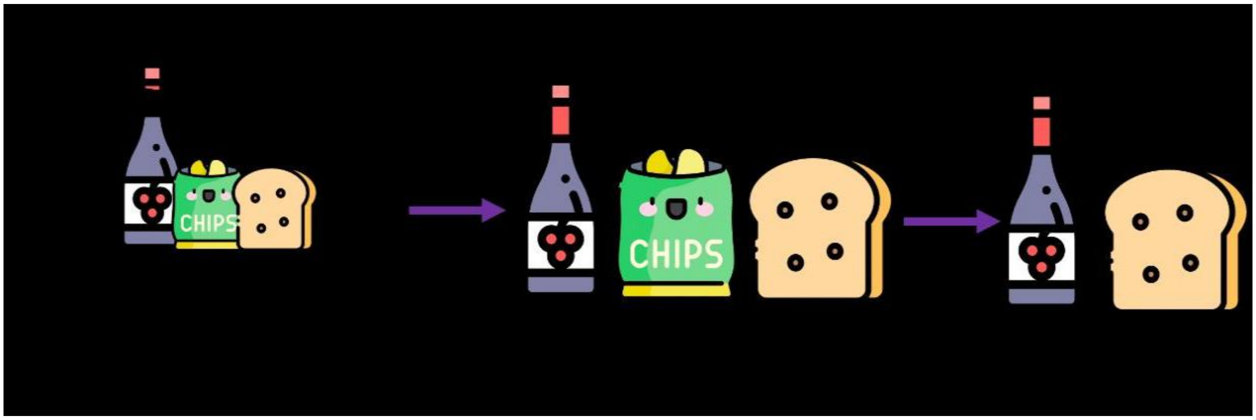
Phase 2: Innovation

Introduction to Innovation

In Phase 1, we established a comprehensive plan for conducting market basket analysis using the Apriori algorithm. Now, in Phase 2, we delve into innovation to address the problem more effectively. We'll explore cutting-edge techniques and approaches that go beyond conventional methods to unveil hidden patterns and associations in the dataset.



Innovative Techniques and Approaches

1.Data Enhancement and Enrichment

One of the core innovations in our approach involves enhancing and enriching the transaction dataset. By incorporating external data sources and cleaning the data rigorously, we aim to improve the quality of our analysis. This data enhancement is a crucial step toward gaining deeper insights.

2. Advanced Analytics and Machine Learning

While the Apriori algorithm is a robust choice for association analysis, we're taking innovation a step further by incorporating advanced analytics and machine learning. These techniques will help us uncover complex patterns that might remain hidden with traditional methods.

3. Predictive Modeling

We're not just stopping at understanding past purchasing behavior. We're innovatively applying predictive modeling to forecast future customer behavior. This will allow us to proactively optimize our strategies based on expected future patterns.

4.Real-time Analysis

In today's dynamic business environment, real-time insights are invaluable. Our approach includes real-time or near-real-time analysis, ensuring that our strategies remain adaptable to changing customer behaviors.

5. Integration of External Data Sources

To gain a holistic understanding of customer behavior, we're integrating external data sources, including demographic data and even real-time weather data. This innovative integration provides a comprehensive view of influencing factors.

6. Scalability and Efficiency

Scalability and computational efficiency are at the forefront of our innovation. We've chosen technologies and tools that ensure our approach can handle large datasets and provide efficient results.

7. Ethical Considerations

Innovation doesn't come at the cost of ethics. We're committed to upholding ethical standards in data usage, privacy, and transparency throughout our project. Our methods adhere to legal and ethical guidelines.

8.Expected Outcomes

With these innovative approaches, we anticipate the following outcomes:

- Deeper insights into customer behavior.
- Identification of cross-selling opportunities with higher accuracy.
- Timely and proactive strategies for business optimization.

9. Challenges and Mitigations

While innovation brings great potential, it also poses challenges. We're aware of potential challenges and have mitigation strategies in place to address them. These include data security, model complexity, and real-time data handling.

10.Timeline and Milestones

We've outlined a timeline for implementing our innovative solutions. Key milestones include data enhancement completion, predictive model development, and the launch of real-time analysis.

