alternative for Market Basket Analysis. *Lecture Notes in Electrical Engineering*, 285 LNEE, 241–248. [https://doi.org/10.1007/978-981-4585-18-7\\_28](https://doi.org/10.1007/978-981-4585-18-7_28)

19. Sirisha, G. N. V. G., Shashi, M., & Raju, G. V. P. (2013). Periodic Pattern Mining – Algorithms and Applications. *Global Journal of Computer Science and Technology Software and Data Engineering*, 13(13).
20. Ozden, B., Ramaswamy, S., & Silberschatz, A. (1998). Cyclic association rules. *Proc. of ICDE '98*, 412–421. <https://doi.org/10.1109/ICDE.1998.655804>
21. Anderson, C.: The Long Tail: Why the Future of Business Is Selling Less of More. Hyperion (2006)
22. MicroStrategy: Business Intelligence in the Retail Industry, MicroStrategy World 2003 Conference, Las Vegas, 2003
23. Toroslu, I. H. (2003). Repetition support and mining cyclic patterns. *Expert Systems with Applications*, 25(3), 303–311. [https://doi.org/10.1016/S0957-4174\(03\)00072-1](https://doi.org/10.1016/S0957-4174(03)00072-1)
24. Yun, C. H., Chuang, K. T., & Chen, M. S. (2006). Adherence clustering: An efficient method for mining market-basket clusters. *Information Systems*, 31(3), 170–186. <https://doi.org/10.1016/j.is.2004.11.008>
25. Agrawal, R., Srikant, R.: Fast Algorithms for Mining Association Rules. In Proc. Of the 20th Int'l Conference on Very Large Databases, Santiago, Chile (1994)
26. Heydari, M., & Yousefli, A. (2017). A new optimization model for market basket analysis with allocation considerations: A genetic algorithm solution approach. *Management & Marketing*, 12(1), 1–11. <https://doi.org/10.1515/mmcks-2017-0001>
27. Kandelaar, J. (2016). A cross-category method comparison for doing market basket analysis, (July), 1–100.
28. Aguinis, H., Forcum, L. E., & Joo, H. (2013). Using Market Basket Analysis in Management Research. *Journal of Management*, 39(7), 1799–1824. <https://doi.org/10.1177/0149206312466147>
29. [Figure 1] Dženan Hamzić (January 19<sup>th</sup>, 2017). Retrieved from <https://dzenanhamzic.com/2017/01/19/market-basket-analysis-mining-frequent-pairs-in-python/>
30. [Figure 3, 4] Hafsa Jabeen (August 21<sup>st</sup>, 2018). Retrieved from <https://www.datacamp.com/community/tutorials/market-basket-analysis-r>
31. [Figure 5, 6, 7, 9] Mohit Yadav. Retrieved from <https://www.geeksforgeeks.org/apriori-algorithm/>

32. *[Figure 8]* Susan Li (September 24<sup>th</sup>, 2017). Retrieved from <https://towardsdatascience.com/a-gentle-introduction-on-market-basket-analysis-association-rules-fa4b986a40ce>
