

Examination of Consumer's buying tendencies

A PROJECT REPORT

for

DATA MINING TECHNIQUES (ITE2006)

in

B.Tech – Information Technology

by

Master Muskan (20BIT0448)

Manas Khandelwal (20BIT0251)

Fall Semester 2022 – 2023

Under the Guidance of

Prof. Dr. B VALARMATHI

Associate Professor(Senior), SITE



VIT[®]
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

School of Information Technology and Engineering

November, 2022

DECLARATION BY THE CANDIDATE

We here by declare that the project report entitled “**Examination of consumer’s buying tendencies**” submitted by us to Vellore Institute of Technology University, Vellore in partial fulfillment of the requirement for the award of the course **Data Mining Techniques (ITE2006)** is a record of bonafide project work carried out by us under the guidance of **Prof. B.Valarmathi**. We further declare that the work reported in this project has not been submitted and will not be submitted, either in part or in full, for the award of any other course.

Place: Vellore

Signature

Date: 12th November, 2022



VIT[®]
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

School of Information Technology & Engineering [SITE]

CERTIFICATE

This is to certify that the project report entitled “**Examination of consumer’s buying tendencies**” submitted by **Manas Khandelwal (20BIT0251)**, **Master Muskan (20BIT0448)**, to Vellore Institute of Technology University, Vellore in partial fulfillment of the requirement for the award of the course **Data Mining Techniques (ITE2006)** is a record of bonafide work carried out by them under my guidance.

Prof. B.Valarmathi

GUIDE

Associate Professor (Senior), SITE

Examination of Consumer's buying tendencies

Manas Khandelwal(20BIT0251)

Master Muskan(20BIT0448)

Department of Information Technology, VIT University, Vellore, Tamil Nadu, India

Abstract

Data mining methods like market basket analysis are used to identify relationships between datasets. A link between a large number of data objects is found through association rule mining. Numerous companies are worried about mining association rules from their databases as a result of the continuous acquisition and storage of vast amounts of data. By finding relationships between various things that buyers place in their shopping baskets, market basket research investigates customer buying habits and trends. Analysing client purchase patterns is beneficial and helps to boost sales. Our system will thus create association rules between itemsets using either the Apriori algorithm, the FP Growth method, or the ECLAT (Equivalence Class Transformation) algorithm in order to design a system for client buying habits of shopping malls, marts, etc. All of these solutions will be compared, and the one with the highest accuracy will be put into practice. This system will assist a market expert in making decisions. In the base paper, they only calculated support, confidence and lift values but we are also calculating conviction

The base paper had a runtime of 382 ms for Apriori algorithm and 288 ms for FP-Growth algorithm.

Keywords – Association, Apriori, FP growth, ECLAT, EDA, support, confidence, lift, leverage, conviction

1. Introduction

Retailers market basket analysis, a methodology for data analysis based on association data mining, to boost sales by better understanding client buying habits. Large data sets, such purchase histories, must be analysed to identify product groups and items that are most likely to be bought together. Market basket analysis reveals the buying habits of consumers by identifying significant relationships between the goods they put in their shopping baskets. In many company organisations, it not only aids in decision-making but also boosts sales. The most popular algorithms for mining frequent itemsets are

apriori and FP growth. Market Basket Analysis is a crucial component of the analytical system used in retail organisations to locate products, create sales promotions for various client categories, and ultimately increase customer happiness and the profitability of supermarkets.

2. Background

In our project, we have utilised the idea of association rule mining.

At its most basic level, association rule mining uses machine learning models to search a dataset for patterns or co-occurrences. It distinguishes frequent if-then relationships, which are the association rules in and of themselves.

A consequent and an antecedent (if) make up an association rule (then). An antecedent is anything that can be located in the data. An item discovered in conjunction with the antecedent is known as a consequent (X).

Measures of an association's rules' efficacy include:

Two key factors—support and confidence—are used to assess an association rule's strength. How frequently a certain rule appears in the database being mined is referred to as support. The frequency with which a particular rule really holds true in practise is referred to as confidence. A rule may exhibit a significant connection in a data set because it occurs frequently, but it may not occur as frequently when put into practise. In this situation, there would be great support but little confidence.

On the other hand, a rule might not seem to stand out much in a data collection, but further investigation reveals that it happens rather frequently. In this situation, there would be enormous confidence but little backing. These metrics enable analysts to distinguish between correlation and causation and to appropriately assess a particular rule.

The ratio of confidence to support is a third value component, sometimes referred to as the lift value. There is a negative correlation between the datapoints if the lift value is negative. A positive correlation exists if the value is positive, and if the ratio is equal to 1, there is no connection.

3. Literature Survey

S. No.	Title of the Paper and Year	Algorithms used	Data Set being used	Performance measures	Scope for future work
1.	Comparison of Apriori, Apriori-TID and FP-Growth Algorithms in Market Basket Analysis at Grocery Stores (2022)	Apriori, Apriori – TID, FP – Growth	SPMF open source	92%	Necessary to compare with other algorithms in similar case
2.	Consumer purchase patterns based on market basket analysis using apriori algorithms (2019)	Apriori, Association Analysis	Shop data set	Support of 0.0567	Accuracy can be improved
3.	A Study of Market Basket Analysis using Item Set Mining (2022)	Frequent Item Sets, Apriori	Seller and consumer data set	Confidence of 32% achieved	Comparison of more data sets

4.	Market basket analysis using apriori algorithm to find consumer patterns in buying goods through transaction data (case study of Mizan computer retail stores) (2020)	Apriori, High frequency pattern analysis	case study of Mizan computer retail stores	100% confidence value achieved	Can increase lift ratio
5.	Clustering Based Approach to Enhance Association Rule Mining (2021)	Apriori, Machine Learning	Market receipt	Support of 0.76	Can increase the percentage of sales
6.	Application of market–basket analysis on healthcare (2021)	Apriori, FP Tree, Naïve Bayes	Patients Electronic Health Records	Minimum support count of 95%	Can make it large to analyse more patient's data

7.	Market Basket Analysis Using Apriori and FP-Growth for Analysis Consumer Expenditure Patterns at Berkah Mart in Pekanbaru Riau (2018)	Apriori, FP Growth Algorithm, Association rule mining	Data set Berkah Mart	Confidence value of 43%	Can work more on accuracy
8.	Market Basket Analysis Using Apriori and FP Growth Algorithm (2019)	Apriori, FP Growth	Analysis using French retail data set	Time for FP Growth was 1798 ms	Reduction of items can bring more accuracy
9.	A Seasonal and Multilevel Association Based Approach for Market Basket Analysis in Retail Supermarket (2021)	Apriori, FP Growth, K – Apriori	Retail Supermarket Data set	74.3%	Alternate techniques to reduce execution time

10.	Structured Critical Review on Market Basket Analysis using Deep Learning & Association Rules (2021)	FP Growth Algorithm, Deep Learning	Offline and Online Retail Data set	Support of 0.67	Developing hybrid framework due to consumer's changing buying patterns
11.	Market Basket Analysis to Identify Customer Behaviors by Way of Transaction Data (2018)	Association rules, Apriori Algorithm	Business Centre Data set	Confidence value average of around 50%	Low association tendency of patterns
12.	Market Basket Analysis Method on Sales Data Using Fp-Growth Algorithm (2021)	FP Growth Algorithm	Data sets on sales of goods	Confidence value average of greater than 50%	Can take monthly data and use more algorithms
13.	Market basket analysis with association rules (2020)	Apriori, FP Growth	Supermarket data set	Confidence of 52%	If correctly interpreted then will add profits for supermarket

