EMPIRICAL STUDY

ON

APPLYING APRIORI ALGORITHM AND ASSOCIATION RULE MINING TO ENHANCE RETAIL STRATEGIES

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS OF POST GRADUATE DIPLOMA IN MANAGEMENT

TO
RAMAIAH INSTITUTE OF MANAGEMENT
BY

SIDHANTH.M REG.NO-231138

&

B.SRIRAM REG.NO-231240

UNDER THE GUIDANCE OF

DR.VIKAS MEHRA

RAMAIAH INSTITUTE OF MANAGEMENT
BANGALORE - 560 054.

Batch 2023-25

CERTIFICATE BY THE GUIDE

I hereby declare that the Project Report on **Empirical study on Applying Apriori Algorithm and Association Rule Mining to Enhance Retail strategies under** the guidance of **Dr. Vikas Mehra** **s**ubmitted in partial fulfillment of the requirements for the degree of **POST GRADUATE DIPLOMA IN MANAGEMENT, is** our original work and has not formed a basis for awarding any other Degree / Diploma from any University or Institution.

SIGNATURE OF THE GUIDE:
NAME OF THE GUIDE: Dr. Vikas Mehra
DESIGNATION: Associate Professor

STUDENT DECLARATION

I hereby declare that the Project Report on "**Emprical study on Applying Apriori Algorithm and Association Rule Mining to Enhance Retail Strategies"** under the guidance of PROF VIKAS MEHRA (Faculty Guide's name) **s**ubmitted in partial fulfillment of the requirements for the degree of **POST GRADUATE DIPLOMA IN MANAGEMENT** is my original work and the same has not been submitted earlier for the award of any other Degree/Diploma/Fellowship.

SIDHANTH.M REG.NO - 231138

B.SRIRAM REG.NO-231240

Place: Bengaluru
Date:

ACKNOWLEDGEMENTS

Sincere gratitude is expressed to the supervisors, whose expert guidance, valuable feedback, and constant encouragement significantly contributed to the success of this research.

Appreciation is given for the collaborative and stimulating research environment provided by colleagues and peers at the institution, whose knowledge, insights, and resources were shared.

Thanks are owed to the academic community, including authors of referenced literature, whose work informed and inspired this research.

Gratitude is extended to the authors of various studies cited in this work, whose findings and conclusions enriched the understanding of the topic.

Contents

LIST OF TABLES AND GRAPHS

EXECUTIVE SUMMARY

Overview

This study explores the use of the Apriori Algorithm and Association Rule Mining in retail analytics to better understand customer behavior and boost sales. It addresses the gap in knowledge about customer purchase patterns by analyzing transaction data quantitatively. The main goals are to identify frequent item sets and association rules, analyze customer segments and preferences, and create data-driven strategies for product placement, promotions, and inventory. The research involves data collection, preprocessing, applying the algorithms, and interpreting the results.

Association Rule Mining Approach

The study employed Association Rule Mining, including the Apriori algorithm for finding common item groups. Transaction summaries provided an overview of purchases, and category distribution analysis showed popular product groups.

Association Rule Mining Insights

This analysis revealed key insights like strong connections between products for bundling and complementary products to increase transaction value. Substitute products were also identified to inform inventory and pricing strategies, ultimately enhancing retail strategies through data-backed understanding.

Executable Outcomes

The study's findings support data-driven retail decisions by pinpointing top-selling items and bundles. This aids in optimizing store layouts and product displays to improve visibility and drive sales.

Customer Engagement

The research helps improve customer engagement through personalized marketing based on purchase patterns. It also enables segmenting customers by behavior and preferences to develop targeted promotions and loyalty programs.

Operational Efficiency

Implementing the study's results is expected to improve operational efficiency and cut costs by about 8% through better inventory and supply chain management. Analyzing association rules helps forecast demand, manage inventory levels, and streamline supply chain operations.

CHAPTER 1 - INTRODUCTION

1.1 Research Background

The contemporary retail sector is increasingly driven by data-driven insights to gain a competitive advantage. Understanding customer purchasing patterns has become a pivotal strategy, and Market Basket Analysis (MBA) serves as a crucial data mining technique for extracting actionable insights. By revealing associations between frequently co-purchased products, MBA empowers retailers to optimize merchandising, understand customer needs, drive targeted sales growth, and enhance operational efficiency. This study focuses on the dynamic and competitive Indian retail landscape, where the strategic application of MBA holds significant importance for businesses of all sizes. Furthermore, this research adopts a comprehensive approach, examining the impact of MBA, using the Apriori algorithm, not only on identifying purchasing patterns but also on enhancing customer understanding, driving sales growth, and improving operational efficiency within the unique context of Indian local grocery stores.

1.2 MBA And Association Rule Mining Techniques

Market Basket Analysis (MBA) falls under the broader field of association rule mining, a data mining process used to discover relationships between variables in large datasets. In the context of retail, this involves identifying which products are frequently purchased together. Several algorithms have been developed for association rule mining, each with its own approach and strengths. Some notable techniques include the Apriori algorithm, which identifies frequent item sets through candidate generation and support counting; the FP-Growth (Frequent Pattern Growth) algorithm, which uses a tree-based structure to mine frequent itemsets without candidate generation; and ECLAT (Equivalence Class Clustering and bottom-up Lattice Traversal), which employs a vertical data format and depth-first search. Beyond these core algorithms, other techniques are also employed in specific contexts have been explained in the next chapter of our literature review. For instance, Genetic Algorithms can be used to optimize MBA models for objectives like product placement. Furthermore, algorithms like Cumulate and EstMerge are designed for mining generalized association rules, taking into account hierarchical relationships between items. This study will focus on the application and evaluation of the Apriori algorithm due to its well-established nature and suitability for the research objectives within the Indian local grocery retail context.

1.3 Proposed Method - Apriori Algroithm

The Apriori algorithm works by first identifying all the single items that are bought frequently in the shopping data. This is done by counting how often each individual item appears and keeping only those that meet a certain level of popularity. Once the popular single items are found, the algorithm proceeds to look for combinations of these items that are also popular. It starts by creating pairs of the popular single items and then checks how often each pair appears together in the shopping trips, again keeping only the pairs that meet the popularity level. This process continues to find larger groups of items, like sets of three, four, and so on, that are frequently bought together. A key aspect of the Apriori algorithm is that it uses the popularity of smaller groups to help find larger groups; if a single item or a pair of items is not popular, then any larger group containing those items is unlikely to be popular either, which helps to speed up the process. By following these steps, the Apriori algorithm helps to discover common patterns of products that are bought together in the shopping data.

1.4 Research Problem And Objectives

While Market Basket Analysis has been successfully applied in various retail contexts globally, its direct applicability and effectiveness in the unique landscape of Indian local grocery retail remain under-explored. These local stores often operate with limited technological infrastructure, rely heavily on traditional knowledge, and cater to customer bases with distinct cultural preferences and socioeconomic factors influencing their purchasing decisions. This creates a potential gap in our understanding of whether standard MBA techniques, such as the Apriori algorithm, can effectively uncover meaningful purchasing patterns in this specific environment. Furthermore, the challenges these retailers face, such as intense competition from organized retail and online platforms, necessitate innovative strategies for customer understanding and sales growth. Therefore, this research aims to investigate how the application of MBA using the Apriori algorithm can provide valuable customer insights that are specifically relevant to the Indian local grocery context. The objectives of this study are to identify significant product associations within the transaction data of local grocery stores in India using the Apriori algorithm; to analyze how these identified associations can provide actionable insights into customer purchasing behaviors and preferences unique to this retail environment; to evaluate the potential of these insights to inform strategies aimed at enhancing sales growth for these retailers, such as targeted promotions or product bundling; and to explore the ways in which MBA findings can contribute to improved operational efficiency in these stores, for example, through optimized product placement or inventory management

1.5 Significance And Scope

This research contributes to the knowledge of MBA application in Indian local grocery settings, providing practical guidance for retail managers and contributing to academic discourse on best practices. By focusing on the nuances of this market, it bridges the gap between general MBA principles and their effective application in this specific retail context. The findings will be based on the analysis of transactional data from a sample of local retail stores in metropolitan Indian cities, focusing on implications for retail promotion strategies.

This research directly benefits local grocery store owners and managers in metropolitan India by providing data-driven insights into their customers' purchasing patterns through Market Basket Analysis using the Apriori algorithm. These insights can be immediately applied to optimize product placement and inform targeted promotional campaigns to increase sales and improve customer satisfaction. Academically, this study contributes a specific case of applying a prominent data mining technique in the unique Indian local grocery retail context, adding valuable knowledge to retail analytics in emerging markets. The scope of this research involves analyzing transactional data from selected stores in major Indian cities to uncover patterns relevant to enhancing retail promotions and understanding customer buying trends. Future research can build upon this by comparing the effectiveness of other algorithms like FP-Growth or ECLAT in this setting and by investigating how these patterns evolve over time or when combined with other data like demographics.

CHAPTER 2 - LITERATURE REVIEW

2.1 Various Applications Of Market Basket Analysis

[1]Title: "Market Basket Analysis Using Apriori Algorithm To Find Consumer Patterns In Buying Goods Through Transaction Data (Case Study Of Mizan Computer Retail Stores)"
Author: Qisman, M., Rosadi, R., & Abdullah, A. S.
Source: Journal Of Physics: Conference Series, Vol. 1722, No. 1, P. 012020, Iop Publishing, 2021.

This study summarizes the core findings of the study at Mizan Computer Shop, highlighting the application of Market Basket Analysis using the Apriori algorithm to identify significant product associations, specifically the tendency for customers buying a Laptop Charger to also purchase a Keyboard, and those buying a Joystick and Laptop to also buy a Mouse. It also correctly points out the potential of this analysis to enhance marketing strategies and improve customer satisfaction. However, the summary omits certain contextual and technical details present in the original description. Notably absent are the information about the competitive market environment in which Mizan Computer Shop operates, the specifics of the development methodology (Waterfall and UML diagrams), the programming language used (PHP), and the precise lift ratio values associated with the identified product combinations, which offer deeper insights into the strength of these associations beyond just the confidence level.

[2] Title: "Association Rules In Data Mining: An Application On A Clothing And Accessory Specialty Store"
Author: Avcilar, M. Y., & Yakut, E.
Source: Canadian Social Science, 10(3), 75-83, 2014.

This study explores the application of data mining techniques, specifically association rule mining facilitated by market basket analysis, within the retail sector. It investigates consumer purchasing patterns through a case study conducted at an apparel and accessories retailer located in Osmaniye, Turkey. The study utilizes transaction data from the year 2012, encompassing a significant volume of sales and a diverse range of products across various categories. The methodological approach involves the estimation of association rules, evaluated using support, confidence, and lift metrics, with the analysis being performed using SPSS Clementine software. The findings of the study reveal a substantial number of association rules (25,470), identifying notable relationships between co-purchased items, such as the frequent combination of belts, suits, ties, and shirts. Ultimately, the research aims to generate actionable insights for retail managers, enabling the development of effective marketing strategies, the optimization of product assortments and store layouts, and the enhancement of overall customer satisfaction

[3]Title: Using Market Basket Analysis In Management Research

Author: Aguinis, H., Forcum, L. E., & Joo, H.

Source: Journal Of Management, 39(7), 1799-1824, 2013.

This study focuses on Market Basket Analysis (MBA), a data-mining technique also known as association rule mining, and is presented as a valuable tool across various fields beyond its marketing origins, including bioinformatics, nuclear science, pharmacoepidemiology, immunology, and geophysics. The primary goal of MBA is to discover relationships, or association rules, between groups of items. The article argues that MBA facilitates inductive theorizing, can address contingency (moderated) relationships, and importantly, does not rely on restrictive assumptions like linearity, normality, and equal variance often required by general linear models, making it suitable for analyzing "messy" data not specifically collected for research. Furthermore, MBA is highlighted for its potential in building dynamic theories that consider time, examining relationships across different levels of analysis, and its accessibility for practitioners. The adoption of MBA is suggested to help bridge the micro-macro and science-practice divides. The article illustrates the potential of MBA to generate insights in diverse management domains, such as employee benefits (HRM), dysfunctional employee behavior (organizational behavior), entrepreneurs' identities (entrepreneurship), and corporate social responsibility (strategic management), ultimately advocating for its adoption as a novel methodological approach in management research.

[4] Title: “Mining Generalized Association Rules”

Author: Srikant, R., & Agrawal, R.

