Individual Coursework Submission Form

Specialist Masters Programme

Surname: Liu	**First Name:** Ching-Wei
MSc in: Business Analytics	**Student ID number:** 180021727
Module Code: SMM799	
Module Title: Applied Research Project	
ARP tutor name: Vali Asimit	**Submission Date:** 24 Aug, 2019

Declaration:

By submitting this work, I declare that this work is entirely my own except those parts duly identified and referenced in my submission. It complies with any specified word limits and the requirements and regulations detailed in the coursework instructions and any other relevant programme and module documentation. In submitting this work I acknowledge that I have read and understood the regulations and code regarding academic misconduct, including that relating to plagiarism, as specified in the Programme Handbook. I also acknowledge that this work will be subject to a variety of checks for academic misconduct.

We acknowledge that work submitted late without a granted extension will be subject to penalties, as outlined in the Programme Handbook. Penalties will be applied for a maximum of five days lateness, after which a mark of zero will be awarded.

Marker's Comments (if not being marked on-line):

Deduction for Late Submission:

Final Mark: %

Savvy Data Insights Project 2
Individual Academic Report of Wholesaler Group

Ching-Wei Liu

August 2019

Supervisor: Dr. Vali Asimit

This report is submitted as part of the requirements for the award of the MSc in Business Analytics.

Abstract

Market basket analysis (MBA) is an analytical technique used to understand customers' purchasing behaviours and uncover associations between products. It is one of the most popular behavioural analytics methods aimed at identifying which items are frequently bought together or placed in the same basket by customers.

This project aims to model customer transactions and compare traditional MBA techniques with network science techniques, and finally investigate the results to provide recommendations for a company. Two MBA models, one based on network algorithms and another based on association rule theory were applied to transaction data provided by a wholesaler. Next, the 10 most significant market baskets identified by each model were analysed for four different markets.

The findings indicate that market baskets vary in different periods and regions, and between the two MBA models. The results can be classified into three different types of item sets based on product categories and features, and thus the company can implement different sales strategies for each one. The rules identified may be relevant to creating competitive advantages for the company.

Table of Contents

List of Tables

List of Figures

1. Introduction

In the last century, data-driven decision making has become more challenging and has played an important role in the strategy development process of companies. Since mining large amounts of data can help to not only understand the sales status of an organisation, but also discover unexplored opportunities, decision-makers are often required to understand the relationship between customers' rationale and behaviour before taking an action.

Market basket analysis (MBA) is a data-mining method that is used to identify frequently-bought item sets to enable the accurate prediction of future purchases. Identifying items frequently purchased together can help optimise marketing strategies regarding which products should be displayed together on a supermarket shelf, or bundled for promotions (Borgelt, 2012). Moreover, it helps to determine the role of certain items such as complementary products.

Association rule mining, a typical application of MBA, focuses on discovering meaningful patterns and interesting relationships between items that exist in the market (Kaur and Kang, 2016). However, Raeder and Chawla (2010) stated that there is no widely accepted measure to assess the usefulness of various rules. Modelling the transaction data as a product network and constructing product communities using network-based techniques can help to isolate useful and actionable rules. Thus, this project's framework involves building two MBA models, one using the association rule mining method and another one using network-science techniques. The results of these models can be used to mine customers' purchasing patterns in different marketplaces and analyse purchase variations in different periods, which may allow companies to improve their sales decisions.

2. Data description

The transaction dataset used in this project was provided by a 'gift and home' wholesaler[1] in the UK. It comprises four sub-datasets[2], namely 'Key accounts', 'The UK', 'France' and 'Rest of the world', corresponding to transaction data collected from its main customers[3], the UK market, the French market and markets in other regions of the world. Each dataset was divided into the four quarters of the year 2018-2019. Table 1 presents information about the variables used in the dataset. To obtain detailed and comparable findings, the two MBA models were applied to each sub-dataset for each quarter.

---

[1] As required by the company's Head of Products, the firms' name is to be kept anonymous.

[2] The transaction data of the firm's US market was not provided.

[3] The Key accounts sub-dataset lists the top 50 customers who have the highest expenditure in the UK.

Table 1: Variables in the wholesaler transaction dataset

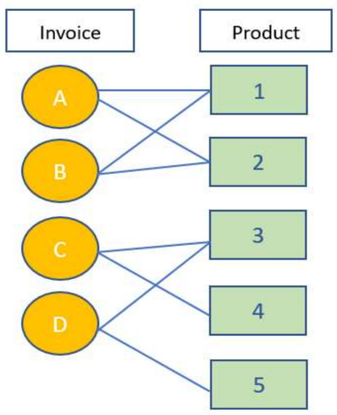
Variable	Description
Item/Product code	Product identification
Customer name	Customer identification UK customers: Customers 1-96, Rest of the world customers: Customers 97-202, French customers: Customer 203-310, Key accounts customers: Customer 311-384
Season	Information about when the product was introduced in the business to be sold (e.g. SS19 indicates spring/summer2019, AW18 indicates autumn/winter 2018, etc.)
Category	Category to which the product belongs (e.g. home accessories, stationery, etc.)
Product type	Type of product sold (e.g. lunch bag, water bottle etc.)
Collection	Trend names within the design world
Colour	Colour of the design
Icon	Icon within the design
Rank	Ranking as per quarterly expenditure
Invoice number	A 6-digit integral number uniquely assigned to each transaction
Invoice date	Date that the invoice was generated There are four periods in each dataset. Q1: April-June 2018, Q2: July-September 2018, Q3: October-December 2018, Q4: January-March 2019