33. *[Table 1]* University of Michigan, John W. Creswell (January 10<sup>th</sup>, 2017). Retrieved from Qualitative Inquiry and Research Design: Choosing Among Five Approaches [Book]

ANNEX Rules	3 Months			April			May			June		
	S	C	L	S	C	L	S	C	L	S	C	L
{'CUCURBITACEAE'} => {'PAPRIKA'}	0.0517	0.2943	3.1094	0.0482	0.3058	3.2096	0.0541	0.2708	2.8865	0.0525	0.3135	3.3030
{'PAPRIKA'} => {'CUCURBITACEAE'}	0.0517	0.5462	3.1094	0.0482	0.5057	3.2096	0.0541	0.5762	2.8865	0.0525	0.5534	3.3030
{'GROENTEN', 'TOMATEN'} => {'CUCURBITACEAE'}	0.0660	0.4910	2.7949	0.0607	0.4569	2.8999	0.0703	0.5230	2.6200	0.0664	0.4891	2.9190
{'CUCURBITACEAE'} => {'GROENTEN', TOMATEN'}	0.0660	0.3756	2.7949	0.0607	0.3854	2.8999	0.0703	0.3520	2.6200	0.0664	0.3965	2.9190
{'TOMATEN'} => {'GROENTEN', CUCURBITACEAE'}	0.0660	0.2602	2.7617	0.0607	0.2479	2.8705	0.0703	0.2728	2.6963	0.0664	0.2579	2.7358
{'GROENTEN', 'CUCURBITACEAE'} => {'TOMATEN'}	0.0660	0.7003	2.7617	0.0607	0.7031	2.8705	0.0703	0.6944	2.6963	0.0664	0.7047	2.7358
{'CUCURBITACEAE'} => {'WORTEL GROENTE'}	0.0531	0.3025	2.4700	0.0518	0.3291	2.5594	0.0578	0.2898	2.2861	0.0494	0.2947	2.6283
{'WORTEL GROENTE'} => {'CUCURBITACEAE'}	0.0531	0.4339	2.4700	0.0518	0.4033	2.5594	0.0578	0.4563	2.2861	0.0494	0.4404	2.6283
{'CUCURBITACEAE'} => {'TOMATEN'}	0.1077	0.6132	2.4182	0.0970	0.6154	2.5126	0.1178	0.5899	2.2908	0.1073	0.6404	2.4864
{'TOMATEN'} => {'CUCURBITACEAE'}	0.1077	0.4248	2.4182	0.0970	0.3959	2.5126	0.1178	0.4573	2.2908	0.1073	0.4166	2.4864
{'PAPRIKA'} => {'TOMATEN'}	0.0577	0.6100	2.4056	0.0542	0.5692	2.3239	0.0584	0.6227	2.4180	0.0603	0.6355	2.4670
{'TOMATEN'} => {'PAPRIKA'}	0.0577	0.2277	2.4056	0.0542	0.2214	2.3239	0.0584	0.2269	2.4180	0.0603	0.2342	2.4670
{'TOMATEN'} => {'SALADES', 'GROENTEN'}	0.0566	0.2232	2.2206	0.0559	0.2283	2.2973	0.0562	0.2183	2.1515	0.0576	0.2237	2.2250
{'SALADES', 'GROENTEN'} => {'TOMATEN'}	0.0566	0.5631	2.2206	0.0559	0.5627	2.2973	0.0562	0.5541	2.1515	0.0576	0.5731	2.2250
{'BOLLEN'} => {'TOMATEN'}	0.0631	0.5512	2.1735	0.0610	0.5285	2.1576	0.0638	0.5674	2.2034	0.0645	0.5558	2.1578
{'TOMATEN'} => {'BOLLEN'}	0.0631	0.2490	2.1735	0.0610	0.2491	2.1576	0.0638	0.2476	2.2034	0.0645	0.2504	2.1578
{'TOMATEN'} => {'WORTEL GROENTE'}	0.0661	0.2609	2.1303	0.0665	0.2716	2.1125	0.0694	0.2695	2.1256	0.0624	0.2421	2.1595
{'WORTEL GROENTE'} => {'TOMATEN'}	0.0661	0.5402	2.1303	0.0665	0.5174	2.1125	0.0694	0.5474	2.1256	0.0624	0.5562	2.1595
