

Smart Retail through Market Basket Analysis

Mohamed Minhaj

Operating in a southern city of India, Shaanz is a home décor retail outlet which was started in 2015 by Shaan after quitting his high paying job in a reputed IT Services Company. The outlet offer hundreds of products related to home décor. The wide range of products and competitive prices coupled with quality customer services have helped Shaanz to retain its position in the fray of home décor business. In early 2018, the store replaced its manual process with automated billing system. As part of this implementation, the store also introduced loyalty cards. The objective of the loyalty cards scheme was to entice customers with discounts for repeat purchases and boost sales, but it met with limited success. Nevertheless, coming from an IT background, Shaan firmly believed in the value of data and wanted to leverage on it. While fifteen months transaction data was available, how to derive value from this trove of business insights was a big question before Shaan.

About the Company

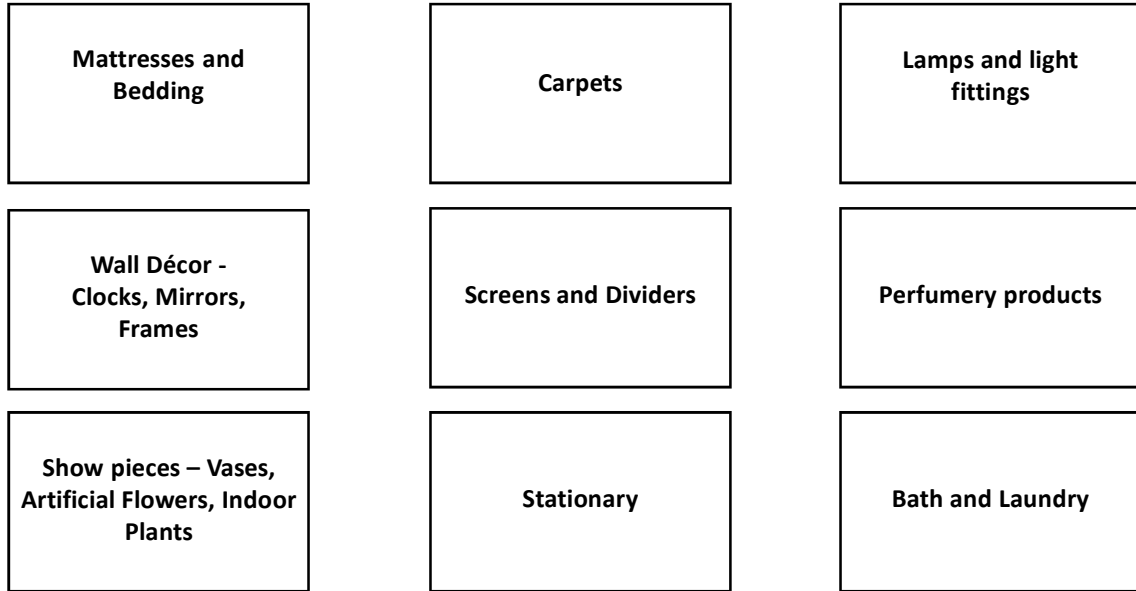
After completing B.E from a top ranked college, Shaan took up a job in an IT services company and in span of three years was promoted as Project Manager. While he was happy with his job both in term of remuneration and professional growth, like many other young executives, Shaan 's desire to be self-employed drove him to the path of entrepreneurship. Using his three year's savings coupled with a bank loan, Shaan was successful in establishing a home décor retail outlet in a rented place. Shaan was concerned about the poor response to his offerings in the first year, but the word of mouth publicity about Shaanz collection, competitive prices and quality customer services facilitated in garnering many customers in the subsequent years. With over thirty different makers and suppliers, it soon became one of the popular destinations for home décor requirements in the region.



Copyright © 2019 Shri Dharmasthala Manjunatheshwara Research Centre for Management Studies (SDMRCMS), SDMIMD, Mysore. This case is published as a part of 'Cases in Management Volume 8 (2019)' with ISBN 978-93-83302-44-4.

