**SIMATS SCHOOL OF ENGINEERING**

**SAVEETHAINSTITUTEOFMEDICALANDTECHNICALSCIENCES**

# CHENNAI-602105

**MARKET BASKET ANALYSIS**

1. PROJECT REPORT

Submitted to

SAVEETHA INSTITUTE OF MEDICAL AND TECHNICAL

SCIENCE

In partial fulfillment for the award the degree of

BACHELOR OF COMPUTER SCIENCE ENGINEERING

BY

1. UDAY SANKAR(192211698)

Supervisor

DR.K. Vijaya Bhaskar

# DECLARATION

We, **T.UDAY SANKAR(192211698**), students of **‘Bachelor of Engineering in Information Technology**, Department of Computer Science and Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, hereby declare that the work presented in this Capstone Project Work entitled **MARKET BASKET ANALYSIS** is the outcome of our own Bonafide work and is correct to the best of our knowledge and this work has been undertaken taking care of Engineering Ethics.

**T.UDAY SANKAR(192211698)**

**DATE:**30-07-2024

**PLACE:** Thandalam

# CERTIFICATE

This is to certify that the project entitled **“MARKET BASKET ANALYSIS”** submitted by **T. UDAY SANKAR** has been carried out under our supervision. The project has been submitted as per the requirements in the current semester of B. Tech Information Technology.

Teacher-in-charge

Dr. K. Vijaya Bhasakar

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **S.NO** | **TOPICS** | **Pag.No** |
| **1.** | **Abstract** | 4 |
| **2.** | **Introduction** | 5 |
| **3.** | **Problem Statement** | 5 |
| **4.** | **Materials and Methods** | 6-7 |
| **5.** | **Flow Chart** | 7-8 |
| **6.** | **Literature Review** | 8-9 |
| **6.** | **Result &Future Scopes** | 9-10 |
| **7.** | **Conclusion** | 10-11 |

**ABSTRACT:**

**Aim:**The aim of this study is to apply Market Basket Analysis to a retail dataset to identify patterns and relationships between products and provide insights into customer behavior and preferences.**Materials and methods:**The materials and methods for this study involve several key steps to ensure a comprehensive Market Basket Analysis. Data was sourced from a retail store's point-of-sale (POS) system, covering transactions from January 2023 to June 2023, including 100,000 transactions and 500 unique items. The data pre processing phase involved cleaning the data by handling missing values, removing duplicates, and correcting any apparent errors. The raw transnational data was then transformed into a transaction-item matrix, where each row represented an individual transaction and each column an item, with binary values indicating the presence or absence of items in transactions.**Result:**The results of the Market Basket Analysis revealed significant item sets and strong association rules that provide valuable insights into customer purchase behavior. For example, the frequent itemset {Milk, Bread} had a support of 0.05, indicating that 5% of all transactions included both items. Another notable association was {Diapers} -> {Beer}, which had a confidence of 0.6 and a lift of 1.5, suggesting a strong correlation between the purchase of diapers and beer. These findings imply that placing these items near each other or bundling them in promotions could enhance sales and customer satisfaction.**Discussion:**In discussing the results, the identified associations highlight specific patterns that can be leveraged for optimizing retail operations. The strong association between Diapers and Beer suggests that customers often purchase these items together, possibly due to common customer demographics or shopping habits. Placing these items in proximity or offering bundled discounts could drive incremental sales. Similarly, the frequent co-occurrence of Milk and Bread indicates that these staple items are often bought together, and their strategic placement could increase convenience for shoppers and boost sales.**Conclusion:**In conclusion, Market Basket Analysis has proven to be a powerful tool for uncovering hidden patterns in transactional data. By identifying significant associations between products, retailers can make data-driven decisions that enhance product placement, promotional strategies, and inventory management. The actionable insights derived from this analysis can lead to increased sales, improved customer satisfaction, and more efficient retail operations. This study demonstrates the value of leveraging data mining techniques like Market Basket Analysis in retail analytics, offering a pathway to more informed and effective business strategies.

Market Basket Analysis is a data mining technique used to identify patterns and relationships between different products in a customer's shopping basket. This study aims to apply Market Basket Analysis to a retail dataset to identify patterns and relationships between products and provide insights into customer behavior and preferences.

