**Myanmar Institute of Information Technology**

****

**II semester 2023-2024**

**CSE 5005 Data Mining**

**Basket Analysis**

**Project Report**

*Submitted By*

**2017-MIIT-ECE-035 Moe Myint Naing**

**2017-MIIT-ECE-036 Naing Khant Win**

**2017-MIIT-ECE-043 Pyae Phyo Tun**

**2017-MIIT-ECE-044 Pyae Sone Khant Aung**

**2017-MIIT-ECE-045 Sai Aung Soe Hein**

**2018-MIIT-CSE-035 Oak Hla Gyi**

**2018-MIIT-ECE-038 Myo Thinzar Kyaw**

# **Abstract**

Basket analysis is a data mining technique used to uncover relationships between items frequently purchased together. In this project, we applied the Apriori algorithm to perform basket analysis on a bakery sales dataset, which includes transactional details such as items purchased, time of purchase, and whether the transaction occurred on a weekday or weekend. The aim was to discover frequent item sets and generate association rules, allowing us to identify purchasing patterns and provide insights for improving business operations, such as product bundling, inventory optimization, and targeted promotions.

The analysis revealed that certain items, such as coffee and pastries, are frequently bought together, particularly during morning hours. This suggests opportunities for creating combo deals or cross-selling strategies. Additionally, differences in purchasing behavior between weekdays and weekends suggest that the bakery could tailor its marketing strategies based on the time of day and day of the week. The recommendations derived from this analysis can help the bakery increase sales and improve customer satisfaction through data-driven decision-making.

**Contents**

[**Abstract** 2](#_Toc178196698)

[**Chapter (1)** 6](#_Toc178196699)

[**Introduction** 6](#_Toc178196700)

[**1.1. Introduction** 6](#_Toc178196701)

[**1.2. Aims and Objectives** 6](#_Toc178196702)

[**1.3. Requirements** 7](#_Toc178196703)

[1.3.1. Dataset 7](#_Toc178196704)

[1.3.2. Programming Language/Tools 8](#_Toc178196705)

[1.3.3. Libraries 8](#_Toc178196706)

[**1.4. System Flow** 8](#_Toc178196707)

[1.4.1. Data Import and Preprocessing 8](#_Toc178196708)

[1.4.2. Transaction Transformation 9](#_Toc178196709)

[1.4.3. Application of Apriori Algorithm 9](#_Toc178196710)

[1.4.4. Analysis of Association Rules 9](#_Toc178196711)

[1.4.5. Recommendation Generation 9](#_Toc178196712)

[**1.5. Group Members** 10](#_Toc178196713)

[**Chapter (2)** 11](#_Toc178196714)

[**Theory Background** 11](#_Toc178196715)

[**2.1 Frequent Item Sets** 11](#_Toc178196716)

[**2.2 Association Rules (Apriori Algorithm)** 11](#_Toc178196717)

[2.2.1. Aprori Algorithm Process 12](#_Toc178196718)

[2.2.2. Importance of Apriori Algorithm 13](#_Toc178196719)

[**2.3. Confidence** 13](#_Toc178196720)

[**2.4. Lift** 13](#_Toc178196721)

[**2.5. Support** 13](#_Toc178196722)

[**2.6. Conviction** 14](#_Toc178196723)

[**2.7. Recommendation** 14](#_Toc178196724)

[**Chapter (3)** 15](#_Toc178196725)

[**Implementation** 15](#_Toc178196726)

[**3.1 Dataset Selection - Bakery Sales Dataset** 15](#_Toc178196727)

[**3.2. Preprocessing** 15](#_Toc178196728)

[**3.3. Exploratory Data Analysis (EDA)** 16](#_Toc178196729)

[3.3.1. Transaction Frequency Analysis 16](#_Toc178196730)

[3.3.2. Item Sales Analysis 16](#_Toc178196731)

[3.3.3. Visualization 16](#_Toc178196732)

[**3.4. Frequent Item Sets- Apriori Algorithm** 17](#_Toc178196733)

[3.4.1. Parameters 17](#_Toc178196734)