3. Models and methodologies

3.1. The network model

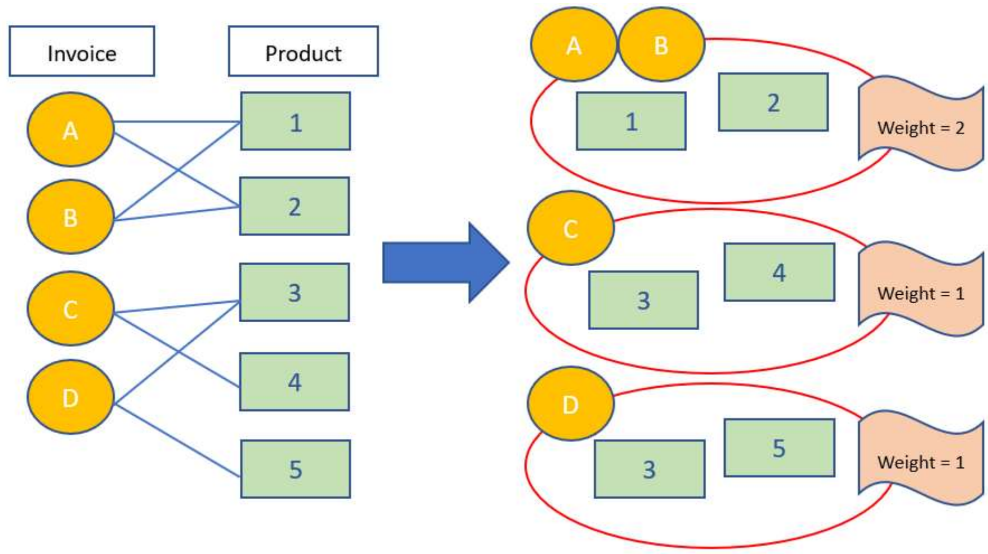
According to Wasserman and Faust (1994), a network is a system comprising nodes and ties. Nodes can represent any element, such as an individual or an item, while ties are connections linking nodes with relations of interest. A network that includes two different sets of nodes wherein ties are only established between nodes in different sets is known as a two-mode or bipartite network (Albert and Barabási, 2002). In this project, the bipartite network concept was utilised to construct a network model of the wholesaler dataset to determine which products are frequently bought together. Since one invoice can be regarded as a transaction, a two-mode network was created using invoices, with products representing two different sets of nodes and links indicating which products are included in an invoice. For example, in Figure 1, Products 1 and 2 are purchased together in Invoices A and B, Products 3 and 4 are purchased together in Invoice C and so on.

Figure 1: Bipartite network of the wholesaler transaction data



Unfortunately, few methods exist for analysing two-mode networks. However, a bipartite network can be converted into a one-mode network for network analysis (Borgatti and Everett, 1997). Zhou et al. (2007) suggested that this conversion can be done through one-mode network projection, stating that a weighted one-mode network is more informative than an unweighted one after projection. Thus, the next step involved creating a weighted one-mode co-purchased product network, with the weight of edges representing the number of times product sets appear in one invoice. Figure 2 illustrates the process of projecting an originally invoice-product bipartite network onto a weighted one-mode product network. It shows that since Products 1 and 2 appear in both Invoice A and Invoice B, the item set $\{\text{Product 1}, \text{Product 2}\}$ has a weightage of two because it appears two times in all the invoices in the wholesaler dataset.

Figure 2: Illustration of the weighted product network projection



In the one-mode product network, products that are bought together and the frequency with which product sets appear in the wholesaler data can be identified. To simplify and conveniently analyse the results, the model was set to only allow two products in a basket. After sorting results by weight value in descending order, product sets with the highest weights, representing the most frequently-bought item sets, were identified and discussed further.

Code implementation:

Step 1: To analyse a complete dataset, data with missing values for any features were ignored.

Step 2: A bipartite network was created using the variable ‘invoice number’ as one set of nodes and ‘item/product code’ as another one.

Step 3: The two-mode (invoice-product) network was projected onto a weighted one-mode (product-product) network and weights were sorted in a descending order. Figure 3 is an example of a result of the co-purchased product network, showing that, among all the transactions, 346Prod and 542Prod were bought together 36 times. The top 10 rules with the highest frequency (largest number of counts) were selected for analysis and comparison with the traditional association rule model.

Figure 3: An example of a network model result

	Prod1	Prod2	count
12881	346Prod	542Prod	36
18089	542Prod	1646Prod	34
12879	346Prod	1646Prod	31
13077	296Prod	2297Prod	28
19696	159Prod	2297Prod	27
13073	296Prod	840Prod	26
13079	296Prod	159Prod	26
16814	840Prod	2297Prod	26
16794	840Prod	159Prod	26
1873	210Prod	286Prod	11

3.2. The association rule model

The concept of association rule mining was introduced by Agrawal, Imieliński and Swami (1993), who stated that every rule is an item set comprising an antecedent or left-hand-side (LHS) and a consequent or right-hand-side (RHS). For instance, a rule can be written as $\{\text{butter}, \text{bread}\} => \{\text{milk}\}$, wherein the former is the antecedent and the latter is the consequent. This rule indicates that when customers buy butter and bread, they also buy milk.

This project utilises Agrawal and Srikant's (1994) Apriori algorithm, which is a popular association rule mining method, to generate rules and select useful rules from all possible rules using specific constraints. Constraints are thresholds for the support and confidence levels. Support is defined as the proportion of transactions involving a specific item set. For example, if an item set $\{\text{butter}, \text{bread}, \text{milk}\}$ has a support level of 0.2, it means that the item set appears in 20% of all transactions in the dataset. Confidence indicates how often a rule holds true. If the confidence level for a rule $\{\text{butter}, \text{bread}\} => \{\text{milk}\}$ is 0.2, it means that 20% of all the transactions involving butter and bread also involve milk. Similar to Agrawal and Srikant's (1994) work, this project mines rules that have a support and confidence level greater than user-defined minimum support and confidence levels.

This project required to the user-defined minimum support and confidence levels to be satisfied at the same time; however, there was no specific guideline on how to determine these thresholds. Thus, in this report, since the goal was to compare the results of traditional association rule mining with those obtained from network science, an association rule model was constructed using the minimum support and confidence levels that can generate as many rules as the network model. Additionally, to make the results of both models comparable, the basket in the association model was set to contain only two items like the network model. After sorting baskets based on the descending order of their support and confidence values, rules with the highest support and confidence levels, which are typically the most important rules, were identified and compared with the network model.

Code implementation:

Step 1: To perform the analysis with a complete dataset, data with missing values for any features were ignored.

Step 2: Transaction data were created using the variables 'invoice number' and 'invoice date' in the dataset. Table 2 provides an example of the transaction data. Each row shows products bought together in one transaction.

Table 2: An example of transaction data in the association rule model

Invoice date	Invoice number	Product
2018/5/23	174366	94Prod, 325Prod, 234Prod
2018/6/29	179000	1719Prod, 1049Prod, 864Prod, 790Prod

Step 3: The transaction data were used to mine rules based on the Apriori algorithm. The support and confidence levels were set to the values that could generate as many rules as the network model. Table 3 shows the support and confidence levels selected and the number of rules found in each sub-dataset.

Since the number of rules found in the association rule model was sensitive to the selected support and confidence values and could not exactly match the number of rules found in the network model, thresholds that could generate an approximate number of rules were chosen.

Step 4: The results were sorted by their support and confidence values in descending order. Figure 4 is an example of a result of the association rule model, which shows that 1646Prod and 542Prod are the LHS and RHS respectively, and that this rule had a support value of 0.0825 and a confidence value of 1. The top 10 rules with the highest support and confidence levels were selected for analysis and comparison with the network model.

