**A PROJECT REPORT**

**ON**

“Sales And Marketing Analysis”

## UNDERTAKEN AT

**“MIT School of Distance Education”**

## IN PARTIAL FULFILMENT OF

**“Post Graduate Certificate in Management - Business Analytics”**

## MIT SCHOOL OF DISTANCE EDUCATION, PUNE.

**GUIDED BY**

**Dr. Jayant Panigrahi**

**Prof. Bonnie Rajesh**

## SUBMITTED BY

**“SRIVANI ARROJU”**

**STUDENT REGISTRATION NO.: MIT2022E00433**

# MIT SCHOOL OF DISTANCE EDUCATION PUNE - 411 038

## YEAR 2023-24

**Exempt Certificate - If you’re not able to provide the Project Executed Certificate**

To

The Director

MIT School of Distance Education, Respected Sir,

This is to request you to kindly exempt me from submitting the certificate for Project Work due to the reason mentioned below:

Tick the right option

1. As per the Rules of the Organization

2. Self Employed

1. Working in Public Sector
2. Full-time Student

Thanking you in anticipation of your approval to my request. Regards

**STUDENT NAME: SRIVANI ARROJU**

**Student ID:MIT2022E00433**



# DECLARATION

I hereby declare that this project report entitled **“Sales and Marketing Analysis”** bonafide record of the project work carried out by me during the academic year **2023-2024**, in fulfillment of the requirements for the award of **“PGCM Business Analytics”** of MIT School of Distance Education.

This work has not been undertaken or submitted elsewhere in connection with any other academic course.

**Sign: Srivani Arroju**

**Name: Srivani Arroju**

**Student ID: MIT2022E00433**



# ACKNOWLEDGEMENT

I would like to take this opportunity to express my sincere thanks and gratitude to **“Dr. JAYANTH PANIGRAHI and Prof. BONNIE RAJESH ”**, Faculty of MIT School of Distance Education, for allowing me to do my project work in your esteemed organization. It has been a great learning and enjoyable experience.

I would like to express my deep sense of gratitude and profound thanks to all staff members of MIT School of Distance Education for their kind support and cooperation which helped me in gaining lots of knowledge and experience to do my project work successfully.

At last, but not least, I am thankful to my Family and Friends for their moral support, endurance, and encouragement during the project.

**Sign: Srivani Arroju**

**Name: Srivani Arroju**

**Student ID: MIT2022E00433**



# ABSTRACT

This is the study on Sales and Marketing Analysis done on an electronic supplier organization in Bangalore. This Analysis helps in reviewing the company sales done and to identify the marketing strategies to enhance the sales of their products. Sales data can help identifying the pattern of the customers purchasing the products and can help you in making decisions about the pricing, marketing strategies, inventories, Promotions and finally the needs of the customers.

Sales is the most important aspect of any business and without the sales no company can operate or survive. That is the reason the sales analysis plays a major role in Business. It considers all aspects of sales including the Raw material to the final output Product. Considering the life cycle of the product many KPI’s are identified and tracked to check whether the resources are being used effectively or not. For a company which is growing in multi-dimensional way it is very much important to ramp up the sales as the business grow and without proper understanding it gets complicated. In this situation sales team needs a mature and clear analysis to guide them.

Now when we discuss the Marketing strategies Understanding, Identification, anticipation and Satisfaction of the end users or client’s requirements and developing them, advertising them, and finally distribution of the products compress of Marketing strategies. Marketing can be classified into tactical and strategic, but both the approaches need analysis done on how the clients are purchasing our products. The analysis done on these strategies can help us in targeting the customers in a better way where we approach them with different advertisements, offers and specifications.

Combining the Sales and Marketing strategies few techniques are followed and based on these techniques we can find the most frequent combinations of items sold together. It is effectively done by developing an effective algorithm that provides the common or frequent pattern algorithms on various data sets. This will help in marketing and sales of the products. This technique is used to bring out the insights on cross selling and frequently bought together products. This technique can be used on larger datasets. We have a large database of the customer transactions. Each transaction consists of the customer information of items purchased in a visit. We present an efficient algorithm that shows all significant association rules between the items purchased in the database. Using this technique can estimate the purchasing patterns and helps business to forecast the sales of the products so that company can plan their raw materials and priorities effectively.

The purpose of this study is to find out the sales and marketing trends and provide the developing sales strategies and providing leads to the company and help them to understand the trends and effective

Crafting a unique write-up on sales and marketing strategies requires a deep understanding of the subject matter. Here are some general strategies:

Customer-Centric Approach: Tailoring products and services to meet the specific needs of the target audience. The customer-centric approach emphasizes placing the customer at the core of all business activities. This strategy involves thoroughly understanding the unique needs and preferences of the target audience. By gathering insights through market research, feedback mechanisms, and data analysis, businesses can tailor their products and services to precisely align with customer demands. This may involve customizing offerings, providing personalized experiences, and ensuring that every touchpoint with the customer reflects a deep understanding of their individual requirements. Ultimately, the goal is to build strong, long-lasting relationships by consistently delivering exceptional value and addressing the specific pain points of the customers.

Digital Marketing: Leveraging social media, content marketing, SEO, and email campaigns to reach and engage with potential customers. Digital marketing encompasses the strategic use of various online channels to connect with potential customers and engage them effectively. This includes leveraging social media platforms, such as Facebook, Instagram, and Twitter, to build brand presence, interact with audiences, and promote products or services. Content marketing involves creating and sharing valuable, relevant content to attract and retain a clearly defined audience. Search engine optimization (SEO) focuses on improving a website's visibility in search engine results, thus driving organic traffic. Email campaigns are used to deliver targeted messages, offers, and updates directly to the inboxes of potential customers. By integrating these digital marketing tactics, businesses can expand their reach, foster meaningful connections, and drive conversions in the online space.

Relationship Building: Focusing on building long-term relationships with customers through personalized interactions and exceptional customer service. Relationship building entails prioritizing the establishment of enduring connections with customers by emphasizing personalized interactions and delivering exceptional customer service. This approach involves understanding the individual preferences, needs, and behaviors of customers to tailor interactions and offerings accordingly. By consistently providing high-quality and personalized experiences, businesses can foster trust, loyalty, and advocacy among their customer bases. This often leads to long-term relationships, repeat business, positive word-of-mouth referrals, and a strong brand reputation. Ultimately, relationship building seeks to create a mutually beneficial dynamic where customers feel valued, understood, and appreciated, thus laying the foundation for sustained business success.

Data-Driven Decision Making: Utilizing data analytics to understand customer behavior and preferences, and to optimize marketing efforts. Data-driven decision making involves the systematic use of data analytics to gain insights into customer behavior and preferences, as well as to refine and enhance marketing strategies. By leveraging various data sources and analytical tools, businesses can uncover valuable patterns, trends, and correlations that provide a deeper understanding of how customers interact with their products or services. This information empowers organizations to make informed decisions, optimize marketing efforts, and allocate resources more effectively. Through data-driven decision making, businesses can adapt their strategies in real time, personalize marketing initiatives, and ultimately improve customer engagement and satisfaction.

Sales Enablement: Equipping sales teams with the right tools, training, and resources to effectively engage with prospects and close deals.

Sales enablement involves the strategic provision of sales teams with tailored tools, training, and resources to enhance their ability to engage with potential customers and successfully secure deals. This encompasses equipping sales professionals with the necessary knowledge, skills, and technology to effectively communicate the value of products or services to prospects throughout the sales cycle. By doing so, organizations can empower their sales teams to deliver compelling presentations, address customer needs, and overcome challenges, ultimately leading to improved sales performance and customer satisfaction. The goal of sales enablement is to optimize the sales process, boost productivity, and drive revenue growth by ensuring that sales representatives are well-prepared and equipped to deliver exceptional experiences to potential clients.

# TABLE OF CONTENTS

|  |  |  |
| --- | --- | --- |
| **Chapter No.** | **Title** | **Page No.** |
| 1 | Introduction |  |
| 2 | Organizational Profile |  |
| 3 | Project Objectives and Scope |  |
| 4 | Data Analysis and Interpretation |  |
| 5 | Conclusion / Findings |  |
| 6 | Suggestions / Recommendations |  |
| 7 | Annexure |  |
| 8 | References / Bibliography |  |

## CHAPTER 1: INTRODUCTION

Data mining represents an advanced technology that reveals concealed predictive findings in extensive databases, offering significant potential for organizations aiming to pinpoint critical data within their data repositories.

Market Basket Analysis is a data mining method employed to reveal purchasing patterns in retail environments. Essentially, it involves analyzing the combinations of products that are frequently purchased together, providing a detailed examination of customer purchases. This analysis identifies the recurring purchase patterns of customers, and it can be utilized by companies to promote deals, offers, and sales. Data mining techniques are instrumental in accomplishing this analytical task.

**History of Market Basket Analysis:**

Market Basket Analysis, also referred to as association rule mining, originated within the retail sector and gained significant traction during the 1990s. Initially developed as a method to reveal correlations between products commonly purchased together by customers, this concept became prominent following the publication of the influential 1993 paper titled "Market Basket Analysis" authored by Agrawal, Imielinski, and Swami. The paper introduced the Apriori algorithm, which revolutionized the mining of frequent itemsets. By leveraging this algorithm, retailers could discern connections between items in transactional data, thereby gaining valuable insights into consumer purchasing behaviors.

The evolution of Market Basket Analysis closely aligns with the surge of data mining and the growing availability of transactional data within retail environments. As businesses endeavored to comprehend customer behaviors and enhance their marketing and sales strategies, Market Basket Analysis emerged as an indispensable tool for uncovering patterns and steering pivotal business decisions.

In contemporary times, Market Basket Analysis has transcended its origins in retail and permeated various industries such as e-commerce, telecommunications, and healthcare. It is now utilized to glean insights into customer preferences, identify cross-selling opportunities, and optimize operational processes. This technique continues to progress alongside advancements in data analytics and machine learning, empowering organizations to extract invaluable associations from extensive transactional data.

Using data mining tools, businesses can anticipate future trends and behaviors, enabling them to make well-informed, proactive decisions. Unlike the traditional retrospective tools present in decision support systems, data mining performs automated, forward-looking analyses. It has the capability to tackle complex business inquiries that were previously too time-consuming to address. By examining databases for hidden patterns, data mining tools reveal predictive insights that might escape human experts due to their biases.

As many companies have already accumulated and enhanced substantial amounts of data, data mining techniques can be seamlessly incorporated into current software and hardware setups, enhancing the utility of existing information resources. When utilized on high-performance client/server or parallel processing computers, these tools can efficiently analyze extensive databases to deliver vital insights. Data mining, also referred to as Knowledge-Discovery in Databases (KDD), encompasses the automated examination of large data volumes through techniques like classification, association rule mining, and clustering.

It connects with multiple fundamental disciplines such as computer science, enhancing established computational methods from statistics, information retrieval, machine learning, and pattern recognition. The development of data mining originates from the initial storage of business data on computers, advancing through improvements in data accessibility, and culminating in real-time data exploration technologies. This progression has now moved beyond retrospective data access and exploration to provide forward-looking and proactive insights. With the advancement of three key technologies—extensive data collection, high-performance multiprocessor computers, and data mining algorithms—data mining is ready for widespread application in the business domain.

This report aims to provide insights on how to boost cross-selling potential and elevate direct mail response rates. Data mining enables the identification of patterns and prediction of potential customer defections, thus aiding in customer retention efforts.

Certainly! Market basket analysis is a data mining technique that uncovers associations between products frequently purchased together. Here's how it helps businesses:

**Identifying Product Associations:** By analyzing transaction data, businesses can discover which items are commonly bought in conjunction with each other. This insight helps in understanding customer behavior and preferences.

**Strategic Product Placement:** Armed with the knowledge of which products are often purchased together, businesses can strategically place these items near each other in physical stores or suggest them together in online stores. This can lead to increased sales as customers find related items more easily.

**Bundled Promotions:** Understanding product associations allows businesses to create bundled promotions, offering related products at a discounted price when purchased together. This entices customers to buy complementary items and increases the average order value.

**Enhanced Cross-Selling:** Utilizing market basket analysis, businesses can recommend complementary products to customers at the point of sale, either in person or online. This can lead to increased cross-selling as customers are made aware of related items they may not have considered.

**Boosting Sales and Customer Satisfaction:** By leveraging the insights gained from market basket analysis, businesses can enhance the overall shopping experience for customers. By offering relevant product recommendations and creating convenient product arrangements, businesses can boost sales and improve customer satisfaction.

In summary, market basket analysis empowers businesses to understand customer behavior, strategically position products, and offer complementary items, ultimately leading to increased sales and improved customer satisfaction.

Structuring the data for Market Basket Analysis is absolutely critical as it directly impacts the quality of insights derived and the effectiveness of leveraging those insights to drive business strategies. Here's why it's so important:

Association Rule Mining: Properly structured data is fundamental for identifying association rules, which reveal which items are frequently purchased together. This requires organizing transactional data in a format that allows for the identification of patterns and relationships between items.

Data Preprocessing: Data structuring involves preprocessing steps such as removing duplicates, handling missing values, and transforming the data into a suitable format for analysis. This is vital for ensuring the accuracy and reliability of the results obtained from the analysis.

