**Market Basket Insights**

**Problem Definition and Design Thinking:**

**Problem Definition:** The problem is to perform market basket analysis on a provided dataset to find the patter of the buying and combination of buying. The goal is to understand customer buying behaviour and identify combinatory products for a retail business. This project involves using association analysis techniques, such as Apriori algorithm, to find frequently co-occurring products and generate insights for business optimization.

**Design Thinking:**

1. **Data Source:** Choose a dataset containing transaction data, including lists of purchased products.

**Dataset Link:** [**https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis**](https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis)

1. **Data Preprocessing**: Prepare the transaction data by transforming it into a suitable format

for association analysis.

1. **Association Analysis**: Utilise the Apriori algorithm to identify frequent itemsets and generate association rules.
2. **Insights Generation**: Interpret the association rules to understand customer behaviour and cross-selling opportunities.
3. **Visualisation**: Create visualisations to present the discovered associations and insights.
4. **Business Recommendations**: Provide actionable recommendations for the retail business

based on the insights.

In today's competitive market, understanding customer behaviour is crucial for business ro make informed decisions. Market Basket insight is a powerful tool for retailers and e-commerce companies to gain a deeper understanding of their customer.

It uses transactional data to answer questions such as:

● Whatproductsarefrequentlypurchasedtogether?  
● Whatisthemostcommonsequenceofpurchases? ● Arethereanytrendsinpurchasebehaviourovertime

These questions can be answered using market basket insight through transactional data

What Is Market Basket Analysis?

Market basket analysis is a strategic data mining technique used by retailers to enhance sales by gaining a deeper understanding of customer purchasing patterns. This method entails the examination of substantial datasets, such as historical purchase records, in order to unveil inherent product groupings and identify items that tend to be bought together.

By recognizing these patterns of co-occurrence, retailers can make informed decisions to optimise inventory management, devise effective marketing strategies, employ cross-selling tactics, and even refine store layout for improved customer engagement.

For example, if customers are buying milk, how likely are they to also buy bread (and which kind of bread) on the same trip to the supermarket? This information may lead to an increase in sales by helping retailers to do selective marketing based on predictions, cross-selling, and planning their ledge space for optimal product placement.

How Does Market Basket Analysis Work?

1. Collect data on customer transactions, such as the items purchased in each transaction, the time and date of the transaction, and any other relevant information.
2. Clean and preprocess the data, removing any irrelevant information, handling missing values, and converting the data into a suitable format for analysis.
3. Use association rules mining algorithms such as Apriori or FP-Growth to identify frequent item sets, sets of items often appearing together in a transaction.
4. Calculate the support and confidence for each frequent itemset, which expresses the likelihood of one item being purchased given the purchase of another item.
5. Generate association rules based on the frequent itemsets and their corresponding support and confidence values. Association rules express the likelihood of one item being purchased given the purchase of another item.
6. Interpret the results of the market basket analysis, identifying which items are frequently purchased together, the strength of the association between items, and any other relevant insights into customer behaviour and preferences.
7. Use the insights from the market basket analysis to inform business decisions such as product recommendations, store layout optimization, and targeted marketing campaigns.

Algorithms Used in Market Basket Analysis

There are multiple data mining techniques and algorithms used in Market Basket Analysis. One of the important objectives is “*to predict the probability of items that are being bought together by customers.”*

* ● **Apriori Algorithm**
* ● **AIS**
* ● **SETM Algorithm**
* ● **FP Growth**

1. **Apriori Algorithm**

Apriori Algorithm is a widely-used and well-known Association Rule algorithm and is a popular algorithm used in market basket analysis. It is also considered accurate and overtop AIS and SETM algorithms. It helps to find frequent itemsets in transactions and identifies association rules between these items. The limitation of the Apriori Algorithm is *frequent itemset generation*. It needs to scan the database many times, leading to increased time and reduced performance as a computationally costly step because of a large dataset. It uses the concepts of Confidence and Support.