Table 3: The number of rules found in both models and the support and confidence levels selected in the association rule model

Sub-dataset	Period	Number of rules found in the network model	Number of rules found in the association rule model	minimum support level	minimum confidence level
Key accounts	Q1	21,872	25,233	0.001	0.5
	Q2	64,986	72,647	0.001	0.5
	Q3	75,364	83,482	0.001	0.5
	Q4	8,531	7,100	0.001	0.55
The UK	Q1	49,065	49,828	0.001	0.15
	Q2	58,310	58,278	0.001	0.25
	Q3	94,371	95,345	0.001	0.2
	Q4	59,145	61,588	0.001	0.2
France	Q1	66,422	72,709	0.002	0.25
	Q2	41,251	49,061	0.002	0.25
	Q3	61,756	60,606	0.002	0.3
	Q4	61,756	65,931	0.002	0.3
Rest of the world	Q1	76,312	76,540	0.0025	0.22
	Q2	89,526	87,585	0.0025	0.25
	Q3	130,748	144,334	0.0025	0.25
	Q4	52,206	49,149	0.0025	0.245

Figure 4: An example of an association rule model result

	lhs	rhs	support	confidence	lift	count
11765	1646Prod	542Prod	0.082524	1.0	8.956522	34
11790	296Prod	2297Prod	0.067961	1.0	13.733333	28
25224	1016Prod	257Prod	0.024272	1.0	29.428571	10
11634	873Prod	571Prod	0.016990	1.0	51.500000	7
25098	513Prod	210Prod	0.016990	1.0	29.428571	7
10964	11Prod	6Prod	0.014563	1.0	68.666667	6
10965	6Prod	11Prod	0.014563	1.0	68.666667	6
10966	11Prod	835Prod	0.014563	1.0	68.666667	6
10967	835Prod	11Prod	0.014563	1.0	68.666667	6
10969	11Prod	168Prod	0.014563	1.0	58.857143	6

4. Analyses

Since consumers' purchase decisions may vary between different regions and periods, the analyses of the two MBA models' results were conducted for four sub-datasets corresponding to the company's key customers and its three market regions. The overall results of these sub-datasets were analysed to provide suggestions for the wholesaler to boost its sales.

4.1. Key accounts

Key account customers are the top 50 customers with the highest expenditure in the UK. In the first quarter (Figure 5), two main patterns appear in the frequently-bought item sets of the two MBA models. First, products that could be used for the same purpose, such as travelling, were usually bought together. For example, both models found that luggage tags (1646Prod) and shopping bags (542Prod) were bought together and were also the most frequently-bought items in this period. Second, products belonging to the same category from different collections or in different colours, such as trinket dishes (296Prod and 2297Prod) and mugs (11Prod, 6Prod, 835Prod and 168 Prod), were usually bought together, indicating that different designs of the same product are viewed as substitutes.

Figure 5: Result for the Key accounts sub-dataset in Q1

Network model

	Prod1	Prod2	count
12881	346Prod	542Prod	36
18089	542Prod	1646Prod	34
12879	346Prod	1646Prod	31
13077	296Prod	2297Prod	28
19696	159Prod	2297Prod	27
13073	296Prod	840Prod	26
13079	296Prod	159Prod	26
16814	840Prod	2297Prod	26
16794	840Prod	159Prod	26
1873	210Prod	286Prod	11

Association rule model

	lhs	rhs	support	confidence	lift	count
11765	1646Prod	542Prod	0.082524	1.0	8.956522	34
11790	296Prod	2297Prod	0.067961	1.0	13.733333	28
25224	1016Prod	257Prod	0.024272	1.0	29.428571	10
11634	873Prod	571Prod	0.016990	1.0	51.500000	7
25098	513Prod	210Prod	0.016990	1.0	29.428571	7
10964	11Prod	6Prod	0.014563	1.0	68.666667	6
10965	6Prod	11Prod	0.014563	1.0	68.666667	6
10966	11Prod	835Prod	0.014563	1.0	68.666667	6
10967	835Prod	11Prod	0.014563	1.0	68.666667	6
10969	11Prod	168Prod	0.014563	1.0	58.857143	6

For the second quarter (Figure 6), the network model shows that consumers who bought keyrings (74Prod, 111Prod and 55Prod) also bought trinket dishes (325Prod), flasks (1671Prod) or mugs (1201Prod), indicating that this group of items may be supplementary goods that can be bundled for sale. Another interesting finding, found by the association rule model, was that Christmas products were purchased together during this period; a wooden suitcase (999Prod) with a reindeer icon was bought with coasters (822Prod), milk bottles (1005Prod) or decorations (2374Prod). These frequently-purchased Christmas products had similar features of wood textures or reindeer icons.

Figure 6: Result for the Key accounts sub-dataset in Q2

Network model

	Prod1	Prod2	count
55633	325Prod	55Prod	43
7675	1671Prod	1201Prod	32
4942	74Prod	111Prod	32
7721	1671Prod	325Prod	32
51779	1201Prod	111Prod	31
4952	74Prod	325Prod	31
4943	74Prod	1671Prod	31
4945	74Prod	55Prod	31
4947	74Prod	1201Prod	31
7904	1671Prod	111Prod	31

Association rule model

	lhs	rhs	support	confidence	lift	count
18431	74Prod	111Prod	0.056140	1.0	17.272727	32
45065	168Prod	6Prod	0.015789	1.0	47.500000	9
72475	2283Prod	1170Prod	0.014035	1.0	57.000000	8
12382	822Prod	2374Prod	0.012281	1.0	43.846154	7
52551	210Prod	286Prod	0.012281	1.0	71.250000	7
72372	2282Prod	1170Prod	0.012281	1.0	57.000000	7
5440	2437Prod	144Prod	0.010526	1.0	51.818182	6
5505	999Prod	1005Prod	0.010526	1.0	47.500000	6
5507	999Prod	822Prod	0.010526	1.0	81.428571	6
5509	999Prod	2374Prod	0.010526	1.0	43.846154	6

In the third quarter (Figure 7), the network model found that products from the same category with different collections or colours, such as mugs (6Prod, 11Prod, 168Prod, 8Prod and 248Prod) were often put in a basket, similar to the findings in the first quarter. The association rule model shows the same pattern for this period as well, such as different kinds of mugs (1163Prod, 2282Prod, 6Prod, 168Prod, 1170Prod and 11Prod). Moreover, the association rule model revealed that Christmas decorations (976Prod) were bought together with plant plots (448Prod) or trinket dishes (553Prod) from the ‘Mandala elephant’ collection. This shows that Christmas decorations can be sold together with plant plots or trinket dishes, and that plant plots and trinket dishes are alternatives to each other.