{'TOMATEN'} => {'KRUIDEN + BLOEMEN'}	0.0502	0.1981	1.9438	0.0481	0.1964	1.8655	0.0519	0.2017	1.9773	0.0504	0.1958	1.9853
{'KRUIDEN + BLOEMEN'} => {'TOMATEN'}	0.0502	0.4929	1.9438	0.0481	0.4569	1.8655	0.0519	0.5092	1.9773	0.0504	0.5114	1.9853
{'GROENTEN', 'TOMATEN'} => {'SALADES'}	0.0566	0.4211	1.9359	0.0559	0.4208	1.9619	0.0562	0.4185	1.8703	0.0576	0.4242	1.9833
{'SALADES'} => {'GROENTEN', 'TOMATEN'}	0.0566	0.2602	1.9359	0.0559	0.2607	1.9619	0.0562	0.2513	1.8703	0.0576	0.2694	1.9833
{'GROENTEN'} => {'CUCURBITACEAE', TOMATEN'}	0.0660	0.2039	1.8927	0.0607	0.1841	1.8990	0.0703	0.2211	1.8780	0.0664	0.2050	1.9105
{'CUCURBITACEAE', 'TOMATEN'} => {'GROENTEN'}	0.0660	0.6125	1.8927	0.0607	0.6262	1.8990	0.0703	0.5967	1.8780	0.0664	0.6191	1.9105
{'SALADES', 'TOMATEN'} => {'GROENTEN'}	0.0566	0.5890	1.8202	0.0559	0.6038	1.8311	0.0562	0.5732	1.8042	0.0576	0.5926	1.8286

{'GROENTEN'} => {'SALADES', 'TOMATEN'}	0.0566	0.1749	1.8202	0.0559	0.1696	1.8311	0.0562	0.1770	1.8042	0.0576	0.1778	1.8286
{'EXOTISCH/TROPISCH'} => {'BESSEN'}	0.0542	0.4363	1.8184	0.0518	0.3986	1.8485	0.0576	0.4328	1.7764	0.0528	0.4831	1.8655
{'BESSEN'} => {'EXOTISCH/TROPISCH'}	0.0542	0.2259	1.8184	0.0518	0.2404	1.8485	0.0576	0.2366	1.7764	0.0528	0.2039	1.8655
{'CUCURBITACEAE'} => {'SALADES'}	0.0692	0.3939	1.8108	0.0619	0.3928	1.8312	0.0777	0.3894	1.7406	0.0671	0.4006	1.8728
{'SALADES'} => {'CUCURBITACEAE'}	0.0692	0.3181	1.8108	0.0619	0.2885	1.8312	0.0777	0.3474	1.7406	0.0671	0.3138	1.8728
{'TOMATEN'} => {'EXOTISCH/TROPISCH'}	0.0557	0.2197	1.7690	0.0561	0.2290	1.7604	0.0598	0.2323	1.7448	0.0511	0.1982	1.8136
{'EXOTISCH/TROPISCH'} => {'TOMATEN'}	0.0557	0.4486	1.7690	0.0561	0.4312	1.7604	0.0598	0.4493	1.7448	0.0511	0.4672	1.8136
{'SALADES'} => {'TOMATEN'}	0.0961	0.4417	1.7417	0.0926	0.4318	1.7628	0.0981	0.4384	1.7023	0.0972	0.4546	1.7650
{'TOMATEN'} => {'SALADES'}	0.0961	0.3789	1.7417	0.0926	0.3781	1.7628	0.0981	0.3809	1.7023	0.0972	0.3775	1.7650
{'KRUIDEN + BLOEMEN'} => {'GROENTEN'}	0.0561	0.5509	1.7023	0.0560	0.5320	1.6132	0.0566	0.5548	1.7463	0.0558	0.5657	1.7456
{'GROENTEN'} => {'KRUIDEN + BLOEMEN'}	0.0561	0.1735	1.7023	0.0560	0.1698	1.6132	0.0566	0.1781	1.7463	0.0558	0.1722	1.7456
{'STENGELGROENTEN'} => {'TOMATEN'}	0.0613	0.4303	1.6969	0.0627	0.4166	1.7007	0.0712	0.4182	1.6237	0.0495	0.4694	1.8223
{'TOMATEN'} => {'STENGELGROENTEN'}	0.0613	0.2416	1.6969	0.0627	0.2560	1.7007	0.0712	0.2763	1.6237	0.0495	0.1922	1.8223
{'SALADES'} => {'GOUDA'}	0.0548	0.2521	1.6896	0.0541	0.2524	1.6800	0.0555	0.2482	1.6867	0.0547	0.2559	1.7030