14.	MARKET BASKET OPTIMIZATION (2022)	Apriori, FP Growth	Retail industry data set	Runtime for Apriori was 2300 ms and FP Growth was 1800 ms approx	More data sets can be collected
15.	MARKET BASKET ANALYSIS: TREND ANALYSIS OF ASSOCIATION RULES IN DIFFERENT TIME PERIODS (2019)	Association Rules, Apriori Algorithm	Consumer's shopping trends	Confidence of 65%	To design and develop an intelligent model
16.	A Comparative Study on Market Basket Analysis and Apriori Association Technique (2017)	Apriori Algorithm	Open source	Support values 0.87 achieved	Algorithms can be changed to get new understandings

17.	Market Basket Analysis using Apriori Algorithm and FPTree Algorithm (2021)	Apriori, FP - Tree Growth Algorithm	layouts on the basis of frequently purchased items in 11supermark et ts and fixing consumer index price as per consumer's	96% accuracy	Accuracy can be improved
18.	Strategic Insights from Optimized Market Basket Analysis in Wordle Game Environment and Its Application in Retail (2020)	Apriori Algorithm	World game scenario data set taken	Confidence of 43% achieved	The subscriptb ased approach in data engineerin g works well when purchase quantities are discrete and low
19.	Temporal Association Rule Mining (2018)	FP Tree	Products sample	66.7%	Can apply new mining techniques
20.	Building Prediction Model using Market Basket Analysis (2018)	Apriori Algorithm	Consumers market transaction data set	Confidence values of 54% achieved	Can increase confidence value

Table 1. Literature Survey

1.. Comparison of Apriori, Apriori-TID and FP-Growth Algorithms in Market Basket Analysis at Grocery Stores –

Author's Name - Andi Ilhamsyah Idris , Eliyah A M Sampetoding , Valian Yoga Pudya Ardhana , Irene Maritsa , Adrisumatri Sakri , Hidayatullah Ruslan , Esther Sanda Manapa

Year of Publication – July 2022

Using the Apriori, Apriori-TID, and FP-Growth algorithms, we will attempt to conduct a market basket analysis in this study to determine the association rules for goods that are often bought at the grocery shop for each season. The primary goal of this study is to plan a search for trends in grocery store shopping during each season. Second, monitor customer transactions for buying food at grocery shops during each season using the Apriori algorithm, Apriori TID, and FP-Growth algorithms. Third, compare Apriori, Apriori-TID, and FP-Growth while examining customer purchases of food made during each season at the grocery shop.

In comparison to the other two methods, Apriori has the longest calculation time based on the average performance evaluation for each season. The performance of FP-Growth is superior than the other two algorithms in terms of computing time, but it uses more memory on average.

2.. Consumer purchase patterns based on market basket analysis using apriori algorithms –

Author's name – A R Efrat , R Gernowo and Farikhin

Year of Publication – 2019

The system will generate rules on the applicability of a product based on the pattern of consumer purchases. By learning which products consumers frequently buy together, minimarket may broaden its marketing techniques by analysing customer purchase habits. Another information source available at the convenience store is transaction data, which is one thing that may be utilised for corporate decision-making. In order to gather consumer purchase patterns for consumer purchasing pattern analysis, we plan to employ the Apriori algorithm approach. The Apriori algorithm may do this by searching databases for high-frequency patterns that exceed a certain threshold.

3.. A Study of Market Basket Analysis using Item Set Mining –

Author's name – Divya V. Walve , Nishigandha Dhanorkar , Vaidehi Date ,
Trupti Sable , Prof. T.G.Ghongade

Year of Publication – May 2022

Organizational mining rules are among the most common difficulties in data mining, which uncovers intriguing patterns from archives such as linkages, sequences, classifications, collections, and much more. The organization's mining rules identify intriguing correlations or connections among a substantial number of data elements. The algorithm currently examines this information to look for user feelings connected with each review after first tallying the amount of positive, negative, and neutral reviews. The algorithm currently collects all reviews for a certain product and counts the amount of good, negative, and neutral reviews.

This offers a sentiment analysis-based automatic product rating system. It leads to a rise in product sales and enables anyone to run a successful business.

4.. Market basket analysis using apriori algorithm to find consumer patterns in buying goods through transaction data (case study of Mizan computer retail stores) –

Author's name – M Qisman, R Rosadi , and A S Abdullah

Year of Publication – 2021

Based on the goods in the transaction, the preprocessing algorithm may evaluate transaction data at Mizan Computer Retail Store by identifying association rules that satisfy the minimal support and minimal confidence requirements. To utilise as an input when creating a marketing plan, it is required to create an application that can apply Market Basket Analysis and the Apriori Algorithm for product purchase patterns at the Cirebon Computer Mizan Retail Store.

5.. Clustering Based Approach to Enhance Association Rule Mining

–

Author's name - Samruddhi Kanhere, Anu Sahni

Year of Publication – May 2021

The market receipt will be the application's input when it is implemented, and the market receipt will be sent to the system for pre-processing. The present characteristics are divided up into the dataset variables from the document. This

continual process of figuring out the relative locations of these features and comparing them to the feature-graph database continues until a match is found. The result will be that person's anticipated market behaviour. In general, market basket analysis is often referred to as association rule mining. Understanding consumer patterns, such as which things buyers are choosing in concert, is helpful for marketing enthusiasts.

6.. Application of market–basket analysis on healthcare –

Author's name – Abishek B. Rao, Jammula Surya Kiran, Poornalatha G

Year of Publication – August 2021

The frequent item sets from the outcome provide the commonly occurring diseases, which have common symptoms that aid doctors in prescribing appropriate drugs in accordance with the disease's symptoms. The medical or health authorities can take preventative steps to deal with the potential often happening diseases as the suggested study can provide specifics about regularly occurring diseases depending on year. Additionally, residents might be made aware of these illnesses, their typical symptoms, etc. so they can take the necessary precautions. The objective of this study is to discover intriguing patterns by taking into account the goods that clients purchased in tandem. Analysis of the combined purchases of the customers' items is done to spot any patterns.

7.. Market Basket Analysis Using Apriori and FP-Growth for Analysis Consumer Expenditure Patterns at Berkah Mart in Pekanbaru Riau

–

Author's name – Mustakim , Della Maulina Herianda , Ahmad Ilham , Achmad Daeng GS , Folkes E. Laumal , Nuning Kurniasih , Akbar Iskandar , Gloria Manulangga , Ida Bagus Ary Indra Iswara and Robbi Rahim

Year of Publication – 2018

In comparison to the Apriori Algorithm, the suggested FP-Growth algorithm is more successful in producing frequent and informative association rules to determine customer spending patterns at Berkah Mart in Pekanbaru. Regarding customer happiness, related products are grouped together to provide customers with a logical way to choose things they would like to purchase. In addition, according to the guidelines of the customer association, the products can be divided into several categories to more cost-effectively cater to the unique demands of clients. The experimental findings indicate that using FP-Growth

and Apriori algorithms to analyse consumer spending patterns at Berkah Mart can boost overall revenue or profit, but FP Growth is advised because it has the fastest rule-form processing times and the best support and confidence values of any a priori algorithm.

8.. Market Basket Analysis Using Apriori and FP Growth Algorithm –

Author's name – Maliha Hossain, A H M Sarowar Sattar and Mahit Kumar Paul
Year of Publication – 2019

According to experimental study, utilising the top-selling items for reduction reduces the time needed for both algorithms compared to using all the products. Once more, after product reduction, it provides the same guidelines and very identical frequent item sets for different support levels. Therefore, from our perspective, using item reduction is advantageous since it requires less processing than previously. Again, both with and without product reduction, FP Growth takes less time than Apriori algorithm. Using sampling without replacement, we have also conducted rule analysis, and the findings indicate that we obtain the same rules with more confidence. We may therefore conclude that the elimination of goods allows for the identification of client buying patterns that need less calculation. The range of percentages for product reduction can be determined in the future using further transactional statistics. It will be interesting to examine individual rules using correlation analysis as well.

