PROBLEM DEFINITION

- In retail and e-commerce industries, the Market Basket Insights problem is a common data analysis and machine learning challenge. It involves analyzing customer purchase data to discover patterns, associations, and insights that can be used to improve various aspects of business operations, including marketing, inventory management, and customer experience. Here is a more detailed problem definition:
- **Problem:** Given a dataset of customer transactions, where each transaction consists of a list of items purchased by a customer, the Market Basket Insights problem aims to uncover meaningful relationships, associations, and patterns among these items. The **goal** is to gain insights into customer behavior and preferences, as well as to identify opportunities for business improvement.

OBJECTIVES

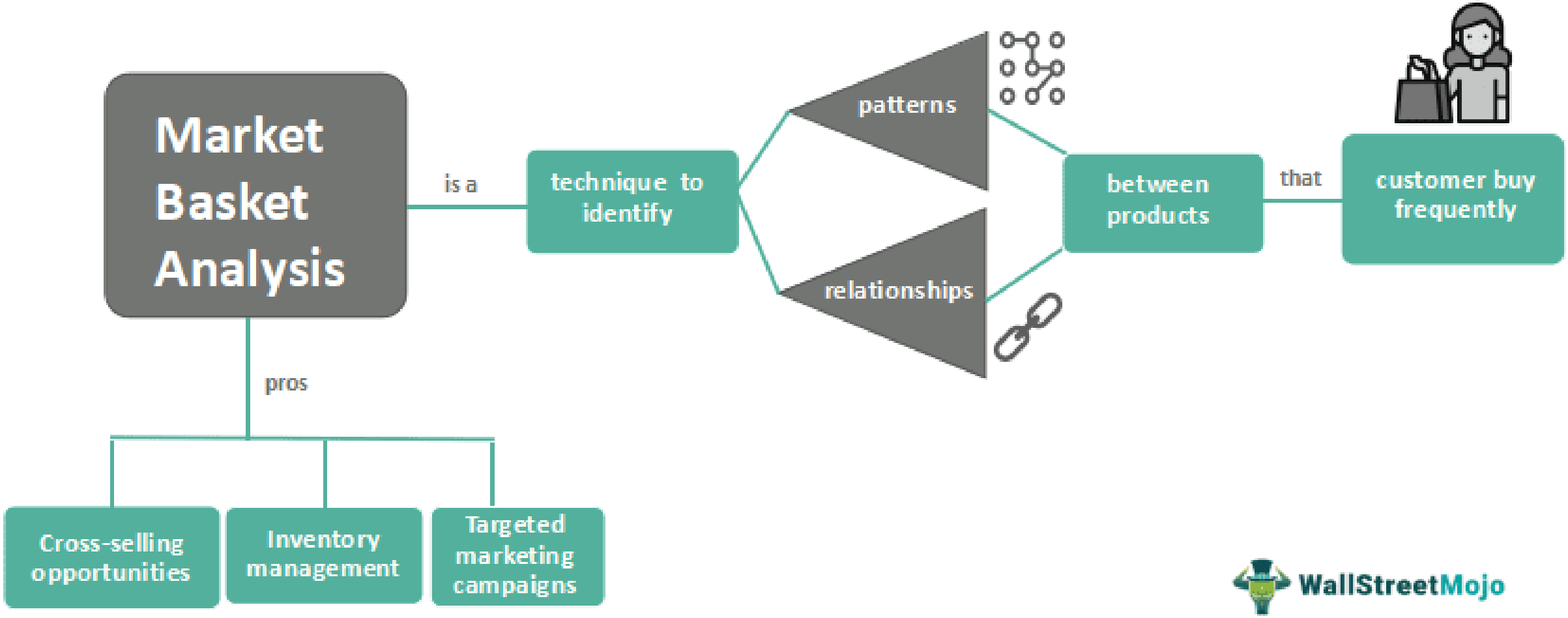
The objectives of the Market Basket Insights are given below:

1. **Association Rule Mining:** Discover frequent itemsets and generate association rules that reveal which items are often purchased together. Association rules typically consist of an antecedent (items in the basket) and a consequent (item likely to be purchased next).
2. **Recommendation:** Use association rules and collaborative filtering techniques to make product recommendations to customers based on their purchase history and the purchasing behavior of similar customers.
3. **Market Basket Analysis:** Analyze the performance of product bundles, discounts, and promotions by examining how items are grouped in customer transactions and how they influence each other's sales.
4. **Inventory Management:** Optimize inventory levels by identifying items that are frequently purchased together or exhibit seasonal trends. This can help reduce stockouts and overstock situations.
5. **Customer Segmentation:** Segment customers based on their purchase patterns, allowing for targeted marketing campaigns and personalized offers.