{'GOUDA'} => {'SALADES'}	0.0548	0.3676	1.6896	0.0541	0.3603	1.6800	0.0555	0.3774	1.6867	0.0547	0.3643	1.7030
{'TOMATEN'} => {'GROENTEN', 'BROOD'}	0.0524	0.2065	1.6865	0.0516	0.2107	1.6723	0.0518	0.2012	1.6939	0.0537	0.2084	1.6942
{'GROENTEN', 'BROOD'} => {'TOMATEN'}	0.0524	0.4277	1.6865	0.0516	0.4096	1.6723	0.0518	0.4362	1.6939	0.0537	0.4364	1.6942
{'PAPRIKA'} => {'GROENTEN'}	0.0512	0.5413	1.6726	0.0509	0.5345	1.6209	0.0499	0.5323	1.6754	0.0529	0.5570	1.7189
{'GROENTEN'} => {'PAPRIKA'}	0.0512	0.1583	1.6726	0.0509	0.1544	1.6209	0.0499	0.1572	1.6754	0.0529	0.1632	1.7189
{'GROENTEN'} => {'CUCURBITACEAE'}	0.0942	0.2911	1.6572	0.0864	0.2619	1.6621	0.1012	0.3185	1.5956	0.0943	0.2909	1.7364
{'CUCURBITACEAE'} => {'GROENTEN'}	0.0942	0.5363	1.6572	0.0864	0.5481	1.6621	0.1012	0.5070	1.5956	0.0943	0.5627	1.7364
{'TOMATEN'} => {'BANANEN'}	0.0505	0.1991	1.6570	0.0544	0.2222	1.5958	0.0493	0.1913	1.6983	0.0481	0.1866	1.6952
{'BANANEN'} => {'TOMATEN'}	0.0505	0.4202	1.6570	0.0544	0.3909	1.5958	0.0493	0.4374	1.6983	0.0481	0.4367	1.6952
{'GROENTEN'} => {'BOLLEN'}	0.0609	0.1882	1.6432	0.0623	0.1890	1.6375	0.0571	0.1796	1.5980	0.0637	0.1965	1.6930
{'BOLLEN'} => {'GROENTEN'}	0.0609	0.5318	1.6432	0.0623	0.5400	1.6375	0.0571	0.5077	1.5980	0.0637	0.5486	1.6930
{'GROENTEN'} => {'TOMATEN'}	0.1344	0.4153	1.6376	0.1329	0.4030	1.6453	0.1344	0.4229	1.6421	0.1358	0.4192	1.6274
{'TOMATEN'} => {'GROENTEN'}	0.1344	0.5300	1.6376	0.1329	0.5426	1.6453	0.1344	0.5217	1.6421	0.1358	0.5274	1.6274
{'WORTELGROENTE'} => {'GROENTEN'}	0.0640	0.5226	1.6148	0.0664	0.5165	1.5662	0.0654	0.5155	1.6225	0.0603	0.5376	1.6590
{'GROENTEN'} => {'WORTELGROENTE'}	0.0640	0.1977	1.6148	0.0664	0.2014	1.5662	0.0654	0.2057	1.6225	0.0603	0.1860	1.6590
{'SALADES'} => {'KIP'}	0.0536	0.2463	1.5982	0.0513	0.2391	1.6024	0.0527	0.2356	1.5621	0.0567	0.2650	1.6332
{'KIP'} => {'SALADES'}	0.0536	0.3477	1.5982	0.0513	0.3437	1.6024	0.0527	0.3495	1.5621	0.0567	0.3493	1.6332

{ 'BESSEN' } => { 'STENGELGROENTEN' }	0.0539	0.2247	1.5779	0.0534	0.2477	1.6458	0.0655	0.2687	1.5788	0.0422	0.1630	1.5453
{ 'STENGELGROENTEN' } => { 'BESSEN' }	0.0539	0.3786	1.5779	0.0534	0.3549	1.6458	0.0655	0.3846	1.5788	0.0422	0.4002	1.5453
{ 'TOMATEN', 'BROOD' } => { 'GROENTEN' }	0.0524	0.5074	1.5678	0.0516	0.5176	1.5696	0.0518	0.5005	1.5754	0.0537	0.5054	1.5595
{ 'GROENTEN' } => { 'TOMATEN', 'BROOD' }	0.0524	0.1618	1.5678	0.0516	0.1565	1.5696	0.0518	0.1631	1.5754	0.0537	0.1656	1.5595
{ 'TOMATEN' } => { 'YOGHURT' }	0.0604	0.2380	1.5552	0.0606	0.2476	1.5173	0.0572	0.2223	1.5643	0.0634	0.2460	1.5868