DATA Preprocessing:

Data preprocessing is done before exploring the specific steps taken to preprocess the data in this study, it's essential to recognize the significance of this phase in the overall data analysis process. Data preprocessing involves a series of operations aimed at converting raw data into a refined dataset. In the context of our research, this involves refining the transactional data from Deli Point to ensure that subsequent frequency analysis accurately reflects customer behavior. These procedures typically involve removing extra spaces, cleaning text data to eliminate unwanted characters, and standardizing text format through case folding. Each step is tailored to enhance different aspects of data cleanliness and structure, both of which are crucial for ensuring the reliability of any subsequent analysis.

The following points provide detailed examples of each preprocessing step implemented:

▪ **Extra Space Removal**:

The purpose of removing extra spaces during data preprocessing is to standardize the format of text data, facilitating accurate data analysis and pattern recognition. Inconsistent spacing can lead to errors or bias in the analysis, especially when dealing with textual data, as space characters can significantly impact string matching algorithms and analytical outcomes. The goal is to ensure uniform formatting of all items in the dataset, thus eliminating variations solely due to spacing errors.

The example in the table below demonstrates the effectiveness of this process by showcasing original text alongside the results after removing extra spaces. These examples illustrate how the removal of leading, trailing, and excessive inter-word spaces can clean and normalize the data, preparing it for further analysis such as frequency counts and the application of data mining algorithms

|  |  |
| --- | --- |
| Extra space removal result example | |
|  |  |
| Original Text | Result |
| SRI VANI | SRIVANI |
| SUPRIYA BENGALURU | SUPRIYA BENGALURU |

▪ **Text Cleaning**:

This process is a crucial component of data preprocessing with the aim to purify the dataset by removing extraneous characters such as (, \*, # were removed that could potentially skew the analysis. The objective of this step is to ensure that the dataset contains only meaningful and relevant characters, thereby enhancing the accuracy of the analysis.

Below example showcases the results of this text cleaning process.

For instance, the original entry RAMA\*KRISHNA ' is stripped of the asterisk to RAMAKRISHNA

Example 2: ARCHANA&SHANKAR has the & symbol and content within removed to become ARCHANA SHANKAR.

These transformations are representative of the broader cleaning process that was applied to the entire dataset. By removing these irrelevant characters, the dataset is standardized, reducing noise, and enabling more precise data mining techniques. This cleaned dataset forms the foundation for subsequent analytical tasks, such as identifying trends, patterns, and associations within the data.

These changes exemplify the comprehensive cleaning process carried out across the entire dataset. By eliminating these extraneous characters, the dataset is standardized, reducing interference, and allowing for more accurate data mining techniques. This refined dataset serves as the basis for subsequent analytical tasks, including the identification of trends, patterns, and associations within the data.

The changes made reflect the overall data cleaning process used for the entire dataset. Removing irrelevant characters standardizes the dataset, reducing noise and enabling more precise data mining techniques. This refined dataset serves as the basis for subsequent analytical tasks, such as identifying trends, patterns, and associations within the data.

Table 2. Text Cleaning Result Example

|  |  |
| --- | --- |
| Text Cleaning Result Example | |
|  |  |
| Original Text | Result |
| RAMA\*KRISHNA | RAMAKRISHNA |
| ARCHANA&SHANKAR | ARCHANA SHANKAR |

**Case Folding**

This is a fundamental data preprocessing technique with the objective to bring about uniformity in textual data. It involves converting all letters in the dataset to lowercase, which is a crucial step for ensuring consistency, especially in cases where text data is case-sensitive.

This uniformity is important because many data processing and text analysis algorithms treat letters with different cases (uppercase vs. lowercase) as distinct characters, which can lead to discrepancies in the results.

Table 3 displays the results of applying the case folding process to the dataset. For instance, the customer’s name 'Srivani’ from the original text remains the same as it does not contain uppercase letters beyond the initial character.

However, 'SriVani’ is transformed to 'Srivani', with all letters standardized to lowercase.

This demonstrates the case folding effect on the dataset, ensuring that all entries are treated equally in subsequent analysis stages.

The process helps remove variations caused by case differences, allowing for more accurate matching and comparison of text strings. This step is crucial for subsequent analytical processes like text mining and pattern recognition, as it improves data integrity and contributes to analysis reliability. For instance, customer names are trimmed to retain only the first name, preventing issues during research when customers are distinguished based on transaction time. This ensures consistency, for example, treating "Anaya January" and "Anaya February" as the same customer in the research.

This step helps to case differences, improving the accuracy of text string matching and comparison. This is crucial for subsequent analytical processes like text mining and pattern recognition, enhancing data integrity and analysis reliability. For instance, trimming customer names to retain only the first name avoids issues when distinguishing customers based on transaction time, ensuring consistency in research.

This step ensures that customer names are consistently represented throughout the transactional data, improving the accuracy of the association rule analysis. In the process of conducting association rule analysis, it is necessary to transform the raw transactional data into a structured format that can be interpreted by computational algorithms. To facilitate this data reshaping, the data is initially transformed into a panda’s data frame, which provides a convenient and flexible framework for data manipulation.

Subsequently, a grouping operation is performed based on the date and invoice number attributes, and the resulting data is subjected to a pivot transformation to arrange it into an appropriate format suitable for input into the "MLX tend" library, which implements the association rule mining algorithm.

The data shape depicted in Table before the application of the pivot transformation reveals that the data is organized on a per-customer basis, with each transaction represented as a separate row, even when multiple transactions are performed simultaneously.

This structure may not be ideal for association rule mining, as the relationships between items in a single transaction may not be effectively captured in this format.

The pivot transformation aims to remedy this by reorganizing the data more suitably for association rule analysis.

After this process, the number of transactions was drastically reduced to 54,744.

The Shape of the Original Data

Example Date Invoice Customer

Archives unsalted butter 10 kg the results in Table demonstrate the new format of the data, where transactions with the same invoice and date values are combined into a single row, and the items are merged into a single line of data.

This transformation enables the analysis of the relationships between items within a single transaction, facilitating a more comprehensive understanding of the underlying associations within the transactional data.

Shape After Group and Pivoting

Remove Duplicate Data consistency is improved by resolving the problems like bad data or duplicate data and instead of finding the whole dataset, focusing on finding associations in the filtered dataset.

Following both processes, data is cleaned, and duplicate data removal is an essential step in data preprocessing aimed at ensuring the uniqueness of each data entry. Duplicate records can occur due to various reasons such as data entry errors or data integration processes, and they can significantly distort statistical analysis by giving undue weight to repeated entries. The goal of this step is to identify and eliminate redundancy in the dataset to maintain the accuracy of the analysis and the integrity of the dataset.

Example of a redundant record where the same transaction appears twice with identical timestamps, invoice numbers, customer names, and purchased items. This repetition does not reflect additional transactions but is rather an unnecessary replication.

The Apriori Algorithm was applied to preprocessed data, with the difference being that the data was not grouped by customer but instead was kept in a single list variable which was fed to the Apriori library. In this process, we used minimum support of 0.01 and minimum confidence of 0.02.

Each rule is evaluated using three key metrics: Lift, Confidence, and Support. Lift is a measure of how much more frequently the items in the rule (antecedent and consequent) are purchased together than would be expected if they were independent; a lift value greater than 1 indicates a positive relationship. Confidence provides an estimate of the probability that the consequent item is bought when the antecedent is purchased, with higher values indicating stronger rules. Support quantifies how common the rule is within the dataset, represented as the proportion of transactions containing both the antecedent and consequent. These metrics collectively help in assessing the strength and relevance of each rule, providing insights into customer purchasing patterns.

For example, the rule 'parsley local → mint flores' with a lift of 2.56, confidence of 0.31, and support of 0.036 suggests that transactions including 'parsley local' are more likely to include 'mint flores' as well, and this occurs in 3.6% of all transactions. Similarly, the rule 'Chicken leg bone in & Orange sunkist → Chicken breast whole boneless' has a very high lift of 27.84, indicating a strong association between the purchase of Chicken leg bone in & Orange sunkist and Chicken breast whole boneless, with this rule being true in 96% of cases where the antecedents are bought, and appearing in 1.2% of all transactions. Meanwhile, the highest lift value was obtained for the transaction “green bell peppers & yellow bell peppers → red bell peppers”. A total of 239 rules were formed. The rules in this table have been selected for their significance in terms of these metrics, indicating potential patterns of customer purchasing behavior. Understanding these patterns can be valuable for strategic decision-making in areas such as marketing, inventory management, and product placement.

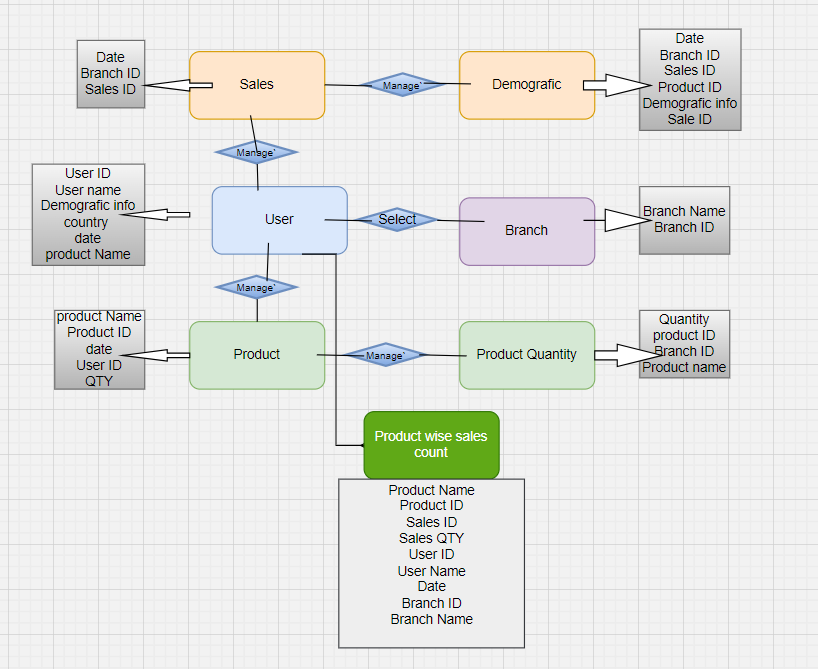
Efficient Analysis: Well-structured data enables efficient and accurate analysis, empowering businesses to uncover actionable insights more effectively. This involves organizing the data into transactional records with clear item associations and customer purchase patterns.

Insightful Decision-Making: Properly structured data empowers businesses to make informed decisions based on the insights gained from Market Basket Analysis. It provides a solid foundation for strategic product placement, cross-selling initiatives, and targeted marketing efforts.

In conclusion, the structuring of data for Market Basket Analysis is of paramount importance, playing a pivotal role in obtaining high-quality insights and leveraging them effectively to drive business strategies.

**Categorizing the data for Market Basket Analysis:**

To conduct Market Basket analysis, the data is structured into transactions, items, and baskets. In this context, a transaction includes a distinct identifier (like a receipt number) and the items bought. The term "items" refers to the individual products purchased, while "baskets" signify the assemblage of items obtained in a single transaction. Organizing the data in this manner facilitates the examination of associations between various items acquired together.



## Flow chart showcasing the data preparation process for Market Basket Analysis

## OVERVIEW

Data mining algorithms incorporate long-standing techniques, which have now evolved into reliable and sophisticated tools surpassing traditional statistical methods. The vast expansion of online information sources necessitates the use of automated tools to locate and analyze desired information, leading to the development of intelligent systems for effective knowledge extraction. Web mining encompasses the discovery and analysis of valuable information from the World Wide Web, including automatic search of online information resources (Web content mining) and the identification of user access patterns from Web servers (Web usage mining).

To create an effective algorithm for locating desired information resources and their usage patterns, as well as a distributed algorithm for geographical data sets that minimizes communication cost and overhead, it is increasingly essential for users to employ automated tools. Association rule mining, a prominent area of data mining research, mostly focuses on centralized environments. While Distributed Association Rule Mining (D-ARM) algorithms have been developed, they assume databases are either horizontally or vertically distributed. However, existing D-ARM algorithms cannot uncover rules based on higher-order associations in distributed textual documents that are a hybrid of both vertical and horizontal distribution, especially when databases are populated with information extracted from textual data.

The Data mining Algorithms can be categorized into:

Association Algorithm

Classiﬁcation

Clustering Algorithm

**Association Algorithm**:

Association algorithms belong to the category of unsupervised learning techniques utilized in data mining to unveil compelling relationships among variables within extensive datasets. Their primary application often lies in market basket analysis, aiming to pinpoint frequently co-purchased items. A well-known algorithm in this domain is the Apriori algorithm, which excels in recognizing patterns like "customers purchasing X also show a tendency to buy Y" Such insights find utility in diverse areas such as optimizing product placement, facilitating cross-selling, and enhancing recommendation systems.

Association algorithms, like the widely used Apriori algorithm, play a fundamental role in Market Basket analysis. These algorithms function by identifying frequent itemsets within the data and subsequently deriving association rules from these itemsets. These rules reveal patterns such as "if item A is purchased, then item B is also likely to be purchased."

In the context of Market Basket analysis, these association rules provide valuable insights into customer purchasing behavior. Businesses can leverage these insights to devise strategies such as effective product placement, cross-selling, and targeted marketing campaigns.

The utilization of Association algorithms empowers businesses to gain a deeper understanding of customer preferences, thereby enabling them to optimize their sales and marketing strategies for enhanced customer satisfaction and improved business performance

Classification:

In market basket analysis, classification refers to the process of identifying connections or patterns between items that are commonly bought together. By recognizing these relationships, businesses can gain insights into customer behavior and preferences. This information can then be leveraged to strategically place products in stores, implement cross-selling tactics, and tailor marketing efforts to specific customer segments. Overall, this method is extensively utilized in retail and e-commerce sectors to boost sales performance and improve overall customer satisfaction.

