

## **MARKET BASKET ANALYSIS FOR A LIFESTYLE STORE, FOCUSING ON THE PURCHASING PATTERNS OF CUSTOMERS**

**Jagruti Sandeep Gijare**

Research Scholar

Department of Management

Sunrise University, Alwar, Rajasthan.

**Dr. Anish Kumar**

Professor

Department of Management

Sunrise University, Alwar, Rajasthan.

### **ABSTRACT**

*The process of identifying connections between various commodities is called market basket analysis. By examining and projecting daily sales of each particular product category, the proposed method may be able to predict the items that have the highest likelihood of being sold. The study area is Mumbai. The typical data set for a shop is one year long. The retail software was used by the business to provide main data. It uses the Microsoft Associative algorithm and the Business Intelligence Development Studio application to produce predictions. That kind of study is called exploratory research. The results might be a helpful tool for creating sales campaigns, upselling, cross-selling, and carefully positioning items within the store.*

*Keywords: Market Basket Analysis, Customers*

### **Introduction**

Market basket analysis is the process of analyzing goods that are purchased individually or in combination in one or more transactions. The relationships between the different components are clarified by this examination of customer behavior. As a result, the strength of these relationships provides valuable information for couponing, up-selling, cross-selling, and other recommendations. A modeling technique called market basket analysis is predicated on the notion that, after purchasing a certain set of goods or contents, a customer is likely to buy a different set of goods or contents. For example, if a customer buys tea plus a bag, he's presumably buying a keepsake. The collection of products a customer buys is called an item set, and market basket analysis may be used to find relationships between item sets. By applying market basket analysis, a retailer may discover what a customer could think about adding to his basket.

### **Literature Review**

Data mining has grown in importance as a subject in marketing literature throughout the last several decades. One of the first uses of data mining is market basket research, which provides the best example of mining association rules. To help users accomplish their objectives, researchers have developed a range of Association Rule Mining and Clustering approaches. Among them are pioneers of data mining, Rakesh Agrawal and Usama Fayyad. They examine several well-established procedures and methods. Rule-generating methods are classified by Shapiro into two groups: those that find quantitative rules and those that find qualitative rules.

Using information from a significant retail company, William Frawley examined the algorithm's effectiveness in 1991. Gregory Piatetsky-Shapiro had first devised it. To extract qualitative rules for categorical data using attributes, they used association rules. Many research domains, such as advertising, recommendation systems, market basket analysis, and strategy development, employ association rules. The marketing literature has separated market basket analysis into two models: explanatory and exploratory. The main idea behind exploratory models is to identify purchasing patterns in point-of-sale data. Exploratory

procedures do not include information on consumer demographics or marketing mix components; instead, they distill a big amount of data into a smaller number of important rules or metrics. These methods are very useful in determining relationships between the data's pieces that were not previously known. These methods may be used to undirected data mining and are also easily computed. However, exploratory approaches are not appropriate for determining or forecasting the fundamental causes of complex problems. Put simply, they are used in the identification of discrete cross-category interdependencies for items or product categories that are purchased in tandem, depending on certain common patterns. These exploratory approaches are often used to investigate relationships between product categories using simple association measures. When evaluating objects in pairs, pairwise associations are utilized to identify which is more desirable or has a higher amount of a certain quantitative quality.

Julander (1992) compared the percentage of customers who buy a certain product to the percentage of total revenues that this product brings in. It is easy to ascertain which products are in demand and what proportion of sales they represent thanks to these comparisons. Since a lot of customers deal with these specific product categories on a regular basis, it's important to look at what the best products are for customers. Since these departments get the greatest traffic in the business, it is imperative to use this knowledge to position additional specialty goods close to them. One of the main areas of research in the field of exploratory analysis is the process of developing association rules.

Few algorithms have been developed that do not depend on the Apriori, according to the results of M.J. Zaki, M. Ogihara, and S. Parthasarathy (1996). For example, Nick Cercone, Howard J. Hamilton, Colin L. Carter, and Robert J. Hilderman established a paradigm for knowledge discovery utilizing market basket data. Attribute Oriented Generalization and Apriori were used to create the methodology. The method was used to explain how to identify trends in consumer purchases and create customer profiles via customer classification. Two essential components of Association Rule Mining (ARM) are support and trust, according to Sourav S. Bhowmick and Qiankun Zhao (2003). They both assess the power of an association rule. Many data scientists have developed their own methods, such as employing the associative algorithm to categorize customers into heavy, moderate, and light users. Customers with greater incomes who visit the firm less often and are less price sensitive likely to be heavy users. On the other hand, frequent shoppers or exceptionally frugal students are usually considered light users. The results showed that families that behave consistently across all product categories tend to be lighter consumers than households that behave inconsistently. Furthermore, it is believed that families with comparable habits are less brand loyal, more price conscious, and less open to in-store ads.