9.. A Seasonal and Multilevel Association Based Approach for Market Basket Analysis in Retail Supermarket –

Author's name – S. Rana and M. N. I. Mondal
Year of Publication – October 2021

In this research, we present a method for mining overall strong regular rules and seasonal association rules in multilevel data sets. Although the majority of the created rules are reasonably predictable for a retailer, tracking season-based product affinities and capitalising on them is quite challenging for a large mega store where the retailer interacts with the enormous number of items on a daily basis.

However, it's critical that the shop understand precisely what products are bought in tandem and at which season. Such data gives a clearer picture of customer needs, which may be used to drive decisions about various merchandising activities including cross- and up-selling campaigns, inventory management, seasonal promotional assistance, and layout design. Even if the

found patterns and principles are common, they are nevertheless helpful and applicable in retail supermarket. In our upcoming work, we are looking at various search techniques to shorten the execution time and decrease the amount of search space needed for huge datasets.

10.. Structured Critical Review on Market Basket Analysis using Deep Learning & Association Rules –

Author's name – Iqra Rehman, Dr. Hamid Ghous

Year of Publication – January 2021

Market basket analysis is currently a major issue in cross-selling and up-selling. Cross-selling involves boosting product sales and incorporating new features into existing products.

Upselling is the practise of offering additional product categories to draw in more customers. Understanding customer purchasing trends is crucial for cross- and upselling. In this case, market basket research is crucial for identifying strong and weak goods that influence profit and loss.

11.. Market Basket Analysis to Identify Customer Behaviors by Way of Transaction Data –

Author's name – Fachrul Kurniawan, Binti Umayah , Jihad Hammad , Supeno Mardi Susiki Nugroho , Mochammad Hariadi

Year of Publication – January 2018

In addition to aiding in the creation of sales promotions, market basket analysis software should be created and installed at supermarkets so that it may be used as a guide for re-managing item stock's incoming and outgoing in the warehouse. Due to the supermarket's incapacity to use transaction data, this project will construct a market analysis application at BC UIN Malang. It is anticipated that this application will perform effectively and produce the intended outcome. The Apriori algorithm is the implemented algorithm in the market basket analysis application in association rule mode. Using a single item as a starting point, an algorithm is utilised to create a frequent itemset, after which the value support of each item is counted. Items with support values greater than the minimum support value were chosen as candidates for the 1-itemset high frequency pattern and the 2-itemset. Recursively, the evolution of the frequent itemset into the 2-itemset, from which the value of confidence would subsequently be derived, was accomplished by the 1-itemset.

12.. Market Basket Analysis Method on Sales Data Using FP - Growth Algorithm –

Author's name – Kenny Marcelino Irawan, Tina Tri Wulansari, Nariza Wanti Wulan Sari

Year of Publication – 2021

As a result of the research that has been done, it has been determined that goods that are frequently purchased simultaneously are those that can be seen from the association rules that meet the minimum support and minimum confidence values with lift. > 1 . The association rule with the highest confidence value is the purchase of cake mats and a cake stand. Seven association rules with a confidence value over 0.5 are also present, and among these seven rules, wheat flour appears 5 (five) times. Thus, it is advised that store owners exhibit wheat flour among products that are frequently bought together, such as sweetened condensed milk, sugar, powdered margarine, and margarine, on the main display. Additionally, store owners might encourage bundling cake mats and boxes.

13.. Market basket analysis with association rules –

Author's name – Yuksel Akay Unvan

Year of Publication – 2020

The purpose of this study is to carry out a market basket analysis utilising association rules. For this, a data set from a supermarket's website at Vancouver Island University was used (Vancouver Island University 2019). This dataset includes information on the 255 distinct goods that consumers have purchased. After the data have undergone the requisite tests and investigations, they will be examined using the Weka software, together with any appropriate interpretations. As a result, the client will find it more appealing to shop due to the growth in market profits. Market operator will undoubtedly benefit and profit from market basket analysis if it is properly evaluated.

14.. MARKET BASKET OPTIMIZATION –

Author's name – K Jaya Laxmi, C Saikiran Reddy, D Vishnu Sai, K Govardhan

Year of Publication – May 2022

Based on Apriori (using the big itemset property), FP-Growth, and market-basket optimization (mainly for faster).

Through the often created itemsets, we will determine whether there is any relationship between the products, which will assist them increase the associated products at their stores in order to increase sales. This will be done by creating frequent itemsets from the data gathered at the stores.

Several of the most recent data mining techniques for evaluating shopping baskets are presented in this study. According to an analysis of current algorithms, using association rule mining algorithms to evaluate market baskets will help to better classify the vast amount of data. In this study, the results of applying the Apriori and FP Growth techniques on the same datasets were equal. FP improves results more quickly than Apriori due to Apriori's candidate set technique.

15.. MARKET BASKET ANALYSIS: TREND ANALYSIS OF ASSOCIATION RULES IN DIFFERENT TIME PERIODS –

Author's name – Sohaib Zafar Ansari

Year of Publication – February 2019

This study demonstrates that time is a crucial consideration that must be made while carrying out any market basket analysis. It may assist offer really intriguing data or insights about the clients that aid in maximising earnings. Cross-selling and promotional possibilities, for instance, might result from the identification of supplemental or complimentary items.

In comparison to the conventional consumer surveys, which are typically quite expensive and time-consuming, this study is better and more affordable. At every stage of the survey investigation, they include inaccuracies as well. In the future, we may use the time series clustering approach to identify association rules as well. because SIM University's School of Business in Singapore referred to it as a better choice for market basket research.

In order to make the feature more practical, future work should also include designing and creating an intelligent prediction model to create the association rules that may be used on a recommendation system.

16.. A Comparative Study on Market Basket Analysis and Apriori Association Technique –

Author's name – Warnia Nengsih

Year of Publication – 2017

One of the data meaning approaches used to create a system of relationships between different items is market basket analysis. The potential proportion of the relationship between the combined components provides fresh knowledge, making it highly helpful for the decision-maker. Some of the available methods can be used with market basket analysis. This study compares market basket analysis using the apriori algorithm with market basket analysis without an algorithm. The same rule is produced by both approaches that are being compared. Both compared procedures provide the same rule as a consequence of comparison analysis in general.

17.. Data Mining –

Author's name - Paramjit Kaur, Kanwalpreet S Attwa

Year of Publication - 2020

International Journal of Computer Science and Information Technologies This paper presents an overview of the data mining, data mining techniques, data mining process, data mining architecture, data mining advantages/disadvantages, applications. Data mining, it is the process of extracting of previously unknown predictive information from large databases. Basically, the main purpose of data mining is to find out unsuspected associations in observational data and to sum up the result in form which is clear, understandable, useful to data owner.

18.. Strategic Insights from Optimized Market Basket Analysis in Wordle Game Environment and Its Application in Retail –

Author's name - Kannan Sekar,Angela Susan Mathew,Sreejith R,Krishnan Chandramohan

Year of Publication - 2022

Market Basket Analysis is considered powerful in retailing. We propose application of Market Basket Analysis (Association Rule Mining) in Wordle game to find association among letters in the word. This study extends the scope of traditional MBA in two critical ways. First, we expand the focus from combinations of alphabets to their meaningful permutations. For instance, the five alphabets A, E, R, S and T can create at least three distinct baskets, namely; STARE, TASER and TEARS. Second, we formulate a workaround to

incorporate the repetition of alphabets in a word in contrast to the original MBA algorithm which is restricted to the presence or absence of an item. The insights from this optimized Market Basket Analysis are then proposed for application in retail environment, which can help businesses to optimize the association rules while accounting for the quantity of the product purchased. This study also features descriptive analytics to identify interesting patterns of letters in Wordle solutions, including frequency distributions by accounting for positions and repetition of alphabets in solutions. In this manner, the results of the study can enable Wordle players to optimize strategies in playing the game. The authors believe that this is the first academically researched attempt at exploring a wordgame using a data mining approach. The research could be further optimized by making changes in the algorithm to account for permutation as well as for repetition for better generalizability as the logic in the current scenario was into data engineering and cleaning for optimization.