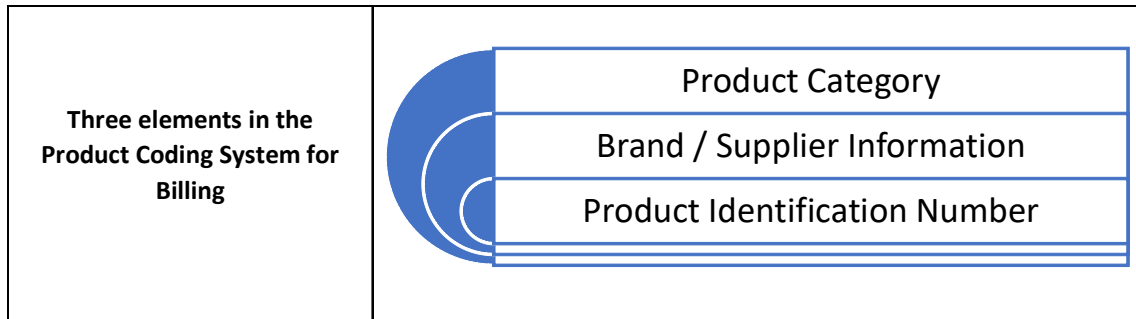
The case writer(s) Mohamed Minhaj, Associate Professor - Systems, SDMIMD, Mysore may be reached at mminhaj@sdmimd.ac.in Author(s) have prepared this case as the basis for class discussion rather than to illustrate either effective or ineffective handling of the situation. The case is related to a sector and a specific company. The case is based on the information available in the public domain. This publication may not be digitized, photocopied, or otherwise reproduced, posted, or transmitted, without the permission of SDMRCMS, SDMIMD, Mysore. For Teaching Notes please contact sdmrcms@sdmimd.ac.in.

Product Categories offered by the Company



“Data Rich, Information Poor” state of the Company

Considering the increase in the foot fall to the store, Shaan embarked on automated billing process in 2018. The objective of implementing a computerized billing process was to reduce the time involved in the billing process. The implementation involved devising a coding system for each product offered in the store, and some mechanism to capture the code of the product at the billing counter with limited intervention from the user. With reference to the coding system, after several deliberations alphanumeric codes were finalized, comprising of three elements, namely – The Product Category, Brand / Supplier Information and a unique identification number. As far as mechanism to capture the Product Code with lesser manual intervention is concerned, easy to implement and inexpensive, Bar Coding System was implemented.



In one of the brainstorming meetings to discuss the implementation of the bar codes and automated billing process, Varun, a senior employee of the store expressed his happiness about the developments but voiced his concern that there was nothing being done about building an element of loyalty with the customers. The other employees also backed Varun's views and there was a collective opinion that introduction of some loyalty scheme was necessary to ensure repeat business, particularly because two more stores in the same vicinity had come up recently. Based on this discussion, it was decided to introduce Loyalty Cards. Under the Loyalty Card Program, the customers were offered discounts, coupons and loyalty point benefits for repeat business.

The loyalty program involved one-time collection of Customer Information, after which unique customer code were assigned. These codes printed on the customer loyalty cards facilitated tagging every transaction with the customer code.

Customers data used for Loyalty Program	
Customer Code	Primary Key
Customer Name	Mandatory
Mobile Number	Mandatory
E-Mail	Optional
Address	Optional
Preferred Payment Method – Card, Cash, eWallet	Optional
Date of Birth	Optional
Favorite Color	Optional

With the implementation of the automated billing system and loyalty program, the store started accumulation a huge amount of data. The billing data was being used for accounting and inventory related analysis and decision making. But the data related to loyalty program did not find any real use in the company. In fact, the loyalty program met with a limited success and did not help in giving impetus to sales. While several resources of the company were being employed for facilitating the loyal program, the company was not getting benefitted in any way and was just accumulating a huge amount of data. The transaction data pertaining to fifteen months was available without resulting in any information or any business insights. The store was in true sense, "Rich in data and poor in information". However, coming from an IT background, Shaan firmly believed in the value of data and explored different avenues of data analytics and finally decided to embark on Market Basket Analysis.

What is Market Basket Analysis ?

Market Basket Analysis, also known as association rule mining or affinity analysis, is one of the techniques used in data mining. Data mining refers to the process of exploring large data sets to discover patterns and extract meaningful information. Frawley define data mining as the process of extracting hidden, previously unknown and potentially useful information from data(Frawley, Piatetsky-Shapiro, & Matheus, 1991). Hand et al., use the term to refer to the process of extracting information from large data sets or databases(Hand, Smyth, & Mannila, 2001). According to Cabena, data mining is an interdisciplinary field bringing together techniques from machine learning, pattern recognition, statistics, databases, and visualization to address the issue of information extraction from large data bases (Cabena, Hadjinian, Stadler, Verhees, & Zanasi, 1998). Data mining basically deals with techniques for data exploration, such as classification, clustering, regression etc.

