Q1 What is lift and why is it important in Association rules ?

**Lift** is a measure of the strength of an association between two items (or itemsets) relative to the expected frequency of those items occurring together by chance.

**Importance of Lift in Association Rules:**

1. **Identifying Strong Relationships**: Lift helps identify strong relationships between items. For example, a high lift value indicates that when item A is bought, item B is more likely to be bought as well, which is a valuable insight for cross-selling or promotions.
2. **Reducing Noise**: When looking at association rules, some pairs may have a high support but are not particularly interesting (e.g., due to the high occurrence of one or both items). Lift helps to filter out such "trivial" rules and focuses on those that have a stronger and more meaningful relationship than what would be expected by chance.
3. **Improving Decision-Making**: By using lift, retailers or businesses can make better decisions about which items to bundle, promote, or stock based on the likelihood of their co-occurrence being above random chance.

Q2 What is support and confidence. How to you calculate them ?

### 1. ****Support:****

* **Definition:** Support measures how frequently an item or a set of items appears in the dataset.

**Formula:**

Support(A) = Number of transactions containing item A​ / Total number of transactions

### 2. ****Confidence:****

* **Definition:** Confidence measures the likelihood that item B is purchased when item A is purchased. In other words, it is the conditional probability of B given A.

**Formula:**

Confidence(A→B) = Support(A→B)​ / Support(A)

Q3 What are some limitation or challenges of Association rules mining ?

Limitation or Challenges of Association rules

 **Scalability**: As the dataset grows larger, the number of possible itemsets increases exponentially. This makes association rule mining computationally expensive and difficult to scale for very large datasets. The algorithm can be slow and resource-intensive.

 **Complexity of Rule Evaluation**: Association rule mining generates many rules, and not all of them are meaningful or interesting. Evaluating and filtering the rules to identify the most valuable ones (in terms of metrics like support, confidence, and lift) is a non-trivial task and can be time-consuming.

 **Redundancy**: Many generated rules may be redundant, meaning they convey the same information or are highly correlated. Identifying and removing such redundant rules is crucial to making the output more interpretable and useful.

 **Sparsity**: In many real-world datasets, especially those with many items, most of the possible itemsets may not appear frequently enough to generate strong rules. This leads to sparse results where only a small number of rules have high support and confidence.

 **Difficulty in Handling Continuous Data**: Association rule mining is typically designed for categorical or discrete data. Dealing with continuous variables requires discretization, which can lead to information loss or inefficiencies in rule generation