

**Market Basket Analysis Using Datamining****Dr.Ramesh Chandavath<sup>\*1</sup>, Dr.V.Surya Narayana<sup>2</sup>, M.Annapurna<sup>3</sup>**<sup>1</sup> Department of MBA, Teegala Krishna Reddy Engineering College, Hyderabad, Telangana, India.**Sarangam Kodati<sup>4</sup>**<sup>4</sup> Department of CSE, Teegala Krishna Reddy Engineering College, Hyderabad, Telangana, India.**ABSTRACT**

Market Basket Analysis is a data processing technique that is used in the discovery of relations among various items. The major goal of market basket analysis in retail is to tell the distributor about a customer's purchasing habits so that the distributor can make the best choices. Market basket analysis may be done using a variety of methodologies. This paper discusses association rule mining, a data mining approach that may be used to investigate consumer purchase patterns and boost sales. Results may be a useful guide for creating promotions, cross-selling products, and setting up inventory in stores.

**Keywords:** Market Basket Analysis, Data Mining, and R.**I. INTRODUCTION**

Nowadays, consumers may choose from a variety of solutions in practically every industry. When a customer needed to make a purchase in the past, he was limited to selecting a product from the store's catalogue. However, the number of possibilities has grown dramatically in the modern technology and globalisation era. Customers may now pick from a wide range of items and variations. Geographical, seasonal, and other restrictions are no longer a problem. Previously seen as luxury items, these objects are now thought of as everyday items. All of this gave the businesses today's seemingly endless opportunities. But because of this boundless opportunity, a vast number of brand-new rivals entered the market. Retail businesses look for marketing tactics to draw in new clients or retain existing ones. Only new marketing approaches, which include effective advertising and sound product planning, could ameliorate the problem[1],[2].

Market basket studies have demonstrated exceptional effectiveness in other nations where they have been used. As a result, international retailers like more mega stores, Metro and walmart have begun employing market basket research to increase profits[4]. But in order to use market basket analysis to get insights, we must have knowledge about our customers' purchases, namely what they buy and when they buy it. As a result, the information on client purchases that is based on their behaviour becomes important [9].

A market basket is a collection of items that a buyer purchases all in one shopping trip. When we go to the supermarket, we frequently purchase a large number of items from various categories and place them all in a single basket. It is regarded as one transaction. The study of all of those baskets is known as market basket analysis.

The term "market basket analysis" refers to a broad range of analytical methods used to identify relationships and connections between particular commodities, as well as consumer behaviours and relationships between products. When applied in retail, it is predicated on the notion that a client is more (or less) likely to buy another group of goods after purchasing one set of goods. For instance, it is commonly known that most of the time when a consumer purchases cool drinks, they also purchase chips. The businesses that sell their items are interested in the purchasing behaviors they induce. In order to develop fresh marketing/sales tactics that can enhance the advantages of the business as well as client experiences, the sellers/supermarkets are interested in assessing which goods are bought in combination. The majority of retail markets place a greater emphasis on what their customers purchase. However, they do not take into account the date of purchase. It is also thought to play a significant role in their purchasing behavior. This thesis is concerned with "when" as well as "what" the client purchases. According to Forbes magazine, marketers are continuously predicting the next big trend and looking into the future, and data driven marketing is the most prominent trend at the moment when timing is very important. The retail firm will have a stronger future as a result of data-driven marketing that takes time into account.

Data mining is increasingly typical for many firms globally. Every day, a significant quantity of data is collected, and this data is used to capture vital information about many areas of every firm. Highly disaggregated data gathering is viewed as the foundation for knowledge extraction. Disaggregated data can expose certain facts immediately, but most of the time we are looking for underlying patterns and principles. Data mining may be used to provide non-trivial insights. Numerous statistical studies that uncover previously hidden characteristics of the data are included in data mining. Mining tools have been shown to be helpful in many firms for locating important information and subsequently giving management answers to challenging challenges. Data mining is commonly seen as a single step of a whole process called Knowledge Discovery in Databases (KDD). "KDD is the nontrivial process of detecting genuine, new, possibly helpful, and ultimately intelligible patterns in data," claim Fayyad et al. Since humans have had the ability to mine massive amounts of data thanks to advancements in computer power, data mining has become increasingly popular[8]. Using a variety of methodologies, knowledge and hidden information may be extracted from data and used in a variety of scenarios. In order to identify and analysis client categories and forecast future behavior, knowledge discovery is frequently employed in marketing.

**II. LITERATURE REVIEW**

Data mining is a useful tool for improving customer service and personalizing offers to suit consumers' wants and motives.