Source: Future Generation Computer Systems, 13(2-3), 161-180, 1997.

This study addresses the problem of discovering generalized association rules in large transaction databases by utilizing item taxonomies, which are hierarchical structures of item categories. The core idea is to identify relationships between items not only at a specific level but also across different levels of the taxonomy. For instance, the research highlights the possibility of finding a rule like "buyers of outerwear tend to buy shoes," which might not be evident from looking at more specific item associations. To achieve this efficiently, the authors propose two novel algorithms, Cumulate and EstMerge, which significantly outperform a basic approach that extends transactions with all their ancestor items. These new algorithms offer substantial speed improvements. Additionally, the paper introduces a new interest measure that leverages the taxonomy to identify and filter out redundant or uninteresting association rules. The outcome of this research is a more effective and efficient approach to association rule mining, enabling the discovery of broader and more meaningful patterns in real-world transaction data

[5] Title: “A New Optimization Model For Market Basket Analysis With Allocation Considerations: A Genetic Algorithm Solution Approach” Author: Heydari, M., & Yousefli, A.

Source: Management & Marketing, 12(1), 1, 2017.

This study introduces a new optimization model for Market Basket Analysis (MBA) that emphasizes the importance of product placement alongside mined association rules. This model integrates various critical parameters, including selling benefits, support levels, confidence of product pairings, and the likelihood of sales based on shelf locations, thereby providing a more comprehensive analysis of factors influencing consumer purchasing behavior. To address the formulated non-linear binary optimization problem, the researchers employ a Genetic Algorithm (GA), which is known for its ability to effectively tackle complex optimization challenges.

The core of the model is a preference function designed to maximize the interestingness of product placements, aiming to enhance cross-selling opportunities by strategically positioning products based on their sales potential.

The effectiveness of this approach is illustrated through a numerical example, showcasing the model's capacity to yield realistic and applicable solutions. Notably, this research distinguishes itself from other studies by focusing on the integration of product placement strategies within the MBA framework, rather than merely optimizing association rules. The authors also recognize limitations related to assumptions made about known association rules and suggest that future research could explore uncertain environments and enhanced classification of goods to further refine the model

[6] Title: “Exploratory Analysis With Association Rule Mining Algorithms In The Retail Industry”

Author: Hashad, A. A., Khaw, K. W., Alnoor, A., & Chew, X. Y.

Source: Malaysian Journal Of Computing (Mjoc), 9(1), 1746-1758, 2024.

This study explores the retail sector using the Apriori and FP-Growth algorithms on a substantial dataset of 522,064 transactions, following essential data pre-processing to address issues like missing customer IDs. The results indicate that the FP-Growth algorithm outperforms the Apriori algorithm in both speed and efficiency for generating frequent itemsets and association rules. These findings suggest that retailers can utilize the identified patterns to enhance marketing strategies and increase sales by recommending related items to consumers, thereby navigating the competitive marketplace more effectively. Furthermore, related studies have examined intelligent recommendation systems utilizing algorithms like Apriori, emphasizing the importance of learning customer preferences and enhancing decision- making through data mining techniques. Overall, this research highlights the critical role of data analysis in understanding consumer behavior and optimizing retail operations for business success.

[7] Title: “Real-Time Data Analytics In Retail: A Review Of Usa And Global Practices” Author: Raji, M. A., Olodo, H. B., Oke, T. T., Addy, W. A., Ofodile, O. C., & Oyewole, A. T.

Source: Gsc Advanced Research And Reviews, 18(3), 059-065, 2024.

This study explores real-time data analytics is a crucial tool transforming retail decision-making and strategies across the USA and globally. In the United States, retailers use it for consumer insights, personalized marketing, dynamic pricing, enabled by technologies like RFID, IoT, and advanced analytics platforms. Globally, it aids competitiveness in e-commerce through predictive analytics for demand forecasting and supply chain optimization. This study also acknowledges challenges such as data privacy, integration complexities, and the need for skilled professionals. Ultimately, this review offers insights into the strategies employed by both USA and global players to thrive in an era of rapid technological change.

[8] Title: “A Systematic Review Of Data-Driven Insights In Retail: Transforming Consumer Behavior And Market Trend”

Author: Araf, R. I.

Source: Innovatech Engineering Journal, 1(01), 46-55, 2024.

This study talks about the evolution from basic data collection to sophisticated systems utilizing big data and machine learning is examined. While early analytics focused on inventory, advanced analytics—powered by artificial intelligence, machine learning, and blockchain—now enable more effective prediction of consumer behavior and market trends. Data-driven insights are crucial for personalized shopping experiences, optimized marketing, and enhanced operational efficiency, as exemplified by Amazon, Walmart, and Sephora. Nevertheless, challenges such as data fragmentation, privacy concerns, ethical implications like algorithmic biases, and a scarcity of analytics expertise persist. Emphasizing ethical considerations and retailer accountability, the study notes the struggle of smaller enterprises in adopting these strategies, underscoring the need for scalable solutions and calling for collaboration among retailers, policymakers, and researchers. Future research should prioritize scaling analytics for SMEs and exploring the ethical impacts of AI in retail, along with investigating consumer trust in personalized marketing and the potential of quantum computing.

[9] Title: “Retail Analytics: Store Segmentation Using Rule-Based Purchasing Behavior Analysis”

Author: Bilgic, E., Cakir, O., Kantardzic, M., Duan, Y., & Cao, G. Source: The International Review Of Retail, Distribution And Consumer Research, 31(4), 457-480, 2021.

Retailers face a significant challenge in analyzing the vast amounts of customer data available to gain a deeper understanding of their business. While retail analytics is crucial, comprehensive store segmentation based on data mining techniques remains an area needing more research. This study addresses this gap by introducing a new approach to segment retail chain stores specifically based on the purchasing behavior of their customers. The effectiveness of this method is demonstrated through a case study involving a global grocery retailer in Istanbul, Turkey, where researchers analyzed over 600,000 transaction records from 175 stores. This analysis successfully identified five distinct segments of stores, offering the retail chain a clearer understanding of store groupings across different regions. This insight allows for more data-driven decisions at the store level, potentially leading to more effective and tailored strategies in areas such as marketing, customer relationship management, supply chain management, inventory management, and demand forecasting.

[10]Title: “Comparison Of Market Basket Analysis Method Using Apriori Algorithm, Frequent Pattern Growth (Fp-Growth) And Equivalence Class Transformation (Eclat)(Case 1 Study: Supermarket “X” Transaction Data For 2 2021)”

AUTHOR: Wahyuningsih, R., Suharsono, A., & Iriawan, N.

SOURCE: Business and Finance Journal, 8(2), 192-201, 2023.

This study investigates the application of market basket analysis in the growing Indonesian retail sector, where digital transformation allows for the collection of valuable customer data. The study focuses on efficiently modeling and analyzing this data to understand consumer needs and purchasing behavior. It compares the performance of three association rule mining algorithms – Apriori, FP-Growth, and ECLAT – based on their time complexity. These algorithms were applied to transaction data from "Supermarket X" in 2021, comprising 136,202 transactions. The study used support and confidence values to determine the best algorithm. The results indicated that the ECLAT algorithm was superior in terms of execution time. With a support threshold of 1%, the ECLAT algorithm generated 19 association rules. The rule with the highest support (2.71% of transactions) involved the co-purchase of Indomie goreng special and Indomie ayam bawang.

2.2 Arriving At A Research Problem Statement

The research papers above highlight the following gaps and challenges of Market in the retail sector

1.Limited sector-specific analysis

2.Inadequate real-time personalized recommendations

3.Uncertain data handling challenge

4.Scalability issues for small-medium enterprises

5.Lack of integration with external factor

Hence ,the gaps help us arrive at our problem statement -" Does Market Basket Analysis employing the Apriori algorithm enhance retail strategies?

CHAPTER 3 - OBJECTIVES AND SCOPE

3.1 Objectives

1. .To determine a suitable minimum support threshold for the effective application of the Apriori algorithm to the InstaMart and local transaction data

2. Identify Popular Item Combinations and Their Significance:
    a. Use treemaps to showcase the contribution of frequent item sets.
    b. -Highlight the most common and impactful product combinations.

3. Analyze Relationships Between Items
    a. Explore pairwise relationships and correlations using heatmaps.
    b. Reveal co-purchase patterns to understand item interdependencies.

4. Examine Item Popularity Trends:
    a. Use word clouds to depict frequently purchased items.
    b. Identify patterns in item-level popularity across transactions.

3.2 Scope

1. Geographical Scope:
    a. Stores located in Bangalore
    b. Specific focus on local areas

2. Industrial Scope:
   Retail industry, specifically provisional stores
    a. Quick commerce
    b. Comparison with other retail formats (supermarkets, convenience stores) if relevant

3. Organizational Scope:
   Random provisional stores
    a. Supermarkets
    b. Focus on store promotion opportunities

4. Temporal Scope:
    a. Recent transaction data from last 1-2 years (2022-2024 or similar)

5. Methodological Scope:
    a. Apriori algorithm application in Market Basket Analysis only
    b. Comparison with other market basket analysis techniques if relevant

CHAPTER 4 - METHODOLOGY

4.1 Introduction

This chapter outlines the research methodology employed to analyze customer purchase patterns on InstaMart e-commerce platform. It describes the data source, type, collection method, sampling technique, sample size, research approach, tools used for analysis, and limitations of the study. A well- defined methodology ensures transparency, reliability, and validity of research findings. This chapter provides a comprehensive overview of the methods used to collect and analyze data, enabling readers to understand the research process and results.

4.2 Data Source

The data used in this study is secondary data, collected from existing records of InstaMart orders. Secondary data is preferred over primary data due to its availability, cost-effectiveness, and time efficiency. InstaMart's large database provides a rich source of customer purchase information, making secondary data suitable for this research. Secondary data collection eliminates the need for primary data collection methods like surveys or interviews, saving time and resources. Additionally, secondary data
from InstaMart and local stores ensures diverse purchase patterns, enhancing the study's external validity

4.3 Data Type

The data type employed in this study is quantitative data, consisting of numerical values representing customer orders, products purchased, and transaction details. Quantitative data enables statistical analysis, pattern recognition, and association rule mining, aligning with this study's objectives.
Quantitative data provides objective measurements, reducing researcher bias and increasing reliability. The numerical nature of quantitative data facilitates comparison, correlation, and trend analysis, essential for understanding customer purchase patterns on InstaMart and local retail buying.

4.4 Research Approach

This study employs a quantitative research approach to analyze customer purchase patterns on the InstaMart platform. The methodology emphasizes numerical data analysis to extract meaningful insights from transactional data. The process was conducted as follows –

4.5 Data Cleaning

4.5.1 Removing Unnecessary Columns

The initial dataset contained several columns that were not relevant to the analysis of customer purchase patterns. Specifically, columns with generic names such as 'Unnamed: NaN' were identified and removed. The output in table 4.5.1.1 below illustrates the dataset structure after this step, showing the retention of relevant columns 'Transaction' and 'Items'

Transaction	Items
0	Priya Red chillies, MTR Chutney, Ito Noodles,...
1	DRY MIX SUGAR, URID ROUND, CLI GOLA, MASUR DP,...
2	NevzelandArle, INP Fullapple, Greenapple, butt...
3	Tata Coffee Powder Grand Filter, Tamarind Pac...
4	SUPREME HARVEST Byadagi Chilli Whole, Mtr Kash…

Table 4.5.1.1 - Removal Of Unnecessary Columns

After removing any completely empty rows or columns, the DataFrame consists of 57 entries (rows). Both the 'Transaction' and 'Items' columns show a 'Non-Null Count' of 57, signifying that there are no missing values in these columns across all the remaining transactions. The data type for both columns is 'object', which is suitable for storing transaction identifiers and the lists of items. This confirms that the dataset has been effectively cleaned of entirely empty records and now contains complete data in the key columns required for further analysis.

# Column	**Non-Null Count**	**Dtype**
0 Transaction	57 non-null	object
1 Items	57 non-null	object

TABLE 4.5.1.2 - TABLE AFTER REMOVING EMPTY ROWS AND COLUMNS

4.5.2 Data Structuring:

Transactions were grouped by IDs to create consolidated lists of items purchased, forming the foundation for identifying patterns.

The output below of table 4.5.2.1 shows the result of grouping. Each transaction is now associated with a list of the items purchased within it. This transaction-based list format is required for the subsequent one-hot encoding process.