Classification involves partitioning a dataset into distinct groups in a way that members within each group are closely related, while groups themselves are distinctly different. This closeness and difference are determined based on specific variable(s) being predicted. For instance, in a typical classification scenario, a database of companies may be divided into groups that are as internally consistent as possible in relation to a creditworthiness variable with "Good" and "Bad" values.

Clustering Algorithm:

In market basket analysis, clustering algorithms play a key role in categorizing items or products according to similarities in customer purchasing patterns. These algorithms are instrumental in identifying correlations and patterns among items that are commonly bought together by various customer segments. This approach enables businesses to gain a deeper understanding of customer preferences, refine product placement strategies, and customize marketing tactics to target specific customer clusters. By leveraging clustering algorithms, businesses can extract valuable insights from customer transaction data, ultimately leading to improved sales performance and heightened customer satisfaction.

Clustering involves the division of a dataset into distinct groups, ensuring that members within each group are closely related, while maintaining significant differences between the groups. This proximity and separation are determined by the measurement of distance across all available variables. With sufficiently large and high-quality databases, data mining technology offers the potential to create new business prospects through:

Automated prediction of trends and behaviours, streamlining the process of extracting predictive insights from vast databases, thereby addressing questions that previously demanded extensive manual analysis. For instance, targeted marketing can be significantly enhanced as data mining identifies the most promising targets for maximizing future returns on promotional mailings and addresses other predictive challenges such as forecasting bankruptcy and recognizing segments of a population likely to respond similarly to specific events.

Automated discovery of previously unknown patterns, as data mining tools uncover concealed patterns within databases in a single operation. While distributed data mining, like DARM, can discover rules from geographically dispersed datasets, communication costs are a consideration due to potentially slower network connections between the datasets in a non-parallel environment.

Proposed solution:

ODAM distinguishes itself from other algorithms through its superior performance achieved by minimizing candidate itemset generation costs. This is accomplished by addressing two key DARM issues: communication and synchronization. Communication, a critical DARM objective, can significantly impact the performance of DARM algorithms, thus reducing communication costs (e.g., message exchange size) is crucial. Synchronization requires each participating site to wait for a specific period until globally frequent itemset generation is complete; longer wait times occur when computing support counts takes more time, hence reducing the computation time of candidate itemsets' support counts. To address communication costs, various message optimization techniques are emphasized, including direct and indirect support counts exchange methods, each with distinct aims, expectations, advantages, and disadvantages. For instance, the first method involves exchanging each candidate itemset’ s support count to generate globally frequent itemsets of that pass (CD and FDM are examples of this approach).

**System requirement specification**

Environmental specification

Within market basket analysis, environment specification commonly denotes the distinct conditions or parameters established for carrying out the analysis. This encompasses elements such as the analysis's timeframe, the types of transactions to be encompassed, the products or categories being assessed, and any other pertinent limitations or criteria. Fundamentally, the environment specification delineates the extent and parameters within which the market basket analysis is executed, guaranteeing that the outcomes are targeted and pertinent to the specific business goals or research inquiries in question.

The environmental specifications list down the hardware’s and software’s required for the project.

Hardware:

4 GB RAM

320 GB HDD

Intel CORE V processor i5

Software:

Windows 10

Tableau

MY SQL Server 19

Data Structure:

The data set is a 2year transaction data of a Energy management company containing 8 columns and 764,836 rows. The number of years was identified using the query below:

“Select Count(Distinct(year(date)))

From transaction data”

The data contains sales transactions information on orders, products, customers, Purchase date, Branch ID and sales details.

Data preparation:

In SQL, preparing data for Market Basket analysis involves structuring transactional data to facilitate association rule mining. Here's a breakdown of the key steps involved:

**Data Extraction:** Initially, you extract transactional data from relevant tables in the database, capturing information about items purchased together in each transaction.

**Data Transformation:** The extracted data is then transformed into a "basket format," where each row represents a single transaction, and the columns denote the different items involved in the transaction.

**Data Aggregation:** Next, the data is aggregated to tally the occurrences of item combinations in transactions. This involves grouping the data by transaction and computing counts for pairs or sets of items that appear together.

**Filtering:** Depending on the analysis objectives, it may be necessary to filter out infrequent or irrelevant items or transactions to focus on meaningful patterns.

**Association Rule Mining:** Using SQL queries, association rule mining is performed on the prepared data. This step entails identifying frequent itemsets and establishing association rules based on predefined support and confidence thresholds.

Here's a simplified example of SQL queries for these steps:

-- Data transformation and aggregation

SELECT transaction\_id, GROUP\_CONCAT(item\_id ORDER BY item\_id) AS items

FROM transaction\_table

GROUP BY transaction\_id;

-- Association rule mining (using hypothetical syntax)

SELECT itemset, COUNT(\*) AS support

FROM transaction\_items

GROUP BY itemset

HAVING support >= min\_support\_threshold;

It's important to note that specific implementations may differ based on the database schema and the SQL dialect being used.

A comma separated file is created by running the query and then it is imported to MY SQL server this step helps us in understand the structure of the data in a better way.

Create database market\_basket\_analysis

Select \* from transaction data

In this step we can do data manipulations like changing the names of the columns to understand them and changing the date format as required.

After completing the data cleaning the data is now ready to perform the analysis.

Problem statement:

1 Identifying the top 3 frequently bought together products? -

To understand and analyze the requirement it is important to calculate the frequency

2 And the association level of these products?

Association rules offer information through "if-then" statements, which are derived from data. Unlike the deterministic if-then rules of logic, association rules are inherently probabilistic in nature.

Apart from the antecedent (the "if" part) and the consequent (the "then" part), an association rule is characterized by two numbers that indicate the level of uncertainty associated with the rule. In association analysis, the antecedent and consequent represent sets of items (referred to as itemsets) that are mutually exclusive (i.e., they do not share any items). The first number, known as the support for the rule, simply denotes the count of transactions containing all items in both the antecedent and consequent parts of the rule. (Sometimes, support is expressed as a percentage of the total number of records in the database.) The other number, referred to as the confidence of the rule, is the ratio of the number of transactions containing all items in both the consequent and the antecedent (i.e., the support) to the number of transactions containing all items in the antecedent.

# CHAPTER 2: ORGANIZATIONAL PROFILE

MIT School of Distance Education (MITSDE) is an esteemed institution dedicated to providing quality distance education in various fields of study. Established under the flagship of the prestigious MIT Group of Institutions, MITSDE has been at the forefront of delivering industry-relevant education through distance learning programs. With a focus on flexibility, accessibility, and excellence, MITSDE aims to empower learners to achieve their educational and professional goals.

Mission:

The mission of MITSDE is to provide affordable and flexible education through innovative distance learning methodologies. It strives to bridge the gap between academic knowledge and practical skills, enabling students to excel in their chosen fields and contribute to society.

Accreditations and Recognitions:

MITSDE is recognized and accredited by several esteemed organizations, ensuring the quality and credibility of its programs. Some of its accreditations and recognitions include:

* Distance Education Council (DEC): MITSDE is approved by the Distance Education Bureau of the University Grants Commission (UGC) and is a member of DEC.
* All India Council for Technical Education (AICTE): MITSDE is recognized by AICTE, which ensures the quality and standards of its technical programs.
* Association of Indian Universities (AIU): MITSDE is a member of AIU, which validates the equivalence of its programs with traditional degrees.

Programs Offered:

MITSDE offers a diverse range of distance learning programs across various disciplines, catering to the educational needs of working professionals, students, and individuals seeking career advancement. The programs include:

1. Postgraduate Diploma in Management (PGDM): Specializations in areas such as Marketing, Finance, Human Resource, Operations, IT, and Supply Chain Management.
2. Postgraduate Diploma in Business Administration (PGDBA): Specializations in Finance, Marketing, HR, Operations, and IT.
3. Postgraduate Diploma in Infrastructure Management (PGDIM): Focuses on the management of infrastructure projects, construction, and urban development.
4. Postgraduate Diploma in Project Management (PGDPM): Equips students with the skills to effectively manage and execute projects in various industries.
5. Postgraduate Diploma in Retail Management (PGDRM): Focuses on retail operations, merchandising, supply chain management, and customer relationship management.
6. Postgraduate Diploma in Financial Management (PGDFM): Concentrates on financial planning, analysis, investment, and risk management.

Learning Methodology:

MITSDE employs a robust and technology-driven learning methodology to ensure an engaging and interactive educational experience for its students. The key features of its learning approach include:

1. Self-Learning Material: MITSDE provides comprehensive study material in print and digital formats, enabling students to study at their own pace.
2. Online Learning: Leveraging advanced technologies, MITSDE offers online lectures, webinars, e-learning platforms, and interactive sessions to facilitate student-teacher interaction and collaborative learning.
3. Industry-Relevant Curriculum: The curriculum is designed to align with industry requirements and to impart practical skills and knowledge to students, ensuring their readiness for the professional world.
4. Student Support: MITSDE offers dedicated academic support to students through faculty interaction, doubt-solving sessions, online discussion forums, and personalized guidance.

Conclusion:

MIT School of Distance Education (MITSDE) stands as a prominent institution in the field of distance education, committed to providing quality programs and holistic learning experiences to students. With its strong emphasis on flexibility, industry relevance, and student support, MITSDE continues to empower learners, equipping them with the knowledge and skills needed to excel in their careers and contribute to society's growth.

It is contributing to the industrial, economic, and social growth of society for over a quarter of a century, Maharashtra Academy of Engineering Education and Research (MAEER)’s MIT Group of Institutions has helped realize the dreams and aspirations of thousands of students. The group has spread its wings across Maharashtra with campuses in Kothrud, Alandi, and Loni- Kalbhor within Pune, along with Latur, Talegaon, Ambejogai, and Pandharpur.

Being the brainchild of its visionary founder, Prof. Vishwanath D. Karad, MAEER established in 1983, managed to craft a niche position for being a one-of-its-kind undertaking that focused on value-based education.

# CHAPTER 3: PROJECT OBJECTIVES AND SCOPE

### OBJECTIVE OF STUDY

The objectives of the report on “Sales and Market Analysis” are as follows:

1. Evaluate a company's sales and marketing strategies with the goal of enhancing their sales performance. The assessment aims to recognize various strategies employed to boost sales and to pinpoint the most popular product combinations.
2. The primary objective of market basket analysis is to discover the items that customers are interested in purchasing. This analysis can assist sales and marketing teams in refining product placement, pricing, as well as cross-selling and up-selling strategies for greater effectiveness.
3. Predictive Market basket Analysis: This type utilizes supervised learning techniques such as regression and classification to simulate the market and study the factors influencing events. Its aim is to understand cross-selling by considering the sequence of items purchased.
4. Differential market Basket Analysis: This type of analysis proves valuable for competitive analysis. It involves comparing purchase histories across different brands, periods, seasons, and days of the week to identify interesting patterns in consumer behavior.
5. Market Basket Analysis assists retailers in comprehending which Stock Keeping Units (SKUs) are favored by customers. SKUs are a method used to track inventory in stores.
6. Predictive Market Basket Analysis utilizes supervised learning to predict future customer behavior. Through identifying cross-selling opportunities based on purchase patterns, it enables the implementation of tailored product recommendations, personalized promotions, and accurate demand forecasting. Predictive market basket analysis is of great significance in understanding and anticipating customer purchasing behavior, offering various advantages:
7. Enhanced Customer Insights: By predicting product combinations, businesses gain valuable customer behavior insights, enabling targeted marketing and improved satisfaction.
8. Optimized Inventory Management: Predictive analysis aids in forecasting demand, leading to improved inventory management and better product placement.
9. Strategic Cross-Selling and Upselling: Understanding product associations facilitates strategic sales opportunities, increasing customer lifetime value. Improved Marketing Strategies: Identifying related products allows for more effective marketing campaigns and promotions.

Operational Efficiency: Anticipating purchasing patterns helps streamline operations, enhancing efficiency and cost savings.

Ultimately, predictive market basket analysis enables data-driven decisions, improved customer experience, and overall business growth.

Market basket analysis is a procedure that seeks connections among entities and items that commonly occur together, such as the assortment of items in a shopper's cart (Source: "Using Information to Develop a Culture of Customer Centricity").

Performing a market analysis can be advantageous in multiple ways, including identifying trends and opportunities in your industry, distinguishing your business from competitors, and minimizing the risks and costs associated with launching a new business (or pivoting an existing one).

1. Through understanding customer behavior and preferences, market basket analysis enables businesses to gain valuable insights into their customers. This insight can be utilized to customize marketing campaigns, personalize communications, and develop tailored offers, ultimately resulting in enhanced customer satisfaction and loyalty.
2. Market basket analysis is valuable for optimizing product placement in both physical stores and e-commerce platforms. It enables businesses to strategically position items by identifying frequently associated products, thereby improving visibility, promoting cross-category purchases, and ultimately boosting overall sales.

In general, the report aims to illuminate the present status of Market Basket Analysis, awareness,

and adoption in the organization, offering valuable insights and recommendations for stakeholders

in the Energy management sector and policymakers to effectively promote digital services.

MEANING OF RESEARCH

Research refers to the systematic investigation and study conducted to gather information, analyze data, and gain insights about the sales and their strategies along with marketing.