In 1997, Gary J. Russell and Wagner A. Kamakura showed that brand preference segmentation could be determined even in the absence of marketing mix data, using long-run market basket data. The estimate of these cross-category preferences is contingent upon the acceptance of several simplifying assumptions. Exploratory models are useful for finding cross-category correlations, but they are not useful for figuring out why they exist. While exploratory market basket research aims to reveal and illustrate hidden relationships between product categories, explanatory models focus on explaining effects. The datasets employed in these models include market basket data, customer attributes, and marketing mix

components.

Andreas Mild and Thomas Reutterer (2003) state that the purpose of explanatory models is to identify and quantify the effects of marketing factors such as price, promotion, and other marketing aspects on cross-category decision making. Regression analysis, logit probit, and multivariate logistic models are the main sources of quantitative data utilized in explanatory models, according to Andreas Mild and Thomas Reutterer (2003). Retailers and managers may utilize household data mining in conjunction with transactional data mining to generate customized target marketing strategies. Supermarkets may cater to a wide spectrum of consumers by using astute pricing techniques via the examination of past transactions. By leveraging the shopping basket, companies may get consumer-focused data instead of using individual goods as the unit of research.

Rakesh Agrawal and Ramakrishnan Srikant expanded on this concept by developing the notion that large relational tables may be mined for quantitative rules. Quantitative rules are constructed using the attributes that are present in these relational databases. Features might be quantitative, like the customer's age, income, etc., or categorical, like the kind of items or car model. Boolean attributes are those that can have two possible values. They are thought to be a singular example of a categorical trait. The authors have named this mining challenge the Quantitative Association Rules problem.

In 2011, R.M.S. Parvathi and S. Prakash proposed a method for qualitatively mining quantitative association rules. The technique is qualitative in character since it converts numerical features into binary traits. Still, determining qualitative norms is the main objective of their research. Dependency tables, decision trees, and patterns are the most often used visual aids for representing these concepts.

### **Objective of the Study**

Retailers need to be able to forecast future customer behavior in addition to understanding current consumer behavior in order to remain competitive. Retailers may increase revenue, strengthen customer relationships, and retain consumers by having a better understanding of customer behavior.

1. To comprehend the purchasing habits of the items that make up the consumers' basket
2. To research the goods that consumers are most likely to have bought.
3. To research a certain product category as well as the most probable items that the buyers bought
4. To forecast and recommend items to specific customers.

### **Research Methodology**

Transactional data obtained from the customers' billing served as the main source of input. Because it comes from a legitimate source of customer purchasing behavior, the data is trustworthy. Retailer software such as Tally Shopper Version 9 was used to extract the data, and it was then linked to the retail store's database. The researcher linked the database to Microsoft tools, including Business Intelligence Development Studio on Server 2008 R2, using structured query language and categorical data from transaction records. The main environment used to construct business solutions, such as Analysis Services and Integration Services, as well as to create the objects needed for business intelligence solutions, is the Business Intelligence Development Studio. The data represents an average over a period of one year for four lifestyle retail establishments located in Mumbai, India. Association Rules, a cutting-edge data mining technique, is used to analyze the data. Using this transactional

data, market basket scenarios are constructed to forecast which goods are most likely to be added by a client to their basket.

### **Building a Market Basket Scenario**

The capacity to make predictions about what a client would wish to buy based on what else is currently in their shopping cart. For this kind of market basket study, data mining is quite helpful, and a data mining model is created. Groups of products from previous client transactions are shown using a mining model. Furthermore, a mining model may be used to forecast potential extra purchases that a consumer could make.

A data source must be constructed before a data source view including customer tables can be added. This process is done in the Business Intelligence Development Software on Server 2008 R2 in order to build the mining model needed to do market basket analysis. For market basket situations, a mining model applies the Microsoft Association Rules algorithm.

The stages for doing a Market Basket Analysis are as follows:

1. Using Nested Tables to Add a Data Source View
2. Developing a Model and Structure for a Market Basket
3. Adapting and Using the Market Basket Framework
4. Examining Models of Market Baskets
5. Assuming Correlations

### **Adding a Data Source View with Nested Tables**

A data source view is made in order to construct the mining model for market basket analysis. A table with many rows of data concerning a single row in the case table is called a nested table. There is a nested table in this data source. For instance, we would normally use a database with a distinct entry for every client as the case table if our model examined how consumers made purchases. Every consumer, however, may make a number of purchases and may choose to examine the things that they commonly buy together or the order in which they are made. We add another table to the data source view that displays the purchases for each client in order to correctly represent these purchases in our model.

There is a many-to-one link between the customer table and this nested purchases table. For every client, the nested table may include several rows, each holding a single product that was bought, along with maybe further details on the order that the purchases were made, the price at the time of the order, or any applicable specials.