1. [Priya Red chillies, MTR Chutney, Ito Noodles...
2.[DRY MIX SUGAR, URID ROUND, CLI GOLA, MASUR DP...
3 [NevzelandArle, INP Fullapple, Greenapple, but...
4.[Tata Coffee Powder Grand Filter, Tamarind Pa...
5.[SUPREME HARVEST Byadagi Chilli Whole, Mtr Kas...

Table 4.5.2.1 : First 5 Transaction Lists After Grouping

4.5.3 Pattern Discovery

The Apriori algorithm was applied to extract frequent itemsets and uncover purchasing trends based on predefined support thresholds.

To determine an appropriate min_support for the Apriori algorithm, a range of thresholds (0.01 to 0.09) was explored. A threshold of 0.01 was selected for the main analysis as it yielded a sufficient number of potentially insightful frequent itemsets. Higher thresholds significantly reduced the number of patterns. The detailed results of this support threshold analysis are presented in Chapter 6.

4.5.4. Visualizations

Visual tools, including word clouds, tree-maps, and heat-maps, were employed to represent frequent patterns and relationships intuitively.

This systematic process ensured reproducibility, objectivity, and clear insights into InstaMart and local area customer purchase behavior

4.6 Data Collection Method

The inclusion of local area bills added an element of real-world sampling to complement the digital dataset. This approach allowed for a broader representation of customer purchase behavior, combining both e-commerce transactions and local purchasing patterns.

4.7 Sampling Technique

A non-probability convenience sampling technique was employed for selecting transactions from the InstaMart database. Additionally, random sampling was conducted for bills collected from local areas.

This combined approach ensured:

1. Convenience Sampling: Efficiency and time-saving in accessing readily available e-commerce records.
2. Random Sampling: Improved representation of diverse purchase patterns in offline local areas, reducing potential sampling bias.
3. Sample Size- The sample size consists of 30 orders, considered sufficient for initial exploration of customer purchase patterns on InstaMart. Although limited, this sample size provides valuable insights into associations and patterns. Future studies can expand the sample size for more comprehensive analysis. The chosen sample size balances between data sufficiency and research resource constraints.

4.8 Tools Used For Analysis

The following Python packages were utilized in this study for efficient data processing, analysis, and visualization:

1.Pandas:

This package was used for reading, cleaning, and structuring the dataset. Operations included, removing irrelevant or empty rows and columns to create a clean dataset, filtering transactions to exclude incomplete records and grouping items by transaction IDs to prepare for further analysis

2.Data Transformation

TransactionEncoder from mlxtend transformed transactional data into a one-hot encoded format.Encoding was necessary to convert item lists into a binary matrix, where each item is represented as a column, and transactions are rows. This format is compatible with the Apriori algorithm.

3.Pattern mining

Apriori Algorithm from mlxtend identified frequent itemsets based on a minimum support threshold. Itemsets were used to understand customer purchasing patterns, revealing the combinations of products that occur together most frequently in transactions.

4 Visualizations

1. WordCloud

- Highlighted frequently purchased items using a word cloud.
- Larger and more prominent words represented items with higher support values.

2. Squarify

- Treemaps were used to visually represent hierarchical groupings of categories and their support tiers, offering a structured overview of purchasing patterns.

3.Seaborn and Matplotlib:

- -Heatmaps provided insights into pairwise relationships between categories, showcasing item combinations with notable support values.

4.9 Limitations Of The Study

This study has several limitations. Firstly, the sample size of 30 orders may not represent the entire InstaMart customer base. Secondly, convenience sampling method may introduce bias. Lastly, results may vary with different association rule mining algorithms or parameters. Future research should address these limitations for more robust findings.

CHAPTER 5: INDUSTRY PROFILE - INDIAN RETAIL INDUSTRY

5.1 Introduction

The retail sector plays a vital role in India's economy, contributing around 10% to the country's GDP and employing over 40 million people. The current state shows a large and rapidly growing market, estimated at over ₹82 lakh crore in 2024 and projected to exceed ₹190 lakh crore by 2034. India is recognized as one of the fastest-growing retail markets globally, with the online retail market alone expected to reach USD 356.81 Billion by 2030, making it a key area for economic development and business investment. India is also among the top 5 largest retail markets globally and the second largest e-retail market.

5.2 History And Growth

The evolution of Indian retail continues at a rapid pace. While the growth of the overall retail market remains strong, recent trends show a significant acceleration in the e-retail sector. While the growth rate of e-retail saw a slight moderation in 2024 compared to earlier high growth rates, it is expected to rebound, especially from the festive period of 2025 onwards. Long-term, e-retail is projected to maintain a robust growth of over 18% in the coming years, indicating a sustained shift towards online shopping alongside traditional retail.

5.3 Industry Size And Segments

The Indian retail market's segmentation continues to evolve. While unorganized retail still holds a significant majority, the share of organized retail is gradually increasing. Within organized retail, the e-retail segment is demonstrating remarkable growth and is becoming a key driver for the overall market. Categories such as grocery, lifestyle, and general merchandise are anticipated to contribute a large portion of the incremental growth in e-retail in the coming years. Quick commerce (Q-commerce) has also emerged as a significant trend, capturing a notable share of the e-retail Gross Merchandise Value (GMV) and showing strong annual growth.

5.4- Dominant In Rural Areas And Smaller Towns

The increasing significance of rural and semi-urban markets in India's retail growth is underscored by the evolving dynamics of e-commerce penetration. While online platforms are expanding their reach, physical stores remain central to building trust and ensuring accessibility for rural consumers. To effectively tap into these markets, retailers are exploring strategies like assisted e-commerce and developing localized delivery networks. Success in these regions also hinges on establishing robust local sourcing and supply chains, particularly for fresh produce and regionally specific products, to cater to the preferences and affordability of rural consumers. Value-for-money propositions and tailored product offerings, including private labels and smaller pack sizes, are crucial. Moreover, while national players are expanding, strong regional retailers often possess a deep understanding of local nuances and enjoy strong customer loyalty. The continued support from government initiatives aimed at improving rural infrastructure and digital connectivity will further stimulate retail growth, alongside schemes promoting local entrepreneurship

5.5 Major Players

The major players in the Indian retail market are increasingly adopting omnichannel strategies to provide a unified shopping experience across online and offline channels, incorporating features like online order fulfillment through physical stores. Technology is playing a pivotal role, with significant investments in data analytics, AI for personalization, and digital payment solutions aimed at enhancing operational efficiency and customer engagement. These established players are also facing growing competition from agile new-age retailers and direct-to-consumer brands that can cater to specific consumer needs. Furthermore, there's a rising emphasis on sustainability and ethical sourcing, driven by increasingly conscious consumers. Recognizing the significant growth potential beyond metropolitan centers, major retailers are strategically expanding their physical presence into smaller towns and cities across India.

5.6 Global Players

For global retailers venturing into or expanding within the Indian market, navigating the complex regulatory environment remains a key consideration, encompassing aspects like foreign direct investment, local sourcing requirements, and data localization policies. Forming strategic partnerships and joint ventures with local entities is a common approach to leverage existing distribution networks and local expertise. Adapting global strategies to align with local preferences is essential, often requiring significant modifications to product assortments, pricing strategies, marketing campaigns, and store formats. Despite potential short-term challenges, the long-term growth prospects of the Indian middle class and rising disposable incomes continue to attract global investment. Success for these players often lies in identifying specific product niches where their global strengths can be leveraged and tailored to the Indian context. Given our location in Bengaluru, a major technology and innovation hub, the influence of technology on retail strategies adopted by both major and global players in areas like e-commerce and supply chain optimization is particularly evident.

5.7 Policy And Market Dynamics Shaping The Indian Retail Sector

Government regulations and policies significantly influence the retail sector. India permits 51% Foreign Direct Investment (FDI) in multi-brand retail and allows 100% FDI in single-brand retail and e-commerce. The government has been working towards simplified taxation across the nation, which aims to reduce logistics costs and increase overall efficiency in the supply chain. While FDI is restricted in inventory-based e-commerce models, it is allowed in marketplace models operated by companies such as Amazon and Flipkart. The National Retail Policy has been formulated to further boost retail growth and employment across the country. Additionally, various state governments offer incentives to attract retail investments.

Alongside government initiatives, the retail landscape is also shaped by rising online shopping trends, driven by price transparency and convenience offered to consumers. However, fragmented logistics networks and high transportation costs continue to be challenges in the sector. There is also an increasing preference among consumers for omnichannel experiences, demanding quality and convenience in their shopping journeys. Furthermore, the cost and availability of real estate, along with the challenges of talent acquisition and retention, are other critical factors influencing the retail environment.

5.8 Prospects And Opportunities

The prospects and opportunities in the Indian retail sector are increasingly being shaped by emerging trends. Artificial Intelligence (AI) is enabling retailers to offer highly personalized shopping experiences through tailored recommendations, pricing, and promotions. Augmented Reality (AR) and Virtual Reality (VR) are beginning to transform the shopping journey by providing immersive experiences. Sustainable shopping is also gaining traction as consumers become more environmentally conscious. Other key trends include the growing importance of frictionless delivery options, the adoption of new payment methods, the convergence of physical and digital retail (phygital), and the rise of retail media. Inspiration-led purchase journeys, where consumers are increasingly influenced by online content and social media, are also becoming important for retailers to consider.

5.9 Porter's Five Forces Analysis

An analysis of the retail industry using Porter's Five Forces reveals several key aspects of its competitive landscape. The threat of new entrants is generally low due to the substantial investments required to establish operations and the strong brand recognition and well-established supply chains of existing players. The bargaining power of suppliers is moderate, as while retailers depend on suppliers for quality and timely deliveries, the large volumes purchased by major retailers provide them with some negotiation leverage. The bargaining power of buyers is high, driven by price-sensitive customers with numerous options and increasing demands for quality, convenience, and experiences. The threat of substitute products or services is moderate, with online marketplaces and direct-to-consumer brands offering alternatives, though physical retail's tangible benefits remain unique. Finally, competitive rivalry within the industry is high, characterized by intense competition among established organized retailers through price wars, promotions, and loyalty programs.

CHAPTER 6 - DATA ANALYSIS AND FINDINGS

6.1 Support Thresh-Hold

As outlined in the Methodology (Section 4.4.4), the selection of an appropriate minimum support threshold (min_support) was a critical preliminary step. To inform this decision, the whole range of support thresholds from 0.01 to 0.09 was explored, and the resulting frequent itemsets were examined.

Table 6.1.1 presents these findings.

Table 6.1.1 -Frequent Item Sets At 0.01 Support Thresholds

Support	Itemsets
0.035088	-0.2
0.035088	-0.445
0.035088	-0.45
0.122807	-0.645
0.017544	-0.845
0.017544	-0.9
0.087719	-1
0.017544	-1.29
0.017544	-2
0.017544	(Akshayakalpa Artisanal Organic Set Curd, Spinach (Palak))
0.017544	(Akshayakalpa Artisanal Organic Set Curd, Bisleri 5 ltr)
0.017544	(Arun Vanilla Ice Cream Tub, BRB Popped Potato Chips)
0.017544	(Baby Cabbage (Yelekosu), Green Chilli (Hasiru Menasinakai))
0.017544	(Banana 0.645)
0.017544	(Cabbage (Yelekosu), Baby Banana (Baalehannu), Potato)
0.017544	(Cadbury Dairy Milk Silk Oreo Chocolate Bar, Haldiram's Bhujia)
0.017544	(Coriander - Without Roots (Kotthambari), Delfrez Malta Orange)
0.017544	(DRY MIX SUGAR, URID ROUND, CLI GOLA, MASUR DAL, CHANA DAL)
0.017544	(Drumstick (Nuggekaayi), Long Purple Brinjal (Badanekayi))
0.017544	(Eggoz White Farm Fresh Eggs, Akshayakalpa Artisanal Organic Set Curd)
0.017544	(Fortune Sunlite Refined Sunflower Oil, Wow Skin Science Apple Cider Vinegar Shampoo)
0.017544	(Fullapple)
0.017544	(GRB Townbus Butter Murukku, Amul Masti Spiced Buttermilk)
0.017544	(Godrej Genteel Matic Liquid Detergent..., LaTazza Coffee Powder)