11.Resources and Budget

Our innovation requires specific resources, both in terms of hardware and software. We've allocated a budget to ensure the project's success.

```
import numpy as np # linear algebra
import pandas as pd # data processing

from apyori import apriori

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

```
/kaggle/input/association-rule-learningapriori/Market_Basket_Optimisation.csv
```

What is Market Basket?

When we go to any supermarket or shop online we try to purchase all the item we need together instead of buying each item seperately. Thus we can define Market Basket as a basket which is used to group together items of a person's interest which he/she will buy in one transaction. Each trip to the market is a single transaction, and in case of e-commerce all items bought in a single login is a transaction

Objective of Market Basket

1. Cross Selling: It is a stratergy where the seller encourages the customer to spend more money by recommending related products that complement what is being bought already by the consumer. This stratergy would encourage the customer to spend more than he/she had actually thought he/she would.
2. Product Placement: When you go to supermarket you may see that the milk items are kept together. Moreover you may see that as you move forward you find the bread making items such as flour, butter, eggs etc kept just after the milk products. This placement of the items on the supermarket shelf follows planogram. A planogram is defined as a “diagram or model that indicates the placement of retail products on shelves in order to maximize sales”.

Assosiation Rules

Association Rule Mining is primarily used when you want to identify an association between different items in a set, then find frequent patterns in a transactional database or relational databases(RDBMS). The applications of Association Rule Mining are found in Marketing, Basket Data Analysis (or Market Basket Analysis) in retailing, clustering and classification. It is most commonly used to analyze customer buying habits by finding associations between the different items that customers place in their “shopping baskets”. The discovery of these associations can help retailers develop marketing strategies by gaining insight into which items are frequently purchased together by customers. The strategies may include:

1. Changing the store layout according to trends
2. Customer behavior analysis
3. Catalog design
4. Cross marketing on online stores
5. What are the trending items customers buy
6. Customized emails with add-on sales etc.

Online retailers and publishers can use this type of analysis to:

1. Inform the placement of content items on their media sites, or products in their catalog
2. Deliver targeted marketing

Advantages of Market Basket Analysis

Market basket Analysis(MBA) is applied to data of customers from the point of sale (PoS) systems.

Advantages for retailers:

1. Increases customer engagement
2. Boosting sales and increasing RoI
3. Optimize marketing strategies and campaigns
4. Identifies customer behavior and pattern

Advantages to customers:

1. In online retail it helps customer by displaying more products related to intended purchase.

Algorithms used in Market Basket Analysis

For each algorithm the important objective is to predict the probability of items that are being bought together by customers. Following are some of the algorithms used for analysis:

1. AIS
2. SETM Algorithm
3. Apriori Algorithm
4. FP Growth

In this article we will focus on **Apriori Algorithm** which is currently the most popular algorithm.

Apriori Algorithm

Objective: It helps to find frequent itemsets in transactions and identifies association rules between these items.

Advantage:

1. It is also considered accurate and overtop AIS and SETM algorithms.
2. It is easy to implement and interpret
3. It can be used on large datasets and can easily be parallelized.

Disadvantage:

1. Calculating support is expensive as it has to go through the entire dataset
2. It is computationally expensive

Key Metrics: It uses the concept of Confidence, Support and Lift.

1. Support - Support of an item or items is the frequency of that or those item(s) appearing out of the total transactions.

   Support = (shampoo + conditioner)/total different baskets = 8/10 = .8

   Both products are bought together in 80% of the transactions. This means that more the value for the support more is the chance of the items purchased together.

2. Confidence - This is the likelihood of an item being purchased given another item is purchased. In other words, it determines how often that the products are purchased together.

   Confidence = (shampoo + conditioner) / shampoo = 5/6 = .83

   This means that 83% of the time shampoo is purchased by the customer, conditioner is also purchased.

3. Lift- It is a metric that tells us by what value the probability of buying Y is leveraged or increased when X is purchased.

4. a. Lift = 1 → There’s no relation between the purchase of the products.

5. b. Lift > 1 → The products are likely to be bought together. Higher the lift, the higher the chances.

   c. Lift < 1 → The products are unlikely to be bought together. They are substitutes.

In [3]:

```python
path = "/kaggle/input/association-rule-learningapriori/Market_Basket_Optimisation.csv"
df = pd.read_csv(path, header = None)
df.head()
```

Out[3]:

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	shrimp	almonds	avocado	vegetables mix	green grapes	whole weat flour	yams	cottage cheese	energy drink	tomato juice	low fat yogurt	green tea	honey	salad	mineral water	salmon	antioxydant juice	frozen smoothie	spinach	olive oil
1	burgers	meatballs	eggs	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	chutney	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
3	turkey	avocado	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	mineral water	milk	energy bar	whole wheat rice	green tea	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

The dataset doesnot have any header. Thus we have used header = None. Otherwise, the first row will be dispalyed as a header row.
Let us check the rows, columns in the dataset along with the null values and type of columns.

In [4]:

```
# number of rows, columns
df.shape
```

Out[4]:

```
(7501, 20)
```

In [5]:

```
# number of null values in each column
df.isnull().sum()
```

Out[5]:

```
0        0
1     1754
2     3112
3     4156
4     4972
5     5637
6     6132
7     6520
8     6847
9     7106
```

```
10    7245
11    7347
12    7414
13    7454
14    7476
15    7493
16    7497
17    7497
18    7498
19    7500
dtype: int64
```

In [6]:

```python
# type of column
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7501 entries, 0 to 7500
Data columns (total 20 columns):
 #   Column  Non-Null Count  Dtype 
---  ------  --------------  ----- 
 0   0       7501 non-null   object
 1   1       5747 non-null   object
 2   2       4389 non-null   object
 3   3       3345 non-null   object
 4   4       2529 non-null   object
 5   5       1864 non-null   object
 6   6       1369 non-null   object
 7   7       981 non-null    object
 8   8       654 non-null    object
 9   9       395 non-null    object
 10  10      256 non-null    object
 11  11      154 non-null    object
 12  12      87 non-null     object
 13  13      47 non-null     object
 14  14      25 non-null     object
 15  15      8 non-null      object
 16  16      4 non-null      object
 17  17      4 non-null      object
 18  18      3 non-null      object
 19  19      1 non-null      object
dtypes: object(20)
memory usage: 1.1+ MB
```

In [7]:

```
df.head()
```

Out[7]:

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	shrimp	almonds	avocado	vegetables mix	green grapes	whole weat flour	yams	cottage cheese	energy drink	tomato juice	low fat yogurt	green tea	honey	salad	mineral water	salmon	antioxydant juice	frozen smoothie	spinach	olive oil
1	burgers	meatballs	eggs	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	chutney	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	turkey	avocado	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	mineral water	milk	energy bar	whole wheat rice	green tea	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

In [8]:

```
#converting dataframe into list of lists
dataframe_list=[]

# values of range is according to the number of rows and columns
for i in range(1,7501):
    dataframe_list.append([str(df.values[i,j]) for j in range(0,20)])
```

In [9]:

```
#applying apriori algorithm
association_rules = apriori(dataframe_list, min_support=0.0045, min_confiden
ce=0.2,
                           min_lift=3, min_length=2)
association_results = list(association_rules)
```

In [10]:

```
association_results[0]
```

Out[10]:

```
RelationRecord(items=frozenset({'chicken', 'light cream'}), support=0.0045
33333333333334, ordered_statistics=[OrderedStatistic(items_base=frozenset(
{'light cream'}), items_add=frozenset({'chicken'}), confidence=0.290598290
5982906, lift=4.843304843304844)])
```

Inferences from the above

1. Support: 0.0045.
   The support is calculated = (chicken + light cream)/(total transactions)

2. Confidence: 0.2905
   Confidence = (chicken + light cream)/(light cream)
   The confidence level for the rule is 0.2905, which shows that out of all the transactions that contain light cream, 29.05% percent contain chicken too.

3. Lift: 4.84
   Lift = ((chicken + light cream)/light cream)/((chicken)/total)
   The lift of 4.8433 tells us that chicken is 4.8433 times more likely to be bought by the customers who buy light cream compared to the default likelihood sale of chicken.

Create a dataframe

In [11]:

```python
Rule = []
Support = []
Confidence = []
Lift = []
for item in association_results:
    pair = item[0]
    items = [x for x in pair]
    Rule.append(items)
    Support.append(str(item[1]))
    Confidence.append(str(item[2][0][2]))
    Lift.append(str(item[2][0][3]))
```

In [12]:

```python
dict = {'Rule': Rule, 'Support': Support, 'Confidence': Confidence,
                                                          "Lift": Lift}
```