INTEGRATION APPROACH

To integrate Market Basket Insights, start by collecting transaction data. Preprocess and clean the data, then apply association rule mining or machine learning algorithms. Analyze the results to identify item relationships and customer preferences. Implement actionable strategies for marketing, inventory management, and enhancing the overall customer experience.

Market Basket Analysis



WallStreetMojo

PROCESS OF MARKET BASKET INSIGHTS

- Data Collection.
- Data Preprocessing
- Association Rule Mining
- Rule Evaluation
- Interpretation and Action

DATA COLLECTION

- It involves gathering the transaction data necessary to analyze customer purchasing behavior and discover meaningful patterns and associations between items.
- The first step is to identify the sources of data that contain information about customer transactions.
- Once the data sources are identified, the relevant transaction data must be extracted.If the data comes from multiple sources, it may need to be integrated into a single dataset.
- The cleaned and transformed data is typically stored in a suitable format or database for analysis.Transaction data is often continually collected and updated.
- Once the transaction data has been collected, cleaned, and prepared, it is ready for the next steps in market basket analysis, including association rule mining, rule evaluation, interpretation, and action planning.
- This analysis helps businesses gain valuable insights into customer behavior and make data-driven decisions to improve their operations and customer satisfaction.

DATA PREPROCESSING

- It involves cleaning and transforming raw transaction data to make it suitable for analysis. Duplicate records, if present in the data, need to be identified and removed.
- If your data comes from multiple sources or systems, integrate it into a single dataset. Ensure that data from different sources is compatible and that unique product and customer identifiers are mapped correctly.
- In cases where there are a large number of items or product categories, dimensionality reduction techniques can be applied to simplify the data without losing important information.
- In some analyses, you may choose to split the dataset into training and testing subsets to evaluate model performance. This is common in machine learning-based market basket analysis.
- Maintain thorough documentation of all the preprocessing steps performed on the data. Documentation helps ensure transparency and reproducibility of the analysis.

ASSOCIATION RULE MINING

- It is employed to discover interesting and meaningful patterns and associations between items that are frequently purchased together in customer transactions.
- Association rule mining requires a dataset of customer transactions.Each transaction consists of a list of items that were purchased together, along with transaction identifiers and timestamps.
- Association rule mining algorithms generate a large number of candidate rules based on the transaction data and user-defined thresholds for support and confidence. Once the association rules have been generated, businesses can interpret them to gain insights into customer behavior.
- Association rule mining is a valuable tool in market basket insights because it helps businesses uncover hidden patterns in customer transaction data, enabling them to make data-driven decisions that can enhance customer satisfaction, increase revenue, and improve overall business operations.