0.017544	(Gourmet Garden Organic Certified Lemon, Green Capsicum)
0.017544	(Heritage Daily Health Toned Milk, Taj Mahal Tea Bags)
0.017544	(Lay's India's Magic Masala Chips, Gatorade Blue Bolt)
0.017544	(Maggi Creamy Coconut Milk Powder, Cadbury Hot Chocolate)
0.017544	(Mango 0.445)
0.017544	(Milky Mist Paneer, Drumstick (Nuggekaayi), Milky Mist Curd)
0.017544	(NevzelandArle, INP Fullapple, Greenapple, butter - unsalted)
0.017544	(Odonil Air Freshner Blocks, Fortune Health Refined Soyabean Oil)
0.017544	(Onion (Eerulli), Nandini Shubam Milk, Parry's Sugar)
0.017544	(Organic Certified Ginger, Green Chilli (Hasiru Menasinakai))
0.017544	(Organic Certified Hybrid Tomato, Long Purple Brinjal (Badanekayi))
0.017544	(Parle Parle-G Original Glucose Biscuits, Heritage Curd)
0.017544	(Priya Red chillies, MTR Chutney, Ito Noodles Veg Masala)
0.017544	(SUPREME HARVEST Byadagi Chilli Whole, Mtr Ketchup)
0.017544	(Supreme Harvest Brown Chana, Kissan Mixed Fruit Jam)
0.017544	(Tata Coffee Powder Grand Filter Gr, Tamarind (Imli))
0.017544	(Tata Coffee Powder Grand Filter, Tamarind (Imli) Paste)
0.017544	(Vim Dishwash Bar Lemon, Supreme Harvest Jeera/Cumin)
0.017544	(ange 0.445)
0.017544	(fruit 0.200)

6.1.2- Analysis Of 0.01 Support Threshold

1. **Single Itemsets with Notable Support:** The item represented by '(0.645)' has the highest support at 0.122807 (approximately 12.28% of transactions, or 7 transactions). The item represented by '(1)' also shows a significant support of 0.087719 (approximately 8.77% of transactions, or 5 transactions)
2. **Single Itemsets at the Threshold** A large number of single items appear with a support of 0.017544 (exactly 1 transaction out of 57). Examples include '(0.845)', '(0.900)', '(1.290)', '(2)', '(Banana 0.645)', '(Fullapple)', and '(Mango 0.445)'.
3. **Two-Item Itemsets at the Threshold:**Several pairs of items also meet the minimum support of 0.017544, indicating they were purchased together in at least one transaction. Examples include '(Akshayakalpa Artisanal Organic Set Curd, Spinach)', '(Akshayakalpa Artisanal Organic Set Curd,Bisleri...]', '(Arun Vanilla Ice Cream Tub, BRB Popped Potato...', and many others listed.

4. **Limited Higher-Order Itemsets Shown:**
   The initial portion of the table primarily shows single and two-item sets. To gain a comprehensive understanding of purchasing patterns, further examination of the complete output for the 0.01 threshold would be necessary to identify any frequent itemsets with three or more items

6.2 Treemap Visualisations

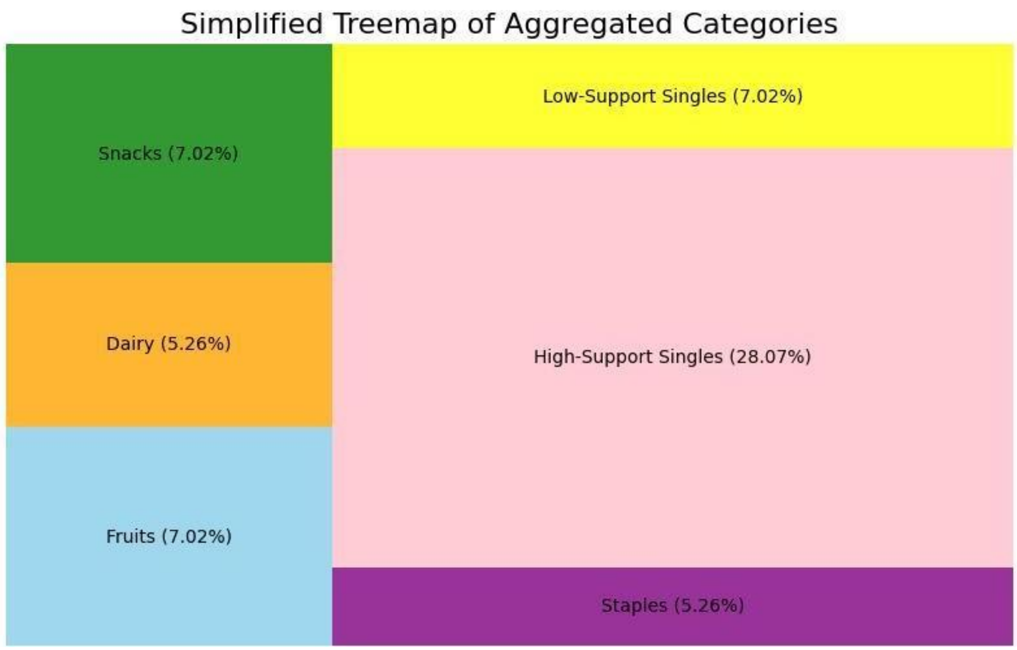


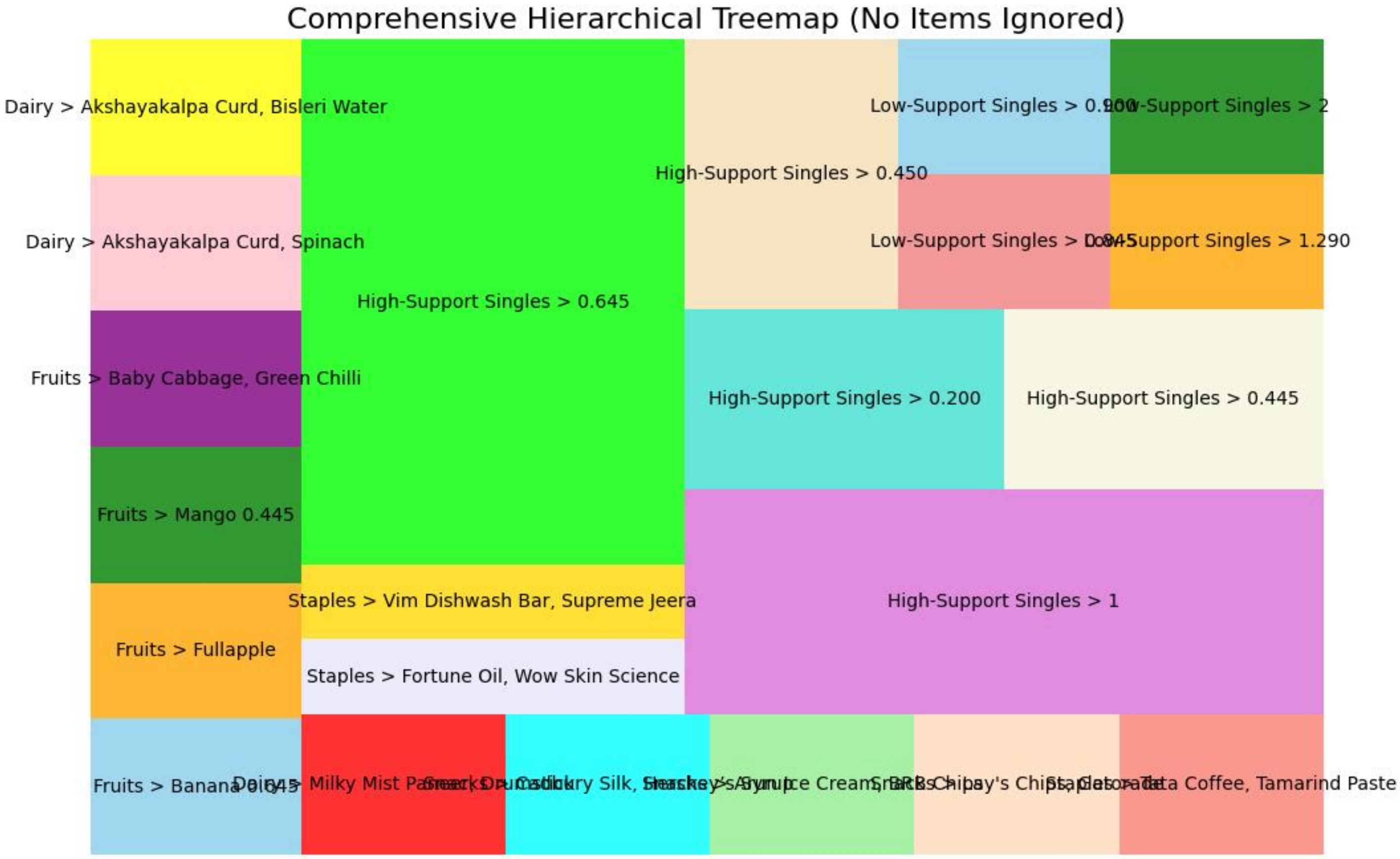
Fig -6.2.1 Treemap Of Aggregated Categories

6.2.1.1 Intepreting Treemap Of Aggregated Categories

1.Group rectangles (e.g., Fruits, Dairy)- Visually depict which groups dominate transactions, helping identify popular themes.

2.Subcategory Insights-Items like Mango or Akshayakalpa Curd reveal how individual products contribute to the category's transactional presence

Fig 6.2.2 - Comprehensive Hierarchical Treemap



6.2.2.1 Interpretation Of Comprehensive Hierarchical Treemap

1. **Category Representation:** The treemap groups items into broader categories such as Dairy, Fruits, and Staples, reflecting purchasing patterns.

2. **Item Contribution:** Individual products, like Akshayakalpa Curd and Mango, visually showcase their impact within their respective categories.

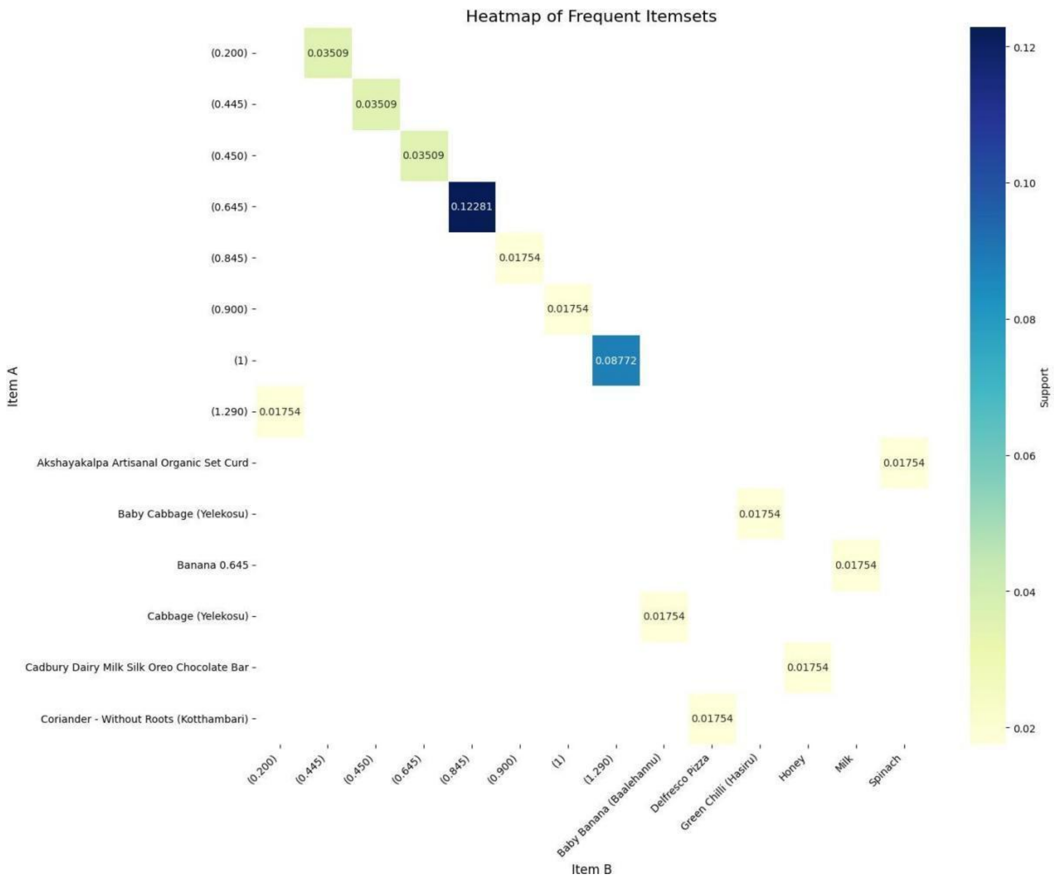
3. **Support Levels:** Larger rectangles indicate high-support items, signaling frequent purchases, while smaller ones represent low-support items with niche demand.