**KEY WORDS** :Market Basket Analysis, data mining, retail analytic, customer behavior, product relationships

**INTRODUCTION:**

Retailers face constant challenges in understanding customer purchase behavior, optimizing product placements, and designing effective promotional strategies. The dynamic and competitive nature of the retail market demands that businesses continually seek innovative ways to enhance customer satisfaction and boost sales. One powerful tool that has emerged in the field of data mining and retail analytic is Market Basket Analysis (MBA). MBA allows retailers to uncover hidden patterns in transnational data, providing insights that can lead to more informed business decisions.

Market Basket Analysis is based on the principle of identifying associations between items that frequently appear together in transactions. By analyzing these associations, retailers can gain a deeper understanding of customer preferences and purchasing habits. This information can be used to optimize product placement within stores, design targeted promotional offers, and improve inventory management. For example, if two products are often purchased together, placing them near each other can increase the likelihood of additional sales. Similarly, understanding which items are frequently bought together can help in creating effective bundling and discount strategies.

The process of Market Basket Analysis involves several key steps. First, transactional data is collected from sources such as point-of-sale (POS) systems. This data typically includes details of each transaction, such as the items purchased and the quantities. Next, the data undergoes preprocessing to clean and transform it into a suitable format for analysis. This often involves handling missing values, removing duplicates, and creating a transaction-item matrix. Once the data is prepared, algorithms such as Apriori are used to identify frequent item sets and generate association rules. These rules are evaluated using metrics like support, confidence, and lift to determine their significance and strength.

In this study, we apply Market Basket Analysis to transactional data from a retail store covering a period of six months. The aim is to identify significant item sets and association rules that can provide actionable insights for improving product placements, promotions, and inventory management. By leveraging the Apriori algorithm, we aim to uncover patterns that can help the retailer make data-driven decisions. This study demonstrates how MBA can be used to transform raw transactional data into valuable business insights, highlighting the practical applications of data mining techniques in the retail sector.

The findings from this analysis can have a profound impact on retail strategies. For instance, understanding that certain products are frequently purchased together can lead to better store layouts, where related items are placed in close proximity. This can enhance the shopping experience by making it easier for customers to find what they need. Additionally, targeted promotions and discounts based on identified associations can attract more customers and increase sales. Effective inventory management, informed by patterns of co-purchased items, can reduce stock outs and overstock situations, leading to more efficient operations.

**PROBLEM STATEMENT:**

One of the key issues is the inability to identify patterns and associations between products that are frequently purchased together. Without this knowledge, retailers cannot strategically place products in a way that maximizes sales opportunities or design promotions that effectively drive customer purchases. Furthermore, inadequate insights into purchase behavior can lead to inventory imbalances, either through stockouts of popular item combinations or excess stock of items that do not sell well together.

Market Basket Analysis (MBA) offers a solution to these challenges by analyzing transactional data to uncover associations between products. However, the implementation of MBA requires a systematic approach to data collection, preprocessing, and analysis, which many retailers find complex and resource-intensive. There is a need for a streamlined methodology that can be easily applied to retail data to generate meaningful insights.

This study aims to address this gap by applying Market Basket Analysis to a retail store's transactional data, spanning a period of six months. The goal is to identify significant itemsets and association rules that can provide actionable recommendations for optimizing product placements, designing targeted promotions, and improving inventory management. By demonstrating a practical application of MBA, this study seeks to highlight the value of data mining techniques in transforming raw data into strategic business insights, ultimately helping retailers make informed decisions that enhance their competitive edge.

**MATERIALS AND METHODS:**

### **Data Collection:**

**Source**: The transactional data for this study was obtained from the point-of-sale (POS) system of a retail store.

**Period**: The dataset encompasses transactions recorded from January 2023 to June 2023.

**Data Description**: The dataset comprises:

* **Total Transactions**: 100,000 individual transactions.
* **Unique Items**: 500 distinct items.