Thus, in this instance, the case table—which includes the Document Number of the goods the customer purchased—as well as the nested table—which includes the Category, Subcategory, and Description of the goods each client purchased—are utilized. The tables are related in a many-to-one manner using the customer bill's Transaction Control Number feature, as shown in screenshot 1.

### **Creating a Market Basket Structure and Model**

A new mining structure is made using the data source view that is produced. A Microsoft Association algorithm-based mining model and a mining structure are developed. It's a method that helps determine which products, in accordance with association regulations, are most likely to be bought jointly. Direct marketers are interested in sets of item purchases that often happen in a certain order. These sets may be found using a predictive market basket analysis. Finding links between goods or making product recommendations to customers are only two uses for associative prediction. An item set is a collection of things in a case. A

collection of item sets and the rules defining how those sets are arranged inside the cases make up an association model. Based on the products that are now in the user's shopping cart, the rules that the algorithm finds may be utilized to forecast the probable future purchases of the consumer.

### **Requirements**

Two essential columns and one predictable column are required for an association model. Variables in the Market Basket Analysis:

1. The study's independent variable was the following:
  2. Doc No: The product's document number that clients have bought
  3. The dependent variable taken into account in this research:
  4. Subcategory: A lifestyle store's product subcategory.
- For instance: As shown in screenshot1, bodycare, haircare, etc.

### **Modifying and Processing the Market Basket Model**

Two of the parameters, Support and Probability, have their default values adjusted as needed before the association mining model is performed.

1. Support establishes the proportion of situations that a rule needs to exist in order to be deemed legitimate. Here, it is stipulated that a rule must be present in 1% or more of the situations.
2. Probability establishes the level of likelihood required for a relationship to be deemed legitimate. Any relationship in this instance that has a probability of at least 10% is to be taken into account.
3. 0.1 is the minimum probability.
4. A MINIMUM OF 0.01 SUPPORT

Adjustments are made to the likelihood and support values. As a result, the Association mining model's structure and parameters are established, and the model is then processed.

### **Exploring the Market Basket Models**

The connection will usually take the shape of a rule: The likelihood that a buyer would purchase wall décor without a memento, for instance, is known as the rule's support. The confidence is the conditional likelihood that a client will buy a body care product.

Once the Association model is constructed, it can be examined using the viewer's rulers tab, which allows us to quickly see which goods often occur together and get a broad understanding of the developing patterns.

Using Microsoft Association Viewer on the Mining Model Viewer Rule tab, one may investigate the association model. This makes it easier to investigate connections between goods, see which ones often show up together, and get a broad understanding of the patterns that are starting to emerge.

The following details are connected to the rules that the algorithm discovers and are shown on the Rules page.

- Rule: The rule's definition. A rule for a market basket model specifies a particular set of products.
- Probability: a rule's possibility, expressed as the right-hand item's probability given the left-hand item.
- Importance: A metric used to assess a rule's utility. An improved rule has a higher value.

Based on the existence of other things, each rule may be used to anticipate whether an item will be included in a transaction. A rule's importance is indicated to assist in determining its

utility since probability on its own might be deceiving. For instance, if a tea is included in every transaction (maybe automatically added to each customer's cart as part of a promotion), the model would provide a rule indicating that the likelihood of tea being a transaction would be 1. This rule is extremely accurate based just on probability; nonetheless, it does not give meaningful information.

### **Predicting Associations**

Predictions are made using the association data that is kept in the models after their processing. The association models serve as the foundation for prediction queries. Associative prediction serves a variety of functions, such as making product recommendations to customers and identifying product associations.

The association model is chosen and the input data is given in terms of the table, that is, the parameter or variable to be used as input and the parameter or variable for which the specific output has to be created, in order to construct a prediction query. Thus, several queries for singleton prediction are made. Next, a batch prediction query is developed so that suggestions based on a customer's recent purchases may be made.

### **Findings**

To Recognize the Product Purchase Trends of the Items in the Customers' Basket Screenshot 2 displays the Rules tab, which provides an overview of developing trends, identifies which particular elements are more closely connected to one another, and highlights the significance of.

Based on the other items that the consumer has previously added to their shopping cart, this is used to suggest things that the user may like to buy.

Thus, a sequence is generated for each category of items using Market Basket Analysis on Server 2008 R2 with the aid of SQL Server Business Intelligence Development Studio.

According to the study, the following pattern has surfaced, which may be utilized for a variety of things, such as strategically placing items on the shelf, increasing sales, improving customer convenience, creating promotions, having salespeople propose products to consumers, etc.