Market Basket Analysis employs methods based on association rules to discover customer purchasing patterns by extracting associations or co-occurrences from store's transactional databases. Discovering, for example, that supermarket customers are likely to purchase milk and bread or that bank customers are likely to use a set of services jointly, can help managers in designing store layout, web sites, product mix and bundling, and other marketing strategies. The methodology was introduced by Agrawal et al. (Agrawal, Imieliński, & Swami, 1993) and can be stated as follows. Given two nonoverlapping subsets of product items, X and Y, an association rule in form of $X \rightarrow Y$ indicates a purchase pattern that if a customer purchases X then he or she also purchases Y.

Market Basket Analysis is one of the key techniques used by retailers to uncover associations between items. It works by looking for combinations of items that occur together frequently in transactions. It allows retailers to identify relationships between the items that people buy using measures of interestingness, based on the concept of strong rules(Li, 2017). One of the popular algorithms used for mining association rules is Apriori. This algorithm proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The association rules computed by Apriori and other algorithms are in the form of if-then statements, such as $A \rightarrow B$, where A and B are sets of the so-called attributes. In an association rule, A is an antecedent and B is a consequent. For association rules like $A \rightarrow B$, it is possible to define three important measures of significance and interest: support, confidence and lift(Berry & Linoff, 2004).

Support = Frequency (X, Y) / N	Support is the number of transactions with both X and Y divided by the total number of transactions.
Confidence (X => Y) = Support (X U Y) / Support (X)	The Confidence value of a rule, X => Y, with respect to a set of transactions T, is the proportion of the transactions that contains X, which also contains Y.
Lift (X => Y) = Support (X U Y) / Support (X) x Support (Y)	The Lift of a rule, X => Y, is the ratio of the observed support to that expected if X and Y were independent.

Significance of "Lift" in a Retail Setup

While "Support" is an indication of how frequently the item sets appear in the market basket, "Confidence" is an indication of how often the association rule has been found to be true. However, "Lift" is probably the most commonly used measure in Market Basket Analysis. As greater lift values indicate stronger associations, the concept of "Lift" enables the retailers to easily identify combinations of items that tend to be purchased together. The insights from Market Basket Analysis and specifically, "Lift" can be valuable for retailers and can help in devising strategies to maximize their profit. Besides, promotion of cross selling and product bundling strategies, Lift can also be instrumental in store's layout design.

Shaanx Approach to Market Basket Analysis

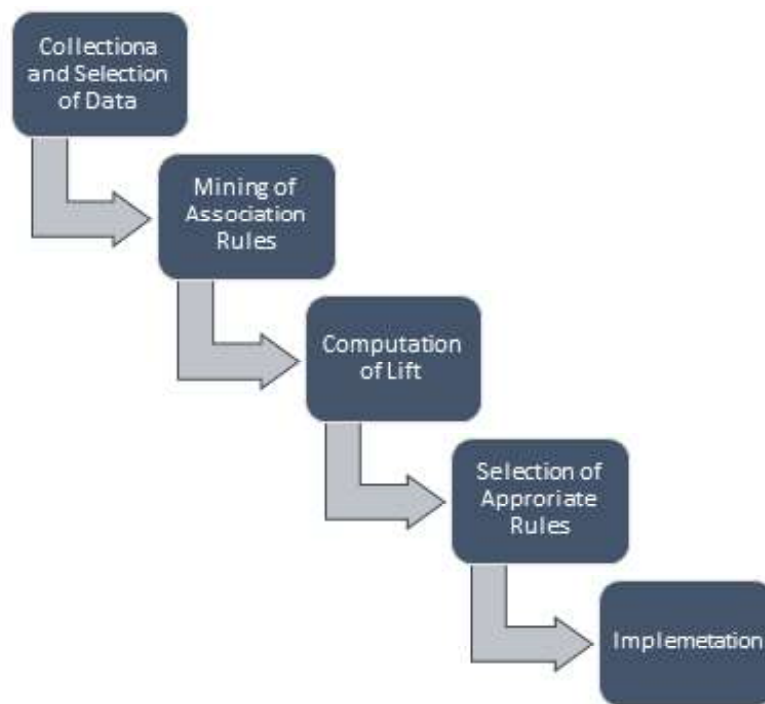
While the company had maintained the paper-records of all the transactions since its inception, it is the implementation of computer-based billing system with bar-codes in 2018, that facilitated the storage of all the customer transactions in an efficient and effective manner. The introduction of loyalty program further aided in capturing new dimensions pertaining to the transactions like the demographic parameters. These parameters were tagged with every transaction indirectly with help of the unique customer number printed on the loyalty cards. With more and more details about the customer transactions getting captured in the organization, Shaanz became really rich in data. However, when it comes to harvesting and harnessing the data that is generated by different processes / technologies used in the store, the outcome was very poor. Except for the regular accounting and inventory related reports, much of the data stored was unused. Therefore, the main objective of embarking on Market Basket Analysis was to transform the unused data into actionable insights, which could be used in decision making related to different