Data mining has taken an important part of marketing literature for the last several decades. Market basket analysis is one amongst the oldest areas within the field of data mining and is the best example for mining association rules.

Aiman Mushtaq in 2015 highlighted how data mining in marketing helps to increase return on investment or net profit. Enhance customer relationship management, do market research, develop better marketing plans, cut back on wasteful spending, etc. As the amount of data increases daily, several strategies are being employed to mine the data for marketing purposes and to further organisational objectives[4].

Solnet et al. in 2016 studied the potential of market basket analysis to grow revenue of hotels. The most alluring services and goods that may draw visitors to hotels, gratify them, and persuade them to make more purchases were investigated and determined by the researchers. This strategy can boost sales without boosting the number of clients [5].

Kaur and Kang (2016) did MBA to identify the changing trends of market data using association rule mining. This study suggested an alternative method of periodic mining that will increase the effectiveness of data mining approaches. This work helped us by revealing intriguing patterns from a sizable database, forecasting future association rules, and providing the correct methods to identify outliers. This study demonstrates progress by offering a fresh perspective on how to think about data changes occurring in addition to mining static data[6].

**III. Methods**

**3.1. Market Basket Analysis Using R**

R is a good tool for statistical and graphical analysis that is ideal for more complex research. With the help of the Arules packages, we can use R to carry out the market basket analysis. The Apriori method, one of the most used techniques for finding connections and correlations between items, is implemented by the Arules packages.

**3.2 Association Rule**

A popular technique for market basket analysis is association analysis, sometimes referred to as affinity analysis or association rule mining. ARM is presently the best approach for analysing huge market basket data, however when there are many goods being sold in a significant volume of transactions[7], the data matrix needed for association rule mining ends up being enormous and sparse, which takes longer to compute. The "what with what" statement is connected to the association rule. This information may take the form of a statement of transaction activity made by supermarket patrons. The parameters listed below can be used to do association rule mining[6].

➤ **Support:**

It checks that how frequently an item is purchased and how frequently the item is occurring in the transactions.

$$\text{Support}(A \rightarrow B) = \frac{\text{Transactions containing item-A and item-B}}{\text{Total no Transactions}}$$

If an item-set qualifies the minimum support value then it is considered for the calculation else item-set is ignored from the calculation. Higher the support value indicates that item is more frequently to occur.

➤ **Confidence:**

The conditional probability of both the items occurring together in a transaction. Probability of purchasing an item-A over item-B and vice-versa.

$$\text{Confidence}(A \rightarrow B) = \frac{\text{Number of Transaction that Contains A and B}}{\text{Total Transactions that contain A}}$$

➤ **Lift:**

Probability of purchasing first item over second item independent of each other.

$$\text{Lift}(A \rightarrow B) = \frac{\text{Confidence of A and B}}{\text{Support}(B)}$$

Lift (A → B) = 1 means that there is no correlation with the given item-set.

Lift (A → B) > 1 means that there is a positive correlation within the item-set, i.e. Products in the item-set, A, and B, are more likely to be bought together.

Lift (A → B) < 1 means that there is a negative correlation within the item-set, i.e. products in item-set, A, and B, are unlikely to be bought together.

➤ **Leverage:**

It contrasts the items occurring together in the item-set and the expected probability of item-set.

$$\text{Leverage } (A \rightarrow B) = P(A \text{ and } B) - [P(A) * P(B)]$$

➤ **Conviction:**

It is expressed as an item-A occurs without B.

$$\text{Conviction } (A \rightarrow B) = \frac{1 - \text{Support}(B)}{1 - \text{Confidence}(A \rightarrow B)}$$

### 3.3. Apriori Algorithm

Apriori algorithm is most widely used for finding frequent items in the transactions. It makes advantage of the item sets' past knowledge. It says that any non-empty subset has to have frequent subsets. The algorithm begins with straightforward principles, adds items to the item-set, and then chooses a subset of the item-set based on a predefined support value. The rule generating procedure only takes into account the things that meet the criteria for the provided support value, and it ignores the rest. The whole dataset is scanned in this repeated process to create every potential item combination.