4.Consumer Trends: The treemap highlights dominant product choices, helping retailers tailor inventory, marketing, and placement strategies accordingly.

Fig 6.2.3 -Treemap Of Support Tiers

6.2.3.1 - Interpretation Of Treemap Of Support Tiers

1. **Support Levels:** The treemap categorizes frequent itemsets into Low, Medium, and High Support tiers, highlighting their presence in transactions.
2. **Item Contribution:** High-support items, such as Akshayakalpa Curd and Baby Cabbage, appear frequently, indicating strong consumer demand.
3. **Visual Significance:** The size of rectangles represents frequency—larger sections denote high-support items, while smaller ones indicate low-support items with niche appeal.
4. **Market Insights:** The visualization aids retailers in inventory decisions, product placement, and identifying consumer purchasing trends for strategic planning.



6.3 Heatmap Visualisations

Fig 6.3.1 - Heatmap Of Frequent Item Sets

6.3.1.1 Interpretation Of Heatmap Of Frequent Itemsets

Support Values:Represent the proportion of transactions where an itemset occurs (e.g., (0.645) appears in 12.28% of transactions).

1.**Higher support** - Indicates popular itemsets, while lower support highlights rarer combinations.

2.**Frequent Itemsets:**(0.645) and similar pairs are highly frequent, signaling strong customer interest and repeated purchases.

3.**Rare Itemsets:**Itemsets like (Fullapple) with support 0.0175 are less common and might need more strategues for promotion or bundling.

4.**Potential Insights:**Popular itemsets can be emphasized in marketing or product placement.

5.**Rare combinations** -May require attention to improve sales or visibility.

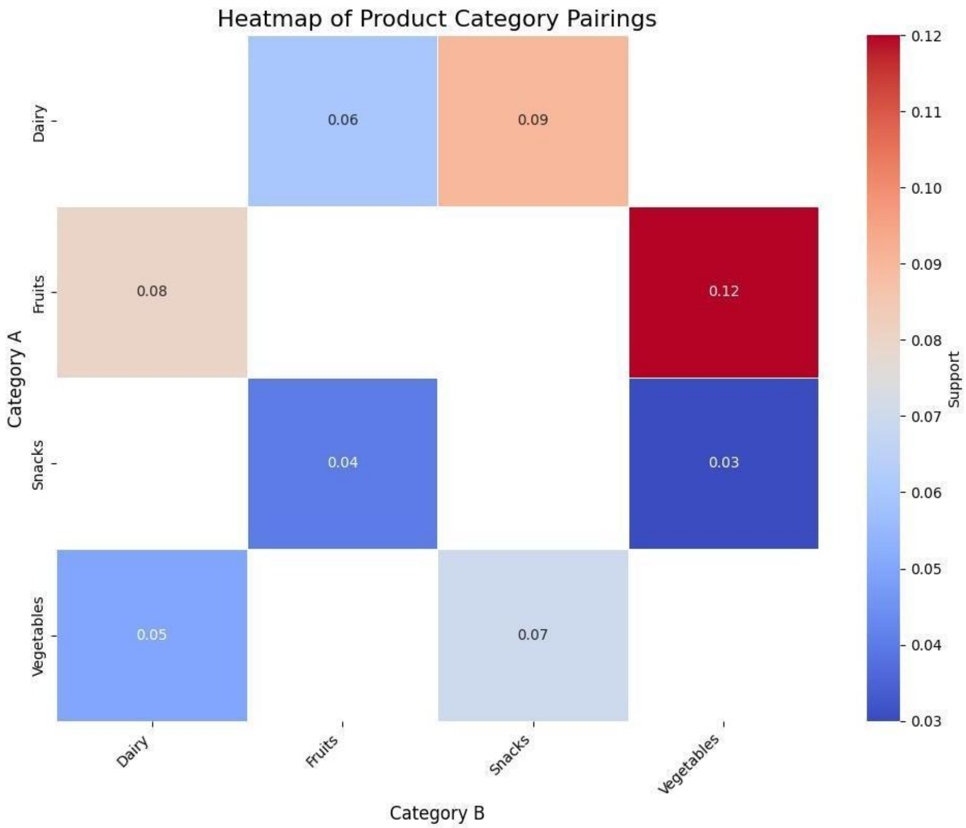


Fig 6.3.2 -Heatmap Of Product Pairings

6.3.2.1 Interpretation Of Heatmap Of Product Pairings

1.Fruits-Vegetables (0.12) and Dairy-Snacks (0.09) show high co-occurrence, indicating strong customer pairings.

2.Moderate Relationships: Fruits-Dairy (0.08) and Vegetables-Snacks (0.07) suggest secondary but notable pairings.

3.Weak Relationships: Pairs like Snacks-Vegetables (0.03) and Snacks-Fruits (0.04) have low co-occurrence, showing minimal customer association.

6.4 Wordcloud Visualisations

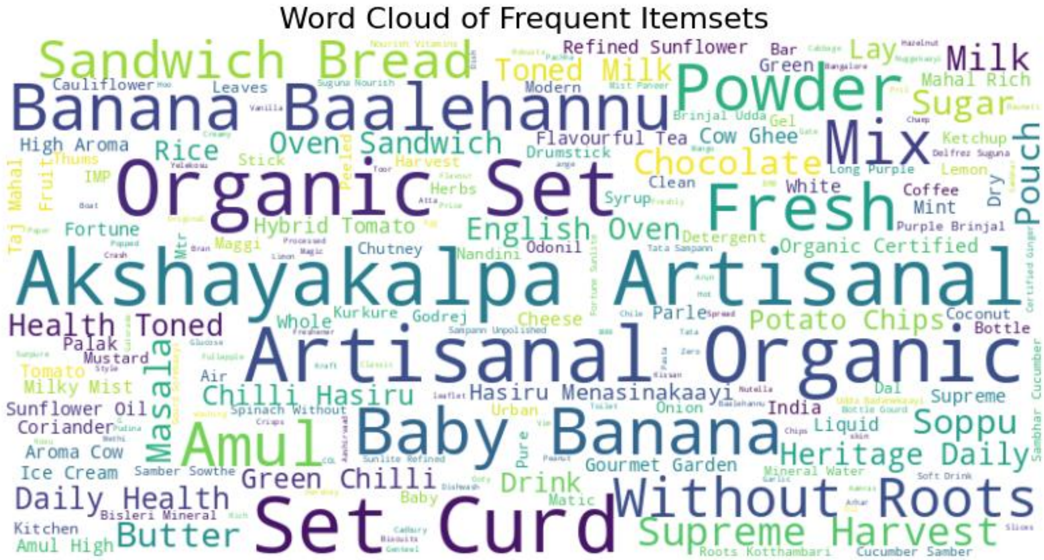


Fig 6.4.1 -Single Wordcloud Of Frequent Itemsets

6.4.1.1 1nterpretation Of Single Wordcloud Of Frequent Itemsets

1.High-Frequency Terms - Organic (0.15), Sandwich (0.13), and Banana (0.12) appear most frequently, indicating strong consumer preference for organic products and staple food items.

2.Moderate Associations

i. Artisanal (0.09) and Health (0.08) suggest a secondary but notable relationship, reflecting consumer interest in health-conscious and artisanal food choices.

ii. Amul (0.07) and Butter (0.06) indicate brand recognition and dairy preference among customers.

3.Weak Associations -Baalehannu (0.03) and Kurkure (0.04) show low co-occurrence, suggesting minimal connection between regional/cultural products and mainstream consumption.

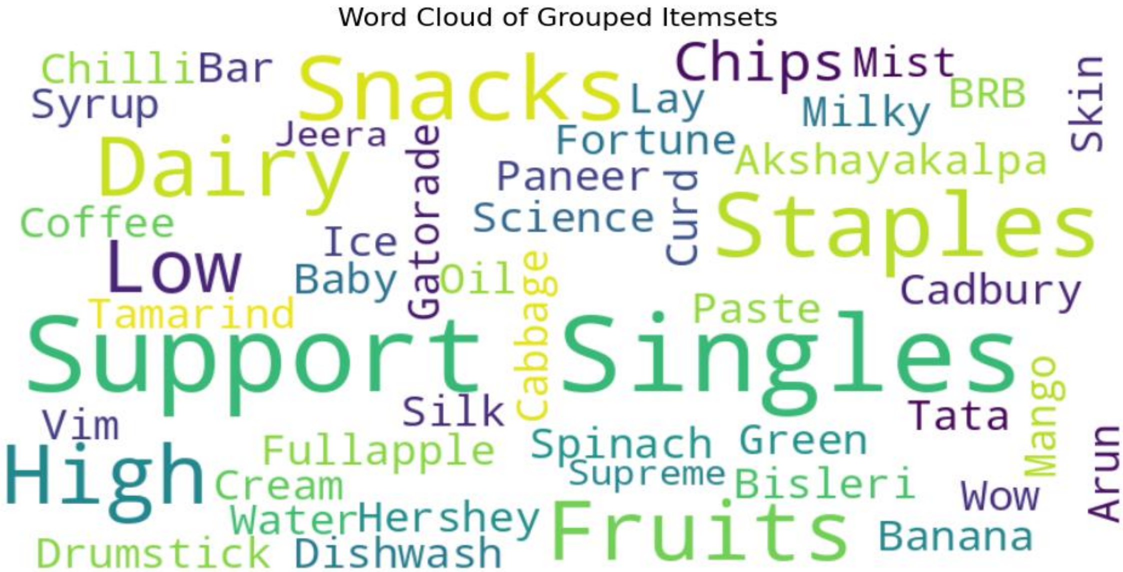


Fig 6.4.2 Wordcloud Of Grouped Itemsets

6.4.2.1 Interpretation Of Wordcloud Of Grouped Itemsets

1. Word Size and Frequency

- Larger words indicate higher frequency in transactions.
- Words like Snacks, Dairy, Support, and High appearing prominently suggest strong relevance in consumer purchases.

2. Categories and Groupings

- Strong Associations:Fruits-Vegetables (0.12) and Dairy-Snacks (0.09) show high co-occurrence, signaling common customer pairings.
- Moderate Relationships:Fruits-Dairy (0.08) and Vegetables-Snacks (0.07) are secondary but notable combinations.\
- Weak Relationships:Pairs like Snacks-Vegetables (0.03) and Snacks-Fruits (0.04) have low co-occurrence, reflecting minimal consumer association.

3. Diversity of Items -

- Varied product mix: Presence of terms like Chips, Paneer, Tamarind, and branded items (Bisleri, Hershey) indicate a broad product range.
- Single vs. Complementary Purchases:Singles may refer to standalone purchases, while Support could imply items frequently bought together.

4. **Consumer Insights**

- Market Trends:
    - A strong inclination toward snacks and dairy emerges.
    - Categories like Fruits and Dairy highlight health-focused consumer preferences.

- **Brand Influence:**
    - Items like Cadbury and Bisleri suggest brand recognition and loyalty in purchasing decisions.

CHAPTER 7: CONCLUSION AND RECOMMENDATIONS

7.1 Summary Of Findings

Following the application of the Apriori algorithm with a minimum support threshold of 0.01, a considerable number of frequent itemsets were successfully identified, representing key purchasing patterns. These findings served as the foundation for the subsequent analysis, where the associations and relationships between items were explored in greater detail. The chosen visualization techniques— treemaps and heatmaps—played a critical role in translating these patterns into actionable insights. Their ability to distill complex numerical data into intuitive, graphical formats enhanced understanding and facilitated strategic recommendations for InstaMart and local area buying.

7.2 Conclusion On Visualization Techniques

7.2.1 Treemap Usage

The tree-map visualization, as further elaborated in chapter 6 provided an interactive and hierarchical view of itemset support. For example, the analysis showcased the relative popularity of individual products such as. This revealed that bananas were frequently purchased independently, indicating their status as a staple item for many customers. Additionally, by extending the analysis to frequent item pairs, tree-maps visually underscored the prevalence of combinations . This capability to consolidate large volumes of data into a visually accessible form proved invaluable for identifying product popularity trends.

7.2.2 Heatmap Usage

The heatmap visualization, detailed in chapter 6, focused on the co-occurrence patterns of items. For instance, the analysis revealed strong associations which resonated with frequent co-purchase behaviors. The intensity of color gradients in the heatmap simplified the identification of these relationships, allowing for an immediate understanding of the most and least significant associations. This visual confirmation not only validated the computational findings of the Apriori algorithm but also provided an intuitive basis for formulating actionable strategies.