19.. Temporal Association Rule Mining –

Author's name - M. Madaio, R. Lasko, J. Cassell, and A. Ogan

Year of Publication - 2018

Using Temporal Association Rule Mining to Predict Dyadic Rapport in Peer Tutoring, This paper is intended to contribute to the detection of interpersonal social states, such as rapport, through nonverbal, task (verbal) and social (verbal) channels, captured through audio and video input. In this paper, we describe a process for using temporal association rule mining to learn patterns of behaviors from an annotated corpus of nearly 60 hours of dyadic peer tutoring interactions. We then use those temporal association rules to predict the “thin - slice” dyadic rapport level for every 30-second time-slice, via a stacked ensemble model.

20.. Building Prediction Model using Market Basket Analysis –

Author's name - R. Gangurde, D. B. Kumar, and D. S. D. Gore

Year of Publication – 2019

MBA is one of the most popular types of data analysis used in the marketing world. The purpose of Market Basket Analysis is to determine what products are most commonly purchased or used by consumers. This MBA is analyzing consumer buying habits by finding associations between different products that consumers place in shopping basket.

4. Existing Systems –

- Some of the existing systems are less accurate.
- The existing systems require more computation power to function.
- More investment on resources were required for the existing systems.

5. Gaps Identified –

Many earlier market basket analyses simply looked at the retail shop's daily or weekly data sets. We may use the monthly data sets to analyse the market's sales, which will enable the shops to highlight the most popular products. Market baskets can also be utilised in a variety of different industries to boost sales, including healthcare, food courts, retail, and shopping malls. When association rules mining is widely used in the field of recommender systems and item pairings are found to have a positive link, suggestions may be issued to customers in order to boost sales. Additionally, it's possible to introduce them to products they otherwise wouldn't have explored.

6. Dataset Description and Sample Data -

The information provided during any sale transaction on this Delhi Mart, including information on the products purchased, the time of the purchase, the overall volume of sales, and the item price. To make strategic decisions that can boost corporate earnings, management at Delhi Mart need more information, such as sales statistics for the most popular items, marginally more popular items, and seldom purchased items. By examining the data of sales transactions that totaled more than 5000 every day, it is possible to maintain inventory by learning the patterns of customer spending that frequently occur at Delhi Mart. Because it is just based on management's opinion by classifying the current items and has not been examined from the perspective of the consumer, the positioning of the product layout is still less precise and optimum.

turkey	avocado									
mineral w	milk	energy bar	whole whe	green tea						
low fat yogurt										
whole whe	french fries									
soup	light cream	shallot								
frozen veg	spaghetti	green tea								
french fries										
eggs	pet food									
cookies										
turkey	burgers	mineral w	eggs	cooking oil						
spaghetti	champagn	cookies								
mineral w	salmon									
mineral water										
shrimp	chocolate	chicken	honey	oil	cooking oi	low fat yogurt				
turkey	eggs									
turkey	fresh tuna	tomatoes	spaghetti	mineral w	black tea	salmon	eggs	chicken	extra dark	chocolate
meatballs	milk	honey	french frie	protein bar						
red wine	shrimp	pasta	pepper	eggs	chocolate	shampoo				
rice	sparkling water									
spaghetti	mineral w	ham	body spray	pancakes	green tea					
burgers	grated che	shrimp	pasta	avocado	honey	white wine	toothpaste			
eggs										
parmesan	spaghetti	soup	avocado	milk	fresh bread					
ground be	spaghetti	mineral w	milk	energy bar	black tea	salmon	frozen sm	escalope		

```
[4] import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
[5] df = pd.read_csv("Market_Basket_Optimisation.csv")
```

```
df.describe()
```

```

olive oil
count      0.0
mean       NaN
std        NaN
min        NaN
25%        NaN
50%        NaN
75%        NaN
max        NaN

```

```
[7] df.columns
```

```
Index(['shrimp', 'almonds', 'avocado', 'vegetables mix', 'green grapes',  
      'whole weat flour', 'yams', 'cottage cheese', 'energy drink',  
      'tomato juice', 'low fat yogurt', 'green tea', 'honey', 'salad',  
      'mineral water', 'salmon', 'antioxydant juice', 'frozen smoothie',  
      'spinach', 'olive oil'],  
      dtype='object')
```

Kaggle Dataset - <https://www.kaggle.com/datasets/devchauhan1/market-basket-optimisationcsv>

7. Proposed method/Algorithms–

We are analysing both the market basket and the customers' purchasing behaviors using the data mining algorithms: the FP Growth Algorithm and the Apriori Algorithm. And we have also used ECLAT algorithm to know the consumer's purchase behavior.

We are calculating support, confidence, lift, leverage and conviction values for our dataset:

- **Support:** Support is an indication of how frequently the items appear in the data.

$$\text{Supp}(X) = \frac{\text{Freq}(X)}{T}$$

- **Confidence:** Confidence indicates how often the rule has been found to be true. Or how often the items X and Y occur together in the dataset when the occurrence of X is already given. It is the ratio of the transaction that contains X and Y to the number of records that contain X.

$$\text{Confidence} = \frac{\text{Freq}(X,Y)}{\text{Freq}(X)}$$

- **Lift:** It is the strength of any rule, which can be defined as below formula

$$\text{Lift} = \frac{\text{Supp}(X,Y)}{\text{Supp}(X) \times \text{Supp}(Y)}$$

- **Leverage:** Leverage computes the difference between the observed frequency of A and C appearing together and the frequency that would be expected if A and C were independent. A leverage value of 0 indicates independence.

$$\text{leverage}(A \rightarrow C) = \text{support}(A \rightarrow C) - \text{support}(A) \times \text{support}(C), \quad \text{range: } [-1, 1]$$

- **Conviction:** A high conviction value means that the consequent is highly depending on the antecedent. For instance, in the case of a perfect confidence score, the denominator becomes 0 (due to $1 - 1$) for which the conviction score is defined as 'inf'. Similar to lift, if items are independent, the conviction is 1.

$$\text{conviction}(A \rightarrow C) = \frac{1 - \text{support}(C)}{1 - \text{confidence}(A \rightarrow C)}, \quad \text{range: } [0, \infty]$$

8. Algorithm –

EDA Algorithm:

EDA (Exploratory Data Analysis) is one of the data analysis methodologies for summarising dataset characteristics with statistical data and graphs. EDA is frequently performed before main data analysis to understand the current status (characters) of the dataset.

Apriori Algorithm:

Association rule mining finds interesting associations and/or correlation relationships among large set of data items. Association rules shows attribute value conditions that occur frequently together in a given dataset. A typical and widely used example of association rule mining is Market Basket Analysis. For example, data are collected from the supermarkets. Such market basket databases consist of a large number of transaction records. Each record lists all items bought by a customer on a single purchase transaction. Association rules provide information of this type in the form of “IF-THEN” statements. The rules are computed from the data, an association rule has two numbers that express the degree of uncertainty about the rule.