**Apriori algorithm:**

- 1:  $k = 1$ .
- 2:  $F_k = \{ i \mid i \in I \wedge \sigma(\{i\}) \geq N \times \text{minsupp} \}$ . {Find all frequent 1-itemsets}
- 3: repeat
- 4:  $k = k + 1$ .
- 5:  $C_k = \text{apriori-gen}(F_{k-1})$ . {Generate candidate itemsets}
- 6: for each transaction  $t \in T$  do
- 7:  $C_t = \text{subset}(C_k, t)$ . {Identify all candidates that belong to  $t$ }
- 8: for each candidate itemset  $c \in C_t$  do
- 9:  $\sigma(c) = \sigma(c) + 1$ . {Increment support count}
- 10: end for
- 11: end for
- 12:  $F_k = \{ c \mid c \in C_k \wedge \sigma(c) \geq N \times \text{minsupp} \}$ . {Extract the frequent k-itemsets}
- 13: until  $F_k = \emptyset$
- 14:  $F_k$

## IV. DATA COLLECTION AND DATA ANALYSIS

Because many facts would be provided in this study, the writers utilised a descriptive research technique. The research project will begin with the formulation of the research topic and go on to a literature evaluation using references from earlier studies. Data will then be gathered through secondary sources and used for R programming analysis. In order to see the patterns of customer purchases that occur, the results of the shopping cart analysis that was performed with R programming will be exhibited in the form of a plot. The next stage is to set up the research flow such that it follows the author's chosen procedures in the hopes that the study will go without a hitch and that the research outcomes may be maximised.

The study used secondary data to further examine the already-existing data using a shopping cart analysis, which allowed for the determination of the support and confidence first. The author's research was conducted in the manner described below: 1. locating secondary data 2. Save the data into a CSV file. 3. Launch R Programming and load the necessary libraries. 4. Run a shopping cart analysis.

5. Results of the analysis are shown.

## V. IMPLICATIONS

In the field of marketing, particularly in retail, there are issues such a lack of product inventory and bad product placement that alter consumer behaviour. With a situation like this as a starting point, a shopping cart analysis can be used to determine the buying trend. Sales Itemset data the following is sales data which is used as a shopping cart analysis.

- citrus fruit semi-finished bread
- tropical fruit yogurt
- whole milk
- pip fruit yogurt
- other vegetables whole milk
- whole milk butter
- rolls / buns
- other vegetables UHT-milk

pot plants  
 whole milk cereals  
 tropical fruit other vegetables  
 citrus fruit tropical fruit  
 beef  
 frankfurter rolls / buns  
 chicken tropical fruit  
 butter sugar  
 fruit / vegetable juice

If we describe it further, it is as follows:

"If someone buys an item in the item set on the left, that person will also buy the item on the right."

{coffee, sugar} => {milk}

When a consumer buys coffee and sugar, he also buys milk. From this it can be seen that there are three important things in association rules, namely:

- Support: shows the percentage of the number of transactions that contain the item.
- Confidence: shows the percentage contained in transactions containing items
- Lift ratio: is an important parameter besides support and confidence. The lift ratio measures how important the rule is based on the value of support and confidence.

**VI. RESULTS**

Market Basket Analysis with R: To perform shopping cart analysis using R Programming, all you have to do is install the necessary libraries, then load the libraries. The packages required are rules, arulesviz and datasets. The process and load results from the three libraries are shown in Figure 1.

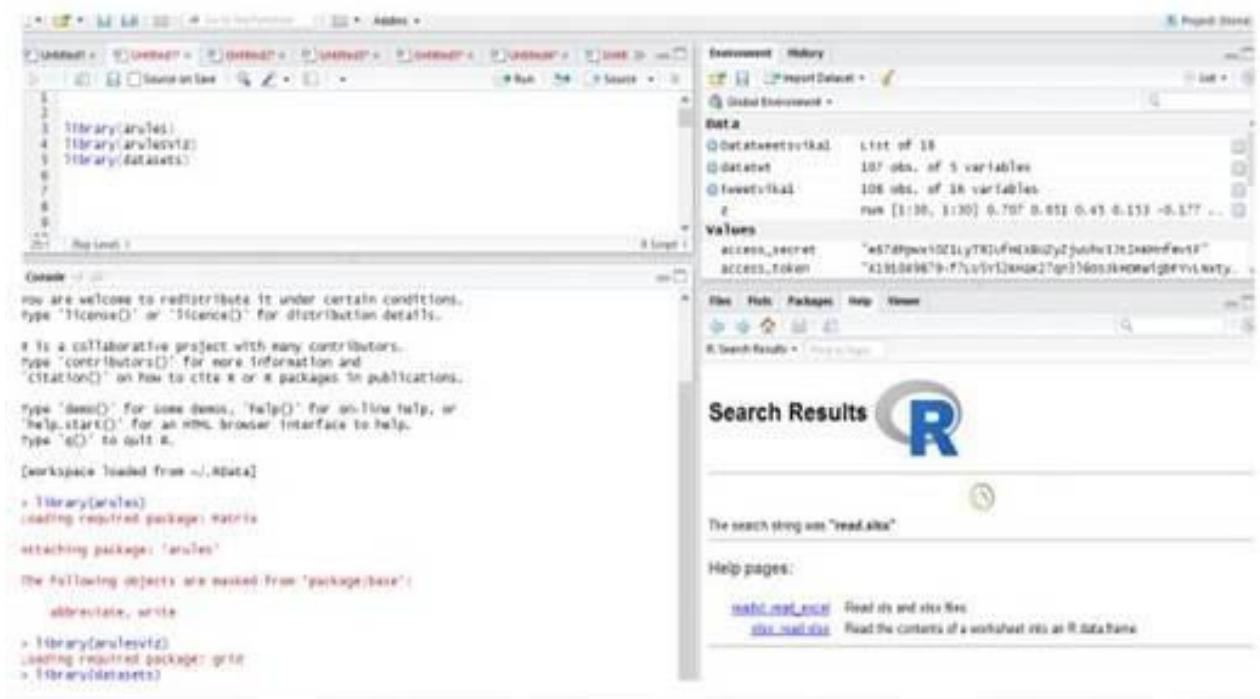


Figure 1. Load Library on R

After the load library process is successful, the next step is to import the data used for shopping cart analysis, how R imports the data shown in Figure 2.

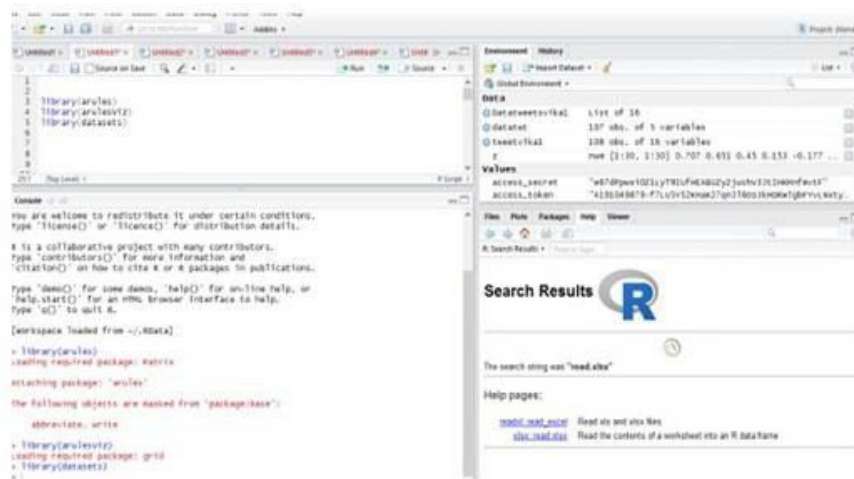


Figure 2. Import Data at R

The results of imported data are depicted in Figure 3.

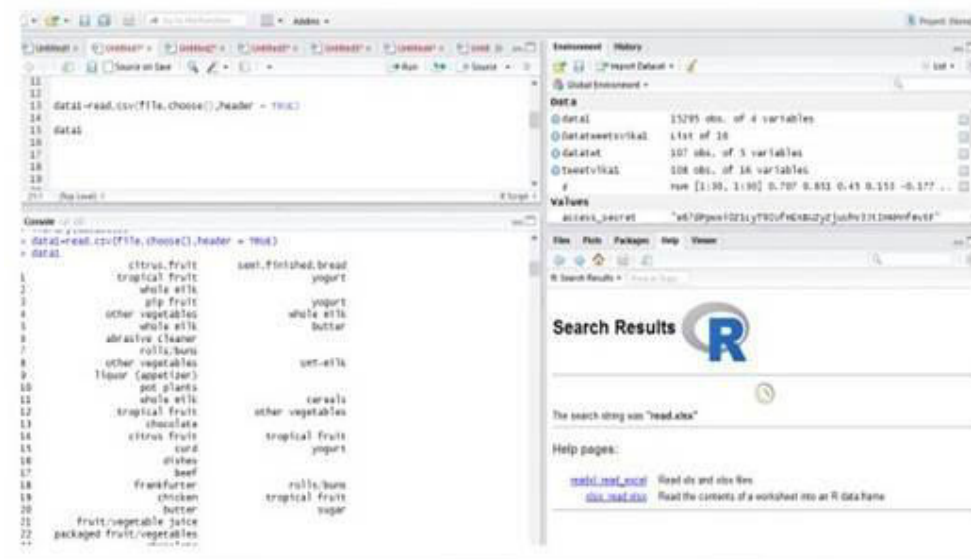
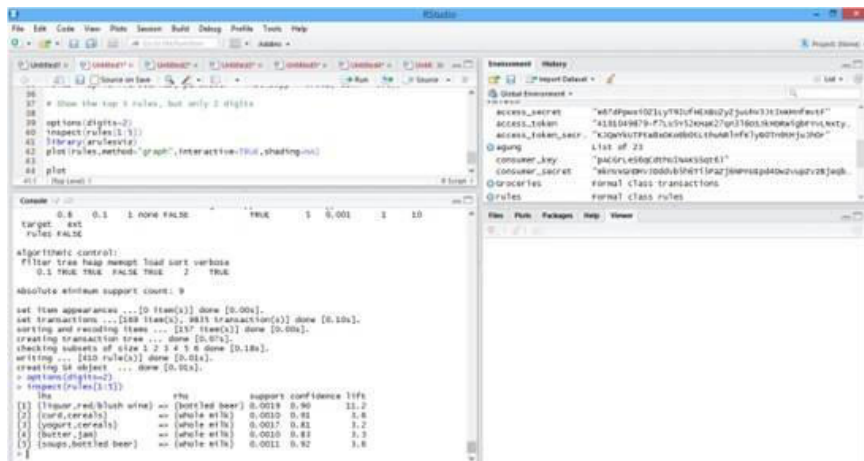


