AUTHOR: Karishma.A

Title: MARKET BASKET INSIGHTS – Unveilling Customer

Behaviour through Association Analysis

Phase 2: Innovation

# OVERVIEW COMPREHENSION

The project aims to analyze transaction data from a retail business to uncover hidden patterns and associations between products. By employing association analysis techniques, such as the Apriori algorithm, the goal is to gain valuable insights into customer purchasing behavior and identify cross-selling opportunities. Market basket analysis with Apriori algorithm, The retailer wants to target customers with suggestions on itemset that a customer is most likely to purchase .

 The given dataset contains data of a retailer; the transaction data provides data around all the transactions that have happened over a period of time.

we will be able increase customer engagement and improve customer experience and identify customer behavior. I will solve this problem with use Association Rules type of unsupervised learning technique that checks for the dependency of one data item on another data item.

# PROJECT PHASE

**Design Thinking:** Employed a human-centered approach to problem-solving, understanding user needs, and defining the scope of the project.

**Problem Definition:** Defined the problem as unveiling customer behavior through association analysis, emphasizing the use of market basket insights.

## NumPy

**What it does:** Helps with numerical operations on arrays.

**In my project:** Useful for handling numerical data, doing calculations, and working with arrays efficiently.

**Pandas:**

**What it does:** Great for handling tabular data.

**In my project:** Helps with loading, cleaning, and analyzing data in a table format.

# NUMPY

1. **Array Operations:**

•**How it Works:** NumPy provides a powerful array object called **NumPy. Array**. You can use NumPy arrays to represent and perform operations on numerical data efficiently.

•**In my Project:** If the dataset involves numerical data (like quantities, prices, etc.), NumPy arrays can be handy for numerical operations and computations.

1. **Efficient Mathematical Operations:**

•**How It Works:** NumPy is optimized for numerical operations. It includes a wide range of mathematical functions that operate on entire arrays efficiently.

•**In my Project:** During data preprocessing or analysis, you might need to perform mathematical operations on your data, and NumPy simplifies this process.

1. **Random Number Generation:**

•**How It Works:** NumPy has a random module for generating random numbers and random arrays.

•**In my Project:** If you need to simulate or generate random data for testing or analysis, NumPy's random functions can be valuable.

1. **Broadcasting:**

•**How It Works:** NumPy allows for broadcasting, which is a powerful way to perform operations on arrays of different shapes.

•**In my Project:** Broadcasting can be useful if you're working with data of different shapes or dimensions and need to perform element-wise operations.

# PANDAS

1. **DataFrames:**

**How It Works:** Pandas introduces the DataFrame, a two-dimensional labeled data structure with columns that can be of different types (similar to a table in a relational database or an Excel spreadsheet).

**In my Project:** The transaction data can be structured as a Pandas DataFrame, making it easy to perform operations on rows and columns.

1. **Data Cleaning:**

**How It Works:** Pandas provides a range of functions for cleaning and preprocessing data, such as handling missing values, dropping duplicates, and filtering data.

**In Your Project:** Use Pandas functions to clean and preprocess your transaction data before running association analysis.

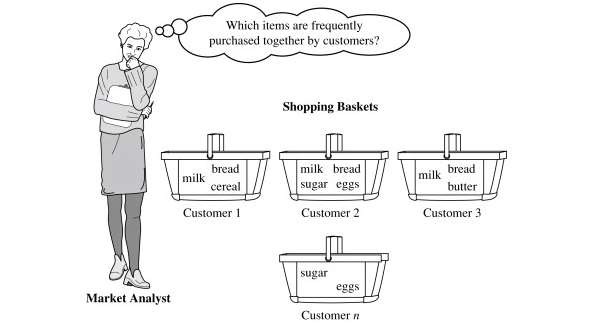
|  |
| --- |
| 1. **GroupBy Operations:**   •**How It Works:** Pandas' **groupby** function allows you to group data based on one or more criteria and perform operations on each group.  •**In my Project:** Grouping by certain criteria (e.g., customer ID) could be helpful in understanding patterns within specific customer segments.   1. **Merging and Joining:**   •**How It Works:** Pandas facilitates combining data from different sources through merging and joining operations.  •**In my Project:** If you have multiple datasets or need to merge information, Pandas can help bring these datasets together.   1. **Time Series Analysis:**   •**How It Works:** Pandas has robust support for time series data, making it easy to work with date and time information.  •**In my Project:** If the dataset includes timestamps or date-related information, Pandas can help analyze temporal patterns. |

.

