MARKET BASKET ANALYSIS

TEAM MEMEBERS

822421106049: PRIYADARSHINI V

PHASE-1 DOCUMENT SUBMISSION

PROJECT: market basket analysis

OBJECTIVE :

The objective of the project involves using association analysis techniques, such as Apriori algorithm, to find frequently co-occurring products and generate insights for business optimization.

Phase 1: Association analysis techniques, such as Apriori algorithm.

1.Data source : A good data source for the masket basket analysis should be accurate dataset containing transaction data, including lists of purchased products.

Dataset link=(https://www.kaggle.com/input/masket-basket-analysis/assignment-1_data.xlsx)

BillNo	Itemname	Quantity	Date	Price	CustomerID	Country	
0	536365	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	536365	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
2	536365	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom
3	536365	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
4	536365	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
...	...	...	...	...	...	...	...
195	536389	CHRISTMAS LIGHTS 10 REINDEER	6	2010-12-01 10:03:00	8.50	12431.0	Australia
196	536389	VINTAGE UNION JACK CUSHION COVER	8	2010-12-01 10:03:00	4.95	12431.0	Australia
197	536389	VINTAGE HEADS AND TAILS CARD GAME	12	2010-12-01 10:03:00	1.25	12431.0	Australia

BillNo	Itemname	Quantity	Date	Price	CustomerID	Country	
198	536389	SET OF 3 COLOURED FLYING DUCKS	6	2010-12-01 10:03:00	5.45	12431.0	Australia
199	536389	SET OF 3 GOLD FLYING DUCKS	4	2010-12-01 10:03:00	6.35	12431.0	Australia

2.Data pre-processing

Before performing association rule analysis, it is necessary to pre-process the data. This involves data cleaning, transformation, and formatting to ensure that the data is in a suitable format for analysis.

Data pre-processing steps may include:

- **Removing duplicate** or irrelevant data
- **Handling missing** or incomplete data
- Converting **data to a suitable forma**t (e.g., binary or numerical)
- Discretizing **continuous variables into categorical** variables
- Scaling or **normalizing** data

3.association analysis

Apriori Algorithm

One of the most popular association rule mining algorithms is the Apriori algorithm. The Apriori algorithm is based on the concept of **frequent itemsets**, which are sets of items that occur together frequently in a dataset.

The algorithm works by first **identifying all the frequent itemsets** in a dataset, and then generating **association rules** from those itemsets.

4. Insights Generation

Association rule analysis generates a large number of **potential rules**, and it is important to evaluate and select the most relevant rules.

The following measures are commonly used to evaluate association rules:

- **Support:**
  - Rules with high support are more significant as they occur more frequently in the dataset
- **Confidence:**
  - Rules with high confidence are more reliable, as they have a higher probability of being true
- **Lift:**
  - Rules with high lift indicate a strong association between the antecedent and consequent, as they occur together more frequently than expected by chance

Conclution:

Market basket analysis is a powerful data mining technique that can be used to gain valuable insights into customer behavior. By identifying patterns in customer purchase data, businesses can use market basket analysis to:

- Optimize their product placement and store layout

- Develop targeted marketing campaigns and promotions
- Create new product bundles and offerings
- Up-sell and cross-sell products to customers
- Improve customer satisfaction and loyalty