[**3.5. Association Rules** 17](#_Toc178196735)

[3.5.1. Parameters 17](#_Toc178196736)

[3.5.2. Rule Interpretation 17](#_Toc178196737)

[**3.6. Recommendation** 18](#_Toc178196738)

[**Chapter (4)** 19](#_Toc178196739)

[**Conclusion and Future Extensions** 19](#_Toc178196740)

[**4.1. Conclusion** 19](#_Toc178196741)

[**4.2. Future Extensions** 20](#_Toc178196742)

[**4.3. Use Cases of Basket Analysis** 20](#_Toc178196743)

[**References** 22](#_Toc178196744)

**Tables**

[Table 1.1 Members Table 9](#_Toc178194649)

**Figures**

[Figure 1.1 Basket Analysis 6](#_Toc178194550)

[Figure 2.1 Aprori Algorithm 11](#_Toc178194551)

# **Chapter (1)**

# **Introduction**

## **1.1. Introduction**

Basket analysis is a widely used data mining technique aimed at identifying associations between items in a dataset, providing valuable insights into customer purchasing patterns. In this project, we focus on applying the **Apriori Algorithm** to the **"Bakery Sales Dataset"**, which contains information on various transactions, including items purchased, the time of day, and whether the transaction occurred on a weekday or weekend. The goal is to uncover **frequent item sets**—combinations of products often bought together—and generate association rules that highlight patterns in customer behavior. This method is particularly useful for businesses, as it can inform strategic decisions in areas such as product bundling, inventory management, and targeted promotions, ultimately helping to enhance customer satisfaction and increase sales. By exploring these associations, we aim to offer actionable insights that can drive operational improvements and marketing effectiveness for the bakery.

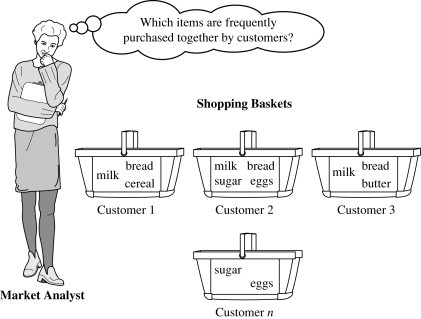


Figure 1.1 Basket Analysis

## **1.2. Aims and Objectives**

The primary aim of this project is to apply advanced data mining techniques, particularly the Apriori Algorithm, to perform a detailed basket analysis on the bakery sales dataset. By leveraging these techniques, the project seeks to identify patterns in customer purchasing behavior and uncover associations between frequently bought items, helping the bakery to optimize its product offerings and marketing strategies.

The Objectives of our basket analysis are -

* **Identify Frequent Item Sets -** One of the key objectives is to discover **frequent item sets**, which are combinations of items that are commonly purchased together in transactions. This will help in understanding the relationships between products and provide insights into which items customers tend to buy together, either during different times of the day or specific days of the week.
* **Derive Association Rules to Uncover Relationships** - Another critical objective is to generate **association rules** based on the frequent item sets. These rules will highlight the conditional dependencies between products—such as "if a customer buys coffee, they are likely to also buy a pastry." By understanding these relationships, the bakery can better anticipate customer needs and preferences.
* **Provide Business Recommendations** - Based on the analysis of frequent item sets and association rules, the final objective is to offer **actionable business recommendations**. These could include suggestions for product bundling, inventory management, and targeted promotions. For example, the bakery could offer combo deals on frequently co-purchased items or adjust stock levels to ensure popular products are always available, especially during peak times like mornings and weekends.

## **1.3. Requirements**

The tools required to build this project are –

* Dataset
* Programming Language/Tools
* Libraries

1.3.1. Dataset

The dataset used for this project is the Bakery Sales Dataset, sourced from Kaggle. It contains transactional data, including variables such as transaction numbers, items purchased, time and date stamps, daypart classifications (morning, afternoon, evening, night), and whether the transaction occurred on a weekend or weekday. This dataset provides a rich source of information for performing basket analysis, allowing us to explore customer purchasing patterns in detail.

