Market Basket Analysis in Polymers Industry: Power BI Case

Ammar Y. ALQAHTANI
King Abdulaziz University, Department of Industrial Engineering, Faculty of Engineering,
P.O. Box 80204, Jeddah, 21589 SAUDI ARABIA
aaylqahtani@kau.edu.sa, +966-12-640-0000 (ext. 69541)
https://orcid.org/0000-0003-3541-9154

Abstract

Nowadays, advanced technology is used in all aspects of business, even small ones. Thus, businesses need to understand the correlations between the products by analyzing consumer behavior throughout the transactions database to enhance the operations process and sales opportunities. This paper's primary objective is to analyze a polymer company's sales database using market basket analysis. The results show that new rules might be applied to control the procurement and sales of goods using association rules as supportive guidance. The correlation between products should be utilized to focus on cross-sales opportunities as it reflects on the revenue stream and attain the highest level of availability for strongly correlated items.

Keywords

Market Basket Analysis; Association Rule Mining; Polymers Industry; Customer Behavior; Operation Strategies.

1. Introduction

Studying the product correlation will help define which two products ordered together more than the others and prevent the delay in supplying them to customers. In addition, it contributes to managing the inventory problems and affects the forecasting of the demand. With advanced technology, businesses can extract useful information from the customer's database to enhance their competitive advantage and customer experience. A well-known tool called Market Basket Analysis (also known as association rule mining) was developed by Agrawal et al., 1993 that is looking for the purchasing pattern of customers in order to design product bundling. For management research, it had been implemented the market basket analysis to study customer behavior (Raorane et al., 2016). For instance, in supermarket transactions, bread and cheese are likely to be purchased together. There are different data mining techniques: classification, clustering, association, prediction, and outlier analysis (Kaur and Shivani, 2016).

Furthermore, it had been discussed different association rule mining algorithms, such as the Apriori Algorithm, Rapid Association Rule Mining (RARM), AIS algorithm, and Frequent Pattern-Tree Algorithm (Zhao and Sourav 2003). The association rule mining approach may apply to many industries where customers tend to purchase a set of products together, such as retails, finance, communication, and many others. Thus, Management needs to identify these product correlations to boost their marketing, sales, and operation strategies (Chen et al., 2005). It has been implemented in the stock market to determine the correlation between stocks through the transaction database and to create a recommender system that helps the investor forecast the stock market price (Paranjape-Voditel and Umesh 2013; Srisawat 2011). Furthermore, in the case of analyzing the transaction data and consumer behavior for the business database, the Apriori Algorithm was proven to be a very useful tool to perform (Kurniawan et al., 2018; Sagin and Berk, 2018). However, in this paper, the association rule mining will be applied in the polymer industry to determine the correlation between a set of products that are more likely to be purchased together to optimize the operation process and sales opportunities.

The rest of the paper is organized as follows: Section 2 describes the methodology of this experiment. Section 3 outlines the results. Finally, section 4 discusses the market basket analysis in the polymer industry.

2. Methodology

In data mining, the classical association rule mining techniques deal with binary attributes; however, real-world data have a variety of attributes (numerical, categorical, Boolean). To deal with the variety of data attributes, the classical association rule mining technique was extended to numerical association rule mining (Kaushik et al., 2021). Initially, the concept of numerical association rule mining started with the discretization method, and later, many other methods, e.g., optimization, distribution are proposed in state-of-the-art. Different authors have presented various algorithms for each numerical association rule mining method; therefore, it is hard to select a suitable algorithm for a numerical association rule mining task (Shahin et al., 2021).

Studying product correlation is one of the project's most important stages, analyzing the product sales using Power PI software to know which two products have demanded more than the others. That will contribute to several stages in the factory processes, such as making a plan for the production line or trading products to prevent the delay in supplying the orders and increasing customer satisfaction, as well as contributing to managing the inventory problems and affecting the demand forecasting.

A strong tool is used to distinguish the correlation between items the customer uses for a certain purpose. Understanding the correlation between items is very important for traders and sellers; for retailers, it helps them to put things together and optimize their space capacity; for pure polymers case, it helps them to know which items have the larger probability of being sold together, to make them available together to avoid lost sales. Analyzing these transactions is known as market basket analysis using the association rule. The method analyzes each transaction and identifies which item is in the transaction, then analyzes thousands of transactions and sees which two or more items are often in one transaction. This could lead to customers' behavior.