{ 'YOGHURT' } => { 'TOMATEN' }	0.0604	0.3944	1.5552	0.0606	0.3716	1.5173	0.0572	0.4028	1.5643	0.0634	0.4087	1.5868
{ 'KIP' } => { 'TOMATEN' }	0.0599	0.3883	1.5314	0.0562	0.3769	1.5389	0.0592	0.3923	1.5235	0.0640	0.3943	1.5307
{ 'TOMATEN' } => { 'KIP' }	0.0599	0.2360	1.5314	0.0562	0.2296	1.5389	0.0592	0.2298	1.5235	0.0640	0.2483	1.5307
{ 'BESSEN' } => { 'YOGHURT' }	0.0562	0.2341	1.5298	0.0532	0.2466	1.5113	0.0529	0.2172	1.5284	0.0624	0.2411	1.5555
{ 'YOGHURT' } => { 'BESSEN' }	0.0562	0.3671	1.5298	0.0532	0.3259	1.5113	0.0529	0.3724	1.5284	0.0624	0.4028	1.5555
{ 'GEBAK' } => { 'BROOD' }	0.0694	0.5153	1.5027	0.0741	0.5223	1.4952	0.0620	0.5116	1.5064	0.0726	0.5119	1.5043
{ 'BROOD' } => { 'GEBAK' }	0.0694	0.2023	1.5027	0.0741	0.2123	1.4952	0.0620	0.1826	1.5064	0.0726	0.2133	1.5043
{ 'TOMATEN' } => { 'GOUDA' }	0.0557	0.2196	1.4720	0.0536	0.2189	1.4570	0.0562	0.2183	1.4835	0.0571	0.2215	1.4740
{ 'GOUDA' } => { 'TOMATEN' }	0.0557	0.3733	1.4720	0.0536	0.3569	1.4570	0.0562	0.3821	1.4835	0.0571	0.3797	1.4740
{ 'SCHARRELKIP' } => { 'GROENTEN' }	0.0519	0.4744	1.4658	0.0543	0.4812	1.4593	0.0524	0.4693	1.4770	0.0491	0.4731	1.4600
{ 'GROENTEN' } => { 'SCHARRELKIP' }	0.0519	0.1603	1.4658	0.0543	0.1646	1.4593	0.0524	0.1648	1.4770	0.0491	0.1516	1.4600
{ 'STENGELGROENTEN' } => { 'GROENTEN' }	0.0666	0.4677	1.4452	0.0715	0.4752	1.4411	0.0773	0.4540	1.4289	0.0507	0.4809	1.4840
{ 'GROENTEN' } => { 'STENGELGROENTEN' }	0.0666	0.2058	1.4452	0.0715	0.2169	1.4411	0.0773	0.2432	1.4289	0.0507	0.1565	1.4840
{ 'KIP' } => { 'GROENTEN' }	0.0718	0.4656	1.4387	0.0705	0.4723	1.4323	0.0694	0.4600	1.4480	0.0755	0.4652	1.4357
{ 'GROENTEN' } => { 'KIP' }	0.0718	0.2218	1.4387	0.0705	0.2137	1.4323	0.0694	0.2184	1.4480	0.0755	0.2329	1.4357
{ 'CUCURBITACEAE' } => { 'BESSEN' }	0.0606	0.3447	1.4365	0.0489	0.3106	1.4404	0.0671	0.3363	1.3803	0.0646	0.3854	1.4881
{ 'BESSEN' } => { 'CUCURBITACEAE' }	0.0606	0.2524	1.4365	0.0489	0.2270	1.4404	0.0671	0.2755	1.3803	0.0646	0.2493	1.4881
{ 'SALADES' } => { 'GROENTEN' }	0.1005	0.4620	1.4276	0.0994	0.4633	1.4051	0.1015	0.4535	1.4275	0.1005	0.4701	1.4506
{ 'GROENTEN' } => { 'SALADES' }	0.1005	0.3106	1.4276	0.0994	0.3014	1.4051	0.1015	0.3194	1.4275	0.1005	0.3103	1.4506
{ 'EXOTISCH/TROPISCH' } => { 'GROENTEN' }	0.0565	0.4547	1.4051	0.0591	0.4541	1.3771	0.0593	0.4455	1.4023	0.0511	0.4672	1.4416
{ 'GROENTEN' } => { 'EXOTISCH/TROPISCH' }	0.0565	0.1745	1.4051	0.0591	0.1791	1.3771	0.0593	0.1867	1.4023	0.0511	0.1575	1.4416