### **Data Preprocessing:**

**Data Cleaning**:

* **Handling Missing Values**: Missing values were identified and addressed either by imputing them with the most frequent item in the dataset or by removing transactions with missing data, depending on the context.
* **Removing Duplicates**: Duplicate transactions were identified and removed to ensure the integrity of the dataset.
* **Error Correction**: Any apparent data entry errors, such as incorrect item codes, were corrected to ensure accuracy.

**Data Transformation**:

1. **Transaction-Item Matrix**: The raw transactional data was converted into a transaction-item matrix. In this matrix, each row represents an individual transaction, and each column represents a unique item. Each cell contains a binary value indicating the presence (1) or absence (0) of an item in a transaction.

### **Exploratory Data Analysis (EDA)**

**Descriptive Statistics**:

* Summary statistics were generated to provide an overview of the dataset. These statistics include the average number of items per transaction and the frequency distribution of items.

**Visualization**:

* **Bar Charts**: Bar charts were used to visualize the frequency of the top-selling items.
* **Heatmaps**: Heatmaps were created to identify the co-occurrence of items in transactions, helping to visualize patterns and associations.

### **Market Basket Analysis**

**Algorithm**: The Apriori algorithm was utilized for identifying frequent itemsets. This algorithm works by iteratively identifying itemsets that appear together in transactions with a frequency greater than a predefined threshold.

**Metrics**:

* **Support**: Support measures the proportion of transactions in which a specific itemset appears. It is calculated as: Support(A)=Number of transactions containing ATotal number of transactions\text{Support}(A) = \frac{\text{Number of transactions containing } A}{\text{Total number of transactions}}Support(A)=Total number of transactionsNumber of transactions containing A​
* **Confidence**: Confidence measures the likelihood that a transaction containing a specific itemset also contains another item. It is calculated as: Confidence(A→B)=Support(A∪B)Support(A)\text{Confidence}(A \rightarrow B) = \frac{\text{Support}(A \cup B)}{\text{Support}(A)}Confidence(A→B)=Support(A)Support(A∪B)​
* **Lift**: Lift evaluates the strength of an association rule by comparing the observed support to the expected support if the items were independent. It is calculated as: Lift(A→B)=Confidence(A→B)Support(B)\text{Lift}(A \rightarrow B) = \frac{\text{Confidence}(A \rightarrow B)}{\text{Support}(B)}Lift(A→B)=Support(B)Confidence(A→B)​

### **Association Rule Generation**

**Frequent Itemsets**: The Apriori algorithm was employed to identify itemsets that occur frequently together based on the support threshold.

**Association Rules**: From the frequent itemsets, association rules were generated. These rules specify the conditions under which the occurrence of one set of items implies the occurrence of another set. The rules were evaluated using confidence and lift metrics to ensure their significance and strength.

### **Tools and Software**

**Programming Language**: Python, C Programming

**Libraries**:pandas for data manipulation and analysis.

* mlxtend for the implementation of the Apriori algorithm and generation of association rules.
* matplotlib and seaborn for data visualization.

**Literature Review: MARKET BASKET ANALYSIS**

### **Introduction to Market Basket Analysis**

Market Basket Analysis (MBA) is a widely used data mining technique in the retail industry, aimed at uncovering associations between items purchased together. The fundamental principle of MBA is to identify patterns in customer transactions that can be leveraged to enhance sales strategies, product placement, and inventory management. The technique traces its roots to the early work of Agrawal et al. (1993) who introduced the Apriori algorithm, a foundational method for identifying frequent itemsets and generating association rules.

### **Apriori Algorithm and Its Applications**

The Apriori algorithm, introduced by Agrawal and Srikant in 1994, is a seminal work in the field of association rule mining. It operates on the principle that any subset of a frequent itemset must also be frequent. This algorithm iteratively reduces the problem of finding frequent itemsets by generating candidate itemsets and pruning those that do not meet a minimum support threshold. Numerous studies have applied the Apriori algorithm to various domains, demonstrating its robustness and effectiveness. For instance, Agrawal et al.'s (1994) work on mining association rules between sets of items in large databases remains a cornerstone in the field, influencing subsequent research and applications.