Research in this report aims to explore and understand the knowledge, perceptions, attitudes, and behaviors sales in the Energy Management Industry. It involves gathering relevant data, analyzing it using appropriate methods and techniques, and interpreting the findings to draw meaningful conclusions.

The purpose of the research in this report is to assess the current state sales happening on the products including the demographic information, identify any gaps or misconceptions, and provide insights and recommendations to enhance the sales and services.

The research process may involve various steps, such as:

Defining the research objectives: Clearly stating the specific goals and objectives of the research, such as understanding the sales and Marketing strategies, identifying factors influencing adoption, or evaluating customer satisfaction with existing Energy Marketing services.

Designing the research methodology: Determining the appropriate research design and methodology, which could include surveys, focus groups, or a combination of methods. Selecting a representative sample of participants from different plants and consumers.

Data collection: Gather information regarding customer transactions, including the purchased items in each transaction, the transaction's time and date, and any other pertinent details.

Data analysis: First, clean and preprocess the data by eliminating irrelevant information, addressing missing values, and formatting the data appropriately for analysis. Next, employ association rules mining algorithms like Apriori or FP-Growth to recognize frequent item sets, which are groups of items commonly bought together in a transaction. Compute the support and confidence for each frequent item set to gauge the probability of one item being purchased alongside another. Then, create association rules using the frequent item sets and their associated support and confidence values. These rules signify the probability of purchasing one item when another item is purchased.

Interpretation and conclusions: Analyze the outcomes of the market basket analysis, recognizing common purchases, evaluating the correlation strength between items, and revealing additional valuable insights into customer behavior and preferences. Utilize the findings from the market basket analysis to guide business strategies like product recommendations, optimizing store layout, and executing targeted marketing campaigns.

Recommendations: Providing recommendations based on the research findings to enhance sales of Energy Marketing giant. These recommendations could include awareness campaigns, educational initiatives, improvements in infrastructure, or targeted marketing strategies.

By conducting research, the report aims to contribute valuable insights and recommendations that can help stakeholders, such as Electric, Energy management industries, policymakers, and individuals, to enhance their services and promote its benefits to the end users.

### NEED OF THE STUDY

The need for the study on “Sales and Market Analysis Project” arises from several factors:

Market Basket Analysis (MBA) plays a pivotal role in Sales and Market Analysis projects as it delves into transactional data to reveal associations and patterns. By identifying products commonly bought together, businesses can strategically place products, devise cross-selling tactics, and create targeted promotions. This comprehensive analysis yields valuable insights into customer behavior, empowering businesses to refine marketing strategies and enhance sales performance. Ultimately, MBA serves as a robust mechanism for grasping customer preferences and driving revenue growth.

1. The rapid expansion of traditional transactional data has created a pressing need for businesses to proactively utilize data to gain fresh perspectives for competitive benefits. To capitalize on these insights, businesses are being prompted to reevaluate their information management practices, dismantle organizational and data silos, and enhance business collaboration. Consequently, this has resulted in the presence of significantly large volumes of transactional data.
2. The task of categorizing this extensive transactional data presents a significant challenge to the data mining community. This is attributed to the continuous nature of transactional data streams, leading to challenges such as concept drift, infinite length, feature evolution, and concept-evolution
3. Infinite length refers to the unending nature of transactional data streams, while concept-evolution occurs with the emergence of new classes in a data stream.
4. To enhance the likelihood of purchasing specific items together and manage stock effectively, superstore managers can strategically place associated item pairs on adjacent shelves.
5. This approach offers businesses improved profit-making opportunities through better inventory management and marketing control. The fundamental concept in MBA aids in identifying associated item pairs within the store.
6. Concept drift within the context of market baskets involves analyzing a customer's purchasing patterns using the vast amount of data stored by businesses to predict changes in their preferences.

The goal of this analysis is to enable businesses to anticipate these changes, thereby satisfying customers and enhancing their revenue. A basic approach would involve generating all item pairs from transactional data streams in each market basket, storing them periodically, and using an appropriate algorithm to identify the most common item pairs. However, this approach is impractical in an environment where millions of market baskets contain billions of different item pairs. Some have suggested using static assumptions by training and testing Association Rule Mining (ARM) algorithms on a transactional dataset from the same population. Nevertheless, this approach does not promote the adaptability needed to address concept drift characteristics.

Market Basket Analysis improves comprehension of the frequent purchase of specific items together. This identification of item associations allows businesses to strategically position these items nearby, increasing their visibility and the probability of combined purchase. Additionally, this analysis empowers businesses to devise tailored promotions and cross-selling strategies, influencing customer purchasing patterns. Ultimately, utilizing the findings from Market Basket Analysis can lead to enhanced sales, customer contentment, and overall business success.

In the realm of market baskets, concept drift signifies the occurrence wherein the correlations between co-purchased items evolve over time. Such changes may stem from alterations in customer inclinations, seasonal fluctuations, or external influences. The ability to detect concept drift holds significant importance for businesses, enabling them to adjust their marketing tactics, product positioning, and promotional efforts to mirror the changing purchasing behaviors of their clientele. Through acknowledging and addressing concept drift, businesses can enhance the effectiveness of their product offerings and uphold the significance of their market basket analysis.

### SCOPE OF THE STUDY

The scope of the report on “Sales and Market Analysis Report” encompasses the following aspects:

Stages involved are following:

Data Collection: Gathering transactional data containing co-purchased items, and optionally, customer details.

Data Preprocessing: Cleaning the data, rectifying inconsistencies, and formatting it appropriately for analysis.

Association Rule Mining: Employing methods like the Apriori algorithm to unearth frequent itemsets and association rules, unveiling correlations between frequently co-purchased products.

Rule Evaluation: Assessing the identified rules using metrics such as support, confidence, and lift to pinpoint the most significant and actionable associations.

Interpretation and Application: Analyzing the outcomes to comprehend buying patterns and leveraging these insights for strategic business decisions like product placement, cross-selling, and targeted marketing.

Monitoring and Adaptation: Continuously monitoring market basket data for concept drift and adjusting strategies to align with evolving customer behaviors.

Throughout the project, collaboration with subject matter experts, effective visualization of findings, and clear communication of insights play pivotal roles in ensuring its success.

1. Geographic Scope: The report focuses specifically on the retail market it can be applied at different levels, such as a specific store, a regional chain of stores, a national market, or even on a global scale for online retail platforms.
2. Target Audience: The target audience for a retail store market basket analysis can include retail managers, marketing teams, and merchandisers. The findings can also be valuable for strategic decision-making, inventory management.
3. Apriori algorithm excelled in association rule mining compared to others. However, it was noted that some limitations include the extended time required for database scanning and the additional workload created by multiple scans for the database search.
4. Adoption Analysis: The research focused on performing association rule mining for a superstore using data sourced from the store's transactional database stored in Microsoft Excel spreadsheets and Microsoft Access databases.
5. To extract rules, the K-Apriori algorithm was applied to the database, generating frequent item-sets and association rules for different categories. The K-Apriori algorithm aims to identify consumer behaviors within groups to selectively cater to specific segments.
6. Factors Influencing Adoption: The report investigates the factors that influence the adoption or resistance towards marketing and sales. It explores the barriers, concerns, and challenges faced by organization when considering the sales and marketing services.
7. Attitudes and Perceptions: The report examines the attitudes, perceptions, and concerns of the Business. It aims to understand their views on the security, reliability, convenience, and benefits of MBA Analysis.
8. Recommendations: Based on the findings of the study, the report provides recommendations to enhance e-banking awareness and adoption in Ahmedabad. These recommendations are tailored to address the identified barriers and promote the benefits of e- banking services among the target audience.

This project aims to create a scalable data analytics market basket model capable of handling data streams of any size and offering real-time visualization of transactions to track concept drift trends. The specific goals are to (1) devise a Scalable Data Analytics Market Basket Model (SDAMBM) for Transactional Data Streams and (2) put the model into operation and assess its performance.

### TYPE OF RESEARCH

This study is DESCRIPTIVE in nature. It helps in breaking vague problems into smaller and more precise problems and emphasizes on discovering new ideas and insights

DATA COLLECTION METHOD

#### Primary Data:

The Project utilizes transactional data obtained from a Sample superstore, which is compatible

with .csv, and .xlsx file formats.

### SAMPLE DESIGN AND SIZE

The Sample data in the dataset contains labels such as Invoice number, quantity of items, item sets, unit price, and transaction date. For this project, the focus is on utilizing the invoice number, unit price, and item labels from the dataset to conduct market basket analysis, which aids in capturing the concept drift of each customer based on transaction instances over time.

### TOOLS AND TECHNIQUES OF ANALYSIS

The various tools and techniques of analysis to examine and interpret the collected data. Here are some commonly used tools and techniques:

1. The data ingestion layer prioritizes data from diverse sources into the database. This includes batch data already in the database and streaming data (such as customer profiles and spending behaviors) generated as customers carry out transactions. Both types of data are in the .xlsx format, which is supported by Tableau.
2. In the data pipelining layer, the collector layer gathers and creates data pipelines to manage the transformation and transfer of data from multiple sources to the ultimate processing destination. Each stage in this process utilized distinct software solutions.
3. Within the data processing layer, the pipelined data were directed to various destinations, and the corresponding data flows were categorized using the data model from the semantic library. For instance, the diverse transactional data obtained from the superstore were leveraged to forecast customer concept drift by merging it with the customer profile data streams at different levels. This routing aims to interface with big data analytics visualization for each transaction instance to predict customer concept drift in market basket analysis.
4. Data Visualization: Presenting the findings through visual representations, such as charts, graphs, and diagrams, can make complex data more accessible and understandable. Visualizations can help illustrate trends, patterns, and comparisons.

The SDAMBM, a Scalable Data Analytics Market Basket Model, was developed to process transactional data streams. Its design involved curating various dimensions from the Elixir and a research conceptual model. This curation process included adding filters to the selected dimensions, forming part of the Business management visualization layer. To address the concept drift problem in transactional data streams for market basket analysis, a combination of Toad Modeler, Corel Draw, and CRAN was used for customer management in the design of SDAMBM.

**Tableau: As a Visualization tool**

Tableau is a data visualization tool that allows users to create interactive and shareable dashboards. It enables users to connect to various data sources, such as databases, spreadsheets, and cloud services, and then transform that data into interactive visualizations and business intelligence reports. Tableau's drag-and-drop interface makes it easy for non-technical users to create customized dashboards, charts, and graphs without requiring extensive programming knowledge.

The software is designed to help businesses and organizations make data-driven decisions by providing a user-friendly platform for exploring and understanding their data. Tableau's visual analytics capabilities make it an ideal tool for uncovering insights, trends, and patterns within large datasets, including those used in market basket analysis.

Overall, Tableau empowers users to create compelling visualizations that can aid in understanding data, identifying trends, and communicating insights effectively. It's widely used across industries for tasks such as market analysis, sales forecasting, and operational analytics.

Tableau is ideal for market basket analysis due to its powerful visualization capabilities and interactive features. It allows users to visually explore large transactional datasets to identify co-occurring items and patterns in customer purchasing behavior. With Tableau, users can create intuitive dashboards and visualizations that make it easy to identify relationships between products and understand customer behavior.

Moreover, Tableau's drag-and-drop interface enables users to quickly analyze large volumes of data and uncover insights without needing extensive programming or statistical knowledge. This makes it particularly well-suited for market basket analysis, as it allows users to explore and analyze transactional data in a user-friendly manner.

Additionally, Tableau's ability to connect to a variety of data sources, including databases and spreadsheets, makes it convenient for accessing and analyzing the diverse data typically involved in market basket analysis.

1. Overall, Tableau's intuitive visual analytics tools, interactive dashboards, and broad data connectivity make it a powerful choice for conducting market basket analysis, providing users with the means to gain valuable insights into customer behavior and purchasing patterns.

### LIMITATIONS OF STUDY

1. Sample Size and Representation: The study's findings may be influenced by the size and representativeness of the sample. If the sample size is small or not diverse enough, it may not accurately reflect the combinations of the analysis made.
2. Self-Reported Data: The study may rely on self-reported data obtained through sample superstore data. This type of data is generic. This can be mapped with existing data from the energy management data but currently due to limitations using sample data to showcase how the analysis is done and models are created.
3. Limited Scope: The study's scope may be limited to assessing awareness and adoption of Market basket analysis without exploring other relevant factors that could influence products analyzed.
4. Time Constraints: The study's timeframe may impose constraints on data collection and analysis. This could limit the ability to capture changes in awareness and adoption levels over an extended period or account for potential fluctuations due to external factors.
5. Lack of Control: As a non-experimental study, there may be limitations in controlling external variables that could impact analysis. Factors such as media influence, marketing campaigns, or economic conditions may have an impact that cannot be directly controlled or measured.

It is important to acknowledge and address these limitations when interpreting the findings of the study. While efforts are made to mitigate these limitations, they should be considered to provide a comprehensive and accurate understanding of analysis done on the data.

### INTRODUCTION OF MARKET BASKET ANALYSIS

Market basket analysis, a data mining technique, examines co-occurrence patterns and evaluates the connection strength between concurrently purchased products. It is also known as frequent itemset mining or association analysis. By recognizing these patterns in any retail environment, it helps understand customer behavior by identifying relationships between purchased items. In essence, market basket analysis enables retailers to identify frequently co-purchased products, ensuring their availability in inventory.