FP – Growth Algorithm:

This algorithm is an improvement to the Apriori method. A frequent pattern is generated without the need for candidate generation. FP growth algorithm represents the database in the form of a tree called a frequent pattern tree or FP tree. This tree structure will maintain the association between the itemset. The database is fragmented using one frequent item. This fragmented part is called “pattern fragment”. The itemset of these fragmented patterns are analyzed. Thus, with this method, the search for frequent itemset is reduced comparatively.

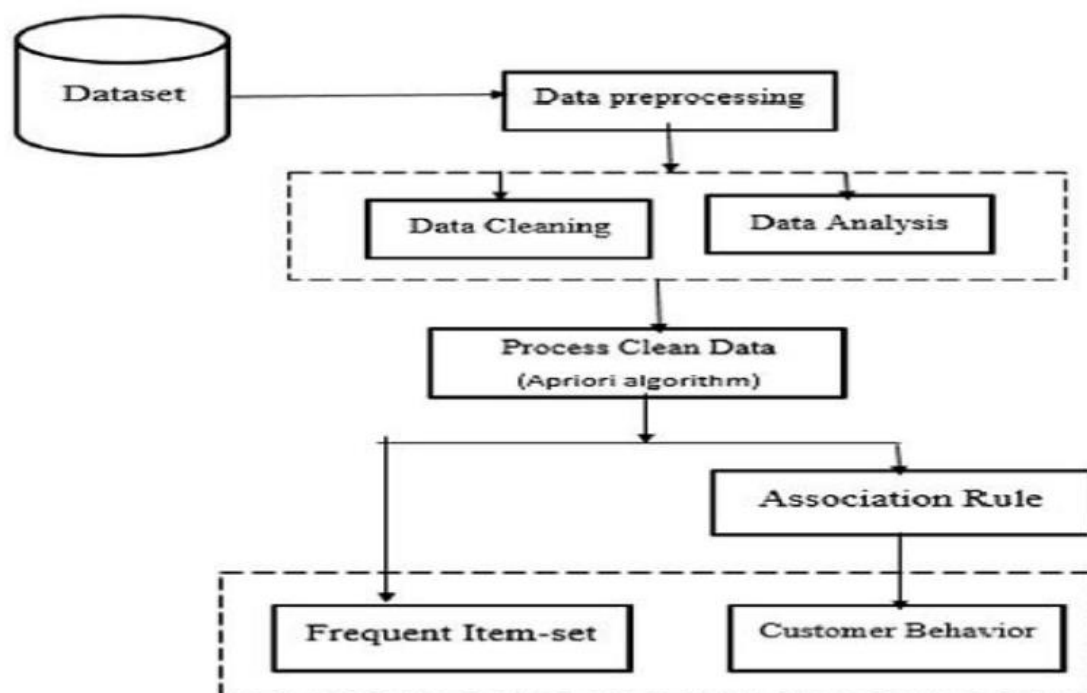


Fig 1. Flowchart of FP Growth Algorithm

ECLAT Algorithm:

The ECLAT algorithm is not the first algorithm for association rule mining. The foundational algorithm in the domain is the Apriori Algorithm. Since the Apriori algorithm is the first algorithm that was proposed in the domain, it has been improved upon in terms of computational efficiency (i.e. they made faster alternatives).

Steps for ECLAT:

1. Get tidlist for each item in the database. Here, we scan the entire database. The tidlist of item {a} is the list of transactions in which item {a} is contained.
2. Intersect the tidlist of item {a} with the tidlist of item {b} and generate a new transaction list whose elements are transactions in which both items {a} and {b} are found.
3. Repeat step 1 on {a}-conditional to other items in the database.
4. For all other items, repeat the above steps.

The Workflow for ECLAT-

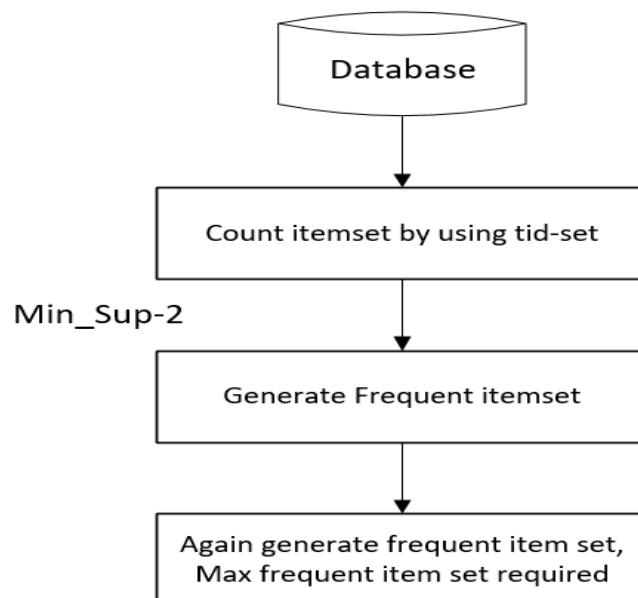


Fig 2. ECLAT Workflow

The Overall Flowchart for the Project:

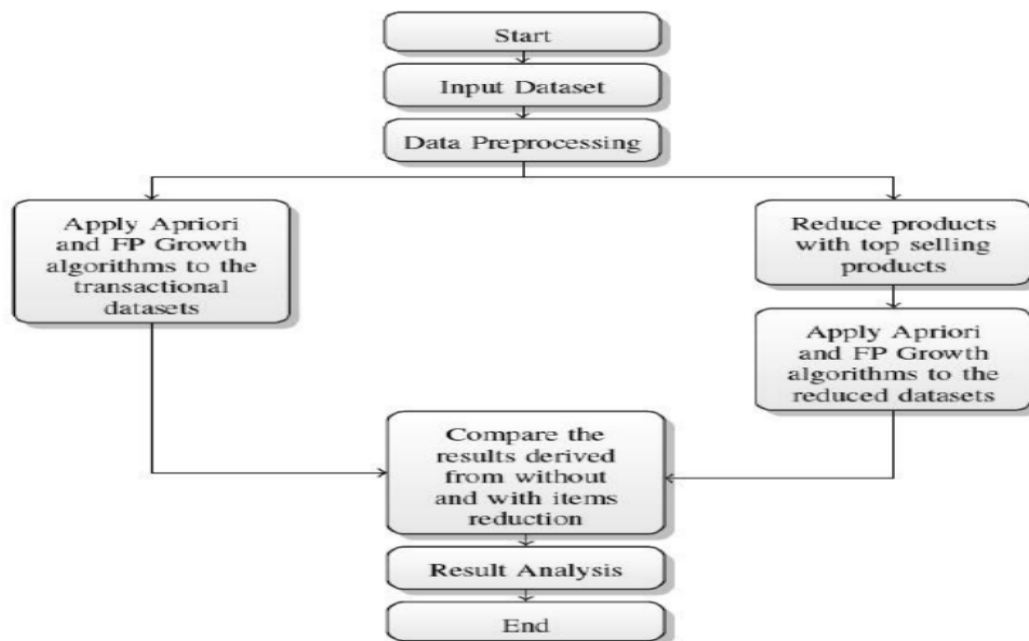


Fig 3. Flowchart for the project

Google Colab Links:

Apriori, Fp-Growth and ECLAT - <https://colab.research.google.com/drive/1-KJHBglCc9bPptsFthen23SDpXPJNUZN?usp=sharing>