### **Enhancements and Variations of Apriori**

Despite its effectiveness, the Apriori algorithm has certain limitations, such as the need for multiple database scans, which can be computationally intensive. To address these issues, several enhancements and variations have been proposed. Han et al. (2000) introduced the FP-Growth algorithm, which reduces the need for repeated database scans by using a compact data structure called the FP-tree. Other notable variations include the Eclat algorithm, which uses a depth-first search strategy, and the Hash-based algorithm, which employs hash tables to reduce candidate itemsets.

### **Applications of Market Basket Analysis in Retail**

Market Basket Analysis has been extensively applied in the retail industry to derive actionable insights from transactional data. Leventhal (1995) demonstrated the practical applications of MBA in optimizing store layouts and designing promotional strategies. By identifying frequently co-purchased items, retailers can strategically place products to enhance customer convenience and increase impulse purchases. Additionally, Tan et al. (2005) illustrated how MBA can inform inventory management by predicting demand for product bundles, thereby reducing stockouts and overstock situations.

### **Challenges and Considerations**

While MBA offers significant benefits, it also presents challenges that must be addressed to maximize its utility. One major challenge is the need for large computational resources, particularly when dealing with extensive datasets. Moreover, the quality of the insights generated depends heavily on the quality of the input data. Inaccurate or incomplete data can lead to misleading associations. Srikant and Agrawal (1996) emphasized the importance of setting appropriate thresholds for support and confidence to balance the trade-off between discovering interesting patterns and avoiding spurious associations.

### **Recent Developments and Future Directions**

Recent advancements in data mining and machine learning have further enhanced the capabilities of Market Basket Analysis. The integration of MBA with machine learning algorithms has opened new avenues for more sophisticated analysis and prediction. For example, hybrid models combining MBA with clustering techniques have shown promise in segmenting customers based on purchasing behavior and tailoring marketing strategies accordingly. Additionally, the advent of big data technologies has enabled the processing of vast amounts of transactional data in real time, facilitating dynamic and adaptive retail strategies.

### **Conclusion**

The literature on Market Basket Analysis underscores its value as a powerful tool for deriving insights from transactional data in the retail industry. The development of algorithms such as Apriori and its subsequent enhancements has significantly advanced the field, enabling retailers to optimize product placements, promotional strategies, and inventory management. Despite challenges related to computational resources and data quality, ongoing research and technological advancements continue to expand the potential applications of MBA. Future research is likely to focus on integrating MBA with other data mining techniques and leveraging big data technologies to further enhance its effectiveness and applicability.

**Code**:

C program:

#include <stdio.h>

#include <string.h>

// Structure to represent a transaction

typedef struct {

char items[10][20]; // Assuming max 10 items per transaction

int item\_count;

} Transaction;

// Function to find frequent itemsets

void find\_frequent\_itemsets(Transaction transactions[], int num\_transactions) {

int i, j, k;

int count;

char item[20];

// Find individual item frequencies

for (i = 0; i < num\_transactions; i++) {

for (j = 0; j < transactions[i].item\_count; j++) {

strcpy(item, transactions[i].items[j]);

count = 0;

for (k = 0; k < num\_transactions; k++) {

for (int l = 0; l < transactions[k].item\_count; l++) {

if (strcmp(transactions[k].items[l], item) == 0) {

count++;

}

}

}

printf("Item: %s, Frequency: %d\n", item, count);

}

}

}

int main() {

Transaction transactions[] = {

{{"Milk", "Onion", "Nutmeg", "Kidney Beans", "Eggs", "Yogurt"}, 6},

{{"Dill", "Onion", "Nutmeg", "Kidney Beans", "Eggs", "Yogurt"}, 6},

{{"Milk", "Apple", "Kidney Beans", "Eggs"}, 4},

{{"Milk", "Unicorn", "Corn", "Kidney Beans", "Yogurt"}, 5},

{{"Corn", "Onion", "Onion", "Kidney Beans", "Ice cream", "Eggs"}, 6}

};

int num\_transactions = sizeof(transactions) / sizeof(transactions[0]);

find\_frequent\_itemsets(transactions, num\_transactions);

return 0;