Before directly jumping into what is Market basket analysis and how it’s done, let’s just talk about why we do MBA (Market basket Analysis) and where it came from. So, MBA is generally a data mining technique that discovers the co-occurrence relationships between items, it is mainly used by business owners to pump up their sales. As Analyst we analyse all transactions and items bought together and measure the probabilities. We then provide these inferences to business user to make critical decisions to boost their sales. How its done, is explained below.

Definition: - Market Basket Analysis is a modelling technique based upon the theory that if you buy a certain group of items, you are more (or less) likely to buy another group of items.Market Basket Analysis can be well explained by answering this simple question.What is the probability of buying product A with product B when bought together?

MBA can only be done when you have bought more than one product in a single transaction. When I say transaction, you can relate it to order or your list of items you want to buy together. For instance, think of a shopping cart or trolley as your basket and we want to analysis the list items you bought.

By far, the most common approach to perform Market Basket Analysis is the Apriori Algorithm. First proposed in 1994 by Agrawal and Srikant, the algorithm has become historically important for its impact on retailers to meaningfully track transaction associations.

Market basket analysis uses association rules

> to forecast the likelihood of specific products being bought together. It involves tallying the co-occurrence of items and identifying associations that happen more frequently than anticipated.

Market analysis theories encompass the principles and structures utilized to assess and comprehend the market landscape and its influence on business prosperity. They play a vital role in recognizing opportunities and threats within the market, essential for fostering innovation and devising impactful business strategies.

1. Benefits and Advantages: Leveraging market basket analysis can lead to boosted sales and heightened customer satisfaction. By utilizing data to identify frequently co-purchased products, retailers can enhance product placement, introduce special promotions, and develop new product bundles, thereby stimulating additional sales of these product combinations.
2. Impact on Banking Industry: Enhance pricing strategies and establish key value item (KVI) pricing strategies and discount guidelines. Pinpoint the most effective promotions to expand market presence. Obtain a deep understanding of primary products for each customer segment, facilitating the development of tailored assortments and pricing for specific customer groups.
3. Enterprises utilize Market Basket Analysis as a tactic for devising store layouts rooted in customers' shopping patterns and buying histories. This concept is also relevant for training machine learning algorithms to assist businesses, particularly within the e-commerce domain.
4. In conclusion Retailers employ market basket analysis, a data mining method, to enhance sales through a deeper comprehension of customer buying behaviors. This technique encompasses the examination of extensive datasets, like purchase history, to unveil product groupings and identify products commonly purchased together.

## Architectural representation of the Data flow-

The platform enhances the relationship between customers and businesses by analyzing

purchasing patterns within the concept drift of customer transactions. It also aids institutional

research by assessing how a customer's consumption impacts their performance known as

extended effects of concept drift, which can be addressed through big data analytics.

Assisting in enhancing the analytics processing of transactional data streams in retail

marketing, the software Luigi version 3.7 was utilized to monitor the concept drift of customers.

This was achieved by studying their purchasing patterns in market basket analysis across a range

of analytics processing methods.

Understanding the architecture of the Market Basket is crucial for gaining insights into the flow of processes within retail and transactional environments. This structure serves as a framework for analyzing the relationships among various items or products. By examining these relationships, businesses can uncover valuable insights into how products are interconnected and how they behave within the context of purchases or transactions.

By delving into the Market Basket architecture, businesses gain the ability to visualize and interpret the intricate web of connections between products. This visualization provides a deeper understanding of how products are linked and how they traverse through the purchasing process. Ultimately, this comprehension empowers businesses to make well-informed decisions regarding product placement, marketing strategies, and inventory management.

In essence, by grasping the nuances of Market Basket architecture, businesses can enhance their strategic decision-making processes, optimize their product offerings, and improve overall operational efficiency.

### ADVANTAGE & DISADVANTAGE OF MARKET BASKET ANALYSIS

**5 MAIN ADVANTAGES OF MARKET BASKET ANALYSIS**

#### Optimization of Campaigns and Promotions

Incorporating Market Basket Analysis into B2B marketing contributes to optimizing B2B promotional campaigns by identifying products or services that are frequently purchased together, thereby increasing the effectiveness of the campaigns.

#### Increased Sales and Optimized ROI

Predictive sales analytics provides a clear route to enhance sales efficiency and boost revenues. Sales leaders consistently assess the challenges their organizations encounter. Being adaptable and responsive to technology enables organizations to promptly address any potential issues and remain agile.

#### Optimization of in store operations

By leveraging a wide range of data points, including first-party and third-party customer data, market basket analysis helps marketers prioritize customers' purchasing patterns, thereby optimizing their in-store applications.

#### Increase in the Market share

Market basket analysis can be utilized to link purchases with demographic and socioeconomic data, enabling the promotion of the most relevant product or service combinations. This, in turn, aids in optimizing customer behavior and ultimately leads to increased sales revenue.

#### Helps in analyzing Customer Behavior

#### A proper analysis help in identification of purchase patterns. This will allow to establish a

#### long term relationship with the customers and provide them better services.

### DISADVANTAGES OF MARKET BASKET ANALYSIS

It offers insights into which products are commonly bought together, yet it doesn't explain the reasons behind these joint purchases. This limitation can reduce the applicability of the analysis for specific uses. Additionally, as it depends on customer data, it may raise privacy issues.

Limited to Transactional Data: Market basket analysis relies solely on transactional data, which may not capture all aspects of customer behavior such as preferences, motivations, or external factors influencing purchasing decisions.

Lack of Causality: While market basket analysis uncovers associations between products, it does not inherently reveal causality, making it challenging to discern whether the relationships are meaningful or coincidental.

Dynamic Market Conditions: Market basket analysis may struggle to adapt to rapidly changing market conditions, consumer trends, or seasonal variations, potentially leading to less accurate insights or outdated recommendations.

Privacy and Ethical Concerns: Analyzing customers' purchase patterns raises privacy concerns, and if mishandled, it can lead to ethical and regulatory issues, especially with the increasing emphasis on data protection laws.

Overlooking Individual Preferences: Market basket analysis focuses on aggregate behavior, potentially overlooking individual customer preferences, leading to generalized strategies that may not resonate with all customer segments.

These drawbacks highlight the need for complementary analytical methods and the importance of interpreting market basket analysis results within the broader context of customer behavior and market dynamics.

Market basket analysis enables industries to understand customer purchasing behaviour and identify associations between products purchased together, aiding in targeted marketing strategies.

It helps industries optimize product placement by uncovering which items are often bought together, thus enhancing the shopping experience and potentially increasing sales.

By identifying cross-selling opportunities, market basket analysis assists industries in boosting revenue and profitability through strategic product bundling or promotions.

It provides insights into customer preferences and trends, empowering industries to tailor their product offerings and improve customer satisfaction.

Industries can utilize market basket analysis to forecast demand for specific products, enabling better inventory management and reduced stockouts.

It supports effective pricing strategies by revealing price sensitivity and understanding how different products' prices impact overall purchasing patterns.

Through understanding customer behavior, industries can personalize recommendations and offers, leading to enhanced customer loyalty and retention.

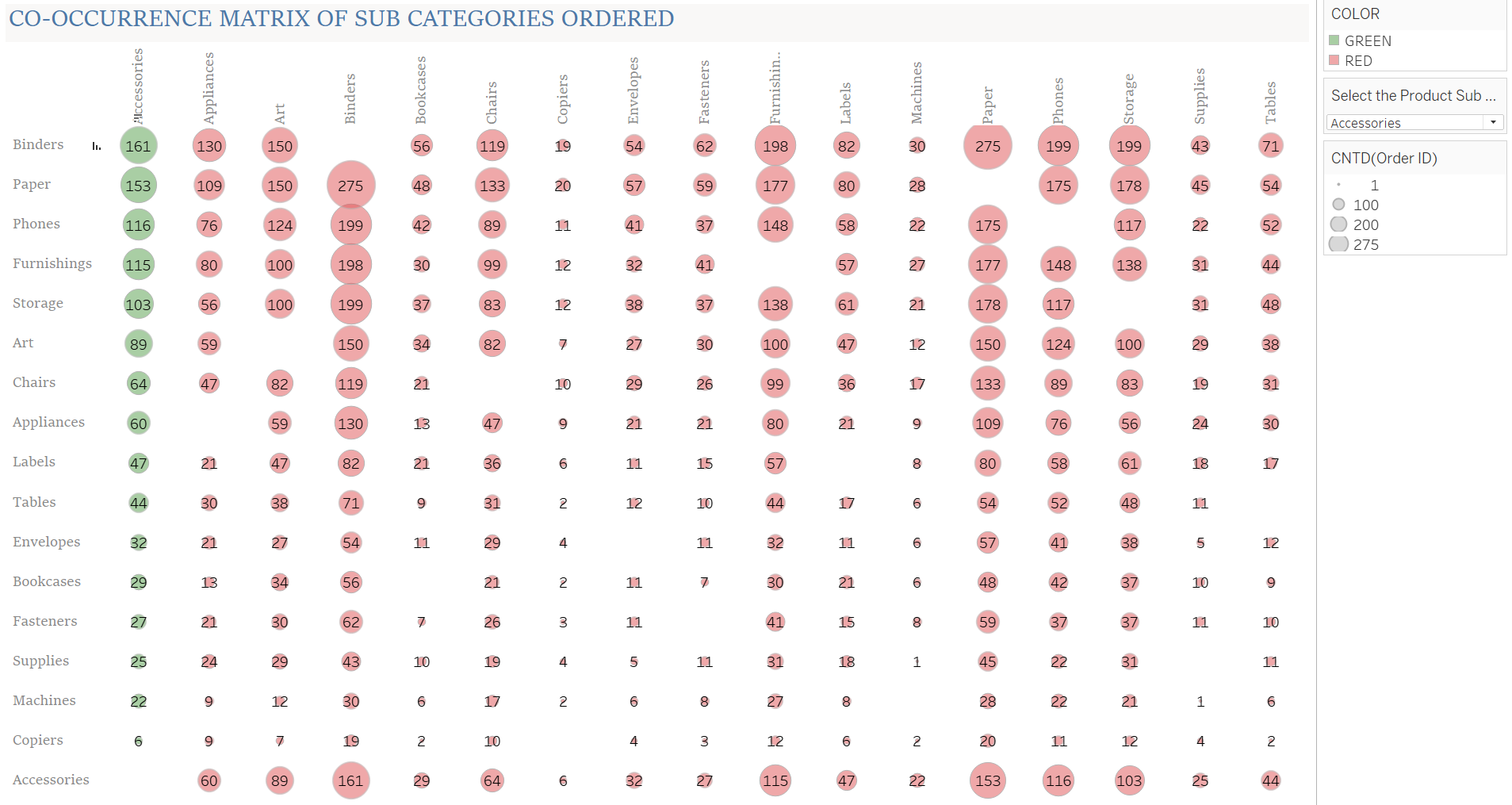
Market basket analysis aids in detecting anomalies or unusual purchasing patterns, which can be crucial in fraud detection and risk management for industries.

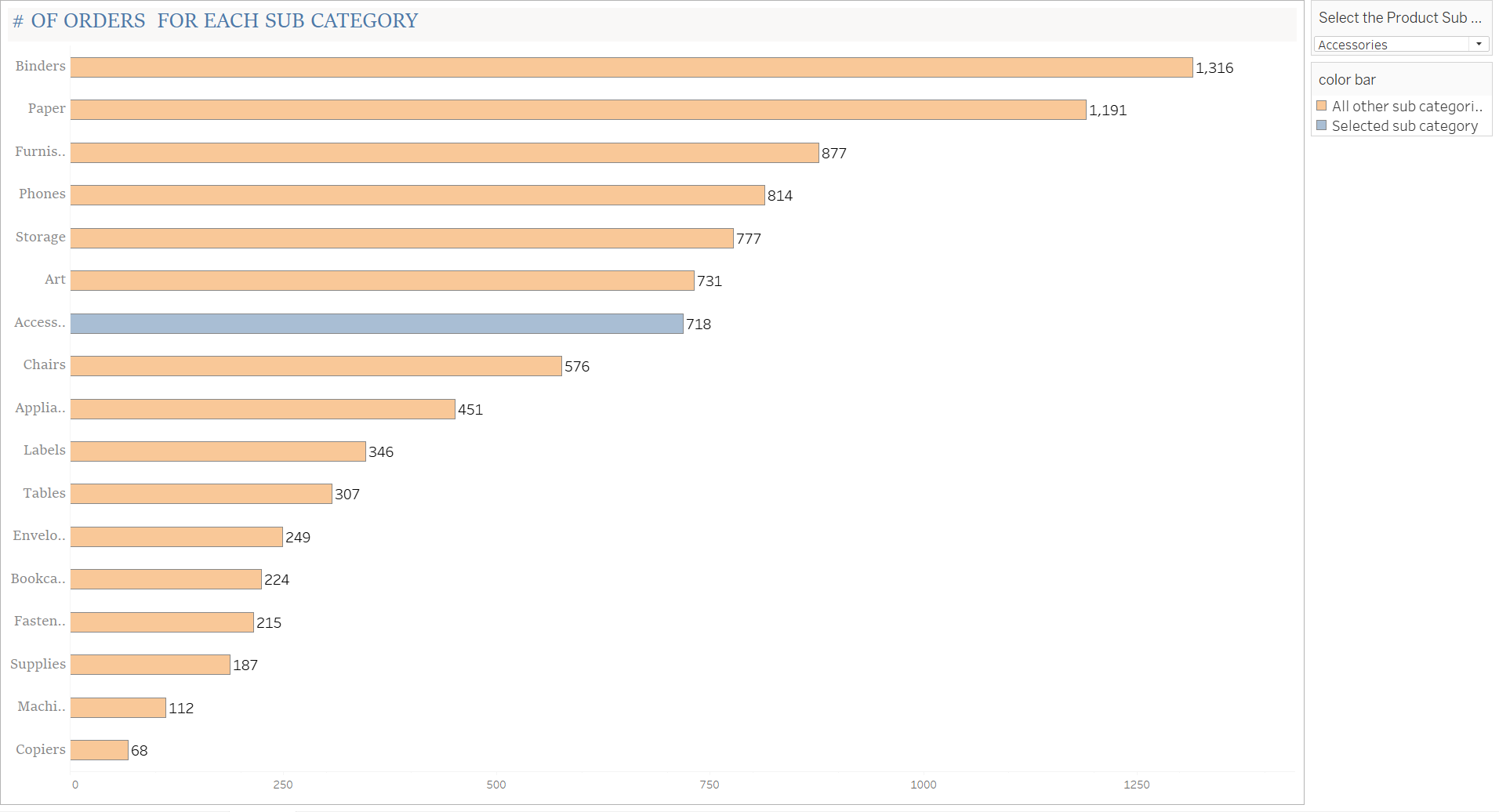
It contributes to efficient supply chain management by optimizing procurement processes and supplier relationships based on customer demand patterns.