2. **AIS Algorithm**

The AIS algorithm creates multiple passes on the entire database or transactional data. During every pass, it scans all transactions. As you can see, in the first pass, it counts the support of separate items and determines then which of them are frequent in the database. Huge itemsets of every pass are enlarged to generate candidate itemsets. After each scanning of a transaction, the common itemsets between the itemsets of the previous pass and the items of this transaction are determined. This algorithm was the first published algorithm which was developed to generate all large itemsets in a transactional database. It focused on the enhancement of databases with the necessary performance to process decision support. This technique is bounded to only one item in the consequent.

* ● **Advantage**: The AIS algorithm was used to find whether there was an association between items or not.
* ● **Disadvantage**: The main disadvantage of the AIS algorithm is that it generates too many candidate sets that turn out to be small. As well as the data structure is to be maintained.

3. **SETM Algorithm**

This Algorithm is quite similar to the AIS algorithm. The **SETM** algorithm creates collective passes over the database. As you can see, in the first pass, it counts the support of single items and then determines which of them are frequent in the database. Then, it also generates the candidate itemsets by enlarging large itemsets of the previous pass. In addition to this, the SETM algorithm recalls the TIDs(transaction ids) of the generating transactions with the candidate itemsets.

* ● **Advantage**: While generating candidate itemsets, the SETM algorithm arranges candidate itemsets together with the TID(transaction Id) in a sequential manner.
* ● **Disadvantage**: For every item set, there is an association with Tid; hence it requires more space to store a huge number of TIDs.

4. **FP Growth**

**FP Growth** is known as Frequent Pattern Growth Algorithm. FP growth algorithm is a concept of representing the data in the form of an FP tree or Frequent Pattern. Hence FP Growth is a method of *Mining Frequent Itemsets*. This algorithm is an advancement to the **Apriori Algorithm**. There is no need for candidate generation to generate a frequent pattern. This frequent pattern tree structure maintains the association between the itemsets.

**A Frequent Pattern Tree** is a tree structure that is made with the earlier itemsets of the data. The main purpose of the FP tree is to mine the most frequent patterns. Every node of the FP tree represents an item of that itemset. The root node represents the null value, whereas the lower nodes represent the itemsets of the data. The association of these nodes with the lower nodes, that is, between itemsets, is maintained while creating the tree.

Advantages of Market Basket Analysis

There are many advantages to implementing Market Basket Analysis in marketing. Market Basket Analysis (MBA) can be applied to data of customers from the point of sale (**PoS**) systems.

It helps retailers in the following ways:

* ● Increases customer engagement
* ● Boosts sales and increases RoI
* ● Improves customer experience
* ● Optimises marketing strategies and campaigns
* ● Helps in demographic data analysis
* ● Identifies customer behaviour and pattern

The main rule that is applied by the market basket analysis algorithm that produces certain output that helps us to understand the basic behaviour and the mind set of the customers who buys product from the store

The result which are produced by the apriori algorithm:

● Antecedents  
● Consequents  
● Antecedentssupport ● Consequentsupport ● Support  
● Confidence  
● Lift  
● Leverage  
● Conviction  
● Zhangs’smetric

**1.Antecendents:**

Antecedents are the items or itemsets that appear before the arrow (=>) in an association rule. They represent the condition or the "if" part of the rule.

**2.Consequents:**

Consequents are the items or itemsets that appear after the arrow (=>) in an association rule. They represent the result or the "then" part of the rule.

**3.Antecedent Support:**

Antecedent support is the support of the antecedent itemset, which measures the frequency of occurrence of the antecedent in the dataset.

**4.Consequent Support:**

Consequent support is the support of the consequent itemset, which measures the frequency of occurrence of the consequent in the dataset.

**5.Support:**

Support is a measure of the frequency of occurrence of the whole itemset (antecedent and consequent) in the dataset. It indicates how often the entire rule is satisfied.

**6.Confidence:**

Confidence measures the strength of the association between the antecedent and the consequent in an association rule. It is defined as the support of the combined itemset (antecedent and consequent) divided by the support of the antecedent.

High confidence indicates a strong relationship between the antecedent and the consequent.