Figure 7: Result for the Key accounts sub-dataset in Q3

Network model

	Prod1	Prod2	count
26007	99Prod	370Prod	17
19584	6Prod	11Prod	17
448	168Prod	6Prod	16
473	168Prod	11Prod	15
57556	8Prod	11Prod	14
19514	6Prod	8Prod	14
16863	553Prod	448Prod	13
26094	173Prod	8Prod	13
19570	6Prod	248Prod	13
26186	173Prod	11Prod	12

Association rule model

	lhs	rhs	support	confidence	lift	count
81512	1170Prod	11Prod	0.017518	1.0	36.052632	12
36877	1163Prod	105Prod	0.014599	1.0	48.928571	10
36895	1163Prod	2282Prod	0.014599	1.0	57.083333	10
36897	1163Prod	6Prod	0.014599	1.0	32.619048	10
36898	1163Prod	168Prod	0.014599	1.0	28.541667	10
36899	1163Prod	1170Prod	0.014599	1.0	57.083333	10
36903	1163Prod	11Prod	0.014599	1.0	36.052632	10
4919	976Prod	448Prod	0.013139	1.0	48.928571	9
4921	976Prod	553Prod	0.013139	1.0	45.666667	9
13812	1688Prod	1642Prod	0.013139	1.0	68.500000	9

In the fourth quarter (Figure 8), customers preferred home/kitchen accessories. For instance, in the network model, bowls (232Prod) were found to be purchased along with bamboo sets (1Prod), cushions (206Prod), cutlery sets (295Prod) or plates (109Prod) in similar colours, such as white, grey, cream or silver. Similarly, in the association rule model, cutlery sets (295Prod) frequently appeared in baskets, and were bought with mugs (281Prod), bowls (232Prod), plates (109Prod), cushions (206Prod), hooks (277Prod) and bamboo sets (103Prod). It is worth noticing that cutlery sets were bought with plates or bowls from the same collection 'sweet dreams' in the white and grey colours, which is recognised in both models.

Figure 8: Result for the Key accounts sub-dataset in Q4

Network model

	Prod1	Prod2	count
3429	103Prod	1Prod	38
6597	1Prod	206Prod	37
5833	277Prod	302Prod	36
6588	1Prod	295Prod	36
2786	281Prod	103Prod	36
8206	109Prod	295Prod	36
4158	232Prod	1Prod	36
4161	232Prod	206Prod	36
4162	232Prod	109Prod	36
4163	232Prod	295Prod	36

Association rule model

	lhs	rhs	support	confidence	lift	count
7077	206Prod	1Prod	0.078390	1.0	11.800000	37
7021	295Prod	281Prod	0.076271	1.0	13.111111	36
7022	281Prod	295Prod	0.076271	1.0	13.111111	36
7023	295Prod	232Prod	0.076271	1.0	13.111111	36
7024	232Prod	295Prod	0.076271	1.0	13.111111	36
7025	295Prod	109Prod	0.076271	1.0	13.111111	36
7026	109Prod	295Prod	0.076271	1.0	13.111111	36
7027	295Prod	206Prod	0.076271	1.0	12.756757	36
7029	295Prod	277Prod	0.076271	1.0	12.102564	36
7031	295Prod	103Prod	0.076271	1.0	10.976744	36

4.2. The UK

In the first quarter (Figure 9), both models showed that products from the same category were more likely to be bought together. For example, money boxes (119Prod, 15Prod and 30Prod) or mugs (1130Prod, 1171Prod and 1499Prod) from different collections or in different colours were frequently bought in pairs. Additionally, there were several other types of items that were usually purchased together, such as money boxes (15Prod) with passport holders (81Prod) or glasses cases (18Prod), and suitcases (88Prod) with letter racks (2142Prod) or lunch boxes (572Prod). In the association rule model, the rules were quite scattered, but it was still evident that kitchen/home accessories were bought frequently together, for instance, coasters (852Prod) formed an item set with trinket dishes (136Prod) or mugs (1805Prod).

Figure 9: Result for the UK sub-dataset in Q1

Network model

	Prod1	Prod2	count
47003	303Prod	81Prod	32
13571	15Prod	81Prod	15
11932	18Prod	15Prod	15
2077	2327Prod	2138Prod	15
45938	27Prod	88Prod	15
38294	119Prod	30Prod	15
13352	15Prod	119Prod	15
30454	56Prod	317Prod	15
31479	2142Prod	119Prod	14
31510	2142Prod	88Prod	14

Association rule model

	lhs	rhs	support	confidence	lift	count
44314	572Prod	88Prod	0.015504	1.0	12.285714	8
11796	903Prod	766Prod	0.013566	1.0	64.500000	7
17346	1890Prod	1573Prod	0.011628	1.0	73.714286	6
42662	1130Prod	1171Prod	0.011628	1.0	73.714286	6
42703	1130Prod	1499Prod	0.011628	1.0	23.454545	6
1093	647Prod	327Prod	0.009690	1.0	23.454545	5
35987	806Prod	834Prod	0.009690	1.0	86.000000	5
3248	226Prod	696Prod	0.007752	1.0	51.600000	4
17212	852Prod	136Prod	0.007752	1.0	73.714286	4
17220	852Prod	1805Prod	0.007752	1.0	57.333333	4

Three clear rules appeared in either the network model or the association rule model in the second quarter (Figure 10). First, customers purchased lunch bags (66Prod), along with glasses cases (18Prod), rugs (28Prod) or other lunch bags (56Prod and 254Prod) from different collections or in different colours. Second, drawer knobs (766Prod, 903Prod and 358Prod) were the most frequently-bought item in this period and were often bought together with other knobs from different collections or in different colours. Third, rugs (28Prod) were frequently purchased with lunch bags (66Prod), bookends (1008Prod) or other types of rugs (175Prod). Aside from these three findings, in the association rule model, it was observed that Christmas decorations (1065Prod, 1191Prod, 749 Prod and 1293Prod) from the same collection 'Wonderland' were usually purchased together.

