

How to use market basket analysis to support retail operations

In this article, we will comprehensively cover the topic of Market Basket Analysis and its various components, then dive deep into the ways of implementing it in machine learning, including how to perform it in Python on a real-world dataset. Machine Learning is helping the Retail Industry in many different ways; you can imagine that from forecasting the performance of sales to identifying the buyers, there are many applications of AI and ML in the retail industry. Market basket analysis is a data mining technique retailers use to increase sales by better understanding customer purchasing patterns. It involves analyzing large data sets, such as purchase history, to reveal product groupings and products likely to be purchased together.

How does Market Basket analysis work?

- 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.
- Clean and preprocess the data, removing any irrelevant information, handling missing values, and converting the data into a suitable format for analysis.
- 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.
- Calculate the support and confidence for each frequent itemset, which expresses the likelihood of one item being purchased given the purchase of another item.
- 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.
- 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 behavior and preferences.
- Use the insights from the market basket analysis to inform business decisions such as product recommendations, store layout optimization, and targeted marketing campaigns.

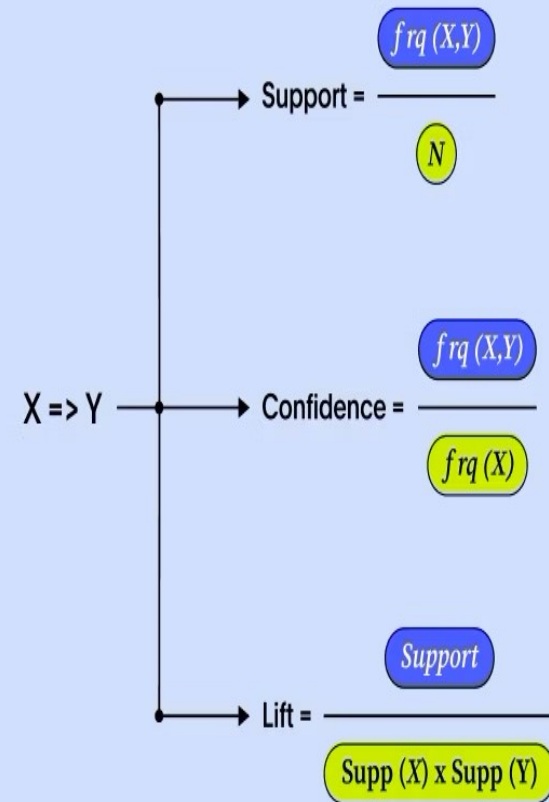
Concepts

There are three main concepts you need to know in order to successfully understand market basket analysis.

1.SUPPORT

2.CONFIDENCE

3.LIFT



What Is Association Rule for Market Basket Analysis ?

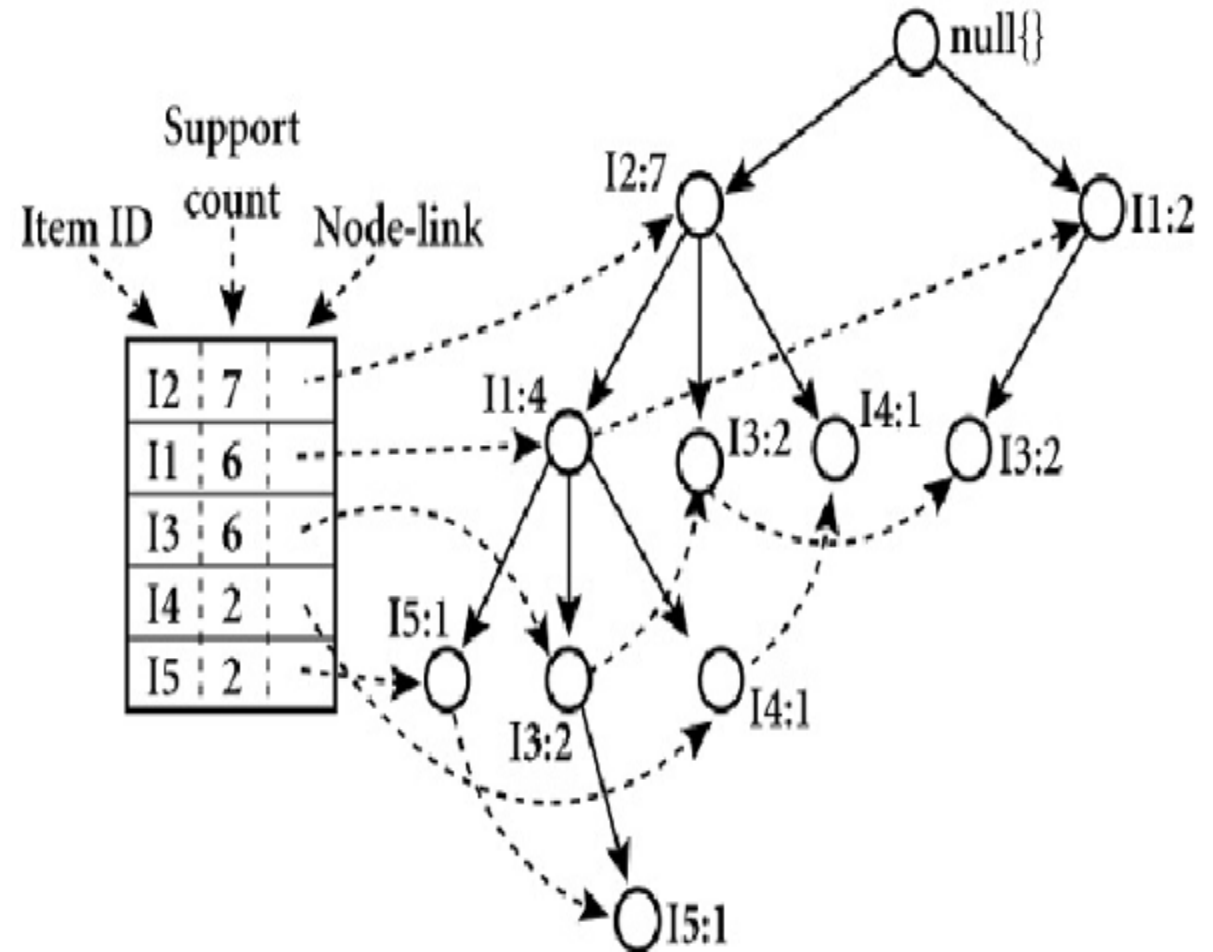
- Let $I = \{I_1, I_2, \dots, I_m\}$ be an itemset. These itemsets are called antecedents. Let D , the data, be a set of database transactions where each transaction T is a nonempty itemset such that $T \subseteq I$. Each transaction is associated with an identifier called a TID(or Tid). Let A be a set of items(itemset). T is the Transaction that is said to contain A if $A \subseteq T$. An **Association Rule** is an implication of form $A \Rightarrow B$, where $A \subset I$, $B \subset I$, and $A \cap B = \varnothing$.
- The rule $A \Rightarrow B$ holds in the data set(transactions) D with supports, where 's' is the percentage of transactions in D that contain $A \cup B$ (i.e., the union of set A and set B , or both A and B). This is taken as the probability, $P(A \cup B)$. Rule $A \Rightarrow B$ has confidence c in the transaction set D , where c is the percentage of transactions in D containing A that also contains B . This is taken to be the conditional probability, like $P(B|A)$. That is,
 - $support(A \Rightarrow B) = P(A \cup B)$
 - $confidence(A \Rightarrow B) = P(B|A)$
- Rules that satisfy both a minimum support threshold (called min sup) and a minimum confidence threshold (called min conf) are called "**Strong**".
- $Confidence(A \Rightarrow B) = P(B|A) =$
- $support(A \cup B) / support(A) =$
- $support\ count(A \cup B) / support\ count(A)$
- Generally, Association Rule Mining can be viewed in a two-step process:
 1. Find all Frequent itemsets: *By definition, each of these itemsets will occur at least as frequently as a pre-established minimum support count, min sup.*
 2. Generate Association Rules from the Frequent item sets: *By definition, these*
 3. *rules must satisfy minimum support and minimum confidence.*

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**

EXAMPLE FOR ALGORITHM

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.



Implementing Market Basket Analysis in python

The Method:

Here are the steps involved in using the apriori algorithm to implement MBA:

1. First, define the minimum support and confidence for the association rule.
2. Find out all the subsets in the transactions with higher support(sup) than the minimum support.
3. Find all the rules for these subsets with higher confidence than minimum confidence.
4. Sort these association rules in decreasing order.
5. Analyze the rules along with their confidence and support.

Implementing Market Basket Analysis Using the Apriori Method

The Apriori algorithm is frequently used by data scientists. We are required to import the necessary libraries. Python provides the *apriori* as an API that is required to be imported to run the Apriori Algorithm.

```
import pandas as pd
import numpy as np
from apyori import apriori
st_df=pd.read_csv("store_data.csv",header=None)
print(st_df)
```

Now we want to read the dataset that is downloaded from Kaggle. There is no header in the dataset; hence, the first row contains the first transaction, so we have mentioned *header = None* here.

OUTPUT FOR THE ABOVE DATASET

0	1	2	...	17	18	19	
0	shrimp	almonds	avocado	...	frozen smoothie	spinach	olive oil
1	burgers	meatballs	eggs	...	NaN	NaN	NaN
2	chutney	NaN	NaN	...	NaN	NaN	NaN
3	turkey	avocado	NaN	...	NaN	NaN	NaN
4	mineral water	milk	energy bar	...	NaN	NaN	NaN
...
7496	butter	light mayo	fresh bread	...	NaN	NaN	NaN
7497	burgers	frozen vegetables	eggs	...	NaN	NaN	NaN
7498	chicken	NaN	NaN	...	NaN	NaN	NaN
7499	escalope	green tea	NaN	...	NaN	NaN	NaN
7500	eggs	frozen smoothie	yogurt cake	...	NaN	NaN	NaN

[7501 rows x 20 columns]

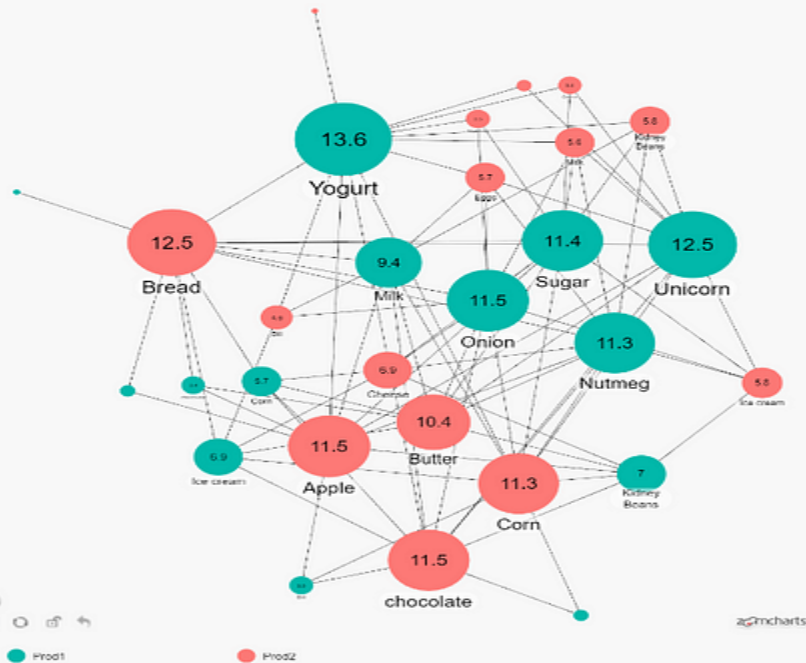
EXAMPLE : TOTAL ANALYSIS FOR DATASET

Market Basket Analysis Dashboard

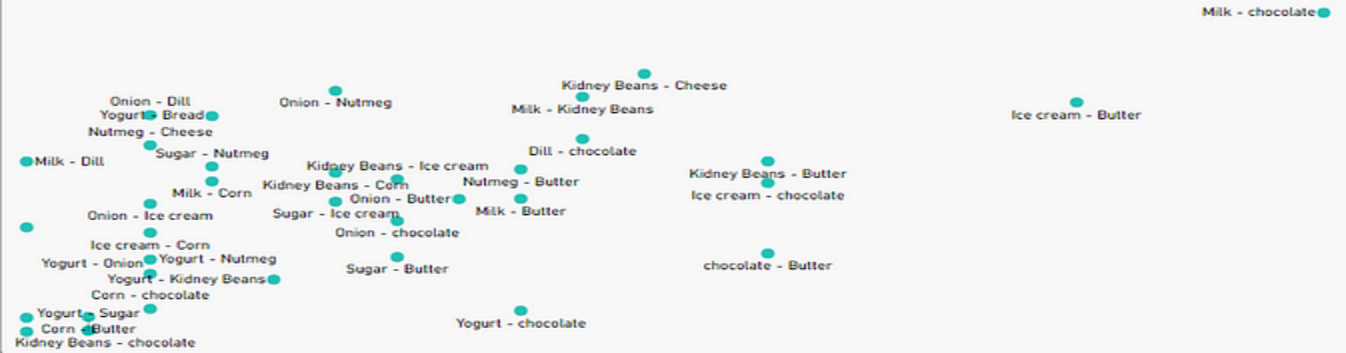
16
Number of Products

999
Total Transactions

Basket Analysis Network



Basket Analysis Map



Market Basket Details

Basket	Sum of Support Basket	Sum of Confidence of Prod1	Sum of Confidence of Prod2	Sum of Lift
Milk - chocolate	21.12%	52.10%	50.12%	1.24
Kidney Beans - Cheese	20.02%	49.02%	49.50%	1.21
Onion - Nutmeg	19.52%	48.39%	48.63%	1.21
Milk - Kidney Beans	19.92%	49.14%	48.77%	1.20
Ice cream - Butter	20.72%	50.49%	49.29%	1.20
Onion - Dill	19.22%	47.64%	48.24%	1.20
Yogurt - Bread	19.32%	45.95%	50.26%	1.20
Dill - chocolate	19.92%	50.00%	47.27%	1.19
Nutmeg - Cheese	19.22%	47.88%	47.52%	1.18
Kidney Beans - Butter	20.22%	49.51%	48.10%	1.18
Milk - Dill	19.02%	46.91%	47.74%	1.18
Sugar - Nutmeg	19.32%	47.19%	48.13%	1.18
Nutmeg - Butter	19.82%	49.38%	47.14%	1.17
Kidney Beans - Corn	19.52%	47.79%	47.91%	1.17
Kidney Beans - Ice cream	19.62%	48.04%	47.80%	1.17
Milk - Corn	19.32%	47.65%	47.42%	1.17
Ice cream - chocolate	20.22%	49.27%	47.98%	1.17

Once we have read the dataset completely, we are required to get the list of items in every transaction. So we are going to run two loops. One will be for the total number of transactions, and the other will be for the total number of columns in every transaction. The list will work as a training set from where we can generate the list of Association Rules.

- `#converting dataframe into list of lists l=[] for i in range(1,7501):
l.append([str(st_df.values[i,j]) for j in range(0,20)])`
- we are ready with the list of items in our training set, then we need to run the apriori algorithm, which will learn the list of association rules from the training set, i.e., list. So, the minimum support here will be 0.0045, which is taken here as support. Now let us see that we have kept 0.2 as the min confidence. The minimum lift value is taken as 3, and the minimum length is considered as 2 because we have to find an association among a minimum of two items.
- `#applying apriori algorithm association_rules = apriori(l, min_support=0.0045,
min_confidence=0.2, min_lift=3, min_length=2) association_results =
list(association_rules)`
- After running the above line of code, we generated the list of association rules between the items. So to see these rules, the below line of code needs to be run.
- `for i in range(0, len(association_results)): print(association_results[i][0])`

OUTPUT

transactions - List (7501 elements)

Index	Type	Size	Value
0	list	20	['shrimp', 'almonds', 'avocado', 'vegetables mix', 'green grapes', 'wh ...
1	list	20	['burgers', 'meatballs', 'eggs', 'nan', 'nan', 'nan', 'nan', 'nan', 'n ...
2	list	20	['chutney', 'nan', 'nan', 'nan', 'nan', 'nan', 'nan', 'nan', 'nan', 'n ...
3	list	20	['turkey', 'avocado', 'nan', 'nan', 'nan', 'nan', 'nan', 'nan', 'nan', ' ...
4	list	20	['mineral water', 'milk', 'energy bar', 'whole wheat rice', 'green tea ...
5	list	20	['low fat yogurt', 'nan', 'nan', 'nan', 'nan', 'nan', 'nan', 'nan', 'n ...
6	list	20	['whole wheat pasta', 'french fries', 'nan', 'nan', 'nan', 'nan', 'nan ...
7	list	20	['soup', 'light cream', 'shallot', 'nan', 'nan', 'nan', 'nan', 'nan', ' ...
8	list	20	['frozen vegetables', 'spaghetti', 'green tea', 'nan', 'nan', 'nan', ' ...
9	list	20	['french fries', 'nan', 'nan', 'nan', 'nan', 'nan', 'nan', 'nan', 'nan ...
10	list	20	['eggs', 'pet food', 'nan', 'nan', 'nan', 'nan', 'nan', 'nan', 'nan', ' ...

Save and Close Close

CONCLUSION

- In this ,we discussed Market Basket Analysis and learned the steps to implement it 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 behavior 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.