# MARKET BASKET ANALYSIS


## TEAM MEMBERS

**822421106079: VIGNESHVAR M**


**PHASE-3 DOCUMENT SUBMISSION**

**PROJECT TOPIC: MARKET BASKET ANALYSIS**

## INTRODUCTION

Market basket analysis is a data mining technique that retailers use to understand customer purchasing patterns. It involves analyzing large datasets of transaction data, such as point-of-sale data, to identify patterns in the products that customers buy together.

## some of the benefits of using market basket analysis:

- **Increased sales:** By understanding customer purchasing patterns, retailers can make better decisions about product placement, product bundles, and marketing campaigns. This can lead to increased sales.

- **Improved customer satisfaction**: By providing customers with the products that they are most likely to want, retailers can improve customer satisfaction.
- **Reduced costs:** By using market basket analysis to reduce inventory waste and improve operational efficiency, retailers can reduce costs.

## ABSTRACT


- ✓ Market basket analysis (MBA) is a data mining technique that retailers use to understand customer purchasing patterns by analyzing large datasets of transaction data, such as point-of-sale data. MBA can be used to identify patterns in the products that customers buy together, which can then be used to improve retail operations, such as product placement, product bundles, and marketing campaigns.
- ✓ MBA is based on the principle that customers are more likely to buy certain products together than others. For example, customers who buy bread are also likely to buy milk and eggs. By identifying these patterns, retailers can make better decisions about how to display and promote their products.
- ✓ MBA is typically performed using a variety of data mining algorithms, such as Apriori and FP-Growth. These algorithms identify frequent itemsets, which are groups of products that occur together in a transaction dataset more often than would be expected by chance.


**PHASE-3 TOPIC:** market basket insights project by loading and pre-processing the transaction data, Load the transaction dataset and pre-process the data for association analysis.

## DATA SOURCE:

**Dataset link:(**https://www.kaggle.com/input/market-basket-analysis/assignment-1_data.xlsx **)**

| BillNo | Itemname | Quantity | Date | Price | Customer ID | Country |
|---|---|---|---|---|---|---|

| | | | | | | |
|---|---|---|---|---|---|---|
| 536365 | WHITE HANGING HEART T-LIGHT HOLDER | 6 | ###### ## | 2.55 | 17850 | United Kingdom |
| 536365 | WHITE METAL LANTERN | 6 | ###### ## | 3.39 | 17850 | United Kingdom |
| 536365 | CREAM CUPID HEARTS COAT HANGER | 8 | ###### ## | 2.75 | 17850 | United Kingdom |
| 536365 | KNITTED UNION FLAG HOT WATER BOTTLE | 6 | ###### ## | 3.39 | 17850 | United Kingdom |
| 536365 | RED WOOLLY HOTTIE WHITE HEART. | 6 | ###### ## | 3.39 | 17850 | United Kingdom |
| 536365 | SET 7 BABUSHKA NESTING BOXES | 2 | ###### ## | 7.65 | 17850 | United Kingdom |
| 536365 | GLASS STAR FROSTED T-LIGHT HOLDER | 6 | ###### ## | 4.25 | 17850 | United Kingdom |
| 536366 | HAND WARMER UNION JACK | 6 | ###### ## | 1.85 | 17850 | United Kingdom |
| 536366 | HAND WARMER RED POLKA DOT | 6 | ###### ## | 1.85 | 17850 | United Kingdom |
| 536367 | ASSORTED COLOUR BIRD ORNAMENT | 32 | ###### ## | 1.69 | 13047 | United Kingdom |
| 536367 | POPPY'S PLAYHOUSE BEDROOM | 6 | ###### ## | 2.1 | 13047 | United Kingdom |
| 536367 | POPPY'S PLAYHOUSE KITCHEN | 6 | ###### ## | 2.1 | 13047 | United Kingdom |
| 536367 | FELTCRAFT PRINCESS CHARLOTTE DOLL | 8 | ###### ## | 3.75 | 13047 | United Kingdom |
| 536367 | IVORY KNITTED MUG COSY | 6 | ###### ## | 1.65 | 13047 | United Kingdom |
| 536367 | BOX OF 6 ASSORTED COLOUR TEASPOONS | 6 | ###### ## | 4.25 | 13047 | United Kingdom |
| 536367 | BOX OF VINTAGE | 3 | ###### ## | 4.95 | 13047 | United Kingdom |

| | | | | | | |
|---|---|---|---|---|---|---|
| | JIGSAW BLOCKS BOX OF VINTAGE | | | | | |
| 536367 | ALPHABET BLOCKS | 2 | ###### ## | 9.95 | 13047 | United Kingdom |
| 536367 | HOME BUILDING BLOCK WORD | 3 | ###### ## | 5.95 | 13047 | United Kingdom |
| 536367 | LOVE BUILDING BLOCK WORD | 3 | ###### ## | 5.95 | 13047 | United Kingdom |

## To load the transaction dataset and pre-process the data for association analysis, you can follow these steps:

1. **Load the transaction dataset**: You can use a programming language such as Python or R to load the transaction dataset from a CSV file.
2. **Remove duplicate rows.** It is important to remove duplicate rows from the transaction dataset, as this can skew the results of the association analysis.
3. **Remove any rows with missing values:** Missing values can also skew the results of the association analysis, so it is best to remove any rows with missing values.
4. **Convert the product_id column to numerics:** This will make it easier to perform the association analysis.
5. **Sort the DataFrame by customer_id and product_id:** This will help to improve the efficiency of the association analysis.