Interpretation: Desktop -> Tableware -> Memorabilia

A customer buying Tableware and Desktop is likely to buy Memento

Similarly, Other Patterns emerged are as below
Tableware, Desk Top -> Memento
Desk Top, Memento -> Tableware
Furniture, Memento -> Hanging
Hanging, Memento -> Furniture
Furniture, Wall Decor -> Memento
Wall Decor, Memento -> Furniture
Tea, Memento -> Bag
Tableware, Furniture -> Memento

### **To Study the Top Most Likely Products Purchased by the Customers.**

For the five goods that clients are most likely to buy, a singleton prediction query is built. As many numbers as needed may be added to it. In order to get the results, a query is constructed in the mining model prediction tab.

As shown in screenshot 3, the expression column displays the five most probable (subcategory) goods that consumers buy. This offers the store an indication of the best-selling goods and attracts consumers' attention.

To Research the Most Frequently Purchased Items by Clients in a Specific Product Category With inputs from nested tables, a singleton prediction query is constructed. As shown in screenshot 5, a query is generated to anticipate goods that are most likely related to the Body Care. Screenshot 6 displays the outcomes of the predictions for goods that are most likely related to body care. This provides the shop with an overview of the body care category and potential goods that the consumers have bought. Using this information, promotions and discounts may be created to entice clients to buy other related items.

### **Creating Multiple Predictions for Suggesting Products to each Customer**

A multiple predictions query may be used to determine which forecasts are most accurate for a given consumer based on their prior purchases.

The creation of the Multiple Predictions Query is depicted in screenshot 7, and the predictions for specific customers are obtained as depicted in screenshot 8. Party ID refers to the contact information of the customers who make the purchases, and expression provides product recommendations based on previous purchases. Making specific recommendations to loyal clients might also benefit from this.

### **Conclusion**

The study looks at how consumers buy products from a lifestyle shop. Because the program offers valuable information on basket building, retailers find it useful in understanding the purchasing behaviors of their customer base. It helps with product assortments, stock replenishment for high-selling items, promotions based on high-selling items in a certain category, product bundling, and providing discounts to draw customers in. The efficiency of the analysis is boosted by the software's greater use of parameters. Retailers may use the data to make strategies and suggest goods to loyal customers. One kind of study that may be carried out using the Business Intelligence Development Studio software using previous sales data is market basket analysis. It helps to maximize the patterns associated with the dynamic behaviors of the purchases that customers make from lifestyle shops. The results of the market basket study might help the company with customer suggestions, cross-selling, and promotion development. The use of this data mining technology may improve the products' thoughtful placement on the shelves. By creating the website such that products that are often purchased together appear together, the retailer's marketing team may also be able to get a deeper understanding of the purchasing patterns of its clientele. It also facilitates the development of advertising and marketing plans. It could be advantageous and help in making decisions that enhance the business and the consumer purchasing experience.

### **References**

1. Rakesh Agrawal, Sirkant Ramakrishnan. (1994). *Fast Algorithms for Mining Association Rules. Proceedings of the 20th VLDB Conference. Santiago.*
2. Robert J. Hilderman, Colin L. Carter, Howard J. Hamilton, and Nick Cercone. (n.d.). *Mining Market Basket Data Using Share Measures and Characterized Item sets.*
3. Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. (1996) *From Data Mining to*



*Knowledge Discovery in Databases. American Association for Artificial Intelligence.*

4. Julander, C.-R. (1992). *Basket Analysis: A New Way of Analyzing Scanner Data. International Journal of Retail and Distribution Management, Volume 20 (7), 10-18.*
5. Gregory Piatetsky-Shapiro, William Frawley. (1991). *Knowledge Discovery in Databases. AAAI/MIT Press.*
6. Garry J. Russel, Wagner A. Kamakura. (1997). *Modeling Multiple Category Brand Preference with Household Basket Data. Journal of Retailing, Volume 73(4), 439-461.*
7. Qiankun Zhao, Sourav S. Bhowmick. (2003). *Association Rule Mining: A Survey. Singapore: CAIS, Nanyang Technological University, No. 2003116.*
8. Rakesh Agrawal, Tomasz Imielinski, Arun Swami. (1993). *Mining Association Rules between Sets of Items in Large Databases. ACM SIGMOD Int'l Conference on Management Data, (pp. 207- 216).*
9. Andreas Mild, Thomas Reutterer. (2003). *An improved collaborative filtering approach for predicting cross-category purchases based on binary market basket data. Journal of Retailing and Consumer Services vol.10, 123-133*
10. <https://msdn.microsoft.com/en-us/library/ms167047.aspx>
11. <https://msdn.microsoft.com/en-us/library/ms173767%28v=sql.105%29.aspx>
12. [https://msdn.microsoft.com/en-us/library/ms173767\(v=sql.105\).aspx](https://msdn.microsoft.com/en-us/library/ms173767(v=sql.105).aspx)