UNIT-4

FREQUENT ITEM SETS AND CLUSTERING

SYLLABUS

- **Frequent Itemsets** :
  - Mining frequent itemsets,
  - market based modelling,
  - Apriori algorithm,
  - handling large data sets in main memory,
  - limited pass algorithm,
  - counting frequent itemsets in a stream,
- **Clustering techniques**: hierarchical, K-means, clustering high dimensional data, CLIQUE and ProCLUS, frequent pattern based clustering methods, clustering in non-euclidean space, clustering for streams and parallelism

Example (Problem Statement)

- You are a data scientist and you get a client who runs a retail store.
- Your client gives you data for all transactions that consists of items bought in the store by several customers over a period of time and asks you to use that data to help boost their business.
- Your client will use your findings to not only change/update/add items in inventory but also use them to change the layout of the physical store or rather an online store.

Solution : (**if this then that**)

- To find results that will help your client, you will use **Market Basket Analysis (MBA)** which uses **Association Rule Mining** on the given transaction data.
- Association Rule Mining is used when you want to find an association between different objects in a set, find **frequent patterns** in a transaction database, relational databases or any other information repository.
- The applications of Association Rule Mining are found in Marketing, Basket Data Analysis (or Market Basket Analysis) in retailing, clustering and classification.
- It can tell you what items do customers frequently buy together by generating a set of rules called **Association Rules**.

- Data Analyst/client can use those rules for numerous marketing strategies:
- Changing the store layout according to trends
- Customer behavior analysis
- Catalogue design
- Cross marketing on online stores
- What are the trending items customers buy
- Customized emails with add-on sales

Frequent patterns

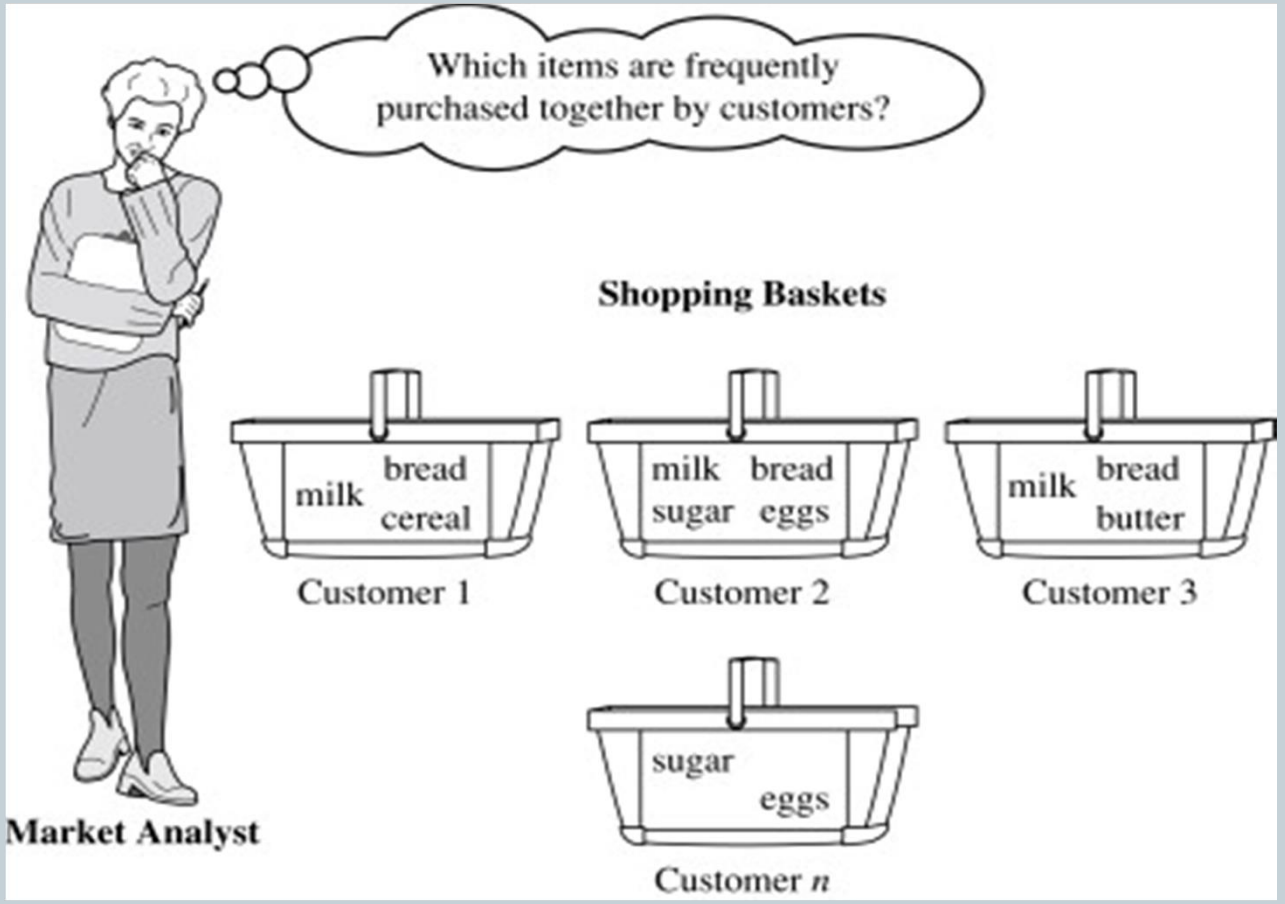
- Frequent patterns are patterns (e.g., itemsets, subsequences, or substructures) that appear frequently in a data set.
- For example, a set of items, such as milk and bread, that appear frequently together in a transaction data set which is a frequent itemset.
- A subsequence, such as buying first a PC, then a digital camera, and then a memory card, if it occurs frequently in a shopping history database, is a (frequent) sequential pattern