**7.Lift:**

Lift is a measure that quantifies how much more likely the consequent is to occur given the antecedent compared to if they were statistically independent. It is defined as the ratio of the support of the combined itemset (antecedent and consequent) to the product of the supports of the antecedent and the consequent.

A lift value greater than 1 suggests a positive association (items are more likely to be bought together), while a value less than 1 suggests a negative association.

**8.Leverage:**

Leverage measures the difference between the observed support of the combined itemset and the expected support if the antecedent and the consequent were independent. It helps identify the deviation from independence.

**9.Conviction:**

Conviction measures the degree to which the consequent relies on the antecedent. It is defined as the ratio of the expected support (assuming independence) of the antecedent and the actual support of the consequent.

High conviction values indicate a strong association between the antecedent and the consequent.

**10.Zhang's Metric:**

It is designed to find rules with high confidence while avoiding rules with very low support.

These metrics are used to assess the quality and significance of association rules discovered by the Apriori algorithm and other similar algorithms. Analysts use these metrics to filter and prioritise rules, selecting those that are most relevant for decision-making, recommendation systems, and other data mining applications.

**Program:**  
**Data Cleaning:**

data = pd.read\_csv("Assignment-1\_Data.csv",sep=';',low\_memory=False) data.head()  
data.columns  
unique\_country = data.Country.unique()

print(unique\_country)

basket\_France = (data[data['Country'] =="France"] .groupby(['BillNo', 'Itemname'])['Quantity'] .sum().unstack().reset\_index().fillna(0) .set\_index('BillNo'))

basket\_UK = (data[data['Country'] =="United Kingdom"] .groupby(['BillNo', 'Itemname'])['Quantity'] .sum().unstack().reset\_index().fillna(0) .set\_index('BillNo'))

basket\_Por = (data[data['Country'] =="Portugal"] .groupby(['BillNo', 'Itemname'])['Quantity'] .sum().unstack().reset\_index().fillna(0)

import numpy as np  
import pandas as pd  
from mlxtend.frequent\_patterns import apriori, association\_rules

data['Itemname'] = data['Itemname'].str.strip() data.dropna(axis = 0, subset =['BillNo'], inplace = True)

data['BillNo'] = data['BillNo'].astype('str') data = data[~data['BillNo'].str.contains('C')]

.set\_index('BillNo'))  
basket\_Sweden = (data[data['Country'] =="Sweden"]

.groupby(['BillNo', 'Itemname'])['Quantity'] .sum().unstack().reset\_index().fillna(0) .set\_index('BillNo'))

def hot\_encode(x): if(x<= 0):

return 0 if(x>= 1):

return 1  
basket\_encoded = basket\_France.map(hot\_encode) basket\_France = basket\_encoded  
basket\_encoded = basket\_UK.map(hot\_encode) basket\_UK = basket\_encoded  
basket\_encoded = basket\_Por.map(hot\_encode) basket\_Por = basket\_encoded  
basket\_encoded = basket\_Sweden.map(hot\_encode) basket\_Sweden = basket\_encoded

**Applying apriori algorithm:**

frq\_items = apriori(basket\_UK, min\_support = 0.02, use\_colnames = True) rules = association\_rules(frq\_items, metric ="lift", min\_threshold = 1)

rules = rules.sort\_values(['confidence', 'lift'], ascending =[False, False]) print(rules.head())

output:

Conclusion

Implement Market Basket Analysis from scratch using Python. We then implemented Market Basket Analysis using Apriori Algorithm. We also looked into the various uses and advantages of this algorithm and learned that we could also use FP Growth and AIS algorithms to implement Market Basket Analysis.

**Key Takeaways**

* ● Market Basket Analysis is a business strategy used to design store layouts based on customers’ shopping behaviour and purchase histories.
* ● This idea is also applicable to machine learning algorithms to teach machines to help businesses, especially in the e-commerce sector.
* ● In this article, we have gone through a step-by-step guide to implementing the apriori algorithm in Python and also looked into the math behind the association rules