174 • CASES IN MANAGEMENT

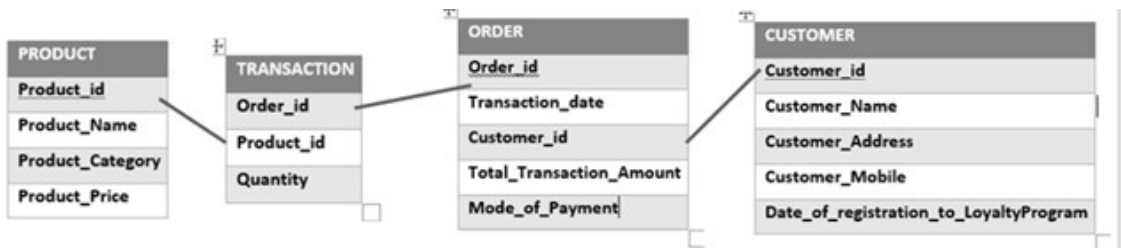
aspects of the retail operations. The key aspects considered while approaching the project were:

- From where to generate the association rules ?
- What techniques to use to generate rules ?
- What metrics are needed to differentiate strong rules vs. weak rules ?

The project involved the following phases:



From the database that had fifteen months transactional data, four tables - ORDER, TRANSACTION, PRODUCT and CUSTOMER were selected for the required analysis. While PRODUCT and CUSTOMER were master tables storing details pertaining to all the products and customers respectively, the ORDER table was the primary table that captured all the orders with details of the products purchased in every order being stored in TRANSACTION table.



For the purpose of mining association rules, after extracting appropriate data from the four tables available for analysis, suitable transformation was performed and market basket data was prepared as per the structure given below :

	A	B	C	D	E	F	G	H	I	J	K
1	Order#	Prod 1	Prod 2	Prod 3	Prod 4	Prod 5	Prod 6	Prod 7	Prod 8	Prod 9	Prod 10
2	1	0	1	1	0	0	0	0	1	0	0
3	2	0	1	0	0	1	0	0	0	0	1
4	3	1	0	0	1	0	0	0	0	1	0
5	4	1	0	0	0	0	1	0	0	1	0
6	5	0	0	0	1	0	0	1	0	0	0
7	6	0	1	0	0	1	0	0	1	0	1
8	7	1	0	0	1	0	0	0	1	0	0
9	8	0	1	0	0	1	0	0	0	1	0
10	9	1	0	0	1	0	0	1	0	0	0
11	10	0	1	0	0	1	0	0	0	1	0

In the above table, while Prod 1, Prod 2 etc., are product ids of different products offered by the home décor agency, the presence of 0 or 1 in the table indicates the absence or presence of a product in a particular market basket (order). The quantity of the product purchased was ignored in the analysis.