{ 'BROOD' } => { 'BANANEN' }	0.0570	0.1661	1.3824	0.0663	0.1898	1.3626	0.0534	0.1573	1.3963	0.0519	0.1525	1.3856
{ 'BANANEN' } => { 'BROOD' }	0.0570	0.4740	1.3824	0.0663	0.4760	1.3626	0.0534	0.4742	1.3963	0.0519	0.4715	1.3856
{ 'TOMATEN' } => { 'BESSEN' }	0.0840	0.3314	1.3810	0.0720	0.2942	1.3640	0.0871	0.3383	1.3885	0.0921	0.3575	1.3803
{ 'BESSEN' } => { 'TOMATEN' }	0.0840	0.3502	1.3810	0.0720	0.3341	1.3640	0.0871	0.3576	1.3885	0.0921	0.3555	1.3803
{ 'EUROPA' } => { 'GROENTEN' }	0.0534	0.4467	1.3802	0.0531	0.4538	1.3761	0.0586	0.4295	1.3518	0.0484	0.4627	1.4280

{'GROENTEN' => {'EUROPA'}}	0.0534	0.1651	1.3802	0.0531	0.1609	1.3761	0.0586	0.1843	1.3518	0.0484	0.1492	1.4280
{'OVERIGE SAUZEN' => {'GROENTEN'}}	0.0527	0.4329	1.3376	0.0516	0.4402	1.3349	0.0532	0.4287	1.3494	0.0532	0.4307	1.3290
{'GROENTEN' => {'OVERIGE SAUZEN'}}	0.0527	0.1629	1.3376	0.0516	0.1566	1.3349	0.0532	0.1674	1.3494	0.0532	0.1642	1.3290
{'BROOD' => {'SALADES'}}	0.0974	0.2841	1.3059	0.0957	0.2739	1.2769	0.0995	0.2931	1.3098	0.0968	0.2845	1.3302
{'SALADES' => {'BROOD'}}	0.0974	0.4478	1.3059	0.0957	0.4461	1.2769	0.0995	0.4448	1.3098	0.0968	0.4527	1.3302
{'GROENTEN' => {'DROGE WORST'}}	0.0564	0.1742	1.3059	0.0554	0.1681	1.3481	0.0602	0.1896	1.2750	0.0532	0.1642	1.3081
{'DROGE WORST' => {'GROENTEN'}}	0.0564	0.4226	1.3059	0.0554	0.4446	1.3481	0.0602	0.4051	1.2750	0.0532	0.4239	1.3081
{'GROENTEN' => {'YOGHURT'}}	0.0646	0.1995	1.3036	0.0690	0.2092	1.2821	0.0593	0.1866	1.3135	0.0659	0.2035	1.3127
{'YOGHURT' => {'GROENTEN'}}	0.0646	0.4219	1.3036	0.0690	0.4228	1.2821	0.0593	0.4173	1.3135	0.0659	0.4254	1.3127
{'GOUDA' => {'GROENTEN'}}	0.0624	0.4186	1.2934	0.0636	0.4231	1.2831	0.0615	0.4179	1.3152	0.0624	0.4151	1.2809
{'GROENTEN' => {'GOUDA'}}	0.0624	0.1930	1.2934	0.0636	0.1928	1.2831	0.0615	0.1935	1.3152	0.0624	0.1925	1.2809
{'GOUDA' => {'BROOD'}}	0.0649	0.4350	1.2686	0.0674	0.4486	1.2843	0.0614	0.4171	1.2282	0.0662	0.4406	1.2947
{'BROOD' => {'GOUDA'}}	0.0649	0.1892	1.2686	0.0674	0.1930	1.2843	0.0614	0.1807	1.2282	0.0662	0.1946	1.2947
{'BESSEN' => {'GROENTEN'}}	0.0978	0.4075	1.2592	0.0895	0.4152	1.2592	0.0985	0.4042	1.2723	0.1048	0.4047	1.2488
{'GROENTEN' => {'BESSEN'}}	0.0978	0.3021	1.2592	0.0895	0.2715	1.2592	0.0985	0.3100	1.2723	0.1048	0.3234	1.2488
{'BROOD' => {'EUROPA'}}	0.0511	0.1489	1.2448	0.0495	0.1417	1.2122	0.0581	0.1712	1.2552	0.0451	0.1324	1.2668
{'EUROPA' => {'BROOD'}}	0.0511	0.4268	1.2448	0.0495	0.4234	1.2122	0.0581	0.4263	1.2552	0.0451	0.4311	1.2668
{'BROOD' => {'DROGE WORST'}}	0.0568	0.1656	1.2413	0.0555	0.1588	1.2736	0.0614	0.1807	1.2156	0.0532	0.1563	1.2447