Frequent patterns

- A substructure can refer to different structural forms, such as subgraphs, subtrees, or sublattices, which may be combined with itemsets or subsequences.
- If a substructure occurs frequently, it is called a (frequent) structured pattern.
- Finding frequent patterns plays an essential role in mining associations, correlations, and many other interesting relationships among data.

- Moreover, it helps in data classification, clustering, and other data mining tasks.
- Thus, frequent pattern mining has become an important data mining task and a focused theme in data research.

What can we define from this picture??



Market Basket Analysis: A Motivating Example

- Frequent itemset mining leads to the discovery of associations and correlations among items in large transactional or relational data sets.
- With massive amounts of data continuously being collected and stored, many industries are becoming interested in mining such patterns from their databases.
- The discovery of interesting correlation relationships among huge amounts of business transaction records can help in many business decision-making processes such as catalog design, cross-marketing, and customer shopping behavior analysis

Market Basket Analysis: A Motivating Example

- This process analyzes 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.
- For instance, if customers are buying milk, how likely are they to also buy bread (and what kind of bread) on the same trip to the supermarket?
- This information can lead to increased sales by helping retailers do selective marketing and plan their shelf space.

Market Basket Analysis: A Motivating Example

- If we think of the universe as the set of items available at the store, then each item has a Boolean variable representing the presence or absence of that item.
- Each basket can then be represented by a Boolean vector of values assigned to these variables.
- The Boolean vectors can be analyzed for buying patterns that reflect items that are frequently associated or purchased together.
- These patterns can be represented in the form of association rules.

Summary of Market Basket Analysis

Market Basket Analysis (MBA), also known as affinity analysis, is a technique to identify items likely to be purchased together.

The analysis can be applied in various ways:

- Develop combo offers based on products sold together

- Organize and place associated products/categories nearby inside a store

- Determine the layout of the catalog of an ecommerce site

- Control inventory based on product demands and what products sell together

Summary of Market Basket Analysis

In order to gain better insights, differentiate Market Basket Analysis based on

- weekend vs weekday sales
- month beginning vs month-end sales
- different seasons of the year
- different stores
- different customer profiles

Summary of Market Basket Analysis

Although Market Basket Analysis reminds pictures of shopping carts and supermarket shoppers, there are many other areas in which it can be applied. These include:

For a financial services company :

- Analysis of credit and debit card purchases.

- Analysis of cheque payments made.

For a telecom operator

- Analysis of telephone calling patterns.