```python
d = pd.DataFrame(dict)
d
```

Out[12]:

	Rule	Support	Confidence	Lift
0	[chicken, light cream]	0.004533333333333334	0.2905982905982906	4.843304843304844
1	[escalope, mushroom cream sauce]	0.005733333333333333	0.30069930069930073	3.7903273197390845
2	[escalope, pasta]	0.005866666666666667	0.37288135593220345	4.700185158809287
3	[herb & pepper, ground beef]	0.016	0.3234501347708895	3.2915549671393096
4	[ground beef, tomato sauce]	0.005333333333333333	0.37735849056603776	3.840147461662528

	Rule	Support	Confidence	Lift
5	[whole wheat pasta, olive oil]	0.008	0.2714932126696833	4.130221288078346
6	[pasta, shrimp]	0.005066666666666666	0.3220338983050848	4.514493901473151
7	[chicken, light cream, nan]	0.004533333333333334	0.2905982905982906	4.843304843304844
8	[chocolate, frozen vegetables, shrimp]	0.005333333333333333	0.23255813953488372	3.260160834601174
9	[ground beef, spaghetti, cooking oil]	0.0048	0.5714285714285714	3.281557646029315
10	[escalope, nan, mushroom cream sauce]	0.005733333333333333	0.30069930069930073	3.7903273197390845
11	[escalope, pasta, nan]	0.005866666666666667	0.37288135593220345	4.700185158809287
12	[frozen vegetables, spaghetti, ground beef]	0.008666666666666666	0.3110047846889952	3.164906221394116
13	[frozen vegetables, olive oil, milk]	0.0048	0.20338983050847456	3.094165778526489
14	[frozen vegetables, shrimp, mineral water]	0.0072	0.3068181818181818	3.2183725365543547
15	[frozen vegetables, spaghetti, olive oil]	0.005733333333333333	0.20574162679425836	3.1299436124887174

	Rule	Support	Confidence	Lift
16	[frozen vegetables, shrimp, spaghetti]	0.006	0.21531100478468898	3.0183785717479763
17	[frozen vegetables, spaghetti, tomatoes]	0.006666666666666667	0.23923444976076555	3.497579674864993
18	[spaghetti, ground beef, grated cheese]	0.005333333333333333	0.3225806451612903	3.282706701098612
19	[herb & pepper, ground beef, mineral water]	0.006666666666666667	0.390625	3.975152645861601
20	[herb & pepper, nan, ground beef]	0.016	0.3234501347708895	3.2915549671393096
21	[herb & pepper, spaghetti, ground beef]	0.0064	0.3934426229508197	4.003825878061259
22	[ground beef, olive oil, milk]	0.004933333333333333	0.22424242424242424	3.411395906324912
23	[nan, ground beef, tomato sauce]	0.005333333333333333	0.37735849056603776	3.840147461662528
24	[spaghetti, shrimp, ground beef]	0.006	0.5232558139534884	3.004914704939635
25	[spaghetti, olive oil, milk]	0.0072	0.20300751879699247	3.0883496774390333
26	[soup, mineral water, olive oil]	0.0052	0.2254335260115607	3.4295161157945335

	Rule	Support	Confidence	Lift
27	[whole wheat pasta, nan, olive oil]	0.008	0.2714932126696833	4.130221288078346
28	[shrimp, pasta, nan]	0.005066666666666666	0.3220338983050848	4.514493901473151
29	[pancakes, spaghetti, olive oil]	0.005066666666666666	0.20105820105820105	3.0586947422647217
30	[chocolate, shrimp, frozen vegetables, nan]	0.005333333333333333	0.23255813953488372	3.260160834601174
31	[ground beef, spaghetti, nan, cooking oil]	0.0048	0.5714285714285714	3.281557646029315
32	[frozen vegetables, nan, ground beef, spaghetti]	0.008666666666666666	0.3110047846889952	3.164906221394116
33	[frozen vegetables, spaghetti, mineral water, ...	0.004533333333333334	0.28813559322033905	3.0224013274860737
34	[frozen vegetables, nan, olive oil, milk]	0.0048	0.20338983050847456	3.094165778526489
35	[shrimp, frozen vegetables, nan, mineral water]	0.0072	0.3068181818181818	3.2183725365543547
36	[frozen vegetables, nan, olive oil, spaghetti]	0.005733333333333333	0.20574162679425836	3.1299436124887174

	Rule	Support	Confidence	Lift
37	[shrimp, frozen vegetables, nan, spaghetti]	0.006	0.21531100478468898	3.0183785717479763
38	[tomatoes, frozen vegetables, nan, spaghetti]	0.006666666666666667	0.23923444976076555	3.497579674864993
39	[spaghetti, nan, ground beef, grated cheese]	0.005333333333333333	0.3225806451612903	3.282706701098612
40	[herb & pepper, nan, ground beef, mineral water]	0.006666666666666667	0.390625	3.975152645861601
41	[herb & pepper, nan, ground beef, spaghetti]	0.0064	0.3934426229508197	4.003825878061259
42	[nan, ground beef, olive oil, milk]	0.004933333333333333	0.22424242424242424	3.411395906324912
43	[shrimp, spaghetti, nan, ground beef]	0.006	0.5232558139534884	3.004914704939635
44	[spaghetti, nan, olive oil, milk]	0.0072	0.20300751879699247	3.0883496774390333
45	[soup, nan, mineral water, olive oil]	0.0052	0.2254335260115607	3.4295161157945335
46	[pancakes, nan, olive oil, spaghetti]	0.005066666666666666	0.20105820105820105	3.0586947422647217

	Rule	Support	Confidence	Lift
47	[frozen vegetables, spaghetti, nan, milk, mine...	0.004533333333333334	0.28813559322033905	3.0224013274860737

Conclusions

1. In this tutorial we went through the market basket analysis using Apriori Algorithm.
2. Study the key metrics used in algorithm
3. Check the assosiation rules and infered the values we got

12.Conclusion

In Phase 2, we're taking our market basket analysis to the next level with innovation. By incorporating advanced analytics, real-time analysis, and ethical considerations, we're poised to generate more valuable insights for the retail business. Our commitment to scalability and efficiency ensures that our innovative approach is not only cutting-edge but also practical.