# Data Mining Lab Assignment-09
## CSE 6th semester
## (Language/Platform: Python)

**Scenario 1:**

***Question:***

**You are tasked with helping a retail store improve its marketing strategy by analyzing customers' purchase patterns. The store has a dataset containing information about customers' transactions, including the items they purchased. Your goal is to identify association rules that could reveal insights into which items are frequently purchased together.**

***Dataset Information:***

- Transactions.csv: A dataset that contains the following columns:
  - Transaction ID: Unique identifier for each transaction.
  - Items: A list of items purchased in each transaction, stored as a comma-separated string (e.g., "bread, butter, milk").

***Task:***

1. **Data Preparation:**
   - Load the dataset into a data structure suitable for association rule mining (e.g., pandas DataFrame).
   - Preprocess the data to prepare it for mining. Ensure that each transaction is represented by a list of items (for example, using a one-hot encoding or transactional data format).

2. **Association Rule Mining:**

   - Apply the Apriori Algorithm to mine frequent itemsets from the dataset. Choose appropriate support and confidence thresholds to ensure that the rules you generate are meaningful.
   - From the frequent itemsets, generate association rules with a given minimum confidence level.

3. **Interpretation:**

   - For each rule generated, interpret the results and explain their potential relevance to the retail business.
   - For example, if the rule is "If a customer buys milk, they are likely to buy bread," explain how the store could use this information in their marketing strategy.

4. **Evaluation:**

- Analyze the rules using metrics such as Lift, Support, and Confidence. Discuss how these metrics can help prioritize the most useful association rules for the retail store.

**Scenario 2: Analyzing Website Clickstream Data for E-commerce Website**

***Question:***

**You are tasked with analyzing the clickstream data from an e-commerce website. The data includes user interactions with the site, such as the pages they visited and the products they viewed. Your goal is to discover associations between products, which can help the company understand which items are often viewed together and may be recommended to users.**

***Dataset Information:***

- Clickstream.csv: Contains the following columns:
  - Session ID: Unique identifier for each session.
  - Viewed Products: A list of products viewed during the session (comma-separated list of product IDs).

***Task:***

1. Preprocess the clickstream data into a format suitable for association rule mining.
2. Apply the Apriori algorithm to mine frequent product associations.
3. Interpret the generated rules and suggest how the company could use these rules for product recommendations.
4. Evaluate the rules based on support, confidence, and lift, and discuss the trade-offs between them.

**Scenario 3: Movie Recommendation System**

***Question:***

**You have been given a dataset containing information about users' movie ratings. The dataset contains records of users who have rated movies. The goal is to mine the dataset to discover frequent movie pairings and generate association rules that can be used to recommend movies.**

***Dataset Information:***

- MovieRatings.csv: Contains the following columns:
  - User ID: Unique identifier for each user.
  - Movie ID: Unique identifier for each movie.
  - Rating: The rating given by the user (e.g., on a scale of 1 to 5).

***Task:***

1. Convert the ratings data into a format suitable for association rule mining.
2. Use the Apriori algorithm to find frequent movie pairs with a given support threshold.
3. Generate association rules based on these frequent itemsets.
4. Evaluate the rules using Lift and Confidence.
5. Based on the rules, suggest possible movie recommendations for users.

**Scenario 4: Market Basket Analysis for a Supermarket**

***Question:***

**A supermarket wants to improve its marketing strategies by analyzing its customers' purchase behavior. The dataset contains transactions from the store, and the goal is to uncover item associations that could help in creating targeted promotions.**

***Dataset Information:***

- SupermarketTransactions.csv: Contains the following columns:
    - Transaction ID: Unique identifier for each transaction.
    - Item List: A list of items bought in each transaction (comma-separated list).

***Task:***

1. Preprocess the data into a transactional format where each transaction contains a list of items purchased.
2. Apply the Apriori algorithm to mine frequent itemsets and generate association rules.
3. Filter the rules based on a minimum Confidence threshold of 0.7 and a minimum Lift of 1.5.
4. Discuss how these association rules can help the supermarket in designing bundled promotions or cross-selling strategies.

**Scenario 5: Analyzing Customer Purchases for a Bookstore**

***Question:***

**A bookstore wants to analyze customer purchase behavior to develop strategies for increasing sales. The goal is to find associations between book categories (e.g., Fiction, Non-Fiction, Mystery, Romance) that can help in recommending books to customers based on their previous purchases.**

***Dataset Information:***

- BookstorePurchases.csv: Contains the following columns:

- Customer ID: Unique identifier for each customer.
- Books Purchased: A list of books purchased by the customer (comma-separated list of book IDs or book categories).

***Task:***

1. Preprocess the dataset into a suitable format for association rule mining (convert book categories into one-hot encoding or use transactional data format).
2. Use the Apriori algorithm to identify frequent book category pairs.
3. Generate association rules with a minimum Confidence of 0.6.
4. Analyze the rules to determine which book categories are commonly bought together.
5. Discuss potential marketing strategies, such as offering discounts or special promotions based on these associations.

**Scenario 6: Social Media Content Engagement Analysis**

***Question:***

**You are working with a social media platform that wants to analyze content engagement data. The platform wants to uncover patterns about which types of posts (e.g., memes, news articles, videos) users tend to engage with together. Your task is to use association rule mining to analyze the data.**

***Dataset Information:***

- SocialMediaEngagement.csv: Contains the following columns:
  - User ID: Unique identifier for each user.
  - Post Type: The type of content the user interacted with (e.g., "memes", "news", "videos").

***Task:***

1. Preprocess the dataset into a transactional format where each transaction is a user's engagement with different post types.
2. Apply the Apriori algorithm to discover frequent itemsets based on user engagements.
3. Generate association rules to identify post types that are frequently engaged with together.
4. Evaluate the results based on Lift, Confidence, and Support.
5. Propose ways the platform could use these insights for content recommendation or targeted advertising.

**Scenario 7: Restaurant Menu Optimization**

### *Question:*

**A restaurant chain wants to analyze customer orders to optimize their menu offerings. The goal is to identify combinations of dishes that customers frequently order together and suggest possible menu pairings.**

### *Dataset Information:*

- RestaurantOrders.csv: Contains the following columns:
    - Order ID: Unique identifier for each order.
    - Items Ordered: A list of items ordered in each transaction (comma-separated list of dish names).

### *Task:*

1. Preprocess the data into a suitable format for association rule mining (either by creating a binary matrix or transactional list).
2. Apply the Apriori algorithm to mine frequent itemsets for dish pairings.
3. Generate association rules with minimum Confidence of 0.7 and Support of 0.05.
4. Analyze the rules to identify the most frequently ordered dish pairings.
5. Discuss how these findings can be used to redesign the menu (e.g., create combo deals or highlight popular pairings).