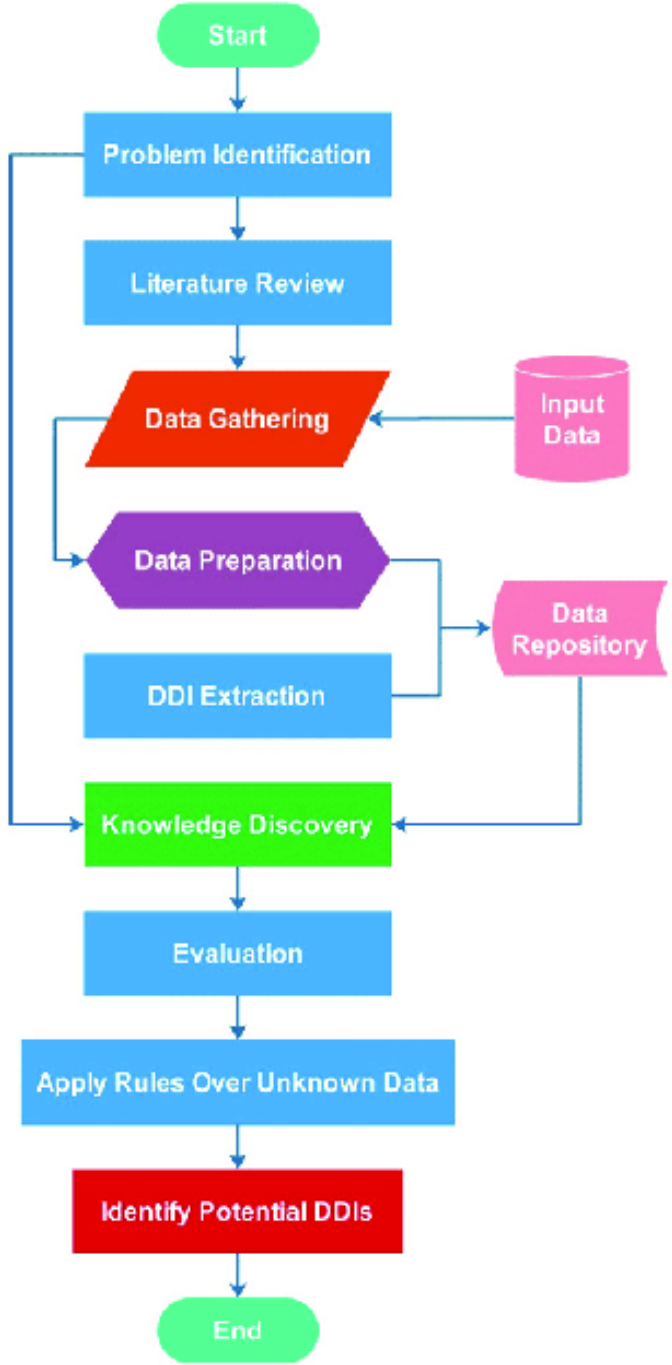
RULE EVALUATION

- Rule evaluation in market basket insights is the process of assessing and quantifying the quality and significance of association rules discovered through association rule mining.
- Users typically set minimum support and confidence thresholds to filter out rules that do not meet their desired level of significance.
- Visual representations, such as scatter plots, heatmaps, or network diagrams, can be used to visualize the relationships between items and the strength of association rules. These visualizations make it easier for analysts to interpret the findings.
- Evaluate whether the discovered rules can be translated into actionable business strategies. Some rules may have high support and confidence but may not be practical to implement due to various constraints.
- After implementing rules, businesses should monitor their impact and gather feedback to continuously refine the association rule mining process and improve the quality of rules generated.

INTERPRETATION AND ACTION

- Interpretation involves making sense of the association rules that have been generated through market basket analysis.
- You may need to segment your customer base based on the discovered patterns.
- Verify the validity of the insights by discussing them with domain experts or stakeholders in your organization.
- They can provide additional context and help assess the practicality of implementing the findings.
- Continuously monitor the results of your actions and gather feedback from customers and employees
- Integrate market basket insights with other data sources, such as customer demographics or purchase history, to create a more comprehensive understanding of your customers and their preferences.
- Keep records of the actions taken and their outcomes. This documentation helps in assessing the long-term impact of market basket insights on your business.

Flowchart


Start
Problem Identification
Literature Review
Data Gathering
Input Data
Data Preparation
Data Repository
DDI Extraction
Knowledge Discovery
Evaluation
Apply Rules Over Unknown Data
Identify Potential DDIs
End

CODE:

```
import pandas as pd
from mlxtend.frequent_patterns
import apriori
from mlxtend.frequent_patterns
import association_rules

# Sample transaction dataset
data = {'TransactionID‘ : [1, 2, 3, 4, 5] , 'Items‘ : [['apple', 'banana', 'chocolate'] , ['banana', 'chocolate'],
['apple', 'banana', 'chocolate', 'diapers'] , ['apple', 'diapers‘] , ['apple', 'chocolate', 'ice cream']]}
df = pd.DataFrame(data)

# Convert items into one-hot encoded format
df_encoded = pd.get_dummies(pd.DataFrame(df['Items'].values.tolist()).stack()).sum(level=0)

# Apply Apriori algorithm to find
frequent itemsets
min_support = 0.4
frequent_itemsets = apriori(df_encoded, min_support=min_support, use_colnames=True)

# Generate association rules
min_confidence = 0.7
rules = association_rules(frequent_itemsets, metric="confidence", min_threshold=min_confidence)

# Display frequent itemsets and association rules
print("Frequent Itemsets:")
print(frequent_itemsets)
print("\nAssociation Rules:")print(rules)
```

OUTPUT

- **Frequent Itemsets:**

	support	itemsets
0	0.6	(apple)
1	0.6	(chocolate)
2	0.4	(banana)
3	0.4	(diapers)
4	0.4	(apple, chocolate)