Ultimately, market basket analysis empowers industries to make data-driven decisions, improve operational efficiency, and better meet the evolving needs of their customer base.

# CHAPTER 4: DATA ANALYSIS AND INTERPRETATION

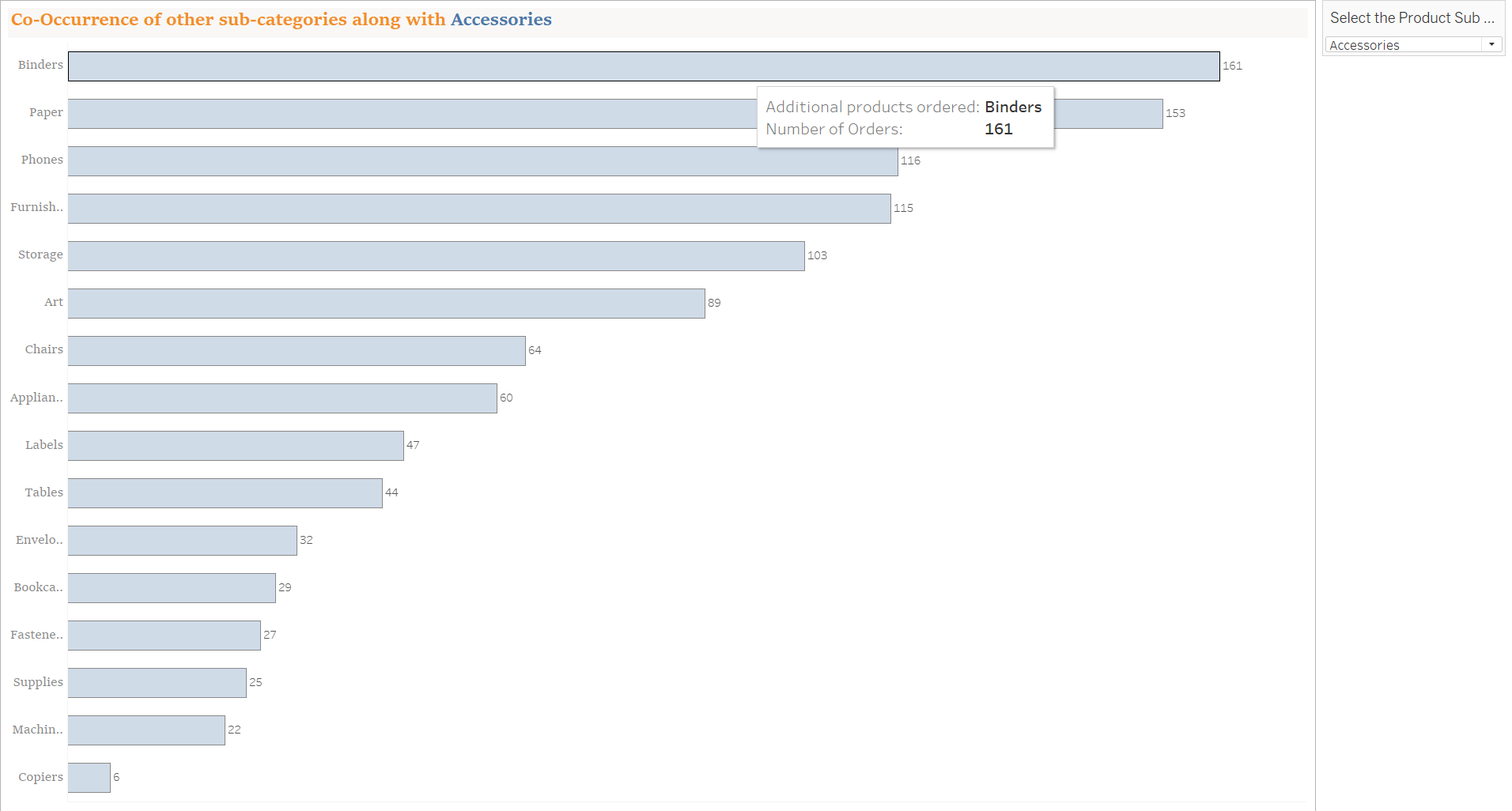
The below graph represents the patterns and no of times these products are bought together. The number indicates the no of times. And the color is the selection made on the subcategory.





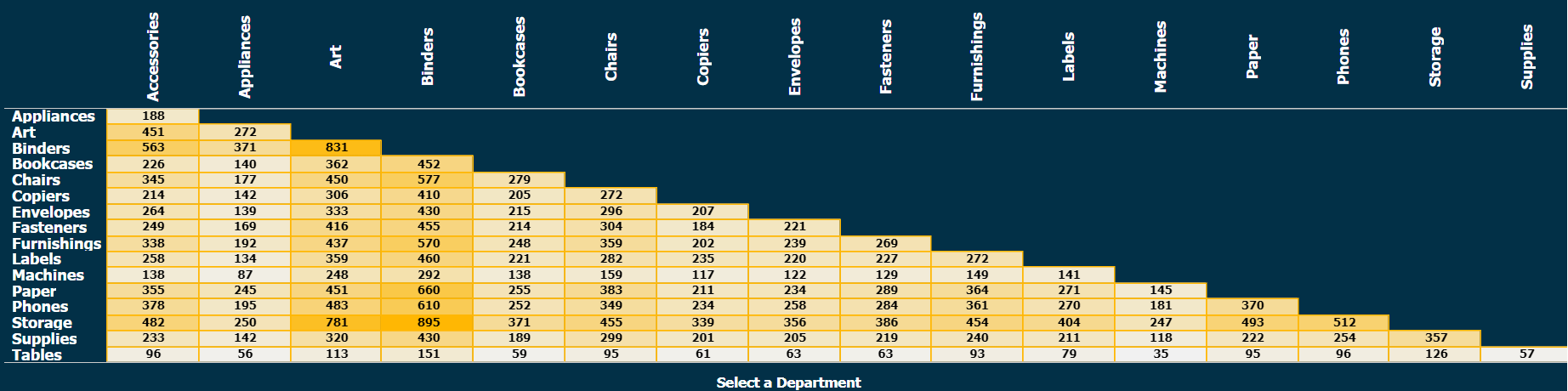
The above graph represents the no of orders placed by the customers subcategory wise. The order count is given on the respective bar graph.

**How many times the products are purchased together?**

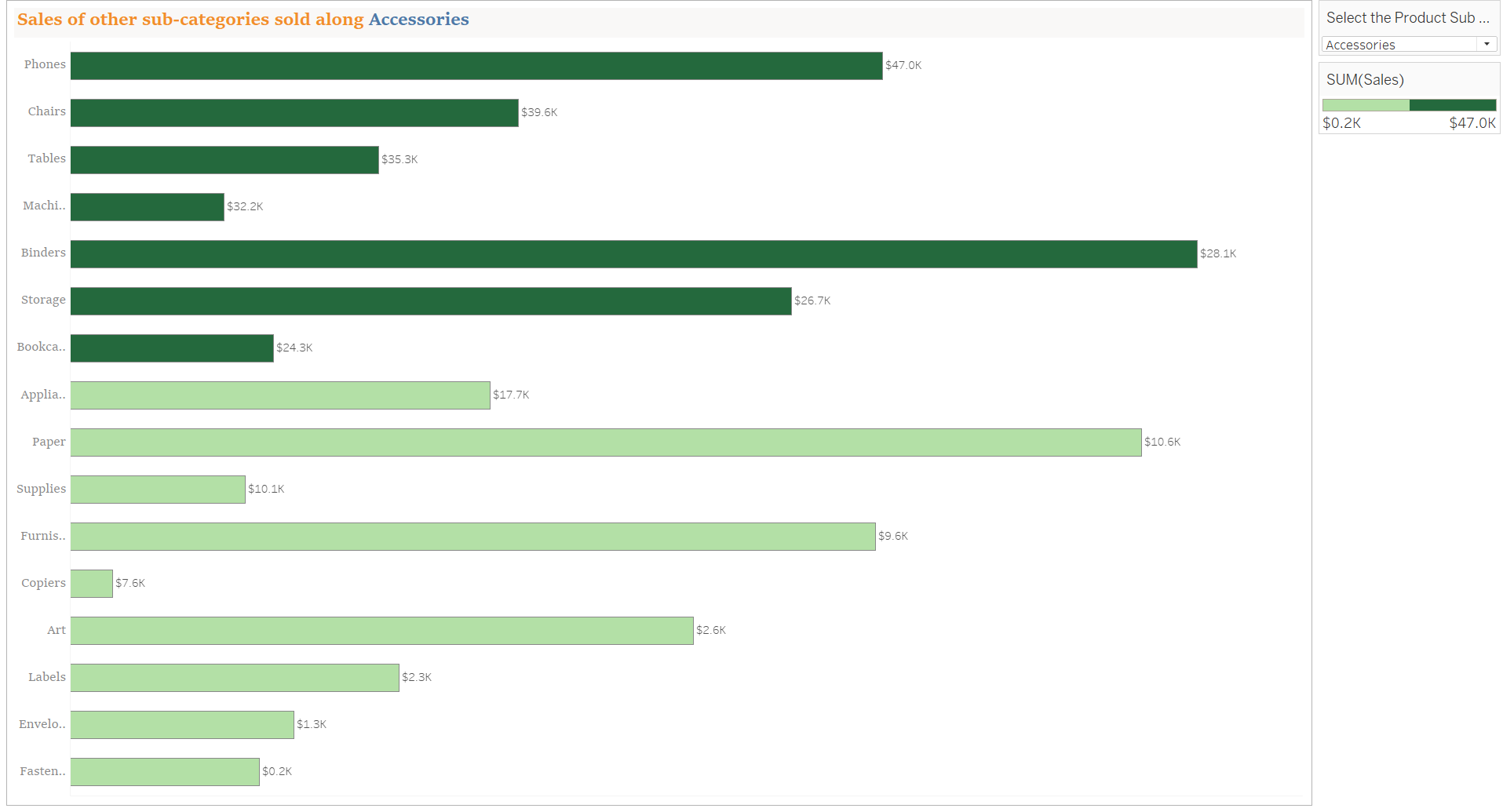


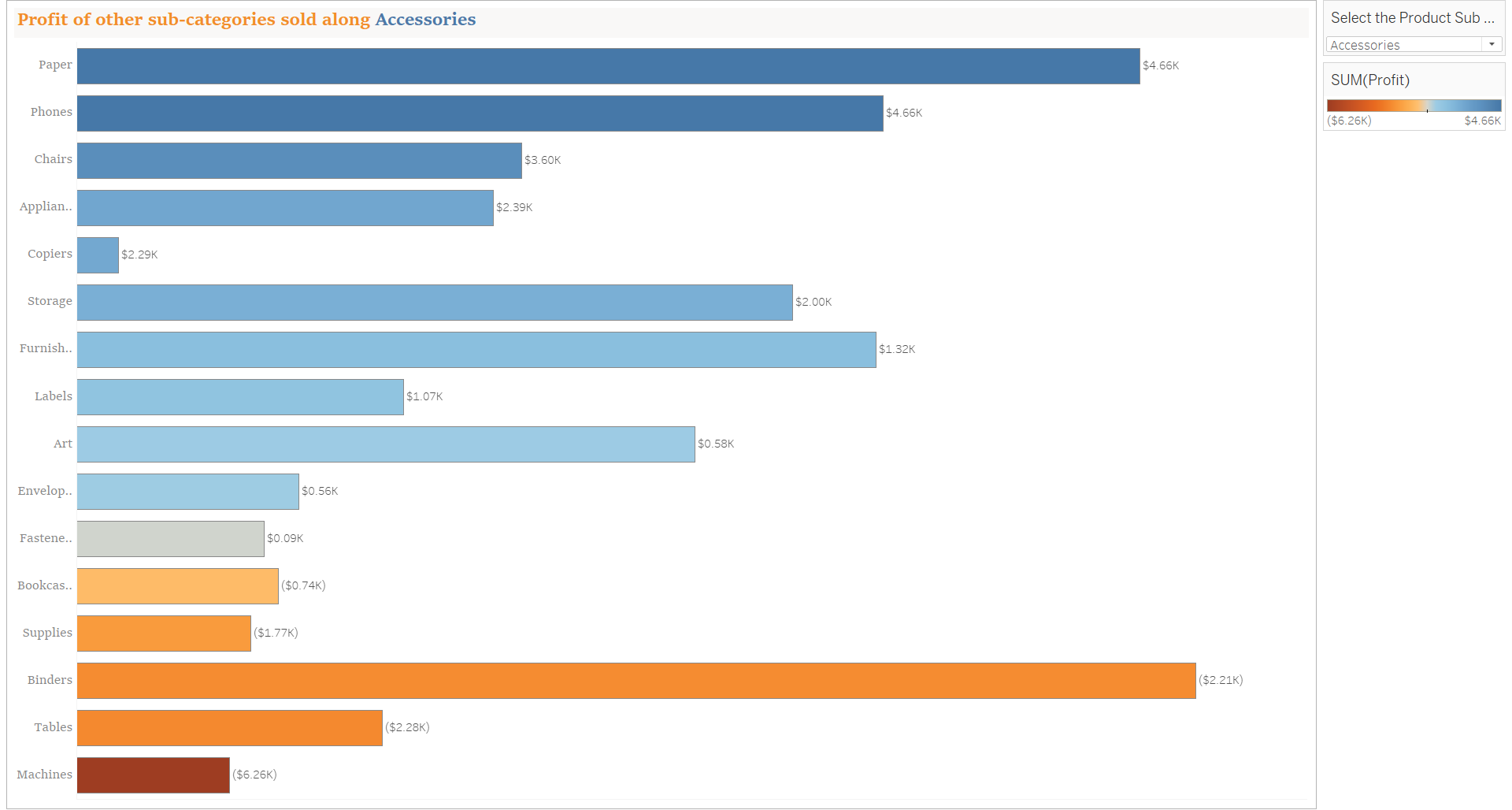
The above chart tells us how many times these products are ordered together in the above example the graph is showing