Power BI was used to execute the method; Power BI is a business analytics service by Microsoft. It aims to provide interactive visualizations and business intelligence capabilities with an interface simple enough for end-users to create their reports and dashboards. The software provides various methodologies for further statistical analysis, association rule was possible to be implemented by Power BI, but the method will be explained in the methodology section.

The many forms of data in Power BI are retrieved using the get data function from a variety of data sources, including Files, Databases, Azure, Online Services, and others, as described in detail in the data sources section. If the data source isn't a file, we'll need to query it or choose it from a folder, depending on the kind of data we've selected (Yumni and Widowati, 2021). To ensure that the data imported into the Power BI tool is error-free, the uploaded data should be analyzed and corrected using Power BI's edit queries feature in the data section of the tool. We can also create linkages between various datasets using the relationships component. To produce a report, click on the report option, which can be found on the tool's left side menu (Box, 2022). A blank white sheet appears on the screen, allowing you to type in any data you want into the datasets you've already created using the tool's modify queries feature. It is then possible to publish the report to Power BI Services when the data has been thoroughly cleaned and rectified, using a variety of visualization methods including Custom Visuals and standard Power BI visualizations such as stack bar charts and stack column charts. Logging into power bi services, we may change the report, if necessary, but the dashboard is created by publishing on the web (Kettner and Geisler, 2022).

After summarizing the process of manufacturing and trade, it needs to simulate that process so that it can study the problems better and know where the bottleneck is. Therefore, the project team decided to use simulation software called Arena. The Arena is simulation software that allows us to make changes without losing money. Moreover, the software gives the ability to make multiple scenarios and then decide which one is the most suitable. The main purpose

of Arena software is to have a clear image of the process, know what disturbs the sequence of work, and increase the utilization of the resources. On the other hand, doing all these actions without using the simulation software will be very costly.

3. Result and Discussions

Association between products has been established for all items, and only high correlation items were considered because it needs more investigation to determine customer behavior. An example of High correlation items, moderate correlation items, and low correlation items was represented and discussed.

3.1 High Correlation Items

In Figure 1 & Figure 2, a correlation between (Blue Masterbatch 2501) and (Blue 2501 PE1). The second column shows the total transaction number for each item, and the third one shows the number of transactions with both items. The last column shows the percentage of both items' transactions from the total transactions for the item. It shows a correlation percentage of 31% between these items, which means that 30% of the transactions for (Blue 2501 PE1) have (Black Masterbatch 2501).

Item Name	all receipts	both products recept	basket percentage
Black Masterbatch 2501	171	171	1.00
Blue 2501 PE1	83	26	0.31
Yellow 35 PE 1	57	11	0.19
White 50	100	10	0.10
White 38	27	7	0.26
PE 280	81	6	0.07
Beige 45292 PE 11	21	4	0.19
White 6012 PE 11	89	4	0.04
Black Masterbatch 2001	7	3	0.43
Green 271715 PE 11	30	3	0.10
Green 353611 PE 11	18	3	0.17
Red 143595 PE 1	41	3	0.07
Black 40	18	2	0.11
Brown CA15I1616 PE 12	3	2	0.67
Desiccant MB	12	2	0.17
PE 05S	42	2	0.05
RED 143595 E PE 11 - (DELETED)	6	2	0.33
REDDSIH Orange 2056 PE 1	14	2	0.14
Beige 5732 PE 11	12	1	0.08
Black 35	31	1	0.03
Blue 1820 PE 11	2	1	0.50
Blue CA 20 PE 1	1	1	1.00
Bv Blue 12525 PE 11	2	1	0.33
Total	**907**	**171**	**0.19**

Figure 1. Items List for High Correlation.

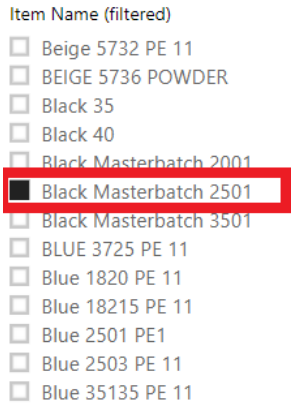

Figure 2. Filtering Tool from Power BI

3.2 Moderate Correlation Items

In Figure 3 & Figure 4, a moderate correlation of 19% was represented between (Blue 2501 PE 1) and (Yellow 35 PE 1). The percentage is not high, but an investigation would drive the company to an answer, and there could be a strong correlation between these items in customer behavior.