- **Association Rules**:

antecedents consequents
antecedent support consequent support

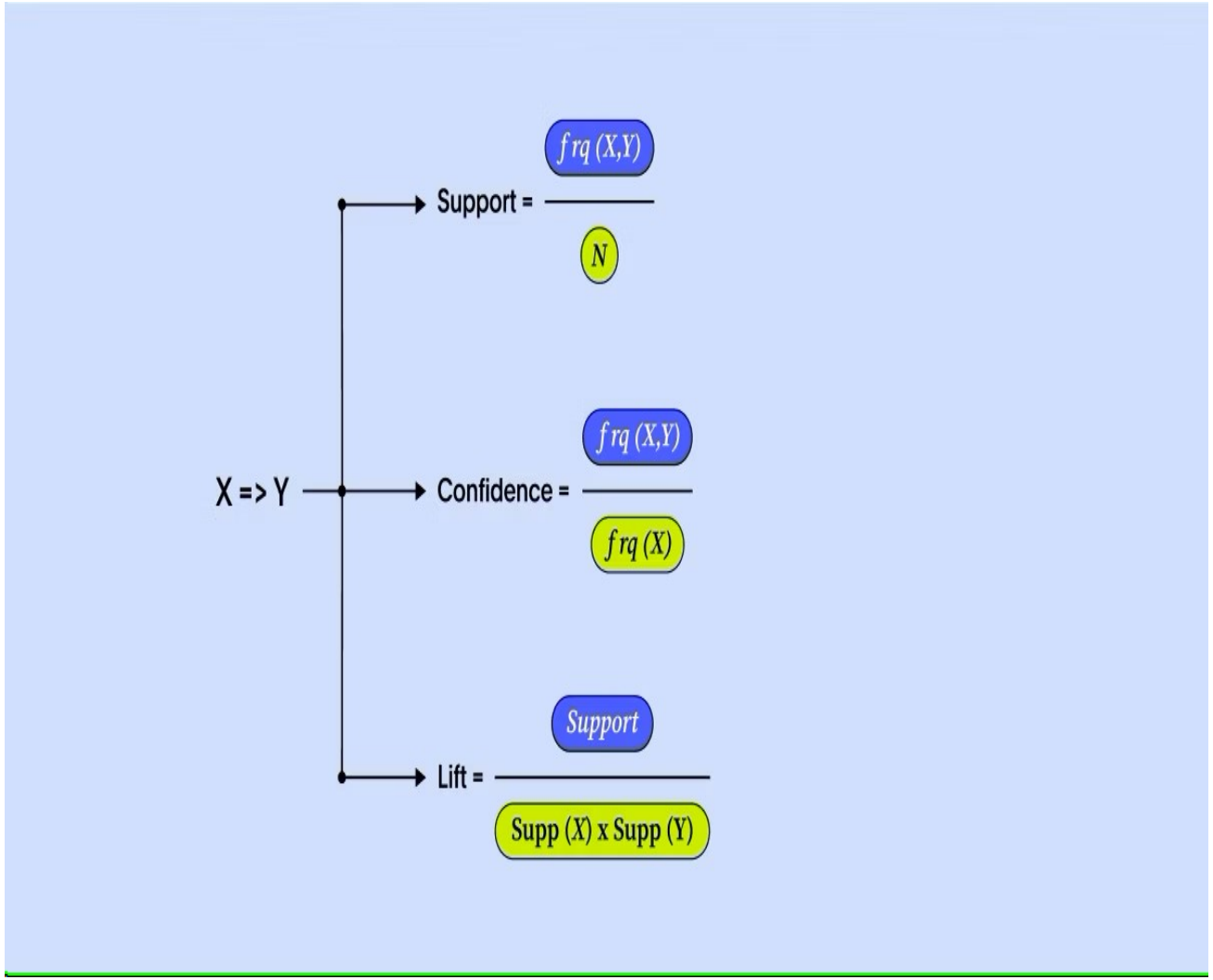
	support	confidence	lift	leverage	conviction
0	(apple)	(chocolate)	0.6	0.6	0.4	0.666667	1.111111	0.04	1.2
1	(chocolate)	(apple)	0.6	0.6	0.4	0.666667	1.111111	0.04	1.2