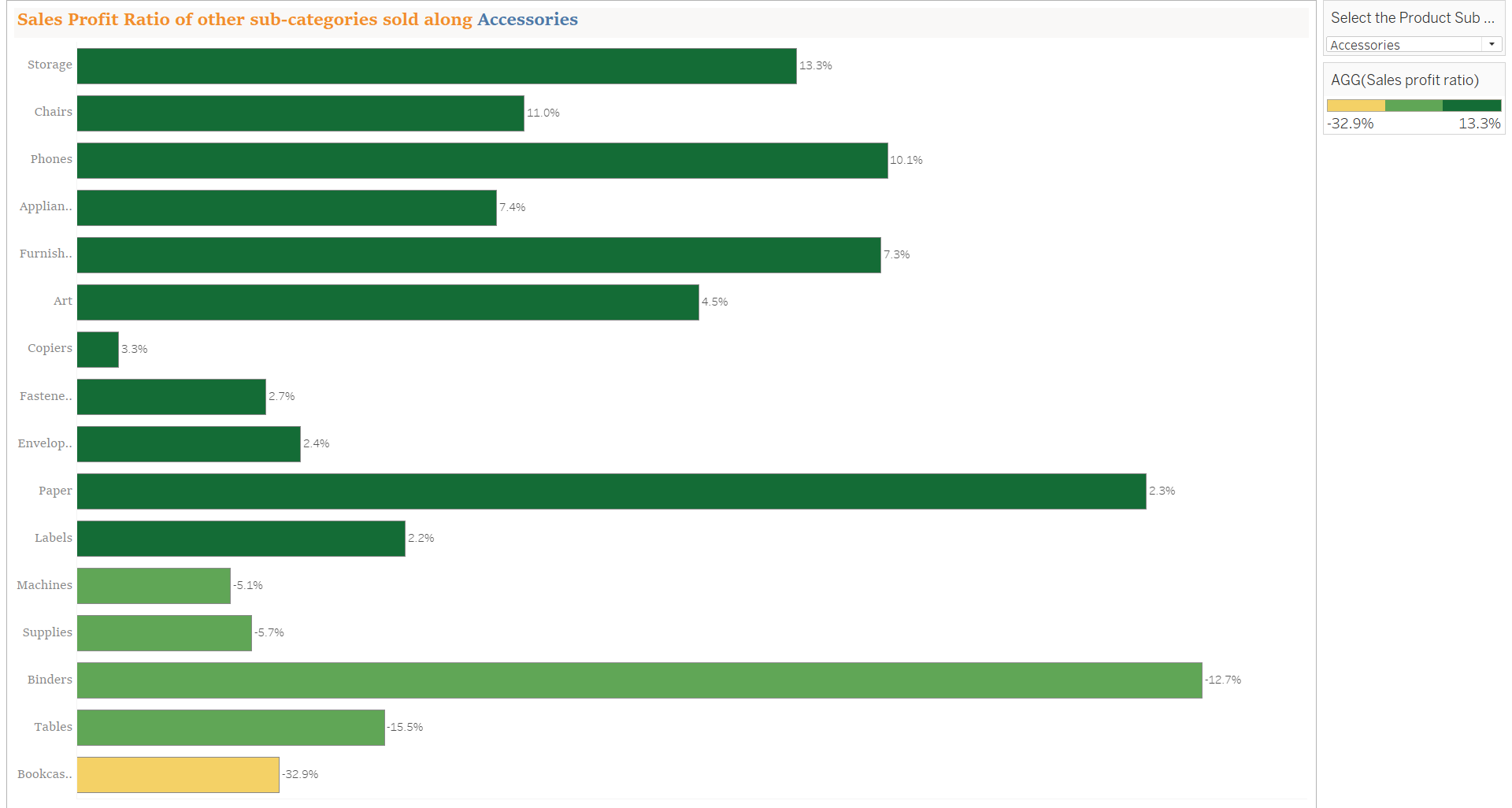
Accessories subcategory and along with Accessories how many times the other subcategories are purchased.



In the above chart the topmost bar is the selected department “Accessories” and the below bars are the associated subcategories of the selected subcategory” Accessories”









The above graph showcases the purchasing behavior and co relation between the various purchased items.

This helps business to introduce new campaigns and loyalty programs and help them understand the picking pattern of the products in the store layout.

# CHAPTER 5: CONCLUSION AND FINDINGS

**CONCLUSION**

In general, this study presented a data analytics strategy to manage shifts in customer purchasing

behavior, aiming to enhance the shopping experience and customer retention for the superstore.

The focus was on implementing a scalable data analytics market basket model for transactional data

streams using big data analytics. The study addressed inherent problems in transactional data streams

and resolved concept drift issues in market basket analysis (MBA) by examining customer buying

patterns and making predictions about potential shifts in purchasing behavior.

Transactional data from the University superstore spanning two years

(January 2, 2017, to December 31, 2018) was utilized, along with simulated streamed data from online

customer transactions. To monitor concept drift while

safeguarding data, a selectivity mechanism was employed to aggregate the data sources for data protection.

Additionally, the data sets were calibrated using a calibration function algorithm before undergoing the

data analysis and evaluation process.

A Scalable Data Analytics Market Basket model (SDAMBM) was developed to engage with the University

superstore for managing its operations. This model offers a broad overview of customer transactions and a

personalized view through the Billy Big data visualization dashboard created with Tableau analytics.

It presents a consumption pattern summary for all customers, including the potential impact of their consumption on health and academic performance, thereby adding significance to big data analysis.

Additionally, it incorporates a restock level to inform the University superstore about inventory levels

for timely restocking.

The customer management model was specifically designed to capture and adapt to changes in customer

consumption patterns, strengthening the bond between customers and the University superstore.

This study enriches the existing knowledge base by introducing a scalable data analytics market basket

model for transactional data streams and an improved approach for managing concept drift in customer

transactional data streams within market basket analysis. These contributions advance the discourse on

developing big data analytics platforms for predicting concept drift in market basket analysis using historical

and current transactions.

In summary, conducting comprehensive sales and marketing analysis plays a pivotal role in gaining deep insights into customer behavior, by giving in-depth sales and marketing analysis is crucial as it allows businesses to gain profound understanding of customer behavior through the examination of various touchpoints such as purchase history, browsing patterns, and engagement metrics. This process involves the systematic review of customer demographics, preferences, and feedback, enabling companies to discern the underlying motivations driving purchasing decisions. By delving into these insights, organizations can tailor their marketing strategies to align with customer needs, thus enhancing customer satisfaction and loyalty. Such a nuanced understanding of customer behavior forms the foundation for creating targeted and effective sales and marketing campaigns, ultimately driving business growth and profitability. discerning evolving market trends, and refining strategies aimed at fostering business expansion. Through the utilization of data-driven insights, organizations are empowered to make well-informed decisions, tailor marketing initiatives to individual customer preferences, and capitalize on a myriad of sales prospects. This analytical process essentially forms the bedrock for attaining a competitive edge and ensuring enduring prosperity in the constantly evolving and competitive contemporary business landscape.

This analytical process essentially serves as the cornerstone for gaining a competitive edge and ensuring sustained prosperity in the perpetually evolving business landscape. By meticulously dissecting market data, consumer trends, and competitor activities, organizations can identify untapped opportunities, anticipate shifts in consumer preferences, and adapt their strategies proactively. This proactive approach enables businesses to stay ahead of the curve, innovate effectively, and maintain relevance in a dynamic marketplace, thereby establishing a solid foundation for long-term success and growth.

# CHAPTER 6: SUGGESTIONS AND RECOMMENDATIONS

Certainly! Market basket analysis can provide valuable insights for businesses. Here are some suggestions and recommendations for conducting market basket analysis:

**Identify Product Affinities**:

Use market basket analysis to uncover products that are frequently purchased together. This can help in optimizing product placement, bundling, and cross-selling strategies.

It involves employing market basket analysis to reveal which products are commonly bought together. This analysis can be instrumental in optimizing how products are placed in stores, creating product bundles, and devising cross-selling strategies. It helps businesses understand customer purchasing behavior and preferences, enabling them to enhance their marketing and sales tactics.

Market basket analysis, often used in retail and e-commerce, examines transaction data to uncover patterns of co-occurrence among products in customers' baskets. By identifying these product affinities, businesses can strategically position related items close to each other in stores or on websites, encouraging additional purchases. Furthermore, this analysis can guide the creation of product bundles, leading to increased sales and customer satisfaction. It also facilitates the development of cross-selling strategies, allowing businesses to recommend complementary products to customers, thereby boosting revenue. This approach ultimately helps businesses tailor their offerings to customer preferences and maximize their sales potential.

**Personalized Recommendations**:

Leverage the insights from market basket analysis to offer personalized product recommendations to customers, enhancing their shopping experience and increasing sales.

**It** involves using the insights gained from market basket analysis to provide tailored product suggestions to customers, thereby improving their shopping experience and driving sales. By understanding which products are frequently purchased together, businesses can offer personalized recommendations that resonate with individual customer preferences. This not only enhances the customer's shopping journey but also increases the likelihood of additional purchases. It's a strategic way to build customer loyalty and boost overall revenue.

Leveraging the insights from market basket analysis allows businesses to understand the purchasing behaviors of their customers. By identifying which products are often bought together, companies can create personalized recommendations for individual customers based on their specific preferences and buying patterns. This level of personalization enhances the overall shopping experience, as customers are presented with suggestions that align closely with their interests.

Moreover, personalized recommendations have a direct impact on increasing sales. By showcasing items that are highly relevant to each customer, businesses can significantly improve the chances of additional purchases. This approach not only drives revenue but also fosters customer satisfaction and loyalty, as it demonstrates an understanding of and care for the customer's individual needs.

Overall, the use of personalized recommendations, informed by market basket analysis, represents a powerful strategy for businesses to connect with their customers on a deeper level, drive sales, and cultivate long-term relationships.

**Inventory Management**:

Use the analysis to optimize inventory management by stocking related products in close proximity, thereby improving operational efficiency and customer convenience.

It involves utilizing the analysis derived from market basket analysis to enhance the efficiency of inventory management processes. By stocking related products in close proximity based on purchasing patterns, businesses can streamline operations and enhance customer convenience.

This approach allows businesses to strategically organize their inventory by placing items frequently purchased together in nearby locations. As a result, it optimizes the operational flow, making it easier to manage and replenish stock. Moreover, by stocking related products in close proximity, businesses can enhance customer convenience, as customers can easily find complementary or related items during their shopping experience.

Ultimately, this optimization of inventory management not only improves operational efficiency within the business but also enhances the overall customer experience. It ensures that the right products are readily available, leading to increased customer satisfaction and potentially higher sales.

By leveraging the insights gained from market basket analysis, businesses can implement a more strategic and customer-centric approach to inventory management, thereby reaping benefits in terms of operational efficiency and customer satisfaction.

By using the insights from market basket analysis to optimize inventory management, businesses can strategically position related products in close proximity within their retail or storage spaces. This strategic placement is based on the frequent co-occurrence of these items in customer transactions.

Optimizing inventory management in this way offers several benefits:

Operational Efficiency: Placing related products together simplifies the stocking and restocking processes. It reduces the time and effort required to manage inventory, as employees can easily identify and restock items that are often purchased together. This streamlines operations and contributes to overall operational efficiency.

Enhanced Customer Convenience: Customers benefit from the convenient arrangement of related products. When complementary or frequently paired items are placed close to each other, it makes it easier for customers to find and purchase all the items they need in one location. This enhances the overall shopping experience and convenience for the customers.