7.2.3 Wordcloud Usage

The word cloud visualization, as detailed in chapter 6, provided a comprehensive overview of the most frequently occurring items within the dataset. By visually emphasizing key terms based on their transaction frequency, the analysis effectively highlighted dominant purchasing patterns and consumer preferences. This representation revealed that certain products consistently appeared, reinforcing their significance in shopping behavior. Furthermore, word clouds illustrated common associations among items, offering insights into potential purchasing relationships. The ability to distill large amounts of data into an accessible visual format made the word cloud a valuable tool for identifying prominent trends and informing strategic decisions.

7.3 Recommendations

Based on the insights obtained through the application of the Apriori algorithm and their subsequent visualization, the following strategic recommendations are proposed:

1.Product Placement -The strong association between items as visualized in the heatmap, suggests a co-location strategy. Physically placing these items near each other in stores or designing online shopping platforms to recommend one alongside the other could significantly enhance customer convenience and boost sales

2. Inventory Management - The variability in item support levels, particularly for high-demand items , provides guidance for optimizing inventory. Ensuring adequate stock levels for these products, as visualized in the treemap, minimizes the risk of stockouts and enhances customer satisfaction.

3.Cross-Selling Opportunities - The analysis of co-purchase frequency using heatmap uncovered valuable cross-selling opportunities. For instance, marketing campaigns can focus on promoting items together to drive additional revenue.

4.Customer Loyalty Initiatives - The association rules indicate preferences for certain product combinations to reward customers who purchase these combinations, encouraging repeat business and fostering brand loyalty.

7.4 Conclusion Of The Study

This study has effectively demonstrated the application of association rule mining in identifying key purchasing patterns within InstaMart's and local unorganized retail buying transaction data. By utilizing quantitative analysis and visualization techniques, such as treemaps, heatmaps, and word clouds, the findings provided valuable insights into consumer behavior and product associations. The word cloud visualization played a significant role in highlighting frequently occurring items, offering a concise representation of dominant purchase trends. These insights enabled the formulation of actionable recommendations to optimize product placement, inventory management, and cross-selling opportunities.

This research directly benefits local grocery store owners and managers in metropolitan India by providing data-driven insights into their customers' purchasing patterns through Market Basket Analysis using the Apriori algorithm. These insights can be immediately applied to optimize product placement and inform targeted promotional campaigns to increase sales and improve customer satisfaction. Academically, this study contributes a specific case of applying a prominent data mining technique in the unique Indian local grocery retail context, adding valuable knowledge to retail analytics in emerging markets. The scope of this research involves analyzing transactional data from selected stores in major Indian cities to uncover patterns relevant to enhancing retail promotions and understanding customer buying trends. Future research can build upon this by comparing the effectiveness of other algorithms like FP-Growth or ECLAT in this setting and by investigating how these patterns evolve over time or when combined with other data like demographics.

CHAPTER 8 - BIBLIOGRAPHY/REFERENCES

[1] Title: “Market Basket Analysis Using Apriori Algorithm To Find Consumer Patterns In Buying Goods Through Transaction Data (Case Study Of Mizan Computer Retail Stores)”
Author: Qisman, M., Rosadi, R., & Abdullah, A. S.
Source: Journal Of Physics: Conference Series, Vol. 1722, No. 1, P. 012020, Iop Publishing, 2021.

[2] Title: “Association Rules In Data Mining: An Application On A Clothing And Accessory Specialty Store”
Author: Avcilar, M. Y., & Yakut, E.
Source: Canadian Social Science, 10(3), 75-83, 2014.

[3] Title: Using Market Basket Analysis In Management Research
Author: Aguinis, H., Forcum, L. E., & Joo, H.
Source: Journal Of Management, 39(7), 1799-1824, 2013.

[4] Title: “Mining Generalized Association Rules”
Author: Srikant, R., & Agrawal, R.
Source: Future Generation Computer Systems, 13(2-3), 161-180, 1997.

[5] Title: “A New Optimization Model For Market Basket Analysis With Allocation Considerations: A Genetic Algorithm Solution Approach” Author: Heydari, M., & Yousefli, A.
Source: Management & Marketing, 12(1), 1, 2017.

[6] Title: “Exploratory Analysis With Association Rule Mining Algorithms In The Retail Industry”
Author: Hashad, A. A., Khaw, K. W., Alnoor, A., & Chew, X. Y.
Source: Malaysian Journal Of Computing (Mjoc), 9(1), 1746-1758, 2024.

[7] Title: “Real-Time Data Analytics In Retail: A Review Of Usa And Global Practices” Author: Raji, M. A., Olodo, H. B., Oke, T. T., Addy, W. A., Ofodile, O. C., & Oyewole, A. T.
Source: Gsc Advanced Research And Reviews, 18(3), 059-065, 2024.

[8] Title: “A Systematic Review Of Data-Driven Insights In Retail: Transforming Consumer Behavior And Market Trend”
Author: Araf, R. I.

Source: Innovatech Engineering Journal, 1(01), 46-55, 2024.

[9] Title: “Retail Analytics: Store Segmentation Using Rule-Based Purchasing Behavior Analysis”
Author: Bilgic, E., Cakir, O., Kantardzic, M., Duan, Y., & Cao, G. Source: The International Review Of Retail, Distribution And Consumer Research, 31(4), 457-480, 2021.

[10] Title: “Comparison Of Market Basket Analysis Method Using Apriori Algorithm, Frequent Pattern Growth (Fp-Growth) And Equivalence Class Transformation (Eclat)(Case 1 Study: Supermarket “X” Transaction Data For 2 2021)”
Author: Wahyuningsih, R., Suharsono, A., & Iriawan, N.
Source: Business And Finance Journal, 8(2), 192-201, 2023

CHAPTER 9 - ANNEXURES

9.1 Raw Data Collected From Instamart And Local Stores For Apriori Algorithm Implementation

Transaction	Items
1	Priya Red chillies, MTR Chutney, Ito Noodles, COL TB, COL GEM, MAX Fresh, Super flexi, Dead line gel
2	DRY MIX SUGAR, URID ROUND, CLI GOLA, MASUR DP, AA , AACH DOMEX, SURF EW 5 SVT BULLET, SUNPURE, Ponds, Santoor
3	NevzelandArle, INP Fullapple, Greenapple, butterfrutt, Indianoi ange, Papaya, IMP Redpears, IMP Chakota, Sitaphal, BlackGrapes, GreenGrapes, goldkivi, kiwi, implenon, IMP Pears, Sendhoor, Mango, hingtonapple
4	Tata Coffee Powder Grand Filter, Tamarind Pack, Everest Kitchen King Masala, Easter Chilly Powder, Home Lites Match Box, Maggi Rich Tomato Ketchup
5	"SUPREME HARVEST Byadagi Chilli Whole, Mtr Kashmiri Chilli, Milky Mist Paneer, Amul Fresh Cream, Klf Coconad Coconut Milk Powder"
6	Odonil Air Freshner Blocks, Fortune Health Rice Bran Refined Oil Pouch, Fortune Sunlite Refined Sunflower Oil, Sundrop Peanut Butter Creamy, Sambhar Cucumber (Samber Sowthe), Nutella Hazelnut Chocolate Spread, Supreme Harvest Coriander Whole (Dhania), Godrej Genteel Matic Liquid Detergent Re
7	Vim Dishwash Bar Lemon, Supreme Harvest Jeera & Methi & Mustard..., Thums Up Soft Drink Bottle, Levista Strong , Amul Processed Cheese Slices
8	Godrej Genteel Matic Liquid Detergent..., LaTranche Butter Croissant (Freshly Made), Veeba Veg Mayonnaise Tandoori, Harpic Power Plus Stain Removal Toilet..., Pril Kraft Dish Washing Liquid (Mint)
9	Lay's India's Magic Masala Chips, Gatorade Blue Bolt Zero Sugar, Lay's Potato Chips - Chile Limon, Drumstick Leaves (Nugge Soppu), Smoor Blueberry Muffins, Theobroma Pound Cake (Contains Egg), Mtr Puliogare Masala Powder, Gatorade Orange Zero Sugar
10	Fortune Sunlite Refined Sunflower Oil, Wow Skin Science Organic Apple Cider..., India Gate Basmati Rice - Classic, Odonil Gel Pocket Mix Air Freshener
11	Cadbury Dairy Milk Silk Oreo Chocolate Bar, Hershey's Caramel Syrup, Thums Up Soft Drink Bottle, Lay's India's Magic Masala Chips, Tata Sampann Unpolished Toor/Arhar Dal, Hershey's Hot Chocolate Drink Powder Mix, Lay's Potato Chips - Chile Limon, Clean Champ Flush Matic Pouch
12	Maggi Creamy Coconut Milk Powder, Cadbury Hot Chocolate Powder Drink Mix..., Godrej Protekt Germ Fight Hand Wash Refil..., Odonil Air Freshener Zipper Mix, Pril Kraft Dish Washing Liquid (Mint), Aashirvaad Superior MP Atta, Nutella Hazelnut Chocolate Spread, Nandini Pure Cow Ghee Pouch, Thums Up Soft Drink Bottle, Amul Processed Cheese Slices, Amul Garlic & Herbs Butter, Fortune Health Rice Bran Refined Oil Pouch, Fortune Sunlite Refined Sunflower Oil

13	"Tata Coffee Powder Grand Filter Gr, Tamarind Pack, Everest Kitchen King Masala , Easter Chilly Powder , Home Lites Match Box, Maggi Ketchup Rich Tomato "
14	"Baby Cabbage (Yelekosu), Green Chilli (Hasiru Menasinakaayi), Cauliflower (Hoo Kosu), Bisleri Mineral Water, English Oven Sandwich Bread, Baby Banana (Baalehannu), Akshayakalpa Artisanal Organic Set Curd, Sambhar Cucumber (Samber Sowthe), Button Roses (Fresh), Carrot"
15	"Supreme Harvest Brown Chana, Kissan Mixed Fruit Jam, Fortune Maida, Wheel Active Clean & Fresh Detergent..., Bournvita Chocolate Nutrition Drink Pouch, Baby Banana (Baalehannu), Milky Mist Cooking Unsalted Butter, Udhaiyam Fried Grams, English Oven Sandwich Bread, Clean Champ Disinfectant Jasmine Toilet..., Supreme Harvest Chilli Guntur Whole Stemless, Stayfree Dry Max Ultra Dry All Night Sanitary..., Tata Sampann Unpolished Chana Dal"
16	"Drumstick (Nuggekaayi), Long Purple Brinjal (Udda Badanekaayi), Tata Sampann Unpolished Toor/Arhar Dal, Akshayakalpa Artisanal Organic Set Curd, Hybrid Tomato, Madhur Pure & Hygienic White Sugar, Amul High Aroma Cow Ghee, Peeled Sambhar Onion by Urban Harvest, Herbs Mix, Gourmet Garden Organic Certified Ginger, Parle Parle-G Original Glucose Biscuits, Nescafe Classic Pure Coffee -Instant..., Taj Mahal Rich and Flavourful Tea, Fresh Price Crash leaflet Bangalore, Baby Bottle Gourd (Sorekaayi), India Gate Basmati Rice Feast Rozzana"
17	"Organic Certified Hybrid Tomato, Long Purple Brinjal (Udda Badanekaayi), Garlic Peeled by Urban Harvest, Baby Banana (Baalehannu), Akshayakalpa Artisanal Organic Set Curd, Supreme Harvest Fenugreek Seeds (Methi), Tender coconut, Coriander - Without Roots (Kotthambari), Heritage Daily Health Toned Milk "
18	Heritage Daily Health Toned Milk, Taj Mahal Rich and Flavourful Tea, Akshayakalpa Artisanal Organic Set Curd, Baby Bottle Gourd (Sorekaayi), Baby Banana (Baalehannu), Supreme Harvest Mustard Big Whole, Sambhar Cucumber (Samber Sowthe), Amul Pasteurised Butter, English Oven Sandwich Bread, Shevanti Flower Mix (Fresh)
19	Gourmet Garden Organic Certified Lemon, Green Cucumber (Hasiru Soutekaayi), Heritage Special Milk, Modern Supreme Sandwich Bread, Robusta Banana (Pachha Baalehannu), Mustard Leaves (Saasive Soppu)
20	Cabbage (Yelekosu), Baby Banana (Baalehannu), Cauliflower (Hoo Kosu), Heritage Daily Health Toned Milk, Ginger (Shunti), Long Purple Brinjal (Udda Badanekaayi), Hybrid Tomato, Green Chilli (Hasiru Menasinakaayi), Heritage Daily Health Toned Milk
21	Akshayakalpa Artisanal Organic Set Curd, Spinach - Without Roots (Palak Soppu), Lay's Potato Chips - Spanish Tomato Tango, Curry Leaves (Karibevu), Kurkure Green Chutney Style Crisps, Paper Boat Aamras Fruit Drink, Green Chilli (Hasiru Menasinakaayi)