ASSOCIATION ANALYSIS(Apriori Algorithm)

|  |
| --- |
| 1. **Finding What's Popular:** Apriori helps us discover what items are frequently bought together in a store. It starts by looking at individual items that are bought often. 2. **Gradually Getting More Complex:** Next, it looks at pairs of items that are frequently bought together. Then, it goes on to triplets, and so on, building up the complexity 3. **Setting Rules:** From these patterns, we create rules like "If someone buys A, they are likely to buy B too. “These rules have a confidence level, indicating how often the pattern holds true. 4. **Filtering Based on Confidence:** We can set a confidence threshold to filter out less reliable rules. This helps focus on the stronger connections between items. 5. **Practical Example:** If many people who buy milk also buy cereal, the algorithm captures this and gives us a rule like "Buying milk => Buying cereal." |

**How it works**

Eg: buys(x,” Bread”)=>buys(X, “eggs”)

Support = 1%, Confidence = 50%

If **Cofidence** 50%, which person buys Bread, they had 50% chances to buy the egg.

If **Support** 1%, which person buys bread at 1% then, they will parallely buy the egg as well as together.

## PROJECT DETAILS

**Data Source URL:** https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis (provided by skillup kaggle dataset)

**Code Structure:** The Jupyter Notebook is structured into well-organized code cells, each serving a specific purpose. The code is documented with comments, explaining the functionality of each cell and the overall logic of the Data Cleaning and further processes.

**Updated Dataset:** As the original dataset cannot also be clean enough in order take that dataset into Data Analysis Phase, I have cleaned the dataset and stored the cleaned dataset as an “updated\_dataset” in xlsx file extension.

**Future plan:**

**1.Try Different Ways:**

Experiment with different methods to find patterns in what people buy.

**2.Watch for Changes:**

Keep an eye on how people's buying habits change over time.

1. **Connect with Online Shops:** Make your analysis work directly on online shops to suggest items when people make purchases.
2. **Make it Shop-Friendly:** Help shops use your system easily, like plugging it into their cash register system.
3. **Ask for Feedback:** Get feedback from users to improve your system.
4. **Predict Future Purchases:** Use smart predict what people might buy next.
5. **Talk to Users:** Guesses to Regularly talk to users to understand their needs.

## Sample code

* import pandas as pd
* from mlxtend.preprocessing import TransactionEncoder
* from mlxtend.frequent\_patterns import apriori, association\_rules
* file\_path = ‘C:\Downloads\’Assignment-1\_data.csv’
* df = pd.read\_excel(‘C:\Downloads\’Assignment-1\_data.csv’,Assignment-1\_data.csv)
* te = TransactionEncoder()
* te\_ary = te.fit(df['Transaction']).transform(df['Transaction'])
* df\_encoded = pd.DataFrame(te\_ary, columns=te.columns\_)
* # Apply Apriori algorithm to find frequent itemsets
* frequent\_itemsets = apriori(df\_encoded, min\_support=0.2, use\_colnames=True)
* # Generate association rules
* rules = association\_rules(frequent\_itemsets, metric="confidence", min\_threshold=0.7)
* print("Frequent Itemsets:")
* print(frequent\_itemsets)
* print("\ Rules:")
* print(rules)

### CONCLUSION

In wrapping up our exploration into market basket analysis, we've uncovered valuable insights into customer purchasing behavior. The journey from problem definition to innovative solutions has provided a deeper understanding of how products are associated, opening doors to strategic opportunities for businesses.