Increased Sales Opportunities: Strategic inventory management based on market basket analysis can lead to increased sales. By making related products more accessible to customers, businesses can capitalize on cross-selling opportunities. Customers may be more inclined to purchase additional items when they find related products conveniently located together.

Overall, leveraging market basket analysis for inventory management optimization is a strategic approach that not only improves internal operational processes but also contributes to a more seamless and satisfying experience for customers.

**Promotions and Pricing**:

Tailor promotional offers and pricing strategies based on the relationships between products identified through the analysis, thus maximizing sales and customer satisfaction. can be customized by leveraging the insights derived from market basket analysis to create targeted promotional offers and pricing strategies. By understanding the relationships between products and customer purchasing behaviors, businesses can optimize their promotions and pricing to maximize sales and enhance customer satisfaction.

Analyzing the relationships between products allows businesses to:

Create Targeted Promotional Offers: Businesses can design promotions and marketing campaigns based on the associations between products frequently purchased together. By offering discounts or special promotions on related items, businesses can encourage customers to purchase complementary products, thus increasing sales and customer satisfaction.

Optimize Pricing Strategies: Pricing strategies can be adjusted based on the analysis of product relationships. For example, businesses can implement bundling strategies where related products are offered at a discounted price when purchased together. Additionally, understanding which products are commonly purchased together can inform pricing decisions, ensuring competitive pricing for related items.

Enhance Customer Satisfaction: Tailoring promotions and pricing based on product relationships can lead to a more personalized and relevant shopping experience for customers. By offering promotions on items that align with customers' purchasing patterns, businesses can enhance customer satisfaction and loyalty.

By leveraging market basket analysis to tailor promotions and pricing strategies, businesses can create a more targeted and customer-centric approach to marketing and sales. This approach not only maximizes sales potential but also contributes to building stronger customer relationships and loyalty.

Tailoring promotional offers and pricing strategies based on the insights from market basket analysis involves a strategic approach to maximizing sales and customer satisfaction.

Targeted Promotional Offers: Businesses can use the analysis of product relationships to design targeted promotional offers that resonate with customer purchasing behaviors. For example, if the analysis reveals that customers frequently purchase cereal and milk together, a targeted promotion could offer a discount on cereal with the purchase of milk, encouraging customers to buy both items together. This not only increases the likelihood of multiple item purchases but also enhances the perceived value for customers.

Dynamic Pricing Strategies: Understanding product relationships can also inform dynamic pricing strategies. Businesses can adjust pricing based on the associations between products. For instance, if certain items are often purchased together, a dynamic pricing strategy could be implemented to offer a slight discount when both items are included in the customer's cart. This approach can incentivize customers to complete the purchase of related items and increase the average order value.

Personalized Recommendations: By tailoring promotions and pricing based on product relationships, businesses can offer personalized recommendations to customers. This can be implemented through targeted marketing communications, personalized promotions, and recommendations based on customers' past purchase behaviors. Providing customers with offers that align with their preferences enhances the overall shopping experience and fosters a sense of personalized attention.

Data-Driven Decision-Making: Leveraging market basket analysis for promotions and pricing enables businesses to make data-driven decisions. By understanding the correlations between products, businesses can make informed choices about which items to include in promotions, how to structure pricing, and when to implement targeted offers. This data-driven approach reduces guesswork and enhances the effectiveness of marketing and sales strategies.

By incorporating the insights from market basket analysis into promotions and pricing strategies, businesses can optimize their marketing efforts, drive sales, and cultivate stronger customer relationships.

**Forecasting and Planning**:

Utilize market basket analysis to forecast demand for related products, aiding in better production planning and supply chain management. It involves leveraging market basket analysis to predict the demand for associated products. This analysis helps in understanding which products are frequently purchased together. By utilizing this insight, businesses can enhance their production planning and optimize their supply chain management by ensuring the availability of related items based on the forecasted demand. This approach ultimately contributes to more efficient inventory management and better customer satisfaction.

Market basket analysis is a data mining technique that identifies the products that are often purchased together by customers. This information can be used to forecast the demand for related products. By understanding the purchasing patterns of customers, businesses can anticipate which items are likely to be in demand together. This insight is valuable for production planning and supply chain management as it allows companies to adjust their inventory levels and production schedules to meet the expected demand. Ultimately, this approach can lead to reduced stockouts, minimized excess inventory, and improved overall operational efficiency.

By implementing these recommendations, businesses can harness the power of market basket analysis to drive sales, improve customer satisfaction, and optimize operational processes.

# CHAPTER 7: ANNEXTURE

Annexture is the supplementary document that provides detailed information about the project. This annexure might include technical details, data sources, methodologies, algorithms used, model performance metrics, and any other relevant information that supports or supplements the main project report. It serves as a reference for stakeholders who require in-depth insights into the project's implementation and outcomes. If you need guidance on creating or reviewing such an annexure

Support — The support showcases the probability in favor of the event under analysis.

Confidence — It expresses the operational efficiency of the rule. It calculated as the ratio of the probability of occurrence of the favorable event to the probability of the occurrence of the antecedent.

Lift Ratio — The lift ratio calculates the efficiency of the rule in finding consequences, compared to a random selection of transactions.

Applications of Market Basket Analysis

— Product Placements  
— Marketing Messages  
— Affinity Promotions  
— Inventory Management  
— Fraud Detections  
— Customer Behaviour  
— Content Placements

# CHAPTER 8: REFERENCES AND BIBLIOGRAPHY

Rick F. van der Lans, (2012) in Data Virtualization for Business Intelligence Systems, A volume in MKSeries on Business Intelligence Published by Elsevier, retrieved fromhttps://www.sciencedirect .com/book/978012394425data virtualization-for-business intelligencesystems.217-229, ISBN: 978-0-12-394425-2 DOI:https://doi.org/10. 1016/C2011-0-07129-6.

Gajbhiye, P. R., & Vaidya, S. G. (2016). Classification and Adaptive Novel Class Detection of Feature-Evolving Data Streams. International Journal of Engineering Research and General Science, 4(3), 616-622.

Masud, M. M., Chen, Q., Khan, L., Aggarwal, C. C., Gao, J., Han, J., Oza, N. C. (2013). Classification and Adaptive Novel Class Detection of Feature-Evolving Data Streams. IEEE Transactions on Knowledge and Data Engineering, 25(7), 1484-1497.

Ojugo, A.A, & Eboka, A.O (2018). Modeling the Computational Solution of Market Basket Associative Rule Mining Approaches using Deep Neural Network. Digital Technologies, 3(1), 1-8.

Dhanabhakyam, M., & Punithavalli, M. (2011). A Survey on Data Mining Algorithm for Market Basket Analysis. Global Journal of Computer Science and Technology, XI(11), 13-21. Retrieved Feburary 8, 2017.

Annie, L., & Kumar, A. (2012). Market Basket Analysis for a Supermaket Based on Frequent item Mining. International Journal of Computer Science, 6(5), 34-43. Retrieved January 10, 2017

Adewole, K., Akintola, A., Ajiboye, A., & Abdulsalam, S. (2014). Frequent pattern and association rule mining from inventory database using apriori algorithm. African Journal of Computing & ICT, VII(3), 35-42.

Deshmukh, H., Ankit, R., Lalit, V., Dhiraj, G., & Rahil, K. (2016). Android Application for Searching and Shopping Based on Cloud. International Journal Of Pure And Applied Research In Engineering And Technology, 4(9), 1368-1374.

Kurniawan, F.; Umayah, B.; Hammad, J.; Mardi, S.;Nugroho, S.; & Hariadi, M. (2018) Market Basket Analysis to Identify Customer Behaviors by Way of Transaction Data, Knowledge Engineering and Data Science (KEDS), Vol 1, No 1, pp. 20–25, ISSN 2597-4602.

Gill, N.S., (2017), Data Ingestion, Processing and Architecture layers for Big Data and IoT, retrieved from https://www.xenonstack .com/blog/big-dataengineering/ingestion processing-big-data-iot stream/, accessed 12/10/2018.

Phani, P., & Murlidher, M. (2013). A Study on Market Basket Analysis Using a Data Mining Algorithm. International Journal of Emerging

Technology and Advanced Engineering, III(6), 361-363. Retrieved from www.ijetea.com.

Woo, J. (2015). Market Basket Analysis using Spark. ARPN Journal of Science and Technology, V(4), 207 209.

[1] M. Y. Ardianto, S. Adinugroho, and I. Indriati, “Penentuan Tata Letak Produk menggunakan

Algoritma FP-Growth pada Toko ATK,” J. Pengemb. Teknol. Inf. Dan Ilmu Komput., vol. 5, no.

9, pp. 3826–3832, 2021, [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-

ptiik/article/view/9741

[2] W. P. Nurmayanti et al., “Market basket analysis with apriori algorithm and frequent pattern

growth (Fp-Growth) on outdoor product sales data,” Int. J. Educ. Res. Soc. Sci., vol. 2, no. 1, pp.

132–139, 2021, doi: https://doi.org/10.51601/ijersc.v2i1.45.

[3] W. W. Ariestya, W. Supriyatin, and I. Astuti, “Marketing strategy for the determination of staple

consumer products using FP-growth and apriori algorithm,” J. Ilm. Ekon. Bisnis, vol. 24, no. 3,

pp. 225–235, 2019, doi: http://dx.doi.org/10.35760/eb.2019.v24i3.2229.

[4] F. Z. Ghassani, A. Jamaludin, and A. S. Y. Irawan, “MARKET BASKET ANALYSIS USING

THE FP-GROWTH ALGORITHM TO DETERMINE CROSS-SELLING,” J. Inform. Polinema,

vol. 7, no. 4, pp. 49–54, 2021, doi: https://doi.org/10.33795/jip.v7i4.508.

[5] E. Hikmawati, N. U. Maulidevi, and K. Surendro, “Minimum threshold determination method

based on dataset characteristics in association rule mining,” J. Big Data, vol. 8, no. 1, pp. 1–17,

2021, doi: https://doi.org/10.1186/s40537-021-00538-3.

[6] L. Shi and Q. Zhu, “Association Rule Analysis of Influencing Factors of Literature Curriculum

Interest Based on Data Mining,” Comput. Intell. Neurosci., vol. 2022, pp. 1–8, 2022, doi:

https://doi.org/10.1155/2022/6866134.

[7] E. R. Kaburuan, Y. Sartika, and I. Agustina, “Sentiment Analysis on Product Reviews from

Shopee Marketplace using the Naïve Bayes Classifier,” LONTAR Komput., vol. 13, no. 3, pp.

150–159, 2022, doi: 10.24843/LKJITI.2022.v13.i03.p02.

[8] A. Sam, B. Denney, C. Haid, and R. Knight, “Simple tools for examining and cleaning dirty data.”

2021. [Online]. Available: https://github.com/sfirke/janitor

[9] J. Han, J. Pei, and Y. Yin, “Mining frequent patterns without candidate generation,” ACM sigmod

Rec., vol. 29, no. 2, pp. 1–12, 2000, doi: https://doi.org/10.1145/335191.335372.

[10] R. Agrawal and R. Srikant, “Fast algorithms for mining association rules,” in Proc. 20th int. conf.

very large data bases, VLDB, 1994, vol. 1215, pp. 487–499. [Online]. Available:

https://www.it.uu.se/edu/course/homepage/infoutv/ht08/vldb94\_rj.pdf

[11] N. K. D. Y. Utami, T. Trianasari, and P. I. Rahmawati, “THE EFFECT OF RELATIONSHIP

MARKETING AND CROSS SELLING ON THE MARKETING PERFORMANCE OF

INSURANCE PRODUCTS PT SUN LIFE INDONESIA SALES OFFICE SINGARAJA,” JMM

UNRAM-MASTER Manag. J., vol. 11, no. 3, pp. 179–191, 2022, doi: 10.29303/jmm.v11i3.722.

[12] F. Purwaningtias, “Strategi Up Selling Pada Website Penjualan,” Simetris J. Tek. Mesin, Elektro

https://doi.org/10.24176/simet.v9i1.1910.

[13] M. Kavitha and S. T. Selvi, “Comparative study on Apriori algorithm and Fp growth algorithm

with pros and cons,” Int. J. Comput. Sci. Trends Technol, vol. 4, no. 4, pp. 161–164, 2016,

[Online]. Available: https://www.ijcstjournal.org/volume-4/issue-4/IJCST-V4I4P28.pdf

[14] J. Heaton, “Comparing dataset characteristics that favor the Apriori, Eclat or FP-Growth frequent

itemset mining algorithms,” in SoutheastCon 2016, 2016, pp. 1–7. doi:

10.1109/SECON.2016.7506659.

[15] P. L. Ginting, N. Dengen, and M. Taruk, “Comparison of Priori and FP-Growth algorithms in

determining association rules,” in 2019 International Conference on Electrical, Electronics and

Information Engineering (ICEEIE), 2019, vol. 6, pp. 320–323. doi:

10.1109/ICEEIE47180.2019.8981438.

[16] R. Ranjan and A. Sharma, “Evaluation of frequent itemset mining platforms using apriori and fp-

growth algorithm,” arXiv Prepr. arXiv1902.10999, pp. 1–6, 2019, doi:

https://doi.org/10.48550/arXiv.1902.10999.

[17] A. Bala, M. Z. Shuaibu, Z. KaramiLawal, and R. I. Y. Zakari, “Performance analysis of apriori

and fp-growth algorithms (association rule mining),” Int. J. Comput. Technol. &Applications, vol.

7, no. 2, pp. 279–293, 2016.

**End of Project Report**