Item Name	all receipts	both products recept	basket percentage
Blue 2501 PE1	83	16	0.19
Green 271715 PE 11	30	7	0.23
Red 143595 PE 1	41	5	0.12
White 38	27	5	0.19
White 50	100	5	0.05
Beige 45292 PE 11	21	4	0.19
Green 353611 PE 11	18	4	0.22
Black 35	31	3	0.10
Green 252026 PE 11	6	3	0.50
White 6012 PE 11	89	3	0.03
Blue 35135 PE 11	2	2	1.00
Green 9595 PE 11	4	2	0.50
REDDSIH Orange 2056 PE 1	14	2	0.14
Sky Blue 4564 PE 11	6	2	0.33
Beige 3035 PE 11	4	1	0.25
Blue 3526 PE 11	1	1	1.00
Brown 151616 PE 11	2	1	0.50
Brown CA15I1616 PE 12	3	1	0.33
Bx Blue 13535 PE 11	3	1	0.33
Dark Pink 401416 PE 11	1	1	1.00
Total	**907**	**57**	**0.06**

Figure 3. Items List for Moderate Correlation.

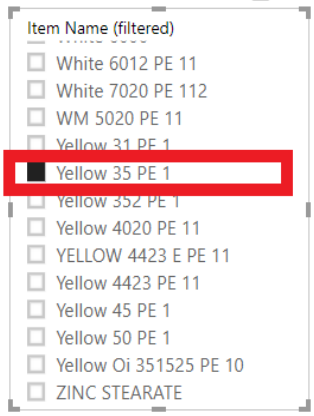

Figure 4. Filtering Tool from Power BI.

3.3 Low Correlation Items

In Figure 5 & Figure 6, it is shown that (PE 175) and (PE 280) has a correlation rate of 0.04, which could be neglected because it is a too-small value. And obviously, there is no correlation between buying these two items together.

Item Name	all receipts	both products recept	basket percentage
PE 175	8	8	1.00
PE 280	81	3	0.04
White 6000	20	2	0.10
Blue 2501 PE1	83	1	0.01
Yellow 4423 PE 11	19	1	0.05
02 PE	1		
80 PPE	1		
Addiblend 66-5	20		
Addiblock 05-3	1		
Addiclear 59	1		
ANTL Block 1694 PE 11	1		
Beige 3035 PE 11	4		
Beige 3530 PE 11	1		
BEIGE 45292 E PE 11	2		
Beige 45292 PE 11	21		
Beige 51922 PE 11	1		
BEIGE 5732 POWDER	3		
Beige 5732 PE 11	12		
BEIGE 5736 POWDER	1		
Black 35	31		
Black 40	18		
Black Masterbatch 2001	7		
Black Masterbatch 2501	171		
Total	**907**	**8**	**0.01**

Figure 5. Item List For Low Correlation.

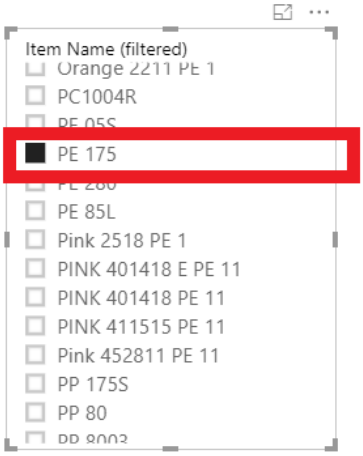

Figure 6. Filtering Item Form Power BI.

3.4 Adding and Removing Correlated Products

When thinking about listing or delisting a new or an existing item. It is very important to know if items are correlated or not. If the items are correlated, then the company should take care that the delisting of an item could affect the demand for another item with a strong correlation in buying decisions. For slow-moving items, it could be offered with correlated fast-moving ones to increase the sales rate for the poor performance one.

4. Conclusions

Market Basket Analysis also known as association rule learning or affinity analysis, is a data mining technique that can be used in various fields, such as marketing, bioinformatics, education field, nuclear science etc. The main aim of Market Basket Analysis in marketing is to provide the information to the retailer to understand the purchase behavior of the buyer, which can help the retailer in correct decision making. There are various algorithms are available for performing Market Basket Analysis. The existing algorithms work on static data, and they do not capture changes in data with time. But proposed algorithm not only mine static data but also provides a new way to take into account changes happening in data. This paper discusses the data mining technique i.e., association rule mining and provide a new algorithm which may be helpful to examine the customer behavior and assists in increasing the sales.