## Here is an example of how to load and pre-process the transaction dataset in Python

```
import pandas as pd

# Load the transaction data

df = pd.read_csv('transaction_data.csv')

# Preprocess the transaction data

# Remove duplicate rows
```

```
df = df.drop_duplicates()

# Remove any rows with missing values

df = df.dropna()

# Convert the `product_id` column to numerics

df['product_id'] = pd.to_numeric(df['product_id'])

# Sort the DataFrame by `customer_id` and `product_id`

df = df.sort_values(by=['customer_id', 'product_id'])

# Calculate the market basket insights

# Get the unique product IDs

unique_product_ids = df['product_id'].unique()

# Create a dictionary to store the market basket insights

market_basket_insights = {}

# Iterate over the unique product IDs

for product_id in unique_product_ids:

 # Get the customers who bought the product

    customers_who_bought_product = df[df['product_id'] ==
product_id]['customer_id'].tolist()

 # Get the other products that these customers bought

    other_products_bought =
df[df['customer_id'].isin(customers_who_bought_product)]['product_id'].t
olist()

# Remove the product ID itself from the list of other products bought
```

```
    other_products_bought.remove(product_id)

  # Add the market basket insight to the dictionary

    market_basket_insights[product_id] = other_products_bought

# Print the market basket insights

print(market_basket_insights)
```

**OUTPUT:**

```
{1: [2, 3, 4], 2: [1, 3, 5], 3: [1, 2, 5], 4: [1, 2, 3], 5: [1, 2, 3]}
```

**Here are some additional examples of how association analysis can be used to improve market basket insights**:

- Identifying the products that are most frequently bought on sale
- Identifying the products that are most frequently bought by customers of a certain demographic
- Identifying the products that are most frequently bought at certain times of the year
- Identifying the products that are most frequently bought in different regions

- Another important application of association analysis in market basket insights is in creating product bundles. Retailers can use association analysis to identify groups of products that are frequently bought together and then create product bundles that include these items. Product bundles can make it easier for customers to find the products that they need and can also lead to increased sales.

- Association analysis can also be used to design store layouts.

## TRANSACTION DATASET:

A transaction dataset is a collection of data that records the products or items that customers purchase together in a single transaction. It is typically stored in a tabular format, with each row representing a transaction and each column representing a product or item.

Here is an example of a transaction dataset:

| Customer ID | Product ID 1 | Product ID 2 | Product ID 3 |
|---|---|---|---|
| 1 | 1 | 2 | 3 |
| 2 | 4 | 5 | 6 |
| 3 | 7 | 8 | 9 |

This transaction dataset shows that customer 1 purchased products 1, 2, and 3 in a single transaction. Customer 2 purchased products 4, 5, and 6 in a single transaction. Customer 3 purchased products 7, 8, and 9 in a single transaction.

Transaction datasets can be pre-processed for association analysis in a number of ways. Some common pre-processing steps include:

- **Removing duplicate rows:** This is important to ensure that each transaction is only counted once.
- **Removing rows with missing values:** This is important to avoid errors when performing the association analysis.
- **Converting categorical variables to numerics:** This is necessary for some association analysis algorithms.
- **Scaling or normalizing the data:** This can improve the performance of some association analysis algorithms.

## CONCLUSION:

- ✓ Association rule mining is a powerful technique for discovering hidden patterns and relationships in large datasets. It is widely used in a variety of domains, including market basket analysis, customer segmentation, and fraud detection.

- ✓ To perform association rule mining, the data must first be pre-processed. This involves cleaning the data, removing any irrelevant or incomplete data, and converting the data into a suitable format for analysis.

**Some common data pre-processing steps for association rule mining include:**

- **Removing duplicate transactions:** If a customer purchases the same items multiple times in a single transaction, these duplicate transactions can be removed.
- **Handling missing values:** Missing values can be handled by imputing them with the most frequent value for the item, or by deleting the entire transaction.
- **Converting categorical variables to binary variables:** Association rule mining algorithms typically require the data to be in binary format. This means that each item in a transaction is represented as a binary variable, indicating whether or not the item was purchased.
- **Discretizing continuous variables:** Continuous variables, such as price, can be discretized by grouping them into bins. This can make the data more efficient to process and can also improve the accuracy of the association rule mining algorithm.

- ✓ Once the data has been pre-processed, it can be used to mine association rules using a variety of algorithms. Some of the most popular association rule mining algorithms include the Apriori algorithm and the Eclat algorithm.
- ✓ Association rule mining can be used to discover a variety of interesting patterns and relationships in transaction data. For example, a grocery store could use association rule mining to discover which items are frequently purchased together. This information could then be used to improve product placement, develop targeted marketing campaigns, and create product bundles.
- ✓ Overall, association rule mining is a valuable technique for discovering hidden patterns and relationships in large datasets. By pre-processing the data carefully, the accuracy and efficiency of association rule mining algorithms can be improved.