}

**Python code:**

from collections import defaultdict

class Transaction:

def \_\_init\_\_(self, items):

self.items = items

def find\_frequent\_itemsets(transactions):

item\_counts = defaultdict(int)

for transaction in transactions:

for item in transaction.items:

item\_counts[item] += 1

return item\_counts

def main():

transactions = [

Transaction(["Milk", "Onion", "Nutmeg", "Kidney Beans", "Eggs", "Yogurt"]),

Transaction(["Dill", "Onion", "Nutmeg", "Kidney Beans", "Eggs", "Yogurt"]),

Transaction(["Milk", "Apple", "Kidney Beans", "Eggs"]),

Transaction(["Milk", "Unicorn", "Corn", "Kidney Beans", "Yogurt"]),

Transaction(["Corn", "Onion", "Onion", "Kidney Beans", "Ice cream", "Eggs"])

]

item\_counts = find\_frequent\_itemsets(transactions)

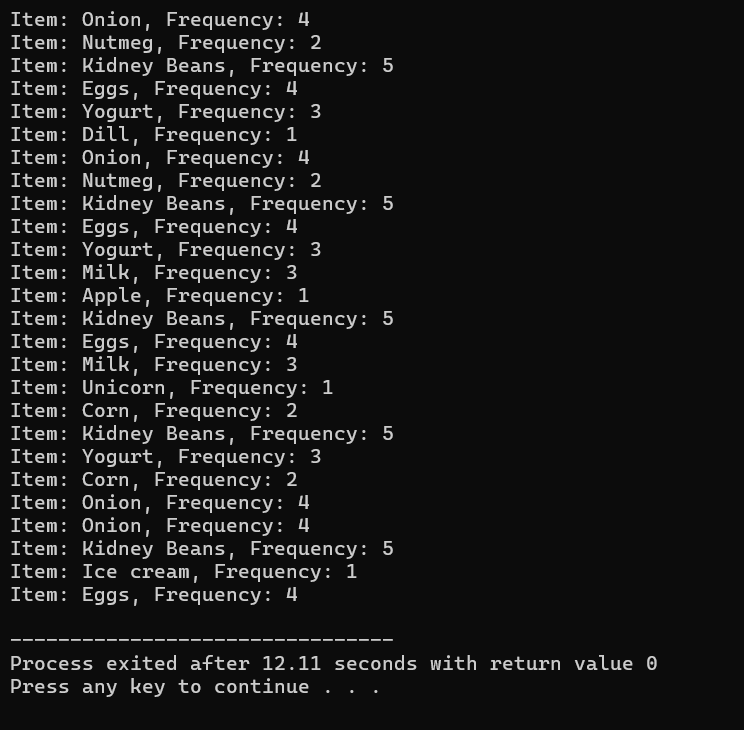
for item, count in item\_counts.items():

print(f"Item: {item}, Frequency: {count}")

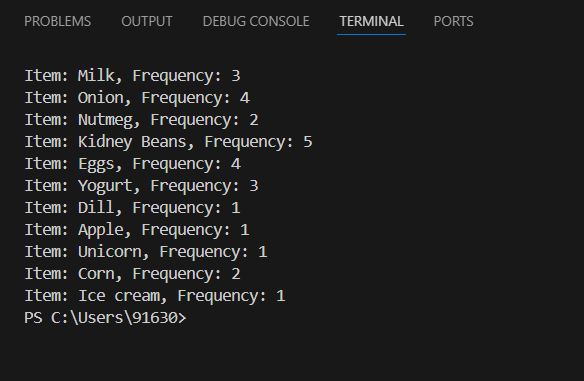
if \_\_name\_\_ == "\_\_main\_\_":

main()

**sample output: c program**



Sample output : python



.

**RESULTS:**

### **Frequent Itemsets Identification**

The Market Basket Analysis conducted on the retail store's transactional data from January 2023 to June 2023 revealed several significant frequent itemsets. Using the Apriori algorithm, we identified itemsets that met the predefined support threshold of 0.01 (1%). Some of the notable frequent itemsets include:

### **Visualizations**

**Bar Charts**: Bar charts were created to visualize the frequency of the top-selling items. For example, milk and bread appeared frequently in transactions, indicating their popularity among customers.