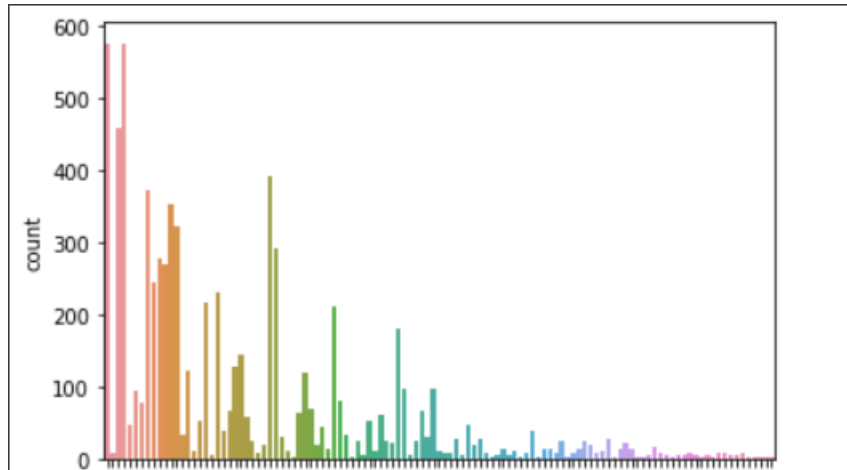
EDA -

<https://colab.research.google.com/drive/1nCEs9qzoqhJ7flpGuyLA1m7eBfVk4enQ?usp=sharing>

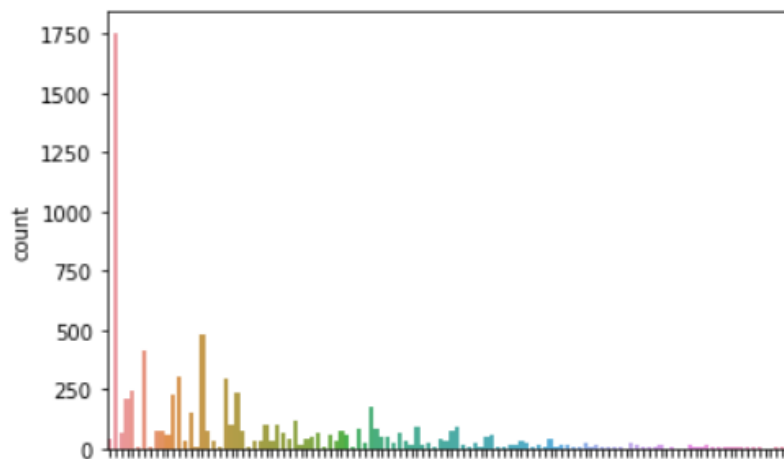
9.. Experiments Results

EDA Results:

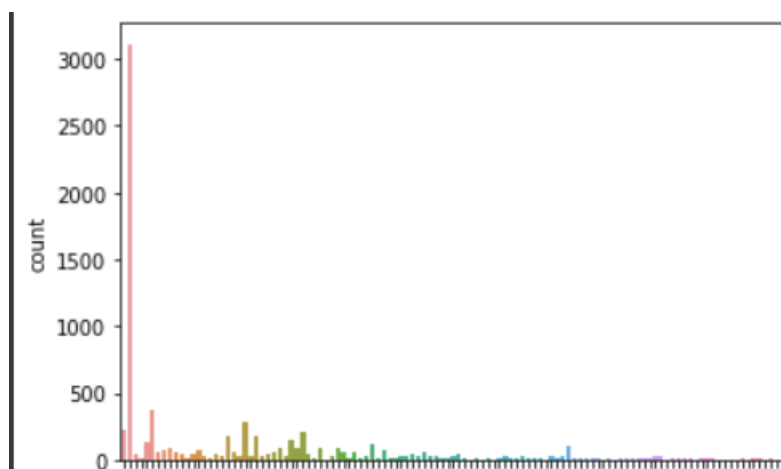
Figure shows count of the products sold every day.



Shrimp

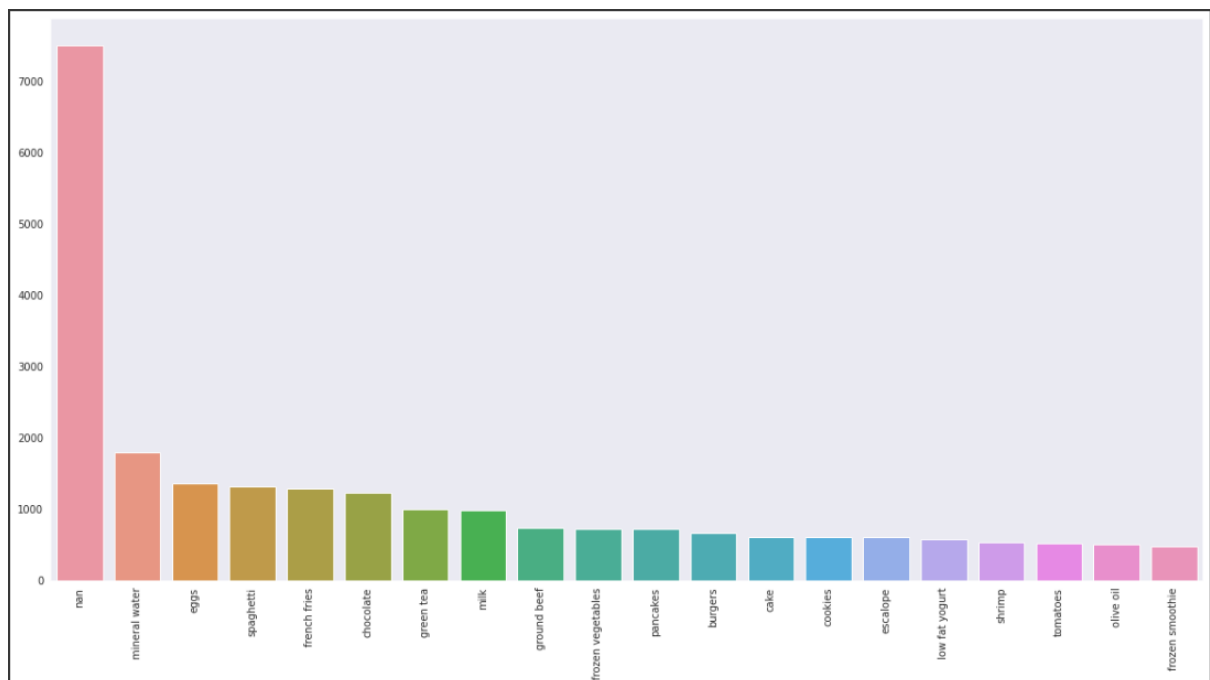
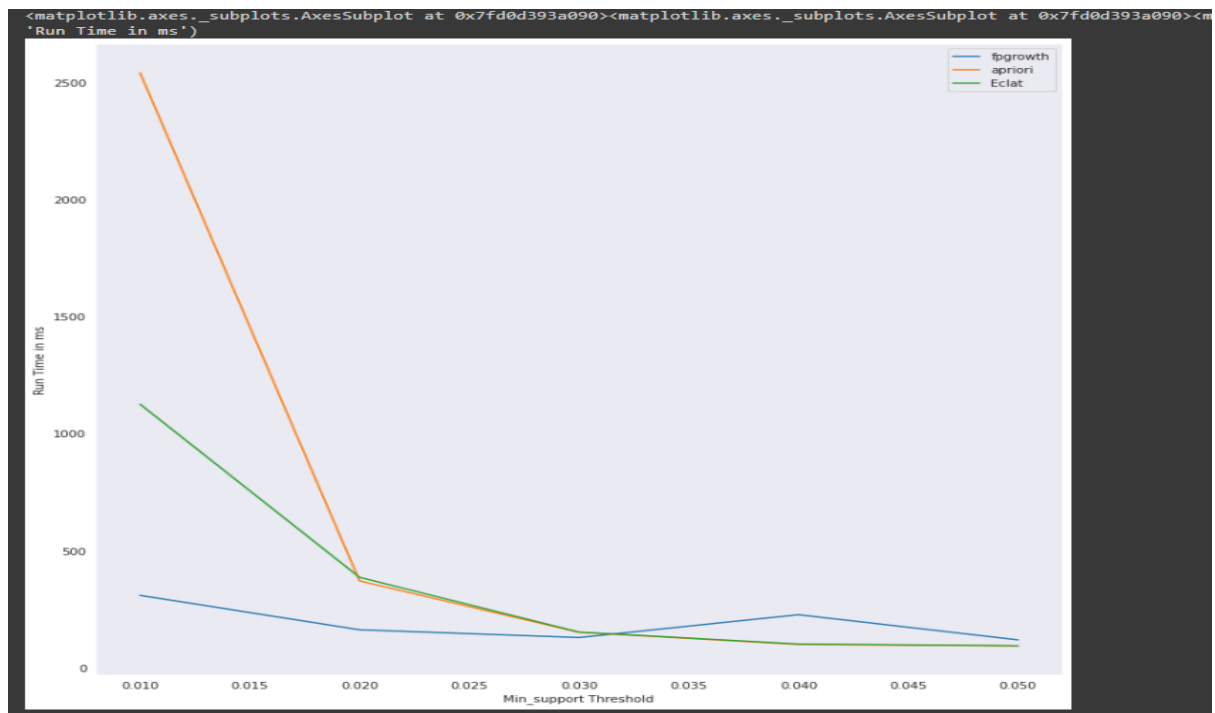


