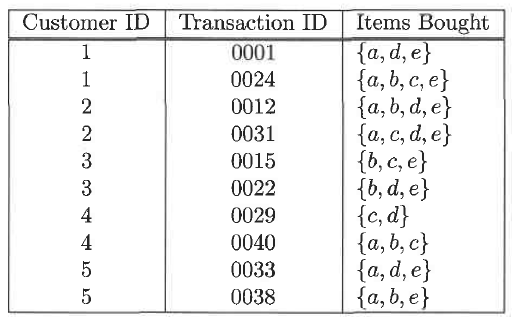
**Lab 7**

**Objectives**

* Be familiar with the association mining concepts
* Practice related algorithms i.e. Aprori Algorithms
* Apply association mining algorithms to solve real world problems

1. Consider the data set shown in the following table and answer the questions below:



1. Computer the support for itemsets {e} {b,d}, and {b, d, e} by treating each transaction ID as a market basket.

Support = 80% for {e}

Support = 20% for {b,d}

Support = 20% for {b,d,e}

1. Use the results in part (a) to compute the confidence for the association rules {b, d} 🡪 {e} and {e} 🡪 {b, d}. Is confidence a symmetric measure?

Confidence = 100% for {b,d} -> {e}

Confidence = 25% for {e} - > {b,d}

Confidence is not symmetric

1. Repeat part (a) by treating each customer ID as a market basket. Each item should be treated as binary variable (1 if an item appears in at least one transaction bought by the customer, and 0 otherwise.)

Support = 80% for {e}

Support = 100% for {b,d}

Support = 80% for {b,d,e}

1. Use the results in part (c) to compute the confidence for the association rules {b, d} 🡪 {e} and {e} 🡪 {b, d}

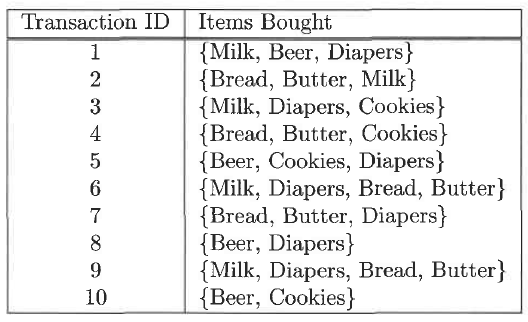
Confidence = 80% for {b,d} -> {e}

Confidence = 100% for {e} - > {b,d}

1. Suppose s1 and c1 are the support and confidence values of an association rule r when treating each transaction ID as a market basket. Also, let s2 and c2 be the support and confidence values of r when treating each customer ID as a market basket. Discuss whether there are any relationships between s1 and s2 or c1 and, c2.

There is a relationship between s1 and s2 and c1 to c2. Each indivudal item within the data set increase so does the support for a rule as well as the confidence. S2 and c2 > s1 and c1.

1. Consider the market basket transactions shown in the following table and answer the questions below:



1. What is the maximum number of association rules that can be extracted from this data (including rules that have zero support)?

602 possible rules

1. What is the maximum size of frequent itemsets that can be extracted (assuming minsup > 0)?

{Milk, Diapers, bread, Butter} has the biggest size = 4, and sup{Milk, Diapers, Bread, Butter} = 2

1. Write an expression for the maximum number of size-3 itemsets that can be derived from this data set.

# of items / 3

1. Find an itemset (of size 2 or larger) that has the largest support.

{bread, butter}, support = 50%

1. Find a pair of items, a and b, such that the rules {a} 🡪 {b} and {b} 🡪 {a} have the same confidence.

{Milk, Butter}, {Milk, Bread}, {Bread, Butter}, {Beer, Cookies}

1. Write a python program to calculate the support count for different combinations of top six frequent items that you generated in the previous lab assignment. (You can use any data structure i.e. you can store transaction as a linkedlist, or you can use Hashtree and etc).

Your output is something like:

{egg, milk} 🡪 support count: 15

{egg, water} 🡪 support count: 10

…………