1.3.2. Programming Language/Tools

The analysis is conducted using Python, a powerful and flexible programming language widely used for data analysis and machine learning tasks. The project is implemented in a Jupyter Notebook environment, which allows for easy code writing, execution, and visualization in an interactive manner.

### 1.3.3. Libraries

* **Mlxtend -** The core library used for implementing the **Apriori algorithm** and generating association rules. This library provides efficient methods for mining frequent item sets and extracting rules that help uncover relationships between products.
* **Pandas** - A fundamental library for data manipulation and preprocessing, **pandas** is used to clean, transform, and organize the transactional data into a format suitable for analysis. It helps in managing the dataset, handling missing values, and converting raw data into the appropriate structure.
* **matplotlib and seaborn** - For data visualization, **matplotlib** and **seaborn** are used to create informative charts and plots. These libraries help in visually representing the results of the analysis, such as displaying item frequencies, transaction distributions, and insights derived from the frequent item sets and association rules.
* **scikit-learn** - This library is utilized for **additional preprocessing tasks**, such as encoding categorical data or standardizing variables when needed. It helps ensure the dataset is well-prepared for efficient and accurate analysis.

## **1.4. System Flow**

1.4.1. Data Import and Preprocessing

The first step in the system is importing the Bakery Sales Dataset using Python libraries like pandas. During preprocessing, the data is cleaned and prepared for analysis. This includes handling missing values, filtering irrelevant data, and transforming the dataset to a more suitable format. Additionally, columns such as transaction time and daypart are carefully examined to extract meaningful insights and ensure the dataset is ready for the next steps.

1.4.2. Transaction Transformation

To apply the Apriori algorithm effectively, the raw transactional data needs to be transformed into a suitable format. Each transaction is converted into a list of items purchased, forming a structured dataset where every row represents a transaction and the corresponding items bought in that transaction. This transformation is essential for identifying frequent item sets and understanding the relationships between different products.

1.4.3. Application of Apriori Algorithm

The Apriori algorithm is then applied to the transformed dataset using the mlxtend library. This algorithm helps in discovering frequent item sets by iterating through the dataset and identifying combinations of items that meet a minimum support threshold. It efficiently narrows down the search space, focusing on high-frequency item combinations, which are the most important for further analysis.

1.4.4. Analysis of Association Rules

Once frequent item sets are identified, the next step is to derive association rules. These rules highlight relationships between different items, expressed in terms of metrics such as support, confidence, and lift. For example, if customers frequently buy coffee with pastries, the rule may indicate that buying coffee increases the likelihood of purchasing a pastry. The analysis of these rules provides valuable insights into customer buying behavior.

1.4.5. Recommendation Generation

Based on the frequent item sets and association rules, the system generates recommendations for business strategies. These include suggestions for product bundling, targeted promotions, and inventory management. For example, if certain items are frequently bought together in the morning, the bakery could create combo offers during that time. Similarly, inventory levels can be adjusted to ensure that high-demand products are always available during peak times.

## **1.5. Group Members**

|  |  |
| --- | --- |
| **Name** | **Roll Number** |
| 2017-miit-ece-035 | Moe Myint Naing |
| 2017-miit-ece-036 | Naing Khant Win |
| 2017-miit-ece-043 | Pyae Phyo Tun |
| 2017-miit-ece-44 | Pyae Sone Khant Aung |
| 2017-miit-ece-045 | Sai Aung Soe Hein |
| 2018-miit-cse-035 | Oak Hla Gyi |
| 2018-miit-ece-038 | Myo Thinzar Kyaw |

Table 1.1 Members Table

# **Chapter (2)**

# **Theory Background**

## **2.1 Frequent Item Sets**

**Frequent Item Sets** in data mining refer to groups of items that often appear together in transactions. For instance, in a grocery store, frequent item sets might include "milk, bread, and eggs" or "barbecue sauce, hot dogs, and buns."