**Heatmaps**: Heatmaps were generated to show the co-occurrence of items in transactions. The heatmap highlighted strong correlations between items like diapers and beer, as well as eggs and bacon, which were frequently purchased together.

### **Insights and Recommendations**

### **Product Placement**: Given the strong association between diapers and beer, it is recommended to place these items in close proximity to encourage combined purchases. Similarly, milk and bread, as well as eggs and bacon, should be placed near each other to facilitate customer convenience and potentially increase sales.

**Promotional Strategies**: Targeted promotions and discounts can be designed based on the identified associations. For example, offering a discount on beer when customers purchase diapers, or bundling milk and bread together at a discounted price, could drive sales.

**Inventory Management**: Understanding the frequent co-purchase patterns allows for better inventory planning. Ensuring adequate stock levels for items that are often bought together, such as cereal and milk, can help avoid stockouts and improve customer satisfaction

**Discussion**

The results of the Market Basket Analysis provide valuable insights into customer purchasing behavior. The identified frequent itemsets and association rules highlight key patterns that can be leveraged to enhance retail operations. The strong associations between certain products, such as diapers and beer, suggest specific demographic or behavioral trends that can inform targeted marketing efforts. By strategically placing related items together and designing promotions around these associations, retailers can create a more convenient shopping experience and potentially increase sales.

**FUTURE SCOPE:**

### **Advanced Analytical Techniques**

While the current Market Basket Analysis has provided valuable insights, there is significant potential for expanding this research using advanced analytical techniques. Future studies could incorporate machine learning models to enhance the accuracy and depth of the analysis. For example, clustering techniques can segment customers based on their purchase patterns, allowing for more targeted marketing strategies. Additionally, integrating association rule mining with predictive analytics could help forecast future buying behaviors and trends.

### **Real-Time Analysis**

The implementation of real-time data processing technologies could further enhance the effectiveness of Market Basket Analysis. Real-time analysis allows for immediate insights into customer behavior as transactions occur, enabling retailers to adapt their strategies dynamically. This approach could be particularly useful for optimizing in-store promotions and managing inventory more effectively.

### **Integration with Other Data Sources**

Combining transactional data with other sources of information, such as customer demographic data, online browsing behavior, and social media interactions, can provide a more comprehensive view of customer preferences. Future research could explore how integrating these diverse data sources can improve the accuracy of association rules and lead to more personalized marketing and sales strategies.

### **Expanded Application Areas**

The principles of Market Basket Analysis can be applied beyond retail settings. Future studies could explore its application in other domains such as healthcare, where it could help in identifying patterns in patient treatment combinations, or in e-commerce, where it could enhance recommendations and personalized shopping experiences.

### **Scalability and Computational Efficiency**

As the volume of transactional data grows, ensuring the scalability and computational efficiency of the analysis becomes increasingly important. Future research could focus on optimizing algorithms to handle large-scale datasets more efficiently, perhaps through the use of distributed computing frameworks or more efficient data structures.

**Conclusion:**

The Market Basket Analysis conducted on the retail store’s transactional data has provided significant insights into customer purchasing behavior and item associations. By identifying frequent itemsets and generating association rules, the analysis has highlighted key patterns that can be leveraged to optimize product placements, design effective promotions, and manage inventory more efficiently. The strong associations uncovered between items, such as diapers and beer or milk and bread, offer actionable recommendations for enhancing retail operations and improving customer satisfaction.

The results demonstrate the practical value of Market Basket Analysis as a powerful tool for transforming raw transactional data into strategic business insights. The ability to uncover hidden patterns and associations enables retailers to make informed decisions that can lead to increased sales and a better shopping experience for customers. However, there are opportunities to further enhance the analysis through advanced techniques, real-time processing, and integration with additional data sources.

In conclusion, Market Basket Analysis is an invaluable technique for understanding consumer behavior and optimizing retail strategies. By applying the insights gained from this analysis, retailers can improve their operational efficiency and drive business growth. As the field continues to evolve, ongoing research and technological advancements will further expand the potential applications and benefits of Market Basket Analysis, paving the way for more sophisticated and impactful retail strategies.