{'DROGE WORST' => {'BROOD'}}	0.0568	0.4257	1.2413	0.0555	0.4449	1.2736	0.0614	0.4129	1.2156	0.0532	0.4236	1.2447
{'BROOD' => {'YOGHURT'}}	0.0651	0.1898	1.2402	0.0698	0.1998	1.2245	0.0595	0.1753	1.2340	0.0665	0.1953	1.2601
{'YOGHURT' => {'BROOD'}}	0.0651	0.4253	1.2402	0.0698	0.4277	1.2245	0.0595	0.4191	1.2340	0.0665	0.4288	1.2601
{'BROOD' => {'KIP'}}	0.0647	0.1885	1.2232	0.0640	0.1833	1.2283	0.0627	0.1847	1.2243	0.0673	0.1977	1.2185
{'KIP' => {'BROOD'}}	0.0647	0.4194	1.2232	0.0640	0.4291	1.2283	0.0627	0.4158	1.2243	0.0673	0.4147	1.2185
{'BESSEN' => {'BROOD'}}	0.0980	0.4083	1.1907	0.0858	0.3980	1.1394	0.0990	0.4065	1.1968	0.1083	0.4182	1.2288
{'BROOD' => {'BESSEN'}}	0.0980	0.2857	1.1907	0.0858	0.2457	1.1394	0.0990	0.2916	1.1968	0.1083	0.3182	1.2288
{'TOMATEN' => {'BROOD'}}	0.1032	0.4071	1.1871	0.0997	0.4071	1.1654	0.1035	0.4020	1.1838	0.1062	0.4123	1.2116
{'BROOD' => {'TOMATEN'}}	0.1032	0.3010	1.1871	0.0997	0.2854	1.1654	0.1035	0.3049	1.1838	0.1062	0.3121	1.2116
{'SALADES' => {'BESSEN'}}	0.0619	0.2848	1.1867	0.0546	0.2545	1.1802	0.0642	0.2869	1.1777	0.0665	0.3110	1.2007
{'BESSEN' => {'SALADES'}}	0.0619	0.2582	1.1867	0.0546	0.2531	1.1802	0.0642	0.2635	1.1777	0.0665	0.2568	1.2007
{'CUCURBITACEAE' => {'BROOD'}}	0.0714	0.4062	1.1846	0.0635	0.4028	1.1530	0.0815	0.4084	1.2026	0.0681	0.4065	1.1945
{'BROOD' => {'CUCURBITACEAE'}}	0.0714	0.2081	1.1846	0.0635	0.1817	1.1530	0.0815	0.2400	1.2026	0.0681	0.2001	1.1945

{'BROOD'} => {'STENGELGROENTEN'}	0.0569	0.1660	1.1657	0.0604	0.1728	1.1480	0.0694	0.2043	1.2006	0.0405	0.1191	1.1295
{'STENGELGROENTEN'} => {'BROOD'}	0.0569	0.3997	1.1657	0.0604	0.4010	1.1480	0.0694	0.4078	1.2006	0.0405	0.3844	1.1295
{'BROOD'} => {'GROENTEN', 'TOMATEN'}	0.0524	0.1527	1.1365	0.0516	0.1477	1.1118	0.0518	0.1526	1.1357	0.0537	0.1577	1.1611
{'GROENTEN', 'TOMATEN'} => {'BROOD'}	0.0524	0.3897	1.1365	0.0516	0.3883	1.1118	0.0518	0.3857	1.1357	0.0537	0.3951	1.1611
{'GROENTEN'} => {'BROOD'}	0.1225	0.3784	1.1036	0.1260	0.3821	1.0938	0.1188	0.3739	1.1010	0.1230	0.3796	1.1153
{'BROOD'} => {'GROENTEN'}	0.1225	0.3571	1.1036	0.1260	0.3607	1.0938	0.1188	0.3498	1.1010	0.1230	0.3614	1.1153
{'COLA KZH'} => {'GROENTEN'}	0.0571	0.3291	1.0169	0.0555	0.3464	1.0505				0.0612	0.3270	1.0092
{'GROENTEN'} => {'COLA KZH'}	0.0571	0.1763	1.0169	0.0555	0.1682	1.0505				0.0612	0.1888	1.0092
{'KOEKJES'} => {'GROENTEN'}	0.0524	0.3264	1.0085				0.0538	0.3353	1.0553			
{'GROENTEN'} => {'KOEKJES'}	0.0524	0.1620	1.0085				0.0538	0.1694	1.0553			