**Why are frequent item sets valuable for businesses?**

* **Product Bundling -** Identifying frequent item sets can help businesses create effective product bundles to increase sales. For example, if customers often purchase both coffee and sugar, bundling them together could encourage more sales.
* **Cross-Selling -** Businesses can use frequent item sets to recommend complementary products to customers. If a customer purchases a new smartphone, suggesting a protective case or screen protector could increase sales.
* **Inventory Management -** Understanding frequent item sets can help optimize inventory levels by predicting which products are likely to be purchased together. This can prevent stockouts or overstocking.
* **Market Basket Analysis -** Frequent item sets are a key component of market basket analysis, which helps businesses understand customer buying behavior and identify hidden relationships between products.

## **2.2 Association Rules (Apriori Algorithm)**

**The Apriori Algorithm** is a popular algorithm used to mine association rules from large datasets. It works by iteratively identifying frequent item sets and using them to generate association rules.

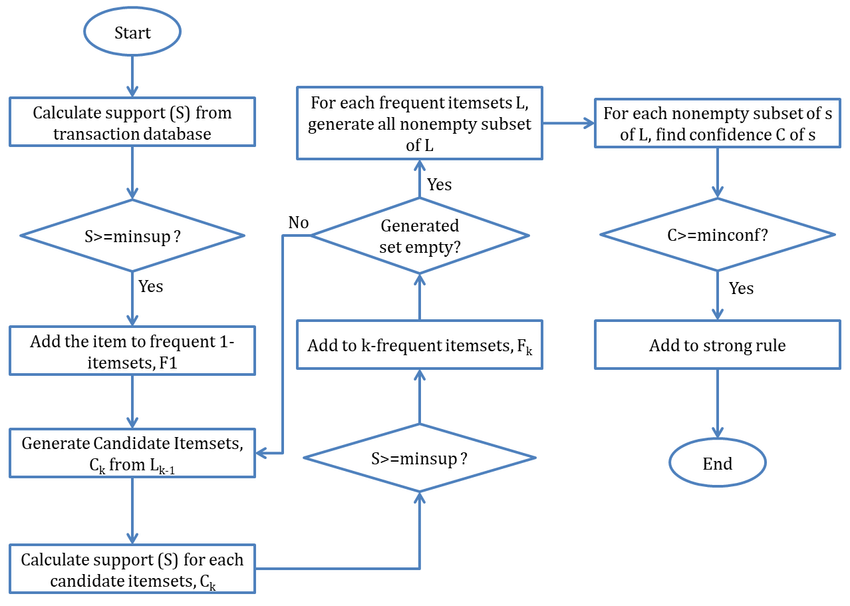


Figure 2.1 Aprori Algorithm

### ****2.2.1. Aprori Algorithm Process****

1. **Generate frequent item sets -** The algorithm starts by finding all frequent item sets of length 1 (single items).
2. **Generate candidate item sets -** Using the frequent item sets from the previous step, the algorithm generates candidate item sets of length 2.
3. **Count candidate item sets -** The algorithm counts the occurrences of each candidate item set in the dataset.
4. **Prune infrequent item sets -** Any candidate item set that does not meet the minimum support threshold is removed.
5. **Repeat steps 2-4 -** The process is repeated until no new frequent item sets can be found.

### ****2.2.2. Importance of Apriori Algorithm****

* **Efficiency -** Apriori is an efficient algorithm for mining association rules, especially for large datasets.
* **Accuracy -** It provides accurate results by ensuring that only frequent item sets are considered.
* **Flexibility -** Apriori can be adapted to various data mining tasks, such as market basket analysis and recommendation systems.

## **2.3. Confidence**

**Confidence** measures the likelihood that item B will appear in a transaction given that item A is already present.

**Note: if 80% of transactions that contain bread also contain milk, then the confidence of the rule "bread -> milk" is 80%.**

## **2.4. Lift**

**Lift** measures the strength of an association rule, considering the independence of the items involved.

**Note: A lift value of 1 indicates that the items are independent. A lift value greater than 1 suggests a positive association between the items, while a lift value less than 1 suggests a negative association.**

## **2.5. Support**

**Support** indicates the frequency of an item set in the dataset. It is calculated as the number of transactions containing the item set divided by the total number of transactions.