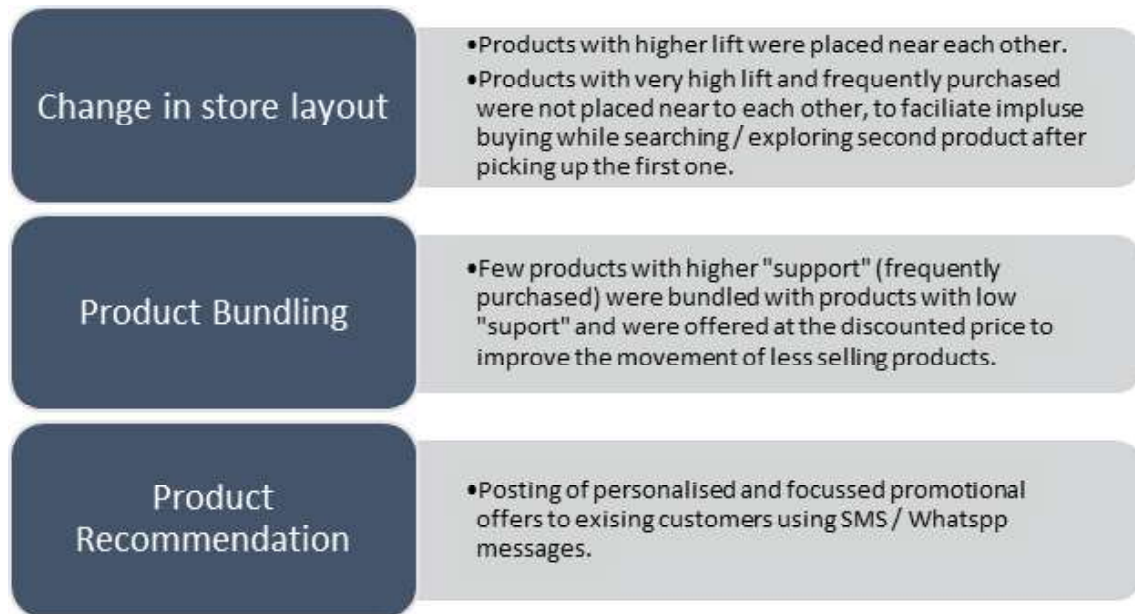
Considering the huge size of the market basket table, appropriate techniques were iteratively employed to optimize the table (reduction in rows and columns). The table after optimization was used to generate the association rules.

Example : Carpet, Screen => Vase.

Further, in an endeavor to differentiate strong and weak association rules, "Lift" between all the frequently purchased products (With high "Support") was computed. The product combinations with "Lift" much greater than 1 were used for understanding the buying patterns.

Use of strong association rules for boosting sales

The insights derived from the Market Basket Analysis in the form of association / affinity rules between frequently purchased products (with high "Lift") were used in multiple ways in the home décor store. While few strategies devised using the association rules were useful in improving the customer experience within the store, other strategies were focused on creating a pull advertising effect.



Is “Lift”, really providing Lift to the Shaanz’s Business ?

Strategic location, good product offerings and competitive pricing, Shaanz had everything that is needed to be successful in a home décor retail business. But for it to continue to be relevant in the market and to constantly entice customers, it is important to improve its customer services and loyalty. Particularly with two more new players in the vicinity and unremitting competition from the online players, it would be challenging for the company to make its presence felt in the market. Under this backdrop, the use of Market Basket Analysis and specifically finding the “Lift” between the products appears to be promising. The three months sales data, post implementation of new strategies based on the data analysis, indicate value addition in terms of foot fall and customer feedback. However, there has not been significant change in the sales records. While time will tell the real impact of changes and the decisions taken based on the customer data, as of now, there are several questions that Shaanz is grappling with:

- The changes made in the store layout and bundling strategies were based on the Market Basket Analysis of fifteen months historical transaction data. Is this a onetime task or continuous activity. In case it is continuous, what should be frequency of performing this analysis?

- Should the association rules be built based on the individual products or the product categories. In case it is individual products, what should be the approach to handle the huge data analysis?
- With the quantity purchased and price of the product missing in the analysis, are the association rules generated, statistically right ? and consequently is "Lift" a good measure to differentiate between the strong and weak rules ?

References

- Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining Association Rules Between Sets of Items in Large Databases. *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, 207-216. <https://doi.org/10.1145/170035.170072>
- Berry, M. J. A., & Linoff, G. S. (2004). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. John Wiley & Sons.
- Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., & Zanasi, A. (1998). *Discovering Data Mining: From Concept to Implementation*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc.
- Frawley, W. J., Piatetsky-Shapiro, G., & Matheus, C. J. (1991). Knowledge Discovery in Databases: An Overview. *Knowledge Discovery in Databases*. <https://doi.org/10.1609/aimag.v13i3.1011>
- Hand, D. J., Smyth, P., & Mannila, H. (2001). *Principles of Data Mining*. Cambridge, MA, USA: MIT Press.
- Li, S. (2017, September 27). A Gentle Introduction on Market Basket Analysis-Association Rules. Retrieved August 13, 2019, from Medium website: <https://towardsdatascience.com/a-gentle-introduction-on-market-basket-analysis-association-rules-fa4b986a40ce>