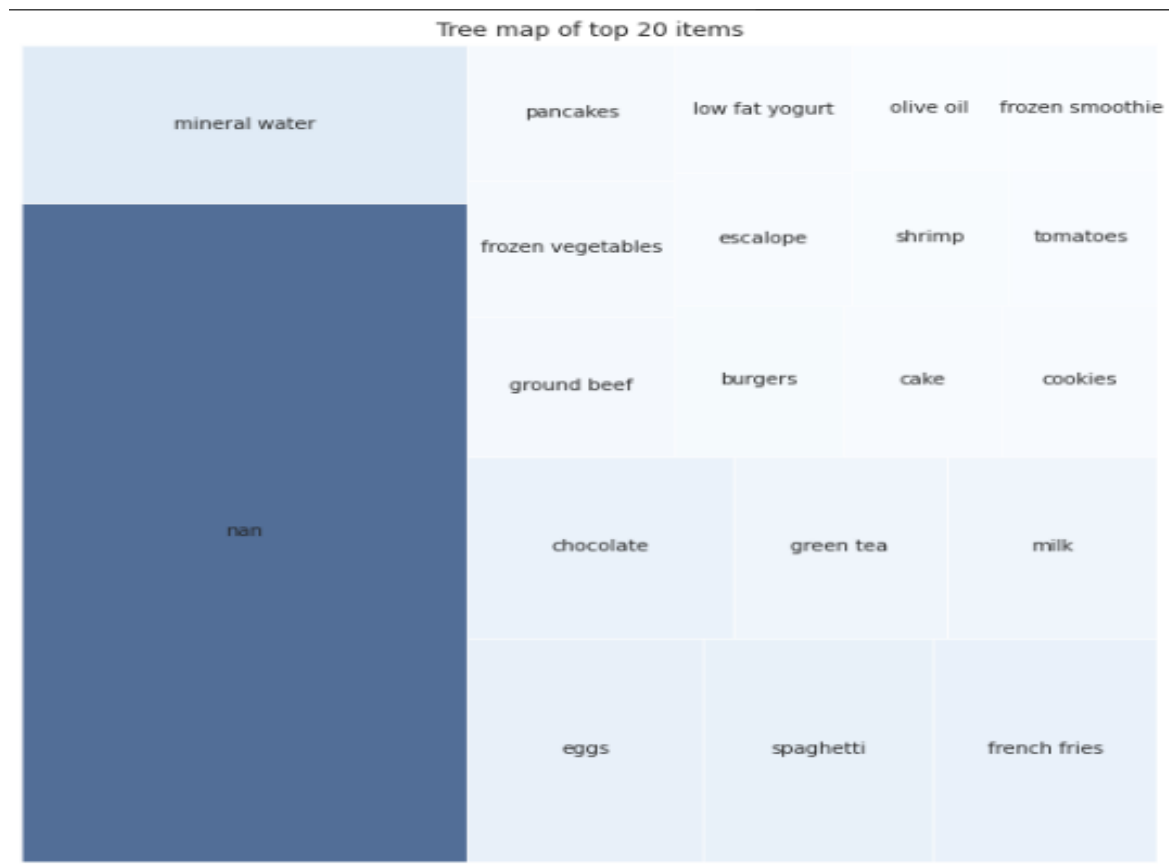
Almonds



Avocado

ECLAT, Apriori, FP-Growth Results:

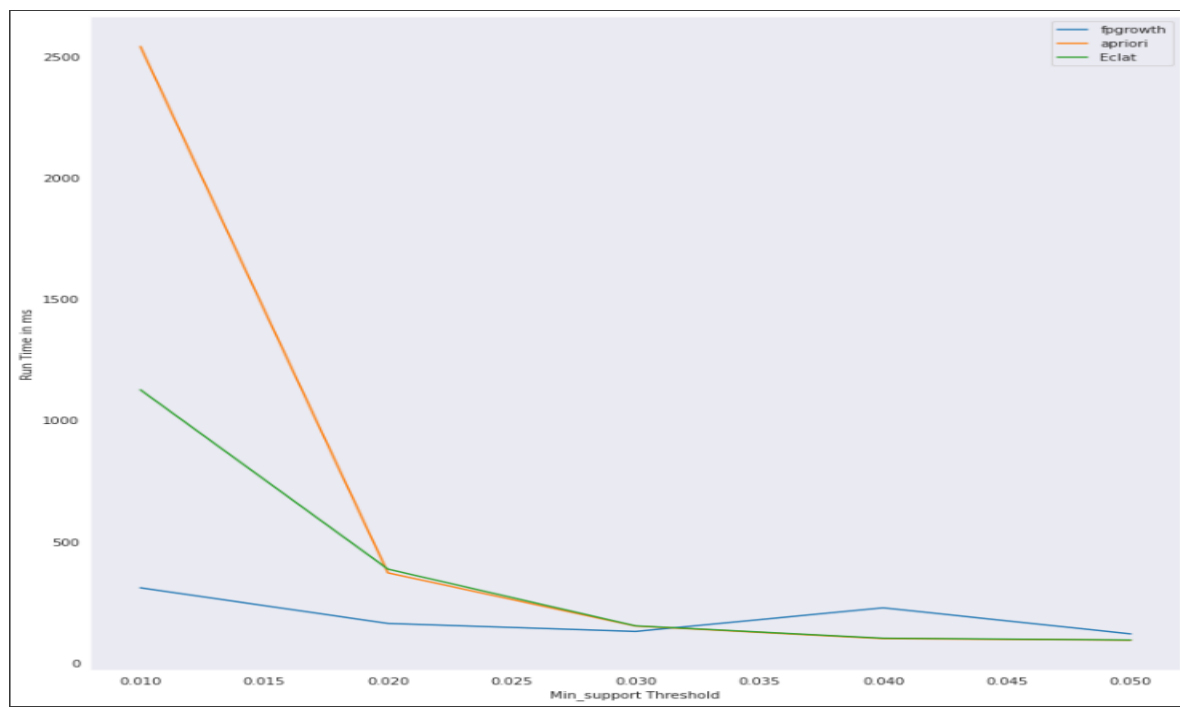




Most Popular Items bought first by the Customers



10.. Comparative study/ Results and Discussions –



- The Apriori algorithm has a greater run time than ECLAT and FP – Growth algorithms respectively.
- ECLAT is faster than Apriori and thus it scans the data faster than Apriori.
- Apriori has serious scalability issues and exhausts available memory much faster than Eclat and FP-Growth.