The generated consumer purchasing behavior from the system may set new rules about product relationships. The analysis implemented in this paper might be applied to different businesses and sales systems to control the procurement activities as well as the sales of goods using association rules as supportive guidance. Correlation between items should be considered a revenue stream to focus on cross-sales opportunities and attain the highest level of availability for strongly correlated items.

References

Agrawal, Rakesh, Tomasz Imieliński, and Arun Swami. "Mining association rules between sets of items in large databases." In *Proceedings of the 1993 ACM SIGMOD international conference on Management of data*, pp. 207-216. (1993).

Box, Alison. "Integrating with Power BI." *Introducing Charticulator for Power BI*. Apress, Berkeley, CA, (2022). 351-362.

Chen, Yen-Liang, Kwei Tang, Ren-Jie Shen, and Ya-Han Hu. "Market basket analysis in a multiple store environment." *Decision support systems* 40, no. 2 (2005): 339-354.

Kaur, Manpreet, and Shivani Kang. "Market Basket Analysis: Identify the changing trends of market data using association rule mining." *Procedia computer science* 85, no. Cms (2016): 78-85.

Kaushik, M., Sharma, R., Peious, S.A., Shahin, M., Yahia, S.B. and Draheim, D. "A systematic assessment of numerical association rule mining methods." *SN Computer Science* 2.5 (2021): 1-13.

Kettner, Benjamin, and Frank Geisler. "Power BI." *Pro Serverless Data Handling with Microsoft Azure*. Apress, Berkeley, CA, (2022). 169-191.

Kurniawan, Fachrul, Binti Umayah, Jihad Hammad, Supeno Mardi Susiki Nugroho, and Mochammad Hariadi. "Market Basket Analysis to identify customer behaviours by way of transaction data." *Knowledge Engineering and Data Science*1, no. 1 (2018): 20.

Paranjape-Voditel, Preeti, and Umesh Deshpande. "A stock market portfolio recommender system based on association rule mining." *Applied Soft Computing* 13, no. 2 (2013): 1055-1063.

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

Sagin, Ayse Nur, and Berk Ayvaz. "Determination of Association Rules with Market Basket Analysis: Application in the Retail Sector." *Southeast Europe Journal of Soft Computing* 7, no. 1 (2018).

Shahin, M., Arakkal Peious, S., Sharma, R., Kaushik, M., Ben Yahia, S., Shah, S.A. and Draheim, D. "Big data analytics in association rule mining: A systematic literature review." 20*21 the 3rd International Conference on Big Data Engineering and Technology (BDET)*. (2021).

Srisawat, Anantaporn. "An application of association rule mining based on stock market." In *The 3rd International Conference on Data Mining and Intelligent Information Technology Applications*, pp. 259-262. IEEE, (2011).

Yumni, Shabrina Zata, and Widowati Widowati. "Implementasi Microsoft Power BI Dalam Memantau Kehadiran dan Transportasi Pegawai." *Jurnal Sains dan Edukasi Sains* 4.1 (2021): 1-8.

Zhao, Qiankun, and Sourav S. Bhowmick. "Association rule mining: A survey." *Nanyang Technological University, Singapore* (2003): 135.

Biography

Ammar Y. Alqahtani, Ph.D., is an associate professor of Industrial Engineering at King Abdulaziz University in Jeddah, Saudi Arabia. He received his BS degree with first honors from the Industrial Engineering Department of King Abdulaziz University, Jeddah, Saudi Arabia, in May 2008. Being awarded a full scholarship by King Abdulaziz University (KAU), he received his MS degree in Industrial Engineering from Cullen College of Engineering, University of Houston. In September 2012, he started his Ph.D. studies in Industrial Engineering at Northeastern University, Boston, Massachusetts. He received his Ph.D. degree in 2017. He has been employed as a faculty member by King Abdulaziz University since December 2008. His research interests are in the areas of environmentally conscious manufacturing, product recovery, reverse logistics, closed-loop supply chains (CLSC), sustainable operations and sustainability, simulation and statistical analysis, and modeling with applications in CLSC and multiple life-cycle products. He has published two books, titled *Warranty and Preventive Maintenance for Remanufactured Products Modeling & Analysis* and *Responsible Manufacturing Issues Pertaining to Sustainability*. He has co-authored several technical papers published in edited books, journals, and international conference proceedings. At Northeastern University, he won the Alfred J. Ferretti research award. He also received the 33rd Quality.