**Note: A high support value indicates that an item set is frequently purchased.**

## **2.6. Conviction**

**Conviction** measures the strength of an implication, considering the possibility that the consequent might occur without the antecedent.

**Note: A higher conviction value indicates a stronger implication.**

## **2.7. Recommendation**

The findings from association rules can be used to make various business recommendations, including

* **Product bundling -** Create bundles of frequently purchased items to increase sales.
* **Cross-selling -** Suggest complementary products to customers based on their purchase history.
* **Inventory management -** Optimize inventory levels by predicting which products are likely to be purchased together.
* **Personalized marketing -** Target marketing campaigns to specific customer segments based on their preferences and buying behavior.

By understanding frequent item sets and association rules, businesses can gain valuable insights into customer behavior and make data-driven decisions to improve their operations and increase sales.

# **Chapter (3)**

# **Implementation**

## **3.1 Dataset Selection - Bakery Sales Dataset**

` The **Bakery Sales Dataset** from Kaggle is a valuable resource for analyzing customer purchasing behavior in a bakery setting. This dataset contains detailed information about individual transactions, including:

* **TransactionNo -** A unique identifier assigned to each transaction, allowing for easy tracking and analysis.
* **Items -** A list of all items purchased in a given transaction, providing insights into the specific products customers are interested in.
* **DateTime -** A precise timestamp indicating the date and time of the transaction, enabling the examination of sales patterns throughout the day and week.
* **Daypart -** A categorization of the transaction into one of four time periods: morning, afternoon, evening, or night. This variable allows for the identification of potential variations in sales activity during different parts of the day.
* **DayType -** A classification of the transaction as either a weekday or weekend, facilitating the study of differences in customer behavior on different days of the week.

This dataset offers a rich source of data for exploring various aspects of bakery sales, such as popular product combinations, peak sales times, and the impact of daypart and day type on customer purchasing habits. By analyzing this dataset, businesses can gain valuable insights into their customers and make data-driven decisions to improve their operations and increase sales.

## **3.2. Preprocessing**

* **Data Cleaning** - Handle missing values and duplicate records.
* **Transformation into Transactions** - Convert the dataset into a format suitable for the Apriori algorithm (e.g., each transaction as a list of items).
* **Encoding** - Convert items into a one-hot encoding if needed, where each row represents a transaction, and each column represents an item.

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

To gain a foundational understanding of the bakery sales data, an exploratory data analysis (EDA) was conducted. This analysis focused on examining transaction frequency across different time periods and analyzing the sales performance of individual items.

### ****3.3.1. Transaction Frequency Analysis****

* **Daily Transactions -** The dataset was analyzed to determine the number of transactions occurring on each day of the week. This analysis revealed potential patterns in customer behavior, such as higher sales volumes on weekends or specific weekdays.
* **Hourly Transactions -** The dataset was further explored to identify the busiest times of day for the bakery. By analyzing the number of transactions occurring during each hour, it was possible to identify peak sales periods and adjust staffing or promotional activities accordingly.

### ****3.3.2. Item Sales Analysis****

* **Product Popularity-** The sales performance of individual items was examined to identify the most popular products in the bakery. This analysis provided insights into customer preferences and helped guide inventory management and product placement decisions.
* **Seasonal Trends-** The dataset was analyzed to detect any seasonal patterns in product sales. This information could be used to anticipate demand fluctuations and adjust pricing or promotional strategies accordingly.

### ****3.3.3. Visualization****

To effectively communicate the findings of the EDA, bar charts and histograms were used to visualize the data. Bar charts were employed to represent categorical data, such as the number of transactions by day of the week or time of day. Histograms were used to visualize numerical data, such as the distribution of item sales. These visualizations provided a clear and concise representation of the data, making it easier to identify trends and patterns.

## **3.4. Frequent Item Sets- Apriori Algorithm**

The Apriori algorithm, implemented using the mlxtend library, was employed to discover frequent item sets within the bakery sales data. These item sets represent groups of products that are commonly purchased together.