Apriori	FP Growth
Apriori generates the frequent patterns by making the itemsets using pairing such as single item set, double itemset, triple itemset.	FP Growth generates an FP-Tree for making frequent patterns.
Apriori uses candidate generation where frequent subsets are extended one item at a time.	FP-growth generates conditional FP-Tree for every item in the data.
Since apriori scans the database in each of its steps it becomes time-consuming for data where the number of items is larger.	FP-tree requires only one scan of the database in its beginning steps so it consumes less time.
A converted version of the database is saved in the memory	Set of conditional FP-tree for every item is saved in the memory
It uses breadth-first search	It uses a depth-first search.

- Apriori is useable with large datasets and Eclat is better suited to small and medium datasets.
- Apriori scans the original (real) dataset, whereas Eclat scan the currently generated dataset.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(chocolate)	(mineral water)	0.163845	0.238368	0.052660	0.321400	1.348332	0.013604	1.122357
1	(mineral water)	(chocolate)	0.238368	0.163845	0.052660	0.220917	1.348332	0.013604	1.073256
2	(spaghetti)	(mineral water)	0.174110	0.238368	0.059725	0.343032	1.439085	0.018223	1.159314
3	(mineral water)	(spaghetti)	0.238368	0.174110	0.059725	0.250559	1.439085	0.018223	1.102008
4	(chocolate, nan)	(mineral water)	0.163845	0.238368	0.052660	0.321400	1.348332	0.013604	1.122357
5	(mineral water, nan)	(chocolate)	0.238235	0.163845	0.052660	0.221041	1.349087	0.013626	1.073426

Algorithms	Runtime	Min Support Threshold
Apriori	380 ms	0.020
FP-Growth	320 ms	0.020
ECLAT	200 ms	0.020

Table showing runtime for the given support values for different algorithms

11. CONCLUSION AND FUTURE WORK -

Conclusion:

From the output above, we see that the top associations are not surprising, with one item for a specific purpose gets purchased along with another flavour required for the same purpose:

Example: If the support is 52% and the antecedent is chocolate and consequent is mineral water then they both are paired together having confidence 32%.

Also, mineral water and spaghetti should be grouped together both having confidence value of 25%.

Future Works:

In the future, once common application of association rules mining is in the domain of recommender systems and item pairs have been identified as having positive relationship, recommendations can be made to customers in

order to increase sales. Also, there is a possibility of introducing customers to items they never would have tried before. We can also work more on reducing the runtime for all the algorithms with some modifications. We can also use SMOT and AIS algorithms.

12. References –

1. ISSN 2548-8384 (online), ISSN 2548-8449 (print) Available Online at <https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index> DOI [10.30865/ijics.v6i2.4535](https://doi.org/10.30865/ijics.v6i2.4535)
2. Efrat, A & Gernowo, R & Farikhin,. (2020). Consumer purchase patterns based on market basket analysis using apriori algorithms. Journal of Physics: Conference Series. 1524. 012109. 10.1088/1742-6596/1524/1/012109.
3. Patil, Bhupal & Khot, Laxmi. (2022). A STUDY ON MARKET BASKET ANALYSIS USING APRIORI ALGORITHM. 10.13140/RG.2.2.19506.48328.
4. M Qisman *et al* 2021 *J. Phys.: Conf. Ser.* **1722** 012020 DOI 10.1088/1742-6596/1722/1/012020
5. S. Kanhere, A. Sahni, P. Stynes and P. Pathak, "Clustering Based Approach to Enhance Association Rule Mining," 2021 28th Conference of Open Innovations Association (FRUCT), 2021, pp. 142-150, doi: 10.23919/FRUCT50888.2021.9347577.
6. Rao, A.B., Kiran, J.S. & G, P. Application of market–basket analysis on healthcare. *Int J Syst Assur Eng Manag* (2021). <https://doi.org/10.1007/s13198-021-01298-2>
7. Mustakim, Mustakim & Herianda, Della & Ilham, Ahmad & GS, Achmad & Laumal, Folkes & Kurniasih, Nuning & Iskandar, Akbar & Manulangga, Gloria & Ary Indra Iswara, Ida Bagus & Rahim, Robbi. (2018). Market Basket Analysis Using Apriori and FP-Growth for Analysis Consumer Expenditure Patterns at Berkah Mart in Pekanbaru Riau. Journal of Physics: Conference Series. 1114. 012131. 10.1088/1742-6596/1114/1/012131.
8. M. Hossain, A. H. M. S. Sattar and M. K. Paul, "Market Basket Analysis Using Apriori and FP Growth Algorithm," 2019 22nd International

Conference on Computer and Information Technology (ICCIT), 2019, pp. 1-6, doi: 10.1109/ICCIT48885.2019.9038197.

9. Rana, Sohel & Mondal, Md. (2021). A Seasonal and Multilevel Association Based Approach for Market Basket Analysis in Retail Supermarket. European Journal of Information Technologies and Computer Science. 1. 9-15. 10.24018/compute.2021.1.4.31.
10. Rehman, Iqra & Ghous, Hamid. (2021). Structured Critical Review on Market Basket Analysis using Deep Learning & Association Rules.
11. Kurniawan, Fachrul & Umayah, Binti & Hammad, Jihad & Nugroho, Supeno & Hariadi, Mohamad. (2017). Market Basket Analysis to Identify Customer Behaviours by Way of Transaction Data. Knowledge Engineering and Data Science. 1. 20. 10.17977/um018v1i12018p20-25.
12. Volume 1, No. 2 (2021) pp. 55-60 ISSN (Online) : 2776-2386 Website: <https://journal.universitasmulia.ac.id/index.php/mst> DOI: 10.47002/mst.v1i2.239
13. S. S. Khedkar and S. Kumari, "Market Basket Analysis using A-Priori Algorithm and FP-Tree Algorithm," 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), 2021, pp. 1-6, doi: 10.1109/AIMV53313.2021.9670981
14. e-ISSN: 2582-5208 Volume:04/Issue:05/May-2022 Impact Factor- 6.752 https://www.irjmets.com/uploadedfiles/paper//issue_5_may_2022/23830/final/fin_irjmets1653394526.pdf
15. Corpus ID: 203432586
<https://www.semanticscholar.org/paper/Market-basket-analysis-%3A-trend-analysis-of-rulesAnsari/0e5884d50bd0ef373d9723e68e3890f833b05a00>
16. W. Nengsih, "A comparative study on market basket analysis and apriori association technique," 2015 3rd International Conference on Information and Communication Technology (ICoICT), 2015, pp. 461-464, doi: 10.1109/ICoICT.2015.7231468.
17. S. S. Khedkar and S. Kumari, "Market Basket Analysis using A-Priori Algorithm and FP-Tree Algorithm," 2021 International Conference on

Artificial Intelligence and Machine Vision (AIMV), 2021, pp. 1-6, doi: 10.1109/AIMV53313.2021.9670981.

18. Vol 6 No 2, July 2022, Page 107–112 ISSN 2548-8384 (online), ISSN 2548-8449 (print) Available Online at <https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index> DOI 10.30865/ijics.v6i2.4535
19. Priyadarshi, Abhishek & Gupta, Chirag & Ganesh, Poornalatha. (2016). Market Basket Analysis using Improved FP-tree. International Journal of Data Mining Techniques and Applications. 5. 110-113. 10.20894/IJDMTA.102.005.002.002.
20. M. Hossain, A. H. M. S. Sattar and M. K. Paul, "Market Basket Analysis Using Apriori and FP Growth Algorithm," 2019 22nd International Conference on Computer and Information Technology (ICCIT), 2019, pp. 1-6, doi: 10.1109/ICCIT48885.2019.9038197.