22	Parle Parle-G Original Glucose Biscuits, Heritage Daily Health Toned Milk, Lady's Finger (Bendekaayi), Akshayakalpa Artisanal Organic Set Curd, Delfrez Suguna Nourish - Vitamins & Minerals..., Baby Banana (Baalehannu), English Oven Sandwich Bread, Spinach - Without Roots (Palak Soppu), Mint - Without Roots (Pudina), Green Chilli (Hasiru Menasinakaayi), Fresh Price Crash leaflet Bangalore
23	GRB Townbus Butter Murukku, Amul Masti Spiced Buttermilk, Kaveri's Kitchen Murmura Mixture (Freshly...), BRB Popped Potato Chips Pasta Cheese Flavour, Amul High Aroma Cow Ghee, Haldiram's Nagpur Salted Peanut POUCH, English Oven Sandwich Bread, Kurkure Green Chutney Style Crisps, Bisleri Mineral Water, Delfrez Suguna Nourish - Vitamins & Minerals..., Akshayakalpa Artisanal Organic Set Curd, Kurkure Namkeen Masala Munch, Paper Boat Aamras Fruit Drink, Kissan Fresh Tomato Ketchup
24	Arun Vanilla Ice Cream Tub, BRB Popped Potato Chips Pasta Cheese Flavour, Baby Banana (Baalehannu), Baskin Robbins Choco Fudge Doublet Bar, Amul Sandwich Vanilla Ice Cream, Amul Gold Frostik Ice Cream Stick, Akshayakalpa Artisanal Organic Set Curd, Coriander - Without Roots (Kotthambari), Arun Chocobar Ice Cream Stick, Grameen Stick Kulfi Desi Malai, Heritage Daily Health Toned Milk
25	Milky Mist Paneer, Drumstick (Nuggekaayi), Mint - Without Roots (Pudina), Ooty Beetroot, Peeled Sambhar Onion by Urban Harvest, English Oven Sandwich Bread, Baby Banana (Baalehannu), Kateri Brinjal (Geeru Gundu Badanekaayi), Ooty Potato (Aloo Gadde), White Radish (Moolangi), Bottle Gourd (Sorekaayi), Gourmet Garden Organic Certified Ginger, Hybrid Tomato, Herbs Mix, Akshayakalpa Artisanal Organic Set Curd
26	"Organic Certified Ginger, Green Chilli (Hasiru Menasinakaayi), Supreme Harvest Crystal Sugar, Gourmet Garden Patented Naturoponic Palak..., Amul High Aroma Cow Ghee, Spinach - Without Roots (Palak Soppu), Green Amaranthus (Harive Soppu), Taj Mahal Rich and Flavourful Tea, Deep Rooted Cauliflower"
27	Onion (Eerulli), Nandini Shubam Milk, Parry's White Label Sugar, Hybrid Tomato, Himalaya Kolfet Cough Syrup , Milky Mist Paneer, Taj Mahal Rich and Flavourful Tea, L'Oreal Paris Moisture Filling Shampoo, With..., Dabur Honitus Syrup, Vim Lemon Dishwash Gel
28	"Coriander - Without Roots (Kotthambari), Delfrez Suguna Nourish - Vitamins & Minerals..., Akshayakalpa Artisanal Organic Set Curd, Sunpure Refined Sunflower Oil, Amul High Aroma Cow Ghee, Baby Banana (Baalehannu), Fortune Sonamasoori Supreme Rice, Aashirvaad Select Sharbati Atta"
29	Eggoz White Farm Fresh Eggs, Akshayakalpa Artisanal Organic Set Curd, Modern Supreme Sandwich Bread, Nandini Shubham Milk
30	"Akshayakalpa Artisanal Organic Set Curd,Bisleri Mineral Water,Veet Profession was strips for dry skin,Modern Supreme sandwhich bread,Robusta Bannana(Pachha Baalehannu), Flyer InstaMag Feb edition, Akshayakalpa Artisanal Organic Set curd"

9.2Apriori Algorithm Coding Implementation In Google Colab

```
# Import necessary modules
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules
import pandas as pd
from google.colab import files
import pandas as pd

# Upload file from your system
uploaded = files.upload()

# Read the uploaded file
df = pd.read_excel(list(uploaded.keys())[0])
print(df.head())

#Dropping columns with 'Unnamed' in their names
df = df.loc[:, ~df.columns.str.contains('^Unnamed', na=False)]

print("After removing unnecessary columns:")

#Viewing first 5 columns
print(df.head())

#Removing rows where all values are NaN
df = df.dropna(how='all')

# Removing columns where all values are NaN
df = df.dropna(axis=1, how='all')

print("After removing empty rows and columns:")

#Viewing meta data
print(df.info())


#Removing rows with NaN in the 'Item' column
transactions_cleaned = df.dropna(subset=['Items'])
print("After removing rows with missing items:")
print(transactions_cleaned.info())

# Grouping items by transaction
transaction_lists = transactions_cleaned.groupby('Transaction')['Items'].apply(list)
print("Transaction lists grouped by Transaction ID:")
print(transaction_lists.head())
```

```
# Converting the transaction lists into a one-hot encoded DataFrame
from mlxtend.preprocessing import TransactionEncoder

# Using TransactionEncoder to transform the transaction lists
te = TransactionEncoder()
te_ary = te.fit(transaction_lists).transform(transaction_lists)
df_encoded = pd.DataFrame(te_ary, columns=te.columns_)

from mlxtend.frequent_patterns import apriori

# Define all thresholds from 0.01 to 0.09
support_thresholds = [round(0.01 * i, 2) for i in range(1, 10)]

# Store results for each threshold
results = {}

for threshold in support_thresholds:
    frequent_itemsets = apriori(df_encoded, min_support=threshold, use_colnames=True)
    results[threshold] = frequent_itemsets

# Display results for each threshold
for threshold, itemsets in results.items():
    print(f"Results for support threshold {threshold}:")
    print(itemsets)
    print("\n")

#Frequent itemsets
from mlxtend.frequent_patterns import apriori

frequent_itemsets = apriori(df_encoded, min_support=0.01, use_colnames=True)
print("Frequent Itemsets:")


print(frequent_itemsets)

#Wordcloud of Frequent itemsets
from wordcloud import WordCloud
import matplotlib.pyplot as plt

# Convert itemsets to a string for word cloud
itemsets_text = " ".join([" ".join(map(str, itemset)) for itemset in frequent_itemsets['itemsets']])

# Generate the word cloud
wordcloud = WordCloud(width=800, height=400, background_color='white',
colormap='viridis').generate(itemsets_text)

# Display the word cloud
plt.figure(figsize=(10, 6))
```

```
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.title('Word Cloud of Frequent Itemsets', fontsize=16)
plt.show()


#Wordcloud of grouped itemsets
from wordcloud import WordCloud
import matplotlib.pyplot as plt

# Define clusters and their items grouped together
grouped_items = {
    "Fruits": ["Banana 0.645", "Mango 0.445", "Fullapple", "Baby Cabbage", "Green Chilli"],
    "Dairy": ["Akshayakalpa Curd", "Spinach", "Bisleri Water", "Milky Mist Paneer", "Drumstick"],
    "Snacks": ["Cadbury Silk", "Hershey's Syrup", "Arun Ice Cream", "BRB Chips", "Lay's Chips", "Gatorade"],
    "Staples": ["Tata Coffee", "Tamarind Paste", "Fortune Oil", "Wow Skin Science", "Vim Dishwash Bar",
"Supreme Jeera"],
    "High-Support Singles": ["0.645", "1", "0.200", "0.445", "0.450"],
    "Low-Support Singles": ["0.845", "0.900", "1.290", "2"]
}


# Combine grouped items into a single string, assigning cluster labels
grouped_text = " ".join([f"{group}: {item}" for group, items in grouped_items.items() for item in items])

# Generate the word cloud
wordcloud = WordCloud(width=800, height=400, background_color='white',
colormap='viridis').generate(grouped_text)

# Display the word cloud
plt.figure(figsize=(12, 8))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')  # No axes for cleaner visualization
plt.title('Word Cloud of Grouped Itemsets', fontsize=16)
plt.show()


#Tremap of support tiers

import matplotlib.pyplot as plt
!pip install squarify
import squarify


# Group itemsets by support tiers
tier_data = {'High Support (>5%)': 0.245614,  # Combined support of high-tier items (e.g., 0.645 + 1)
          'Medium Support (3–5%)': 0.105264,  # Combined support of mid-tier items (e.g., 0.200, 0.445, 0.450)
          'Low Support (<3%)': 0.649122}  # Combined support of all low-tier items
```

```
labels = [f"{tier} ({support:.2%})" for tier, support in tier_data.items()]
sizes = list(tier_data.values())

# Generate treemap
plt.figure(figsize=(8, 6))
squarify.plot(sizes=sizes, label=labels, alpha=0.8, color=['lightblue', 'orange', 'pink'])
plt.title("Treemap of Support Tiers", fontsize=16)
plt.axis('off')
plt.show()



#Simplified treemap of aggregated categories
import matplotlib.pyplot as plt
import squarify

# Simplified grouping for categories
categories = {
    "Fruits": 0.070176,  # Combined support of all fruit-related items
    "Dairy": 0.052632,   # Combined support of all dairy-related items
    "Snacks": 0.070176,  # Combined support of all snack-related items
    "Staples": 0.052632, # Combined support of all staple-related items
    "High-Support Singles": 0.280702,  # Combined high-support individual items
    "Low-Support Singles": 0.070176    # Combined low-support individual items
}

# Prepare labels and sizes for the treemap
labels = [f"{category} ({support:.2%})" for category, support in categories.items()]
sizes = list(categories.values())

# Generate the treemap
plt.figure(figsize=(10, 6))
squarify.plot(sizes=sizes, label=labels, alpha=0.8, color=[
    'skyblue', 'orange', 'green', 'purple', 'pink', 'yellow'
])
plt.title("Simplified Treemap of Aggregated Categories", fontsize=16)
plt.axis('off')
plt.show()


#Comprehensive Hierarchical treemap
import matplotlib.pyplot as plt
import squarify
```

```
# Hierarchical data grouped by themes and subcategories
itemsets = [
    # Fruits
    ("Fruits > Banana 0.645", 0.017544),
    ("Fruits > Fullapple", 0.017544),
    ("Fruits > Mango 0.445", 0.017544),
    ("Fruits > Baby Cabbage, Green Chilli", 0.017544),


    # Dairy
    ("Dairy > Akshayakalpa Curd, Spinach", 0.017544),
    ("Dairy > Akshayakalpa Curd, Bisleri Water", 0.017544),
    ("Dairy > Milky Mist Paneer, Drumstick", 0.017544),

    # Snacks
    ("Snacks > Cadbury Silk, Hershey’s Syrup", 0.017544),
    ("Snacks > Arun Ice Cream, BRB Chips", 0.017544),
    ("Snacks > Lay's Chips, Gatorade", 0.017544),

    # Staples
    ("Staples > Tata Coffee, Tamarind Paste", 0.017544),
    ("Staples > Fortune Oil, Wow Skin Science", 0.017544),
    ("Staples > Vim Dishwash Bar, Supreme Jeera", 0.017544),

    # High-Support Singles
    ("High-Support Singles > 0.645", 0.122807),
    ("High-Support Singles > 1", 0.087719),
    ("High-Support Singles > 0.200", 0.035088),
    ("High-Support Singles > 0.445", 0.035088),
    ("High-Support Singles > 0.450", 0.035088),

    # Low-Support Singles
    ("Low-Support Singles > 0.845", 0.017544),
    ("Low-Support Singles > 0.900", 0.017544),
    ("Low-Support Singles > 1.290", 0.017544),
    ("Low-Support Singles > 2", 0.017544)
]

# Prepare labels and sizes for the treemap
labels = [f"{item}" for item, support in itemsets]
sizes = [support for _, support in itemsets]
```