Figure 10: Result for the UK sub-dataset in Q2

Network model

	Prod1	Prod2	count
42695	28Prod	175Prod	23
16049	18Prod	66Prod	18
36727	13Prod	2Prod	17
37199	56Prod	66Prod	16
42676	28Prod	66Prod	16
56285	303Prod	81Prod	16
48564	52Prod	127Prod	16
54742	254Prod	66Prod	16
27591	1008Prod	28Prod	15
37124	56Prod	52Prod	14

Association rule model

	lhs	rhs	support	confidence	lift	count
56163	380Prod	192Prod	0.013769	1.0	44.692308	8
57457	1293Prod	1065Prod	0.012048	1.0	58.100000	7
57852	1191Prod	1065Prod	0.012048	1.0	58.100000	7
58090	749Prod	1065Prod	0.012048	1.0	58.100000	7
56829	2389Prod	1507Prod	0.010327	1.0	44.692308	6
57705	1489Prod	534Prod	0.010327	1.0	72.625000	6
19614	903Prod	766Prod	0.008606	1.0	116.200000	5
19615	766Prod	903Prod	0.008606	1.0	116.200000	5
19621	903Prod	358Prod	0.008606	1.0	72.625000	5
19629	766Prod	358Prod	0.008606	1.0	72.625000	5

In the third quarter (Figure 11), the two models recognised several bundle sets of different types of products. First, toys (126Prod) were bought with plant pots (77Prod, 146Prod, 702Prod and 465Prod) from different collections or in different colours, clothing accessories (266Prod), mugs (105Prod) or money boxes (186Prod). Second, passport holders (81Prod) were bought together with luggage tags (303Prod) and glasses cases (18Prod). Nonetheless, some item sets with similar features were also bought together. For instance, cushions (23Prod, 313Prod and 206Prod) from the 'Sweet dreams' or 'Stars' collections and in the colours of silver, white or beige were likely to be purchased together. Additionally, plant pots (465Prod and 480Prod) from the 'Scandi boho' collection and in the black and cream colours were frequently bought together.

Figure 11: Result for the UK sub-dataset in Q3

Network model

	Prod1	Prod2	count
18985	13Prod	2Prod	18
37888	81Prod	303Prod	16
29843	290Prod	1008Prod	16
88972	480Prod	465Prod	15
44118	23Prod	313Prod	14
84880	1184Prod	1049Prod	14
44158	23Prod	206Prod	14
37793	81Prod	18Prod	14
79905	1170Prod	1163Prod	14
46683	18Prod	315Prod	13

Association rule model

	lhs	rhs	support	confidence	lift	count
69868	465Prod	480Prod	0.018916	1.0	41.736842	15
41046	2064Prod	1956Prod	0.011349	1.0	44.055556	9
57625	266Prod	26Prod	0.011349	1.0	31.720000	9
44984	126Prod	266Prod	0.010088	1.0	88.111111	8
44986	126Prod	77Prod	0.010088	1.0	66.083333	8
44988	126Prod	105Prod	0.010088	1.0	56.642857	8
44990	126Prod	702Prod	0.010088	1.0	61.000000	8
44992	126Prod	465Prod	0.010088	1.0	52.866667	8
44994	126Prod	186Prod	0.010088	1.0	56.642857	8
44996	126Prod	146Prod	0.010088	1.0	41.736842	8

Home/kitchen accessories were popular in the fourth quarter (Figure 12). Product sets often comprised products from the same category with similar characteristics, such as money boxes (119Prod, 30Prod and 15Prod) from the ‘Gold’ or “Vintage map” collection and mugs (1573Prod, 2024Prod, 2111Prod and 2064Prod) from the ‘Bohemian’ collection and in multiple colours. Nevertheless, different types of products were purchased together as well. For instance, suitcases (27Prod) were bought with rugs (19Prod), and hooks (1285Prod) were bought with bookends (1601Prod), meaning that these products could be bundled for promotions.

Figure 12: Result for the UK sub-dataset in Q4

Network model

	Prod1	Prod2	count
56690	303Prod	81Prod	28
55177	27Prod	88Prod	22
55155	27Prod	19Prod	17
36217	13Prod	40Prod	16
46665	119Prod	30Prod	16
47428	130Prod	66Prod	15
28586	1008Prod	2142Prod	15
17901	15Prod	119Prod	15
16674	18Prod	13Prod	15
3320	821Prod	15Prod	15

Association rule model

	lhs	rhs	support	confidence	lift	count
43430	2024Prod	2111Prod	0.011952	1.0	31.375000	6
43440	2024Prod	1573Prod	0.011952	1.0	27.888889	6
32365	495Prod	1014Prod	0.009960	1.0	45.636364	5
52203	1776Prod	1601Prod	0.009960	1.0	33.466667	5
53282	1351Prod	1028Prod	0.009960	1.0	71.714286	5
20377	2064Prod	2111Prod	0.007968	1.0	31.375000	4
30367	417Prod	272Prod	0.007968	1.0	100.400000	4
41999	1387Prod	748Prod	0.007968	1.0	20.080000	4
45893	879Prod	1499Prod	0.007968	1.0	33.466667	4
47633	1285Prod	1601Prod	0.007968	1.0	33.466667	4

4.3. France

For the France sub-dataset, the two models show that home/kitchen accessories were quite popular in the first quarter (Figure 13) and the item sets that customers bought had a clear preference in terms of design. Cutlery sets (44Prod), recognised as the most frequently-bought products in the network model, is an example. They were often bought together with bowls (32Prod), mugs (49Prod), plates (21Prod) or water bottles (41Prod), all belonging to the same ‘Bear camp’ collection and having the same ‘bear’ icon. Moreover, the association rule model shows that four types of items (245Prod, 150Prod, 164Prod and 191Prod) from the same ‘fruit & veg’ collection formed pair sets and even shared the same colour (cream).

Figure 13: Result for the France sub-dataset in Q1

Network model

	Prod1	Prod2	count
2278	44Prod	32Prod	43
18382	49Prod	32Prod	38
2251	44Prod	21Prod	37
2316	44Prod	49Prod	37
37016	193Prod	113Prod	36
8319	41Prod	21Prod	36
2235	44Prod	41Prod	36
10625	21Prod	32Prod	36
62592	60Prod	2Prod	35
8609	41Prod	16Prod	35

Association rule model

	lhs	rhs	support	confidence	lift	count
66885	732Prod	195Prod	0.031579	1.0	28.500000	9
42264	605Prod	661Prod	0.024561	1.0	35.625000	7
61639	150Prod	245Prod	0.024561	1.0	35.625000	7
66731	164Prod	245Prod	0.024561	1.0	35.625000	7
67107	191Prod	245Prod	0.024561	1.0	35.625000	7
33847	391Prod	267Prod	0.021053	1.0	40.714286	6
35831	1031Prod	127Prod	0.021053	1.0	19.000000	6
48286	1297Prod	732Prod	0.021053	1.0	31.666667	6
48295	1297Prod	195Prod	0.021053	1.0	28.500000	6
63098	377Prod	376Prod	0.021053	1.0	40.714286	6

In the second quarter (Figure 14), the network model shows similar findings as those in the first quarter. Customers tended to buy cutlery sets (44Prod) with bowls (32Prod), plates (21Prod), mugs (49Prod) or lunch bags (16Prod) from the ‘Bear camp’ collection and with the ‘bear’ icon. In the association rule model, the most common rules were that hooks (525Prod) were bought together with home/kitchen accessories or bags, such as plant pots (57Prod), bowls (485Prod), vases (394Prod), lunch bags (123Prod) or suitcases (34Prod), but there was no similar feature representing these item sets. Finally, it was observed that a mug (72Prod) with a bunny icon, introduced during Easter, appeared in both models and was one of the most frequently-purchased items in this period, indicating that consumers may buy Easter products in the second quarter and that this bunny mug could be bundled with other mugs (166Prod) or hooks (525Prod) for festive promotion according to the rules identified.