### ****3.4.1. Parameters****

* **min\_support -** This parameter specifies the minimum support threshold for an item set to be considered frequent. A higher min\_support value results in fewer frequent item sets being identified.
* **Threshold Selection -** The choice of the min\_support threshold depends on the desired level of granularity and the specific goals of the analysis. A lower threshold may reveal more frequent item sets, but it can also increase the computational cost and generate a larger number of rules.

## **3.5. Association Rules**

Once the frequent item sets were identified, association rules were generated. These rules describe relationships between items, such as "If a customer purchases item A, they are likely to also purchase item B."

### ****3.5.1. Parameters****

* **min\_confidence -** This parameter specifies the minimum confidence threshold for an association rule to be considered significant. A higher min\_confidence value ensures that the rules have a strong likelihood of holding true.
* **min\_lift -** This parameter measures the strength of an association rule relative to the independence of the items involved. A higher min\_lift value indicates a stronger association between the items.

### ****3.5.2. Rule Interpretation****

By examining the generated association rules, key patterns in customer purchasing behavior were identified. For example, a strong association between "coffee" and "pastry" might suggest that customers often enjoy these items together. This information can be leveraged to implement targeted promotions or create product bundles.

## **3.6. Recommendation**

Based on the insights gained from the analysis, the bakery can implement various business strategies to enhance sales and customer satisfaction.

* **Product Bundling -** By identifying frequently purchased item sets, the bakery can create attractive product bundles to encourage customers to buy more. For example, offering a discount on a coffee and pastry bundle could increase sales of both items.
* **Targeted Promotions -** The bakery can tailor its promotional activities to specific customer segments based on their purchasing behavior. For instance, customers who frequently purchase healthy items might be offered discounts on salads or fruit.
* **Inventory Management -** Understanding customer purchasing patterns can help the bakery optimize its inventory levels, ensuring that popular products are always in stock while minimizing the risk of stockouts or overstocking.

By effectively utilizing the insights derived from the data analysis, the bakery can make data-driven decisions to improve its operations, enhance customer satisfaction, and drive sales growth

# **Chapter (4)**

# **Conclusion and Future Extensions**

## **4.1. Conclusion**

The basket analysis conducted on the Bakery Sales Dataset provided meaningful insights into customer purchasing patterns and identified key product associations that can influence business strategies. Using the **Apriori algorithm**, we were able to uncover frequent item sets that reveal combinations of products often purchased together. These insights have the potential to significantly impact how the bakery approaches product bundling, promotions, and overall operational efficiency.

Some of the key takeaways from this analysis include-

* **Popular Product Combinations** - The identification of frequent item sets is invaluable for creating **product bundles** that align with customer preferences. By offering popular combinations, such as coffee and pastries, as combo deals, the bakery can encourage customers to buy more items, boosting sales.
* **Peak Sales Times** - Analysis of the transactional data also highlighted peak sales times, particularly in the morning and during weekends. Understanding these patterns allows the bakery to optimize staffing levels and ensure that sufficient resources are available during busy periods. It also helps in adjusting promotional strategies to target these high-traffic times more effectively.
* **Seasonal Trends** - The dataset also provided insights into potential seasonal fluctuations in purchasing behavior. By identifying **seasonal trends**, the bakery can plan for changes in customer demand, adjust inventory, and implement season-specific pricing or promotional campaigns to better cater to customer needs.

The **Apriori algorithm** proved to be highly effective in identifying these frequent item sets. By leveraging these insights, the bakery can make informed, data-driven decisions that enhance customer satisfaction, improve operational efficiency, and ultimately, drive higher sales and profitability. The analysis demonstrates the value of using basket analysis to optimize product offerings and promotions in a retail environment.

## **4.2. Future Extensions**

While the current analysis provides a comprehensive understanding of customer purchasing behavior, there are several avenues for future exploration and enhancement. These extensions can add depth to the insights gained and help the bakery further refine its business strategies.