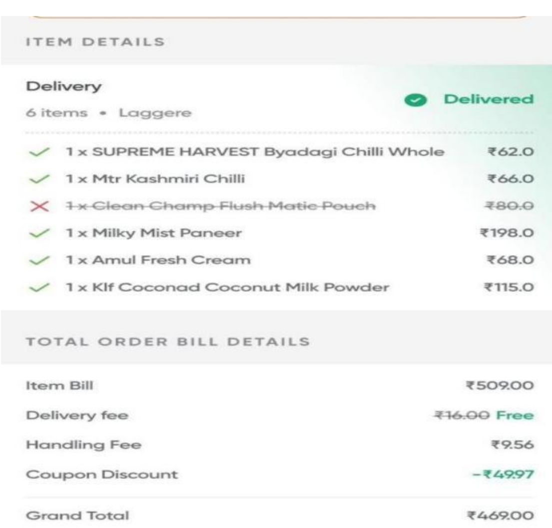
```
# Generate the treemap
plt.figure(figsize=(12, 8))
squarify.plot(sizes=sizes, label=labels, alpha=0.8, color=[
    'skyblue', 'orange', 'green', 'purple', 'pink', 'yellow', 'red',
    'cyan', 'lightgreen', 'peachpuff', 'salmon', 'lavender', 'gold',
    'lime', 'orchid', 'turquoise', 'beige', 'wheat', 'lightcoral'
])
plt.title("Comprehensive Hierarchical Treemap (No Items Ignored)", fontsize=16)
plt.axis('off')  # Hide axes for a cleaner look
plt.show()

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# input data, all arrays have equal length
data = {
    'Item A': [
        "(0.200)", "(0.445)", "(0.450)", "(0.645)", "(0.845)", "(0.900)", "(1)", "(1.290)",
        "Akshayakalpa Artisanal Organic Set Curd", "Baby Cabbage (Yelekosu)", "Banana 0.645",
        "Cabbage (Yelekosu)", "Cadbury Dairy Milk Silk Oreo Chocolate Bar", "Coriander - Without Roots 
(Kotthambari)"
    ],
    'Item B': [
        "(0.445)", "(0.450)", "(0.645)", "(0.845)", "(0.900)", "(1)", "(1.290)", "(0.200)",
        "Spinach", "Green Chilli (Hasiru)", "Milk", "Baby Banana (Baalehannu)", "Honey", "Delfresco Pizza"
    ],
    'Support': [
        0.035088, 0.035088, 0.035088, 0.122807, 0.017544, 0.017544, 0.087719, 0.017544,
        0.017544, 0.017544, 0.017544, 0.017544, 0.017544, 0.017544
    ]
}

# Convert to DataFrame
df = pd.DataFrame(data)

# Pivot for heatmap
pivot = df.pivot(index="Item A", columns="Item B", values="Support")

# Generate heatmap
plt.figure(figsize=(15, 12))
sns.heatmap(pivot, annot=True, cmap='YlGnBu', fmt=".5f", linewidths=0.5, cbar_kws={'label': 'Support'})
plt.title("Heatmap of Frequent Itemsets", fontsize=16)
plt.xlabel("Item B", fontsize=12)
plt.ylabel("Item A", fontsize=12)
plt.xticks(rotation=45, ha="right")
plt.yticks(rotation=0)
plt.tight_layout()
```

```
plt.show()


#Heatmap of frequent itemsets
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt


data = {
   'Item A': [
      "(0.200)", "(0.445)", "(0.450)", "(0.645)", "(0.845)", "(0.900)", "(1)", "(1.290)",
      "Akshayakalpa Artisanal Organic Set Curd", "Baby Cabbage (Yelekosu)", "Banana 0.645",
      "Cabbage (Yelekosu)", "Cadbury Dairy Milk Silk Oreo Chocolate Bar", "Coriander - Without Roots
(Kotthambari)"
   ],
   'Item B': [
      "(0.445)", "(0.450)", "(0.645)", "(0.845)", "(0.900)", "(1)", "(1.290)", "(0.200)",
      "Spinach", "Green Chilli (Hasiru)", "Milk", "Baby Banana (Baalehannu)", "Honey", "Delfresco Pizza"
   ],
   'Support': [
      0.035088, 0.035088, 0.035088, 0.122807, 0.017544, 0.017544, 0.087719, 0.017544,
      0.017544, 0.017544, 0.017544, 0.017544, 0.017544, 0.017544
   ]
}

# Convert to DataFrame
df = pd.DataFrame(data)

# Pivot for heatmap
pivot = df.pivot(index="Item A", columns="Item B", values="Support")

# Generate heatmap
plt.figure(figsize=(15, 12))
sns.heatmap(pivot, annot=True, cmap='YlGnBu', fmt=".5f", linewidths=0.5, cbar_kws={'label': 'Support'})
plt.title("Heatmap of Frequent Itemsets", fontsize=16)
plt.xlabel("Item B", fontsize=12)
plt.ylabel("Item A", fontsize=12)
plt.xticks(rotation=45, ha="right")
plt.yticks(rotation=0)
plt.tight_layout()
plt.show()
```

```
#Heatmap of product category pairings
import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt

# Categorized data with support values
data = {
    'Category A': [
        "Fruits", "Fruits", "Vegetables", "Vegetables", "Dairy", "Dairy", "Snacks", "Snacks"
    ],
    'Category B': [
        "Vegetables", "Dairy", "Dairy", "Snacks", "Fruits", "Snacks", "Fruits", "Vegetables"
    ],
    'Support': [0.12, 0.08, 0.05, 0.07, 0.06, 0.09, 0.04, 0.03]
}

# Convert to DataFrame
df = pd.DataFrame(data)

# Pivot for heatmap
pivot = df.pivot(index="Category A", columns="Category B", values="Support")
```

```python
# Generate heatmap
plt.figure(figsize=(10, 8))
sns.heatmap(
    pivot, annot=True, cmap='coolwarm', fmt=".2f", linewidths=0.5,
    cbar_kws={'label': 'Support'}, annot_kws={"size": 10}
)
plt.title("Heatmap of Product Category Pairings", fontsize=16)
plt.xlabel("Category B", fontsize=12)
plt.ylabel("Category A", fontsize=12)
plt.xticks(rotation=45, ha="right", fontsize=10)
plt.yticks(fontsize=10)
plt.tight_layout()
plt.show()
```

9.3 Sample Bills Used To Collect Raw Data



ITEM DETAILS

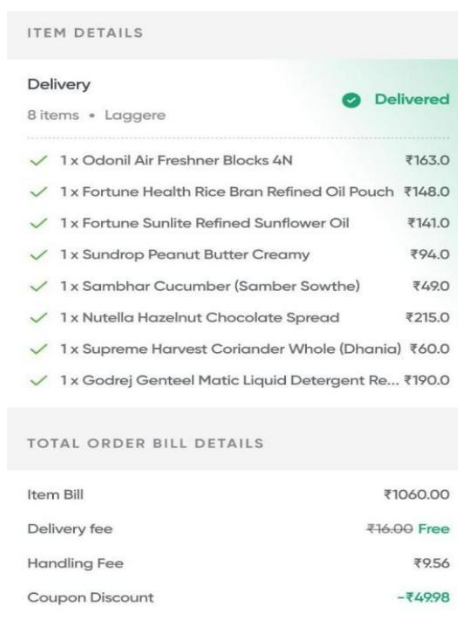
Delivery — Delivered

6 items • Laggere

1 x SUPREME HARVEST Byadagi Chilli Whole	₹62.0
1 x Mtr Kashmiri Chilli	₹66.0
~~1 x Clean Champ Flush Matic Pouch~~	~~₹80.0~~
1 x Milky Mist Paneer	₹198.0
1 x Amul Fresh Cream	₹68.0
1 x Klf Coconad Coconut Milk Powder	₹115.0

TOTAL ORDER BILL DETAILS

Item Bill	₹509.00
Delivery fee	~~₹16.00~~ Free
Handling Fee	₹9.56
Coupon Discount	-₹49.97
Grand Total	₹469.00



ITEM DETAILS

Delivery — Delivered

8 items • Laggere

1 x Odonil Air Freshner Blocks 4N	₹163.0
1 x Fortune Health Rice Bran Refined Oil Pouch	₹148.0
1 x Fortune Sunlite Refined Sunflower Oil	₹141.0
1 x Sundrop Peanut Butter Creamy	₹94.0
1 x Sambhar Cucumber (Samber Sowthe)	₹49.0
1 x Nutella Hazelnut Chocolate Spread	₹215.0
1 x Supreme Harvest Coriander Whole (Dhania)	₹60.0
1 x Godrej Genteel Matic Liquid Detergent Re...	₹190.0

TOTAL ORDER BILL DETAILS

Item Bill	₹1060.00
Delivery fee	~~₹16.00~~ Free
Handling Fee	₹9.56
Coupon Discount	-₹49.98

PHONE :9964307800,8050736800

SIDDALINGESHWARA RICE CORNER

#447, 22nd Main Road, Nandini Layout, B'lore-96
BILL NO : 18920 DATE : 05/03/2025 12:22
Name : CASH SALES

#	Product	MRP	Qty	Rate	Amount
1	DRY MIX 250	190.00	2.000	145.00	290.00
2	SUGAR	48.00	2.000	43.00	86.00
3	URID ROUND	165.00	0.500	140.00	70.00
4	CLI 1RS	1.00	20.000	0.92	18.40
5	MASUR GOLA	150.00	1.000	120.00	120.00
6	DOMEX T 500	102.00	0.500	86.00	43.00
7	AA DP 100G	39.00	1.000	30.00	30.00
8	AACH CP 100	42.00	1.000	37.00	37.00
9	25RS SOYA S	25.00	1.000	24.00	24.00
10	10RS PONDS	10.00	5.000	9.50	47.50
11	SANTOOR WHI	36.00	2.000	37.00	74.00
12	SURF EW 500	76.00	1.000	74.00	74.00
13	SVT BULLTE	75.00	26.000	66.00	1716.00
14	5L SUNPURE	825.00	1.000	740.00	740.00

E & O.E. Total : 3370.00

YOUR SAVINGS : 488.60

. ***THANK YOU VISIT AGAIN***

#	Product	MRP	Qty	Rate	Amount
1	5RS MUNCH	5.00	3.000	5.00	15.00
2	AVALAKKI	70.00	0.500	60.00	30.00
3	CHIK 1RS	1.00	16.000	0.90	14.40
4	COL SALT 10	76.00	1.000	74.00	74.00
5	COMFORT 860	235.00	1.000	225.00	225.00
6	DHEEPAM 1L	206.00	1.000	175.00	175.00
7	DHOOP 12	12.00	1.000	12.00	12.00
8	EXO 125	10.00	4.000	9.50	38.00
9	F GRAM	160.00	0.250	115.00	28.75
10	GRAM DALL	140.00	0.250	105.00	26.25
11	GN OIL 1L	160.00	2.000	149.00	298.00
12	KARPUR 50G	80.00	1.000	50.00	50.00
13	KISSAN S&S	80.00	1.000	78.00	78.00
14	LEVISTA 200	190.00	1.000	180.00	180.00
15	METHI 100 G	25.00	1.000	20.00	20.00
16	MUSTARD 100	25.00	1.000	20.00	20.00
17	P.CHUTE 250	110.00	1.000	106.00	106.00
18	PATTA	750.00	0.050	500.00	25.00
19	PEPPER 50G	75.00	1.000	50.00	50.00
20	RED LBL 250	155.00	1.000	150.00	150.00
21	RICE FLOUR	65.00	1.000	55.00	55.00
22	SANTOOR 100	39.00	1.000	38.00	38.00
23	SHIVALING	200.00	0.500	140.00	70.00
24	SOMPU	96.00	0.500	70.00	35.00
25	SUGAR	48.00	2.000	44.00	88.00
26	SURF BAR 22	20.00	4.000	19.00	76.00
27	SURF LIQ FL	199.00	1.000	190.00	190.00
28	ULLAS BIG	110.00	1.000	95.00	95.00
29	URID DALL	175.00	0.250	130.00	32.50
30	WHEEL POW 3	39.00	1.000	38.00	38.00
31	YIPPEE 14	15.00	4.000	14.50	58.00

O.E. Total : 2391.00

YOUR SAVINGS : 296.35

*THANK YOU VISIT AGAIN***