Figure 14: Result for the France sub-dataset in Q2

Network model

	Prod1	Prod2	count
958	44Prod	32Prod	54
940	44Prod	21Prod	51
5715	21Prod	32Prod	47
982	44Prod	49Prod	44
11387	49Prod	32Prod	41
23948	4Prod	3Prod	40
5742	21Prod	49Prod	39
32077	166Prod	57Prod	33
1030	44Prod	16Prod	33
32029	166Prod	72Prod	32

Association rule model

	lhs	rhs	support	confidence	lift	count
42371	525Prod	485Prod	0.109023	1.0	8.580645	29
42381	525Prod	394Prod	0.109023	1.0	7.388889	29
42387	525Prod	57Prod	0.109023	1.0	6.650000	29
42393	525Prod	123Prod	0.109023	1.0	6.820513	29
42395	525Prod	34Prod	0.109023	1.0	7.388889	29
42399	525Prod	72Prod	0.109023	1.0	6.045455	29
39766	118Prod	111Prod	0.037594	1.0	16.625000	10
41029	42Prod	36Prod	0.037594	1.0	20.461538	10
38598	83Prod	58Prod	0.033835	1.0	13.300000	9
39178	164Prod	245Prod	0.033835	1.0	24.181818	9

In the third quarter (Figure 15), the network model shows results similar to those in the last two quarters. The model revealed that products, such as water bottles (396Prod), bamboo sets (115Prod), lunch boxes (318Prod) or umbrellas (157Prod) from the ‘Puppy dog playtime’ collection were often purchased together. Item sets from the ‘Bear camp’ collection were still popular in this period. The association rule model shows that lunch bags (52Prod) were often bought with mugs (288Prod), photo frames (562Prod), home accessories (340Prod) or sippy cups (509Prod); however, these items had no common features. In the end, both models recognised the rule that water bottles (396Prod) and lunch boxes (318Prod) from the ‘Puppy dog playtime’ collection were bought together, meaning that this was a strong rule and it was quite reasonable to bundle these two items.

Figure 15: Result for the France sub-dataset in Q3

Network model

	Prod1	Prod2	count
1694	44Prod	32Prod	43
1555	44Prod	21Prod	42
9413	21Prod	32Prod	39
7478	41Prod	29Prod	37
28304	396Prod	115Prod	36
28348	396Prod	318Prod	35
60784	318Prod	157Prod	34
28310	396Prod	157Prod	34
30989	115Prod	157Prod	34
30995	115Prod	158Prod	34

Association rule model

	lhs	rhs	support	confidence	lift	count
58648	318Prod	396Prod	0.101156	1.0	9.351351	35
60135	288Prod	52Prod	0.072254	1.0	8.871795	25
59952	562Prod	52Prod	0.066474	1.0	8.871795	23
59966	562Prod	110Prod	0.066474	1.0	8.238095	23
59362	340Prod	52Prod	0.046243	1.0	8.871795	16
59323	611Prod	110Prod	0.040462	1.0	8.238095	14
57727	509Prod	208Prod	0.031792	1.0	11.533333	11
57735	509Prod	23Prod	0.031792	1.0	10.484848	11
57737	509Prod	52Prod	0.031792	1.0	8.871795	11
51708	144Prod	209Prod	0.026012	1.0	24.714286	9

In the fourth quarter (Figure 16), the network model shows that water bottles (396Prod) were popular, and were usually bought with lunch boxes (318Prod), travel mugs (157Prod) or bamboo sets (115Prod). Besides, the 'Bear camp' collection was still popular, and cutlery sets (44Prod), bowls (32Prod) and plates (21Prod) were the three products from this collection that were commonly bought together. The association rule model indicates that stationery was a best-selling product in the last quarter. Luggage tags (495Prod) were purchased frequently with stationery, such as different types of sticky notes (404Prod, 604Prod, 557Prod and 444Prod) or pens (361Prod), meaning that these items could be bundled. However, different types of stationery could also be purchased together, for example, sticky notes (604Prod) with notebooks (289Prod and 260Prod).

Figure 16: Result for the France sub-dataset in Q4

Network model

	Prod1	Prod2	count
1694	44Prod	32Prod	43
1555	44Prod	21Prod	42
9413	21Prod	32Prod	39
7478	41Prod	29Prod	37
28304	396Prod	115Prod	36
28348	396Prod	318Prod	35
60784	318Prod	157Prod	34
28310	396Prod	157Prod	34
30989	115Prod	157Prod	34
30995	115Prod	158Prod	34

Association rule model

	lhs	rhs	support	confidence	lift	count
63125	294Prod	298Prod	0.140426	1.0	6.714286	33
65492	557Prod	323Prod	0.140426	1.0	5.595238	33
65584	361Prod	323Prod	0.140426	1.0	5.595238	33
65409	604Prod	289Prod	0.136170	1.0	5.222222	32
65417	604Prod	260Prod	0.136170	1.0	4.700000	32
65192	495Prod	404Prod	0.123404	1.0	7.343750	29
65194	495Prod	604Prod	0.123404	1.0	7.343750	29
65198	495Prod	557Prod	0.123404	1.0	7.121212	29
65200	495Prod	444Prod	0.123404	1.0	6.714286	29
65202	495Prod	361Prod	0.123404	1.0	7.121212	29

4.4. Rest of the world

In the first quarter (Figure 17), the network model reveals that home/kitchen accessories were popular. Money boxes (40Prod) were the most frequently-bought items in the identified rules and were often purchased with other types of money boxes (15Prod, 730Prod and 119Prod) or water bottles (262Prod). The association rule model shows that consumers often purchased stationery during this period. Customers who bought photo frames (1940Prod) also bought planners (886Prod), sticky notes (1295Prod), notepads (1182Prod), pouches (696Prod) or mugs (441Prod) from the same 'Paint splash' collection. It is worth noting that both models found that pouches (961Prod) were bought together with coin purses (652Prod) from the same 'Paint splash' collection, indicating that this pair was a frequently-bought item set.

Figure 17: Result for the Rest of the world sub-dataset in Q1

Network model

	Prod1	Prod2	count
31628	40Prod	15Prod	44
41433	262Prod	15Prod	44
31562	40Prod	262Prod	42
31588	40Prod	730Prod	42
19392	587Prod	664Prod	41
31467	40Prod	119Prod	41
63718	119Prod	15Prod	41
41370	262Prod	730Prod	41
51792	664Prod	105Prod	41
56984	652Prod	961Prod	40

Association rule model

	lhs	rhs	support	confidence	lift	count
32291	961Prod	652Prod	0.156863	1.0	5.666667	40
34993	468Prod	262Prod	0.129412	1.0	4.322034	33
35131	1295Prod	1182Prod	0.062745	1.0	15.000000	16
34549	886Prod	1295Prod	0.058824	1.0	15.937500	15
34551	886Prod	1182Prod	0.058824	1.0	15.000000	15
29305	1940Prod	886Prod	0.039216	1.0	17.000000	10
29315	1940Prod	1295Prod	0.039216	1.0	15.937500	10
29317	1940Prod	1182Prod	0.039216	1.0	15.000000	10
29319	1940Prod	696Prod	0.039216	1.0	6.375000	10
29324	1940Prod	441Prod	0.039216	1.0	11.590909	10

In the second quarter (Figure 18), the results of the network model vary from those of the association rule model. In the network model, home/kitchen accessories were frequently-bought items whereas in the association rule model, frequently-purchased items were quite diverse. For instance, LED lights (2Prod) were often purchased with mugs (7Prod), water bottles (213Prod) or other types of LED lights (13Prod and 5Prod) in the network model. Moreover, it also shows that customers preferred to buy different products with the same design, such as bowls (354Prod) and plates (284Prod) from the 'Woodland friends' collection. This trend appears in the association rule model as well. Desk accessories (1251Prod) often bought with notebooks (722Prod) from the 'Paint splash' collection is a good example of this trend. It is interesting to note that Christmas decorations from the 'Wonderland' collection (1065Prod and 805Prod) were popular item sets during this period.

Figure 18: Result for the Rest of the world sub-dataset in Q2

Network model

	Prod1	Prod2	count
30116	2Prod	7Prod	13
70334	354Prod	284Prod	13
29971	2Prod	13Prod	13
38865	6Prod	7Prod	13
47605	11Prod	7Prod	12
30229	2Prod	5Prod	12
30192	2Prod	213Prod	11
33367	327Prod	452Prod	11
74012	13Prod	175Prod	11
38599	6Prod	11Prod	11

Association rule model

	lhs	rhs	support	confidence	lift	count
41161	1065Prod	805Prod	0.040816	1.0	21.777778	8
44575	722Prod	1251Prod	0.040816	1.0	17.818182	8
31746	252Prod	97Prod	0.035714	1.0	14.000000	7
45131	196Prod	175Prod	0.035714	1.0	9.333333	7
34646	73Prod	343Prod	0.030612	1.0	12.250000	6
38956	806Prod	669Prod	0.030612	1.0	32.666667	6
38957	669Prod	806Prod	0.030612	1.0	32.666667	6
40302	554Prod	100Prod	0.030612	1.0	14.000000	6
42318	708Prod	135Prod	0.030612	1.0	21.777778	6
43633	849Prod	956Prod	0.030612	1.0	28.000000	6

In the third quarter (Figure 19), it was observed that frequently-bought item sets in the network model were home/kitchen accessories and bags, while those in the association rule model were home/kitchen accessories and Christmas products. In the network model, it was observed that customers who bought LED lights (20Prod) tended to buy lunch boxes (366Prod and 747Prod), coin purses (996Prod), money boxes (15Prod) or lunch bags (143Prod) together, and preferred the ‘Little llama’ collection. In the association rule model, first, home/kitchen accessories were purchased together, such as different types of rugs (576Prod and 207Prod), cutlery sets (44Prod) with mugs (49Prod), etc. Additionally, Christmas baubles (2382Prod, 1833Prod and 974Prod) of different colours from the ‘Christmas fun’ collection were popular in this quarter and were often purchased together.

Figure 19: Result for the Rest of the world sub-dataset in Q3

Network model

	Prod1	Prod2	count
55191	366Prod	143Prod	60
55172	366Prod	747Prod	60
55078	366Prod	20Prod	59
78139	143Prod	747Prod	57
93080	747Prod	20Prod	57
124586	20Prod	996Prod	56
55198	366Prod	15Prod	56
120907	15Prod	20Prod	55
78164	143Prod	15Prod	55
78053	143Prod	20Prod	55

Association rule model

	lhs	rhs	support	confidence	lift	count
71378	576Prod	207Prod	0.028571	1.0	13.333333	8
74467	44Prod	49Prod	0.028571	1.0	28.000000	8
70887	1857Prod	1591Prod	0.021429	1.0	23.333333	6
72041	1327Prod	1884Prod	0.021429	1.0	31.111111	6
72716	21Prod	49Prod	0.021429	1.0	28.000000	6
44742	121Prod	29Prod	0.017857	1.0	25.454545	5
71996	772Prod	646Prod	0.017857	1.0	17.500000	5
72356	2231Prod	2382Prod	0.017857	1.0	40.000000	5
72451	1833Prod	2382Prod	0.017857	1.0	40.000000	5
72464	1833Prod	974Prod	0.017857	1.0	28.000000	5

In the fourth quarter (Figure 20), items frequently bought together in both models were home/kitchen accessories and bags. The network model found that passport holders (81Prod) were the most frequently-purchased items and were usually bought together with different kinds of mugs (1497Prod, 218Prod and 479Prod), money boxes (15Prod) or lunch bags (130Prod). It also reveals that customers who purchased water bottles (369Prod) bought lunch boxes (515Prod), coin purses (987Prod) or passport holders (1014Prod) at the same time. In the association rule model, however, the most frequently-bought items were photo frames (2183Prod) and were purchased together with different mugs (479Prod, 1497Prod and 218Prod), passport holders (81Prod), or pouches (961Prod).

Figure 20: Result for the Rest of the world sub-dataset in Q4

Network model

	Prod1	Prod2	count
37276	515Prod	369Prod	64
25281	987Prod	369Prod	57
44535	1014Prod	369Prod	57
12594	1497Prod	81Prod	56
9971	218Prod	81Prod	56
51911	15Prod	81Prod	55
9970	218Prod	961Prod	55
25359	987Prod	515Prod	55
35410	479Prod	81Prod	55
4426	130Prod	81Prod	55

Association rule model

	lhs	rhs	support	confidence	lift	count
23273	938Prod	1163Prod	0.162651	1.0	6.036364	54
23649	652Prod	417Prod	0.162651	1.0	5.928571	54
23661	652Prod	515Prod	0.162651	1.0	4.811594	54
23169	546Prod	938Prod	0.159639	1.0	6.148148	53
23173	546Prod	1163Prod	0.159639	1.0	6.036364	53
23130	2183Prod	479Prod	0.156627	1.0	5.724138	52
23132	2183Prod	1497Prod	0.156627	1.0	5.442623	52
23134	2183Prod	218Prod	0.156627	1.0	5.533333	52
23136	2183Prod	81Prod	0.156627	1.0	5.269841	52
23144	2183Prod	961Prod	0.156627	1.0	3.018182	52

4.5. Overall analysis and suggestions

From the four sub-datasets, it is evident that customer preferences in different regions are quite diverse; however, similar characteristics have been observed in customers' purchasing behaviours. The first trend is that different types of items tended to be bought together. For example, LED lights were found to be purchased together with mugs, water bottles or other types of LED lights in Q2 for the Rest of the world dataset. As observed for the Key accounts dataset in Q1, some product sets have specific purposes for purchase, such as luggage tags and shopping bags, which are likely to be bought together for travelling. Nonetheless, most baskets did not provide such clues. As Wilson and Schooler (1991) stated, people often do not know why they like something and analysing their reasons reduces people's satisfaction with their choices. Therefore, although there may be no evident reason why certain products are bought together, the identified rules can still be viewed as hints to bundle these items for sale.

The second phenomenon is that products with the same design are more likely to be bought together. The most obvious instance of this trend is that customers in France preferred to buy kitchen/home accessories that belonged to the same 'Bear camp' collection and with the same 'bear' icon. This suggests that the wholesaler can change the way products presented based on their designs rather than functions. Moreover, customers' purchasing preferences have some seasonal influences. For the Key accounts, the UK and the Rest of the world sub-datasets, customers began purchasing Christmas related products in Q2, indicating that the wholesaler can introduce Christmas item sets bundling products according to identified rules.

Finally, it was observed that products from the same category with different designs tended to be bought at the same time. For example, for the Key accounts sub-dataset, mugs from different collections or in different colours were usually bought together in Q3. These products could be substitutes for each other, meaning that if a product with one design is out of stock, a product with another design could serve as an alternative for the original one. Additionally, these rules could be linked to the profit that specific products provide to increase benefits. To illustrate, supposing that customers who purchase a mug from Collection A are likely to buy it together with a mug from Collections B or C, the wholesaler can give priority to recommending mugs from Collections B or C that are more expensive for customers thereby gaining more profit.

5. Limitations

In this report, several assumptions were made to ensure the analysability of the dataset. First, since there are values provided without details in the wholesaler's dataset, data with missing values have been ignored to ensure the completeness of the dataset. Thus, it is probable that the perspectives elaborated on in this report may have limited universality. Second, to simplify the implementation and analysis processes, the results of the two MBA models were restricted to having only two items per rule. Further study could focus on expanding the number of items in a market basket to obtain more interesting findings.

Moreover, the wholesaler dataset only comprises product-related data and lacks information about the company, its customers and its markets. If the company provides further information regarding the presentation of its products, the sales promotions introduced, etc., further discussion of the MBA results could be had regarding the company's present status. Otherwise, the lack of knowledge regarding the company's customers and the background of its markets may have caused limited perspectives, making the analysis of customers' purchasing behaviour more difficult. For example, understanding the operating structures of customers could help to determine the reasons why they buy certain products together. In the end, having more details regarding the different markets, such as sales status and promotions in the UK and France markets, and the regional components of the Rest of the world sub-dataset could be useful to investigate changes in the rules for different regions and could provide wide-ranging interpretations.

6. Conclusions

Data-driven analysis has become a major technique for obtaining an in-depth understanding the status of an organisation and provides a foundation for decision-makers to develop strategies and make precise predictions. MBA is a data-mining method used to understand customers' purchasing behaviour by inspecting the products that are bought together, which can provide a reasonable basis for sales decisions.

This report applies an MBA algorithm to transaction data provided by a UK wholesaler by constructing two different models. One model was based on network science and involves creating a bipartite invoice-product network and subsequently projecting it onto a one-mode product network to obtain the market baskets of co-purchased products. The second model was based on an association rule mining method that uses the Apriori algorithm to discover market basket rules exceed the thresholds of certain support and confidence levels. Finally, we analysed and compared the 10 most noteworthy market baskets extracted from the four sub-datasets of the wholesaler data to provide suggestions for the company.

Based on the results for each sub-dataset, namely ‘Key accounts’, ‘The UK’, ‘France’ and ‘Rest of the world’ for four quarters, it was observed that market baskets vary between periods and regions, and even between the two models. Nevertheless, there are some findings that could help the wholesaler to boost sales. First, different types of items are likely to be bought together and can be bundled for discounts or promotions. Secondly, different products with similar features tend to be bought at the same time, meaning that the company can present its products based on the similarity of their designs. Additionally, festival-related products follow the identified rule that similar collections must be displayed together and introduced in the appropriate season. Finally, products from same category with different designs can be bundled together as well. Product sets that are consistent with the above rule can act as substitutes and the wholesaler can recommend to customers products that provide the highest profit to establish a competitive advantage.

In conclusion, the above analyses and suggestions can be utilised as reference to improve the wholesaler’s sales strategies. However, there are some limitations associated with the lack of detailed information regarding customers and markets and the restrictions of the MBA models. These drawbacks may limit the comprehensiveness of the findings and more facts may be required for further study. Although this report has its limitations, the wholesaler can still utilise the proposed rules and perceptions to understand its customers' purchasing behaviours before making sales decisions.

References

Agrawal, R., Imieliński, T. and Swami, A. (1993). Mining association rules between sets of items in large databases. *ACM SIGMOD Record*, 22(2), pp.207-216.

Agrawal, R. and Srikant, R. (1994). Fast algorithms for mining association rules in large databases. *Proceedings of the 20th VLDB Conference*, pp.487-499.

Albert, R. and Barabási, A. (2002). Statistical mechanics of complex networks, *Reviews of Modern Physics,* 74(1), pp.47-97.

Borgatti, S. and Everett, M. (1997). Network analysis of 2-mode data. *Social Networks*, 19(3), pp.243-269.

Borgelt, C. (2012). Frequent item set mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(6), pp.437-456.

Kaur, M. and Kang, S. (2016). Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining. *Procedia Computer Science*, 85, pp.78-85.

Raeder, T. and Chawla, N. (2010). Market basket analysis with networks. *Social Network Analysis and Mining*, 1(2), pp.97-113.

Wasserman, S. and Faust, K. (1994). *Social network analysis: methods and applications,* Cambridge University Press, New York.

Wilson, T. and Schooler, J. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology*, 60(2), pp.181-192.

Zhou, T., Ren, J., Medo, M. and Zhang, Y. (2007). Bipartite network projection and personal recommendation. *Physical Review E*, 76(4).