* **Advanced Algorithms** - The Apriori algorithm is effective for small to medium-sized datasets, but for larger datasets, more advanced algorithms like FP-Growth could be used to improve efficiency and scalability. FP-Growth can handle larger datasets more effectively by compressing the data and allowing for faster identification of frequent item sets without the need to generate candidate sets like Apriori does. Applying this algorithm could enable the bakery to analyze even more data over time, providing a richer set of insights.
* **Temporal Patterns** - Incorporating an analysis of temporal patterns could provide additional insights into how customer purchasing behavior evolves over time. For instance, identifying shifts in product associations during different parts of the day, across weekdays versus weekends, or during seasonal peaks can allow the bakery to create dynamic promotions and adjust inventory accordingly. Understanding these temporal changes can be crucial for real-time marketing and inventory management.
* **Real-time Data Integration** - Another potential extension would involve integrating real-time data into the analysis. By applying association rule mining to real-time transactions, the bakery could make timely decisions and offer personalized recommendations to customers as they shop. Real-time analytics could also be used to respond quickly to sudden changes in customer behavior, such as during special events or holidays, enabling the bakery to be more adaptive and customer-focused.

## **4.3. Use Cases of Basket Analysis**

The application of basket analysis extends far beyond the bakery industry. Its versatility and ability to uncover meaningful relationships between products or services make it valuable across numerous sectors. Below are some of the prominent use cases-

* **Retail** - In traditional retail, basket analysis is essential for optimizing product placement, creating targeted promotions, and enhancing personalized recommendations. By understanding which products are frequently purchased together, retailers can rearrange store layouts to place complementary products nearby, making it easier for customers to find related items, thereby increasing sales.
* **E-commerce** - In the world of online retail, basket analysis plays a crucial role in improving website navigation, recommendation systems, and personalized marketing campaigns. E-commerce platforms use basket analysis to suggest related products during the shopping experience, increasing the likelihood of customers adding more items to their cart. This technique also helps in crafting personalized email campaigns based on a customer’s purchase history, driving engagement and repeat purchases.
* **Healthcare** - Basket analysis can be applied to patient data to identify patterns of co-occurring diseases or treatments. This information can be used to improve patient care by suggesting treatment plans based on common disease associations or highlighting potential complications. Healthcare providers can also use basket analysis to optimize inventory management for medical supplies based on frequently co-prescribed drugs or procedures.
* **Financial Services** - Banks and insurance companies can use basket analysis to detect fraudulent activity, identify high-value customers, and tailor product offerings. By analyzing customer transactions, financial institutions can find unusual patterns that may indicate fraud or identify trends in customer spending behavior to develop personalized financial products, such as loan offerings or investment plans.

Basket analysis is a powerful tool for businesses across industries, enabling them to uncover hidden patterns, optimize operations, and make more informed decisions. As businesses continue to adopt data-driven strategies, the potential for basket analysis to enhance customer experience, streamline operations, and increase profitability is immense. With future extensions, such as advanced algorithms and real-time data integration, the applications of basket analysis will continue to evolve and expand, providing even deeper insights and opportunities for growth.

# **References**

1. **Hahsler, M., Chelluboina, S., Hornik, K., & Buchta, C. (2011)**. *The arules R-Package Ecosystem: Analyzing Interesting Patterns from Large Transaction Datasets*. *Journal of Machine Learning Research*, 12, 1977-1981.
2. **Tan, P., Steinbach, M., & Kumar, V. (2005)**. *Introduction to Data Mining*. Pearson.
3. **Rakesh Agrawal, Tomasz Imielinski, & Arun Swami (1993)**. *Mining Association Rules Between Sets of Items in Large Databases*. ACM SIGMOD Record.
4. **Borgelt, C. (2005)**. *An Implementation of the FP-growth Algorithm*.
5. **Raschka, S. (2018)**. *mlxtend: Providing Machine Learning and Data Science Utilities and Extensions to Python's Scientific Computing Stack*. *Journal of Open Source Software*.
6. **Kaggle (n.d.)**. *Bakery Sales Dataset*. Kaggle.
7. **Scikit-learn Developers (2023)**. *Scikit-learn: Machine Learning in Python*.
8. **Witten, I.H., Frank, E., Hall, M.A., & Pal, C.J. (2016)**. *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier.