JANUARY 2022

MARKET BASKET ANALYSIS AND PRODUCT SALES ANALYSIS TO INCREASE SALES FOR SMALL AND MEDIUM-SIZED BUSINESSES

AI PRODUCT/SERVICE PROTOTYPE IDEATION (AI IN ECONOMY)

Prepared by: Vaasu Bisht
ML Intern
Feynn Labs

02

Table of content

03

INTRODUCTION

The goal of business is to increase profit and it cannot be done by just pitching one product at a time to the customers. Hence big businesses have begun mining data relating to frequently bought items. This is called Market Basket Analysis. Its integration of business with technology mainly Artificial Intelligence. Using this these enterprises try to find out associations between different items and products that can be sold together which gives them assistance in the right product placement. This can be used by small/medium businesses to help them retain their customers, increase profit and do inventory management by analyzing the transactional database. This would not only help them in expanding their operations but also help them compete with big and already established firms.

PROBLEM STATEMENT

There are many small/medium businesses in Indian cities, which sell similar products and at similar prices, which leads to the problem of customer retention. Many shops which are present in smaller cities have not faced the immense challenges given by big giants like Amazon and big basket as they are not present there. So the businesses in these cities lack technological integration and are not aware of techniques that can help them boost the business and also prepare them to compete against big giants when they enter their cities.

04

BUSSINESS NEED ASSESSMENT

Every street in India has a shop. That's why there is so much competition for customer retention, especially for small-level stores. They need a tool that can uncover meaningful correlations between different products that can show small/medium-level retailers how to locate products together and how to cross-promote and recommend items that customers often put in their shopping carts at the same time. This can be achieved with the use of a market basket and product sales analysis. With the help of the market basket and product sales analysis, we can change the arrangement and product display patterns to show customers products that they might add to their shopping carts.

TARGET SPECIFICATIONS AND CHARACTERIZATION

CUSTOMER RETENTION- By analyzing the association's rules and coming up with combinations of items that occur in transactions frequently. It can help vendors better anticipate the purchasing habits of their customers and keep them satisfied. By giving group discounts and keeping items which are bought in combination together for a convenient shopping experience. Market Basket Analysis would also help businesses in small Indian cities to give tough competition when big giants enter their city.

INCREASE SALES AND CUSTOMER SATISFACTION:-
Using machine learning and data analysis vendors can visualize which products are being sold more frequently and which products customers buy with those products. So they can focus on which items to buy more for their inventory, so their management of loadout and finances are improved.

05 EXTERNAL SEARCH

- Market basket analysis in e-commerce business explained (A Case study)

- Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining

- Market Basket Analysis: Increasing Sales and Improving ROI by Optimizing Campaigns and Promotions

BENCH MARKING ALTERNATE PRODUCTS

Market Basket Analysis is one of the key techniques used by a lot of E-commerce stores like Amazon, Bigbasket to uncover nonobvious, usually hidden, and counterintuitive associations between items, products, or categories. Currently, it is only used by big companies. These companies are only present in metropolitan cities which opens an opportunity for local vendors to fill up that void efficiently. The association rule mining would help local shops and vendors to discover associations and correlations between items in the transactional and relational dataset. The discovery will be helpful for the local vendor to gain insight into which items are frequently bought together by customers

06

APPLICABLE PATENTS

1)Retail Data Mining Using Co-Occurrence Consistency.

2)Enhanced market basket analysis

3)Method and system for researching product dynamics in market basket in conjunction with aggregate market basket properties

4)A fast method for renewal and associated recommendations for market basket items

APPLICABLE CONSTRAINTS

1) License for the open-source codes that might be used in the model
implementation.

2) Data protection and privacy regulations

3) Laws related to privacy for collecting data from user

4) Government Regulations for small scale businesses

07 APPLICABLE REGULATIONS

1. Average tend to lie. This leads to iffy correlations Analyzing the market basket helps to discover a trend, but once the action is taken it becomes difficult to assess the integrity of the relationship.

2. Lack of initial data to perform algorithms.

3. The Algorithm won't be able to detect rarely bought items, thus the analysis won't be able to tell which items are sold frequently with that item. The rarely bought item might be very expensive and thus may bring a huge profit to the vendor and giving offers with the items which are sold with it might also increase the sale of that item, eventually increasing profit.

4) Most local vendors and Shopkeepers lack technical knowledge, so it can be a little tough to Convince them to use machine learning.

5) Transparent use of the data obtained from the customer.

BUSINESS OPPORTUNITY

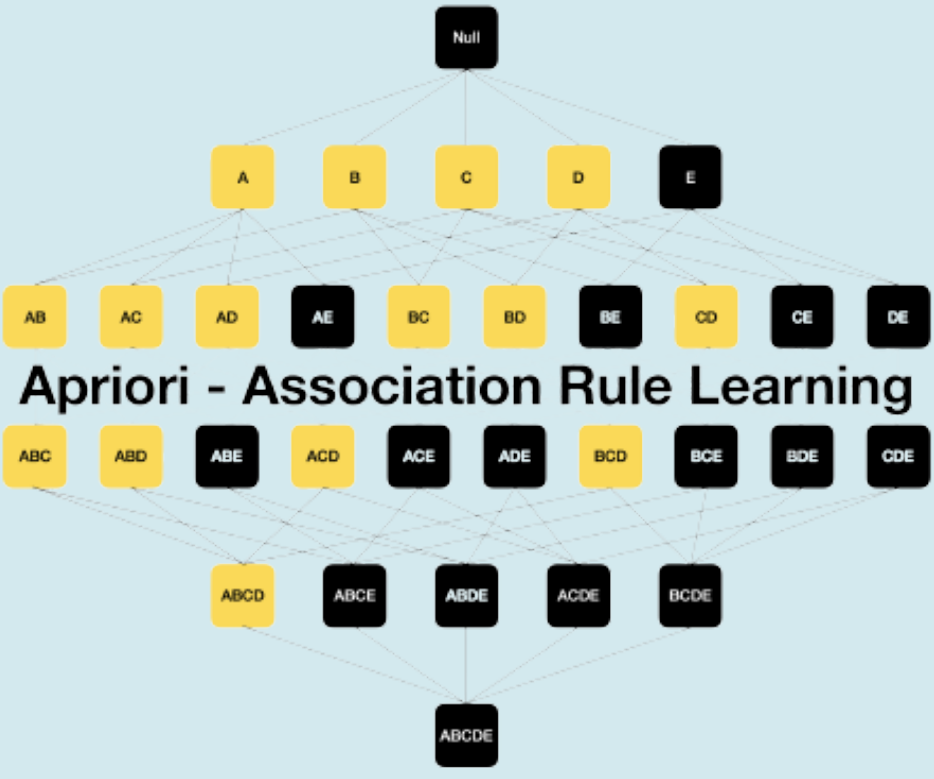
A lot of shops and retail stores have not integrated technology into their business yet. There is a lot that Artificial intelligence can do in this field and the opportunities are limitless. Market basket analysis is used by many big companies to group products and provide recommendations. Applying Market Basket Analysis to small and medium-sized businesses will give them an in-depth analysis of what things customers are buying and what they are not buying, which would, in turn, help them to increase their sales thus increasing and will help in maintaining their inventory. This will also align with the long-term goal of any business which is expanding and increasing its reach. If the model is successfully implemented for the small local businesses, the model can be expanded for medium-level retail stores, those with multiple stores. This can also be used by shops in malls.

08 CONCEPT GENERATION

First and foremost we began by gaining an understanding of the problem. There is cut-throat competition in the market as there are many local stores in Indian streets, but this is nothing when compared with challenges that will be thrown by big giants like Amazon and Bigbasket when they enter in small and medium cities of India. A local-level business that already lacks technological integration will start to lose customers as they won't be able to understand customers as clearly as their competitors. To avoid this, businesses should start analyzing their customer buying patterns and take advantage of local knowledge to expand their business.
This product/service is about equipping businesses with Artificial Intelligence. So vendors can understand the purchasing pattern of their customers and using that they can change, update, add, remove, or group items and can efficiently manage their inventory. They can also update the layout of physical stores or online to give customers a better shopping experience.

CONCEPT DEVELOPMENT

For the development most important thing we need is data. It can be collected from the vendors or can be collected by surveying. Then we will be applying association data mining which is an unsupervised machine learning technique that checks for the dependency of one data item on another data item and maps accordingly. As the Apriori algorithm deciphers the association rules, it would be used by our client to increase its sales and this would lead to a collection of more data which would be again running through the algorithm giving more precise association rules and improving the efficiency even more.



09 FINAL PRODUCT PROTOTYPE (ABSTRACT)

This product will be a mobile application that vendors can access. The vendors will need to create an account on it and connect it with Billing Machines via Wi-Fi or Bluetooth which will help in collecting the data and provide the analysis after running through the algorithm, which will help the vendor understand the buying patterns.

PRODUCT DETAILS

How does it work?

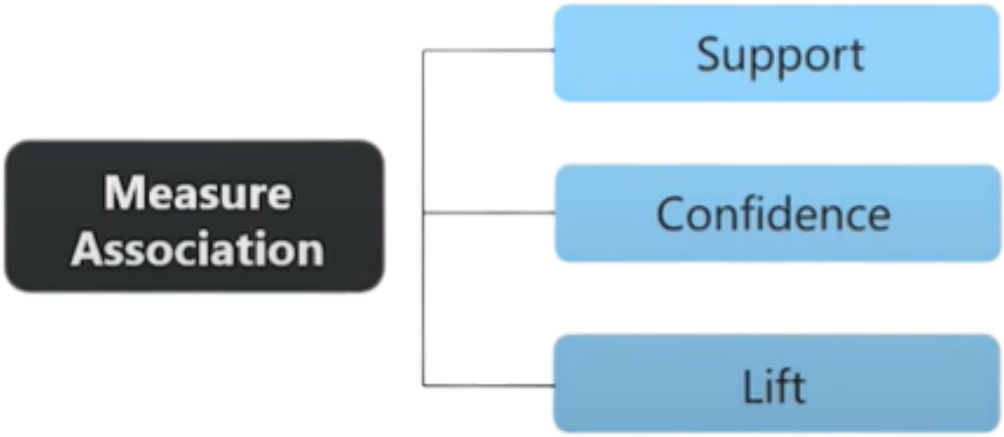
First of all the vendor needs to create its account on the APP. Then upload the past purchases if they are in digitized format, if not then he/she can take pictures and upload the data, Machine Learning engineering with help of NLP will convert the data into a CSV file. Then the vendors need to link their billing machine and inventory with the APP. The unsupervised machine learning algorithm will tell the association rules, which will be used by them to increase their sales and retain customers. The application will also provide an in-depth analysis of the sales in a graphical format so the vendor cab understand the monthly buying patterns, most favorite brands of its customers, etc. which would really help in inventory management.

Data Sources

For this service, there cannot be a common dataset as shopping patterns in India vary from place to place. So the best dataset for each city will come from the past purchases from the small/medium business of that city or thought surveys and later when APP is connected with Billing machine new dataset will be continuously collected and efficiency of Artificial intelligence model will be improved significantly. But for the prototype, I will be using the dataset available on Kaggle.

Algorithm:-

This product will use the Apriori algorithm which will mine frequent itemsets and devise association rules from the transactional databases. This algorithm uses three types of **matrices that will help us to measure association.**



a)SUPPORT- Frequency of combinations of items bought together. Helps to filter out items bought less frequently.

b) CONFIDENCE- Gives frequency of how often items A and B are bought together given the frequency of item

c) LIFT - Refers to the strength of any rule. Life will give a final verdict on whether we have to spend time on that particular rule or not.

$$Support = \frac{freq(A,B)}{N}$$

$$Confidence = \frac{freq(A,B)}{freq(A)}$$

$$Lift = \frac{Support}{Supp(A)\ x\ Supp(B)}$$

Association Rule Mining can be thought of as an IF-THEN relationship. It is not causality but a co-occurrence pattern that comes into force. IF also called ANTECEDENT these are groups of items that are typically found in itemset. THEN is called CONSEQUENT it comes along as an item with an antecedent group. This means that IF a person buys item A he/she will most probably buy item B.

Team required to develop

For development, I will require a machine learning engineer, full-stack developers, IOT engineering and, some people in the field for data collection.

CODE IMPLEMENTATION (PROTOTYPE)

How does it work?

First of all the vendor needs to create its account on the APP. Then upload the past purchases if they are in digitized format, if not then he/she can take pictures and upload the data, Machine Learning engineering with help of NLP will convert the data into a CSV file. Then the vendors need to link their billing machine and inventory with the APP. The unsupervised machine learning algorithm will tell the association rules, which will be used by them to increase their sales and retain customers. The application will also provide an in-depth analysis of the sales in a graphical format so the vendor cab understand the monthly buying patterns, most favorite brands of its customers, etc. which would really help in inventory management.

Data Sources

For this service, there cannot be a common dataset as shopping patterns in India vary from place to place. So the best dataset for each city will come from the past purchases from the small/medium business of that city or thought surveys and later when APP is connected with Billing machine new dataset will be continuously collected and efficiency of Artificial intelligence model will be improved significantly. But for the prototype, I will be using the dataset available on Kaggle.

14 CODE IMPLEMENTATION (PROTOTYPE)

Importing Algorithm

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Importing other libraries

```
!pip install apyori
from apyori import apriori
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules
```

```
# IMPORTING DATA SET
df = pd.read_csv('https://raw.githubusercontent.com/vaasu2002/Market-Basket-Analysis/main/dataset.csv')
df.head()
```

	memberNumber	Date	itemDescription
0	1808	21-07-2015	tropical fruit
1	1808	21-07-2015	candy
2	2300	19-09-2015	pip fruit
3	1187	12-12-2015	other vegetables
4	3037	01-02-2015	whole milk

```
print(f"Number of rows:- {df.shape[0]}")
```

```
Number of rows:- 38765
```

```
[5] df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 38765 entries, 0 to 38764
Data columns (total 3 columns):
 #   Column          Non-Null Count  Dtype 
---  ------          --------------  ----- 
 0   memberNumber    38765 non-null  int64 
 1   Date            38765 non-null  object
 2   itemDescription  38765 non-null  object
dtypes: int64(1), object(2)
memory usage: 908.7+ KB
```

```
[29] df.isnull().sum().sum()
```

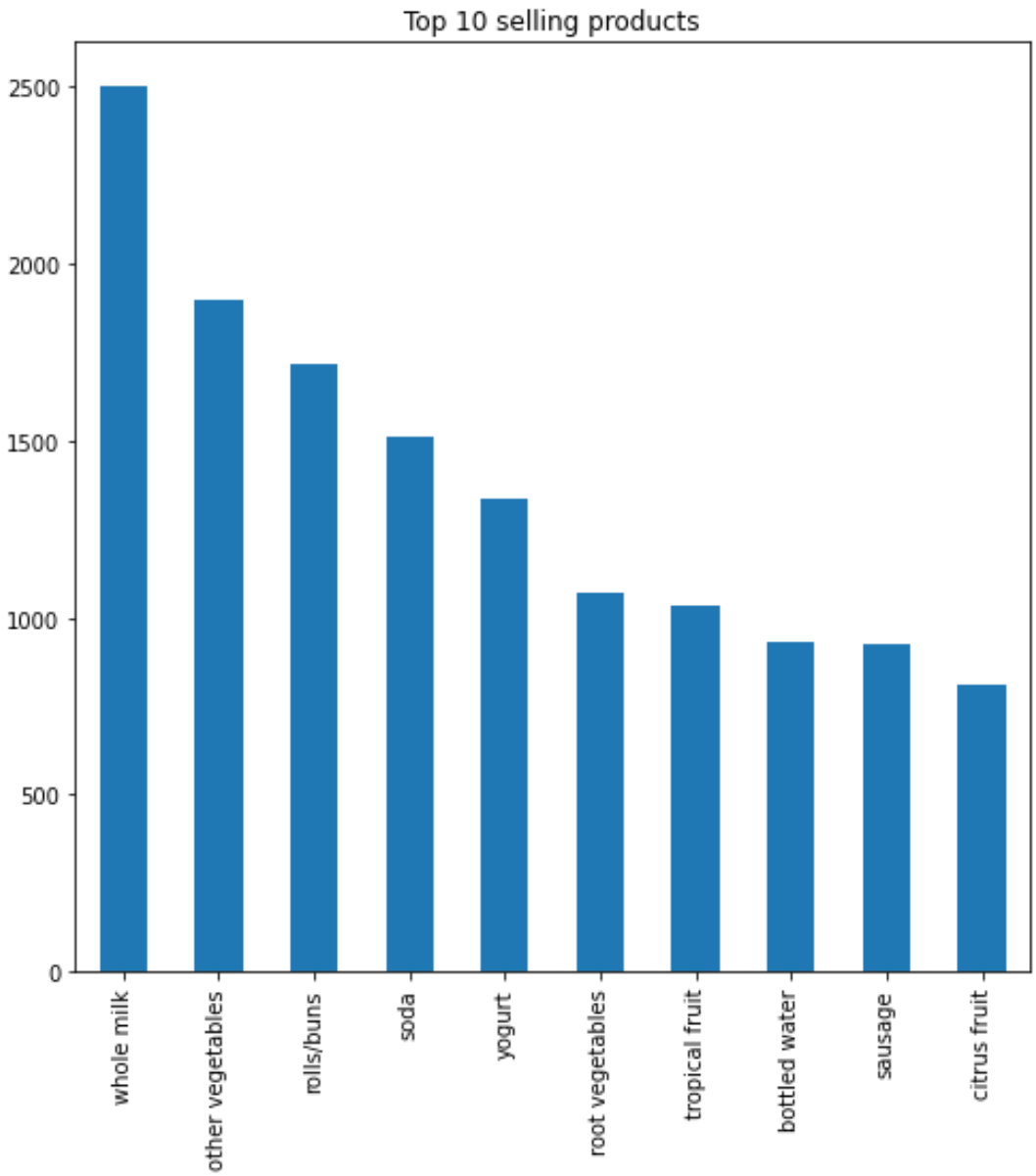
```
0
```

PERFORMING EDA

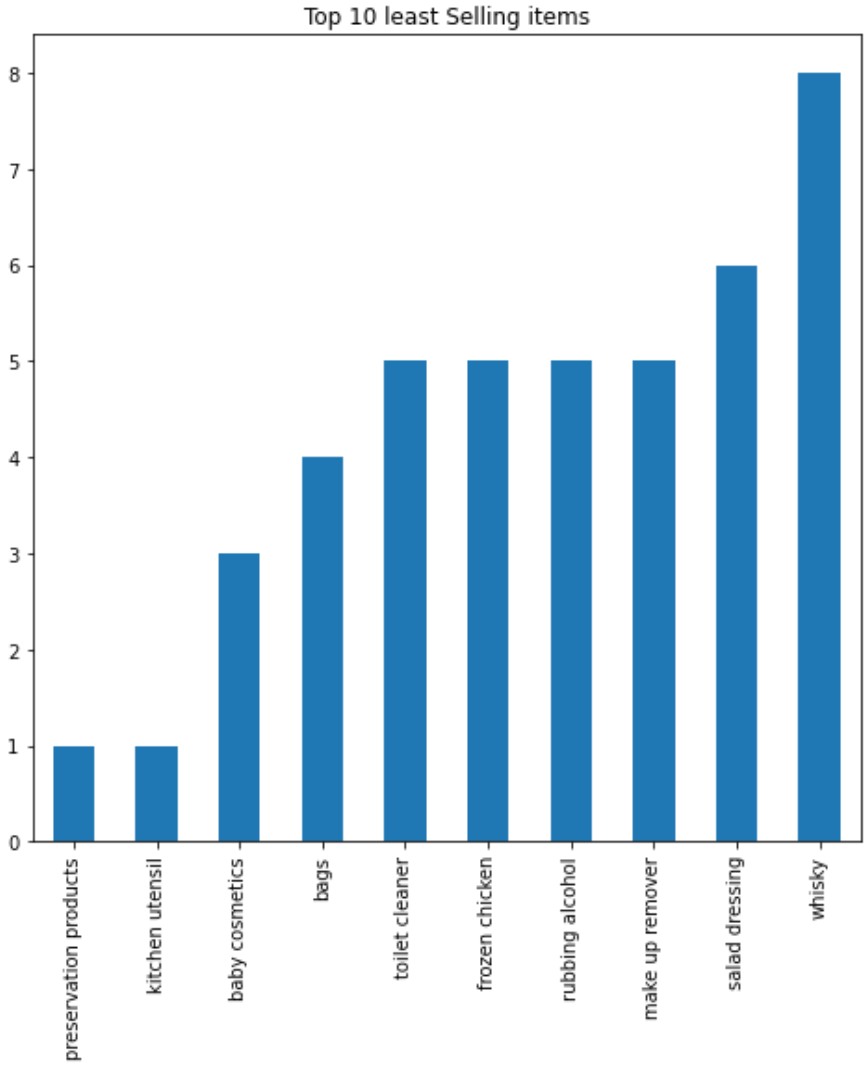
The APP will be showing a graph that would help vendors dynamically understand their sales which would help in inventory management. These are some sample questions for the prototype:

- Which of the items is the top-selling item?
- Which of the items are the 10 least selling items?
- Top 10 customers
- Year Of majority transactions?
- The number of transactions in 2015(month-wise)?

```
plt.figure(figsize=(8,8))
df.itemDescription.value_counts().head(10).plot.bar()
plt.title('Top 10 selling products')
```



```
plt.figure(figsize=(8,8))
df.itemDescription.value_counts().tail(10).sort_values().plot.bar()
plt.title('Top 10 least Selling items')
```




```
# TOP 10 CUSTOMERS
df.memberNumber.value_counts().head(10)
```

```
3180    36
3050    33
2051    33
3737    33
2433    31
3915    31
2625    31
2271    31
3872    30
2394    29
Name: memberNumber, dtype: int64
```

Year Of majority transactions

```
[15] df.Year.value_counts()
```

```
2015    20488
2014    18277
Name: Year, dtype: int64
```

Number of transactions in 2015(month wise)

```
[16] df[df.Year == 2015]['Month'].value_counts().sort_index()
```

```
1     1829
2     1485
3     1792
4     1666
5     1710
6     1791
7     1645
8     1963
9     1613
10    1663
11    1777
12    1554
Name: Month, dtype: int64
```

FEATURE ENGINEERING

```
data = df.copy()
data = pd.get_dummies(data['itemDescription'])
data1 = df.copy()

data1.drop(['itemDescription'],axis=1,inplace=True)
data1 = data1.join(data)
```

```
products = df['itemDescription'].unique()
print(f"Number of products:- {len(products)}")
```

```
Number of products:- 167
```

```python
data2 = data1.groupby(['memberNumber','Date'])[products[:]].sum()
data2 = data2.reset_index()[products]

def naming(data):
    for i in products:
        if data[i]>0:
            data[i]=i
    return data

data2 = data2.apply(naming,axis=1)

newdata = data2.values
newdata = [i[i!=0].tolist() for i in newdata if i[i!=0].tolist()]
# ARRAY OF ITEMS BOUGHT TOGETHER
newdata
```

```
[['whole milk', 'yogurt', 'sausage', 'semi-finished bread'],
 ['whole milk', 'pastry', 'salty snack'],
 ['canned beer', 'misc. beverages'],
 ['sausage', 'hygiene articles'],
 ['soda', 'pickled vegetables'],
 ['frankfurter', 'curd'],
 ['whole milk', 'rolls/buns', 'sausage'],
 ['whole milk', 'soda'],
 ['beef', 'white bread'],
 ['frankfurter', 'soda', 'whipped/sour cream'],
 ['other vegetables', 'frozen vegetables'],
 ['whole milk', 'butter'],
 ['tropical fruit', 'sugar'],
 ['butter milk', 'specialty chocolate'],
 ['rolls/buns', 'sausage'],
 ['root vegetables', 'detergent'],
 ['frozen meals', 'dental care'],
 ['rolls/buns'],
 ['dish cleaner', 'cling film/bags'],
 ['canned beer', 'frozen fish'],
 ['other vegetables', 'hygiene articles'],
 ['tropical fruit', 'pip fruit', 'whole milk'],
 ['rolls/buns', 'chocolate', 'red/blush wine'],
 ['other vegetables', 'shopping bags'],
 ['whole milk', 'rolls/buns', 'packaged fruit/vegetables', 'chocolate'],
 ['whole milk', 'root vegetables', 'pastry'],
 ['rolls/buns'],
 ['margarine', 'whipped/sour cream'],
 ['rolls/buns', 'bottled water', 'softener'],
 ['whole milk', 'rice'],
 ['shopping bags', 'skin care'],
 ['whole milk',
  'rolls/buns',
  'frankfurter',
  'chicken',
  'chocolate',
  'flour',
```

Aprori Implementation

```
from mlxtend.frequent_patterns import apriori
x = apriori(data, min_support=0.2, use_colnames=True)
```

	antecedents	consequents	antecedent support	consequent support
0	brown bread	frozen vegetables	0.054452	0.02437
1	yogurt	butter milk, canned beer	0.075762	0.0462
2	rolls/buns	frankfurter, coffee	0.046202	0.0725
3	soda	chewing gum, sausage	0.055452	0.0367
4	soft cheese	yogurt, other vegetables	0.28857	0.0412

```
from apyori import apriori
association = apriori(newdata ,min_support = 0.02 ,min_confidence = 0.01 ,max_length=3,min_lift =3)
```

Association Rules:-

```
st(association)

elationRecord(items=frozenset({'baking powder', 'mustard'}), support=0.00020049455323130388, ordered_statistics=[OrderedStatistic(items_base=frozenset({'baking powder'}), items_add=frozenset({'mustard'}), confidence
elationRecord(items=frozenset({'specialty fat', 'beef'}), support=0.00020049455323130388, ordered_statistics=[OrderedStatistic(items_base=frozenset({'beef'}), items_add=frozenset({'specialty fat'}), confidence=0.005
elationRecord(items=frozenset({'soap', 'berries'}), support=0.00020049455323130388, ordered_statistics=[OrderedStatistic(items_base=frozenset({'berries'}), items_add=frozenset({'soap'}), confidence=0.009202453987730
elationRecord(items=frozenset({'butter milk', 'zwieback'}), support=0.00026732607097507187, ordered_statistics=[OrderedStatistic(items_base=frozenset({'butter milk'}), items_add=frozenset({'zwieback'}), confidence=0
elationRecord(items=frozenset({'chewing gum', 'candles'}), support=0.00020049455323130388, ordered_statistics=[OrderedStatistic(items_base=frozenset({'candles'}), items_add=frozenset({'chewing gum'}), confidence=0.0
elationRecord(items=frozenset({'candles', 'oil'}), support=0.00020049455323130388, ordered_statistics=[OrderedStatistic(items_base=frozenset({'candles'}), items_add=frozenset({'oil'}), confidence=0.04545454545454549
elationRecord(items=frozenset({'canned beer', 'cookware'}), support=0.00020049455323130388, ordered_statistics=[OrderedStatistic(items_base=frozenset({'canned beer'}), items_add=frozenset({'cookware'}), confidence=0
elationRecord(items=frozenset({'canned fruit', 'onions'}), support=0.00020049455323130388, ordered_statistics=[OrderedStatistic(items_base=frozenset({'canned fruit'}), items_add=frozenset({'onions'}), confidence=0.1
elationRecord(items=frozenset({'chewing gum', 'photo/film'}), support=0.00026732607097507187, ordered_statistics=[OrderedStatistic(items_base=frozenset({'chewing gum'}), items_add=frozenset({'photo/film'}), confiden
elationRecord(items=frozenset({'chocolate', 'fish'}), support=0.00020049455323130388, ordered_statistics=[OrderedStatistic(items_base=frozenset({'chocolate'}), items_add=frozenset({'fish'}), confidence=0.00849858356
elationRecord(items=frozenset({'meat', 'cleaner'}), support=0.00020049455323130388, ordered_statistics=[OrderedStatistic(items_base=frozenset({'cleaner'}), items_add=frozenset({'meat'}), confidence=0.096774193548387
elationRecord(items=frozenset({'sausage', 'cocoa drinks'}), support=0.00020049455323130388, ordered_statistics=[OrderedStatistic(items_base=frozenset({'cocoa drinks'}), items_add=frozenset({'sausage'}), confidence=0
elationRecord(items=frozenset({'tidbits', 'coffee'}), support=0.00020049455323130388, ordered_statistics=[OrderedStatistic(items_base=frozenset({'coffee'}), items_add=frozenset({'tidbits'}), confidence=0.006342494713
elationRecord(items=frozenset({'liver loaf', 'condensed milk'}), support=0.00020049455323130388, ordered_statistics=[OrderedStatistic(items_base=frozenset({'condensed milk'}), items_add=frozenset({'liver loaf'}), co
elationRecord(items=frozenset({'cookware', 'oil'}), support=0.00020049455323130388, ordered_statistics=[OrderedStatistic(items_base=frozenset({'cookware'}), items_add=frozenset({'oil'}), confidence=0.176470588235294
elationRecord(items=frozenset({'newspapers', 'cream'}), support=0.00020049455323130388, ordered_statistics=[OrderedStatistic(items_base=frozenset({'cream'}), items_add=frozenset({'newspapers'}), confidence=0.25, lif
elationRecord(items=frozenset({'sausage', 'cream'}), support=0.00020049455323130388, ordered_statistics=[OrderedStatistic(items_base=frozenset({'cream'}), items_add=frozenset({'sausage'}), confidence=0.25, lift=4.14
elationRecord(items=frozenset({'nuts/prunes', 'cream cheese '}), support=0.00020049455323130388, ordered_statistics=[OrderedStatistic(items_base=frozenset({'cream cheese '}), items_add=frozenset({'nuts/prunes'}), co
elationRecord(items=frozenset({'mustard', 'curd cheese'}), support=0.00020049455323130388, ordered_statistics=[OrderedStatistic(items_base=frozenset({'curd cheese'}), items_add=frozenset({'mustard'}), confidence=0.0
elationRecord(items=frozenset({'dishes', 'detergent'}), support=0.00026732607097507187, ordered_statistics=[OrderedStatistic(items_base=frozenset({'detergent'}), items_add=frozenset({'dishes'}), confidence=0.0310077
elationRecord(items=frozenset({'detergent', 'soft cheese'}), support=0.00026732607097507187, ordered_statistics=[OrderedStatistic(items_base=frozenset({'detergent'}), items_add=frozenset({'soft cheese'}), confidence
elationRecord(items=frozenset({'dish cleaner', 'frozen meals'}), support=0.00026732607097507187, ordered_statistics=[OrderedStatistic(items_base=frozenset({'dish cleaner'}), items_add=frozenset({'frozen meals'}), co
elationRecord(items=frozenset({'hygiene articles', 'dog food'}), support=0.00020049455323130388, ordered_statistics=[OrderedStatistic(items_base=frozenset({'dog food'}), items_add=frozenset({'hygiene articles'}), co
elationRecord(items=frozenset({'frankfurter', 'ketchup'}), support=0.00026732607097507187, ordered_statistics=[OrderedStatistic(items_base=frozenset({'frankfurter'}), items_add=frozenset({'ketchup'}), confidence=0.0
elationRecord(items=frozenset({'frozen fish', 'specialty chocolate'}), support=0.0003341575887188398, ordered_statistics=[OrderedStatistic(items_base=frozenset({'frozen fish'}), items_add=frozenset({'specialty choco
elationRecord(items=frozenset({'frozen meals', 'instant coffee'}), support=0.00020049455323130388, ordered_statistics=[OrderedStatistic(items_base=frozenset({'frozen meals'}), items_add=frozenset({'instant coffee'})
elationRecord(items=frozenset({'frozen potato products', 'meat'}), support=0.00026732607097507187, ordered_statistics=[OrderedStatistic(items_base=frozenset({'frozen potato products'}), items_add=frozenset({'meat'})
elationRecord(items=frozenset({'frozen potato products', 'processed cheese'}), support=0.00020049455323130388, ordered_statistics=[OrderedStatistic(items_base=frozenset({'frozen potato products'}), items_add=frozens
```

Git Hub Code - https://github.com/vaasu2002/Market-Basket-Analysis

Conclusion:-

In this way, machine learning can be used to recognize customer purchasing patterns, which would help businesses to earn more profit and help them grow. The data analysis would also help businesses to maintain their inventory.