Concepts

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

1.SUPPORT
2.CONFIDENCE
3.LIFT

DIAGRAM



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**

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 ***apyori*** 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)
```

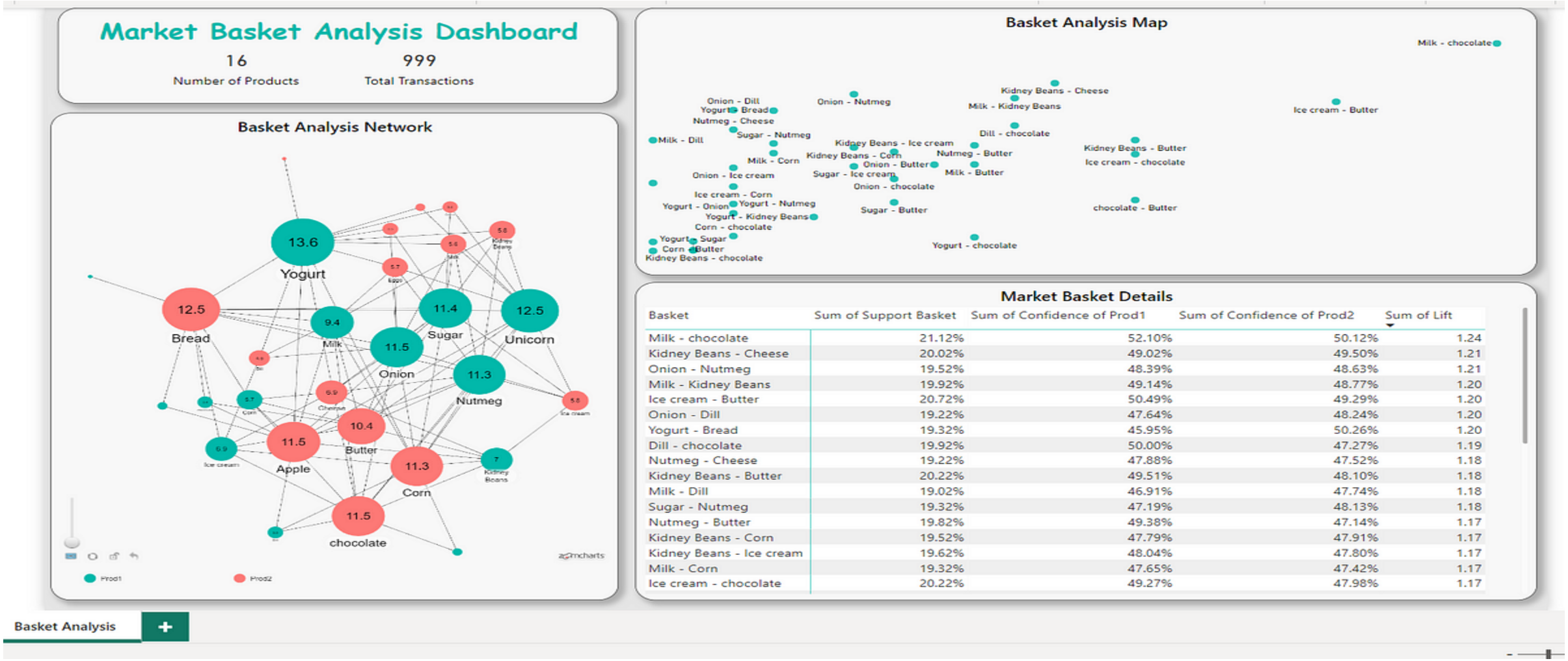
Now we want to read the dataset that is downloaded from Kaggle. There is no header in the dataset; hence, the first row contains the first transaction, so we have mentioned *header = None* here.

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



APPLICATIONS

- Telecommunication
- Bioinformatics
- Affinity promotion
- Fraud detection
- Cross selling

BENEFITS

- Cost effective and Flexible
- Store Layout
- Recommend products based on customer purchase patterns
- Identifies sales influencers

CONCLUSION

Market Basket Insights represent a pivotal tool in the arsenal of modern businesses, particularly those in the retail and e-commerce sectors. By leveraging data analytics, machine learning, and real-time processing, businesses can gain a competitive edge in a rapidly evolving marketplace. The ability to extract actionable intelligence from customer transaction data is a game-changer, offering a multitude of benefits ranging from enhanced customer experiences and increased sales to more efficient inventory management and targeted marketing. Market Basket Insights not only illuminate customer preferences and behaviors but also empower organizations to make data-driven decisions, adapt to changing market dynamics, and ultimately thrive in an era where understanding and meeting customer needs is paramount.