Figure 3. Results of Import Data at R

The output is as follows:



Illustrated by the table can be seen in table 1.

Table 1. Example Output

No	Lhs	Rhs	Support	Confidence	Lift
1	{liquor.red/blush wine}	=>{bottled beer}	0.0019	0.90	11.2
2	{curd,cereals}	=>{whole milk}	0.0010	0.91	3.6
3	{yogurt,cereals}	=>{whole milk}	0.0017	0.81	3.2
4	{butter,jam}	=>{whole milk}	0.0010	0.83	3.3
5	{soups.bottled beer}	=>{whole milk}	0.0011	0.92	3.6

**VII. CONCLUSION**

Marketing decision-makers are under additional strain as a result of the fierce competition between retail establishments and the expanded selection of products offered to clients. Long-term client relationship management has become necessary. When using data mining software to search databases for intriguing patterns, association rules are crucial. The simplest algorithm, apriori, is used to identify recurring patterns in consumer transaction databases. To uncover the hidden patterns that predominate in the transactional database, frequent pattern mining has been widely employed for market basket analysis of consumer transaction data.

The article analyzes the patterns of client product-purchasing behaviour in a retail setting. The programme helps merchants better understand the shopping habits of their clients and provides insightful data on how baskets are created. It aids in product assortments, the creation of promotions based on goods likely to sell with a certain category, the bundling of goods, and the provision of discounts to encourage people to buy goods. The research may be used by retailers to develop plans and offer advice to devoted consumers.

**REFERENCES**

[1] Chiu, S., & Tavella, D. 2008. Data mining and market intelligence for optimal marketing returns. Oxford, UK: Butterworth-Heinemann.

[2] Saurkar Anand V, Bhujade V, Bhagat P, Khaparde A. A Review Paper on various Data Mining Techniques. International Journal of Advanced Research in Computer Science and Software Engineering 2014;4( 4):98-101.

[3] Raorane AA, Kulkarni RV, Jitkar BD. Association Rule – Extracting Knowledge Using Market Basket Analysis. Research Journal of Recent Sciences 2012;1(2):19-27.

[4] Mushtaq, A. (2015). Data Mining For Marketing. Journal of Marketing and Consumer Research, 9, 17–29. [5]. Solnet, D., Boztug, Y., & Dolnicar, S. (2016). An untapped gold mine? Exploring the potential of market basket analysis to grow hotel revenue. International Journal of Hospitality Management, 56, 119–125. <https://doi.org/10.1016/j.ijhm.2016.04.013>

[6] Kaur, M., & Kang, S. (2016). Market Basket Analysis: Identify the changing trends of market data using association rule mining. Procedia - Procedia Computer Science, 85(Cms), 78–85. <https://doi.org/10.1016/j.procs.2016.05.180>

[7] Trupti , A. K., & Prof. Santosh, V. C. (2014). An Overview of Association Rule Mining Algorithms. International Journal of Computer Science and Information Technologies vol 5(1). 927-930

[8] K.Pradeep Reddy, Sheshan R, Sarangam Kodati and Nara Sreekanth, “Data mining and big data analysis of algorithms”, International Journal of Engineering Research-Online, Volume 6, Issue 6.

[9] Dhanabhakyam, M. and Punithavalli, M. 2011. A Survey on Data Mining Algorithm for Market Basket Analysis, Global Journal of Computer Science and Technology, 11(11), pp. 23-28.