**Association Rules Questions**

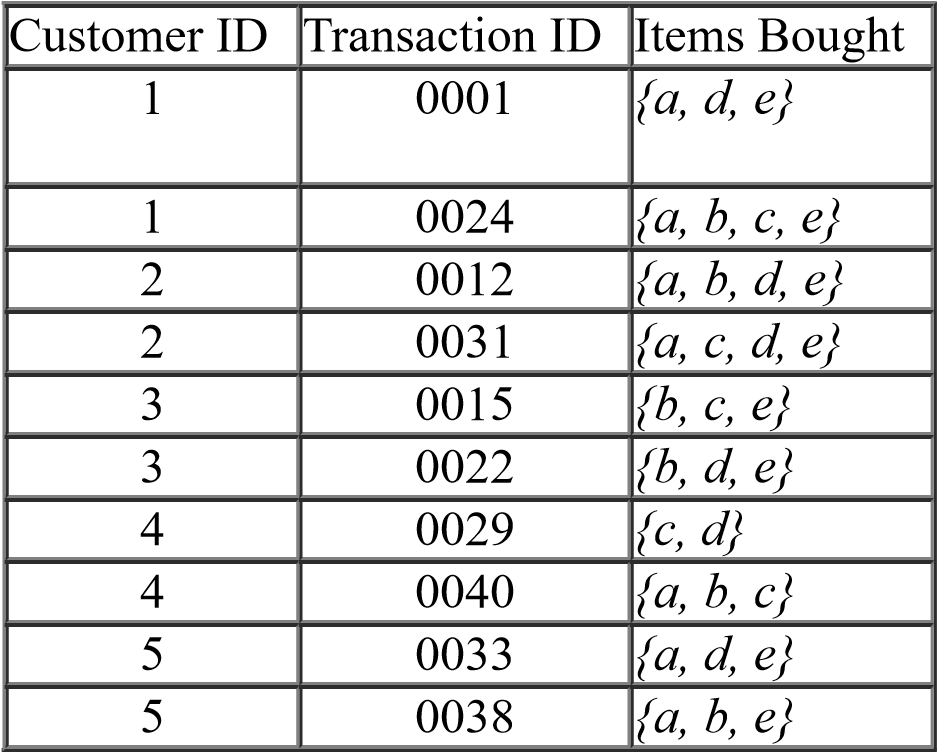
# Objectives

**Explore the decision rule generation** with a relatively simple, and clean, dataset.

**Experiment with the different support and confidence thresholds** of the association rule algorithm to identify a set of rules.

**Experiment with the Apriori** algorithm.

**Q1.** Consider the data set



1. Compute the support for itemsets *{a}*, {*b, c}*, and {*a,b,e}* by treating each transaction ID as a market basket. You have 10 transactions
2. Use the results in part (a) to compute the confidence for the association rules *{b, c} → {a}* and *{a} → {b, c}*.
3. Repeat part (a) by treating each customer ID as a market basket. You have 5 customers. Eachitem should be treated as a binary variable (1 if an item appears in at least one transaction bought by the customer, and 0 otherwise.)
4. Use the results in part (c) to compute the confidence for the association rules *{b, c} → {a}* and *{a} → {b, c}*.

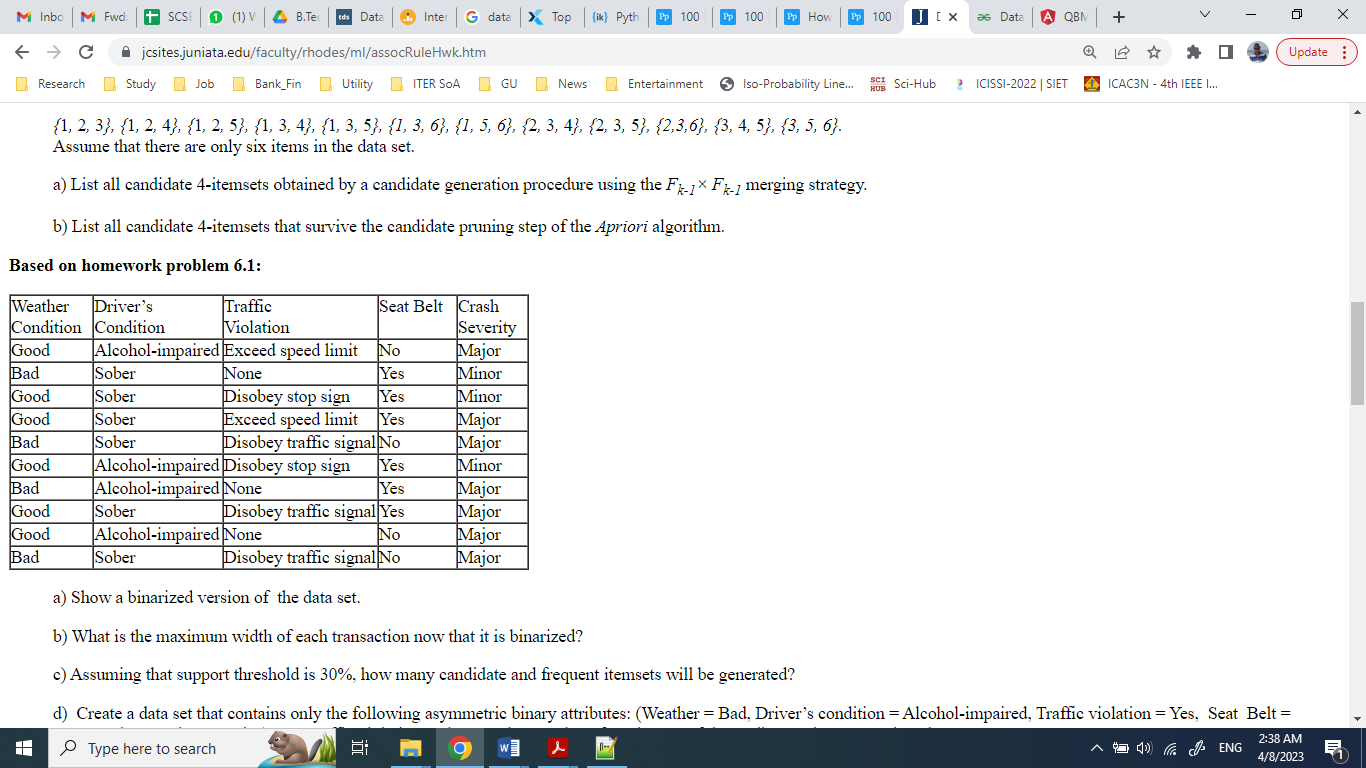
**Q2.** Consider the following set of frequent 3-itemsets:

*{*1*,* 2*,* 3*}, {*1*,* 2*,* 4*}, {*1*,* 2*,* 5*}, {*1*,* 3*,* 4*}, {*1*,* 3*,* 5*}, {1, 3, 6}, {1, 5, 6}, {*2*,* 3*,* 4*}, {*2*,* 3*,* 5*}, {2,3,6}, {*3*,* 4*,* 5*}, {3, 5, 6}.*

Assume that there are only six items in the data set.

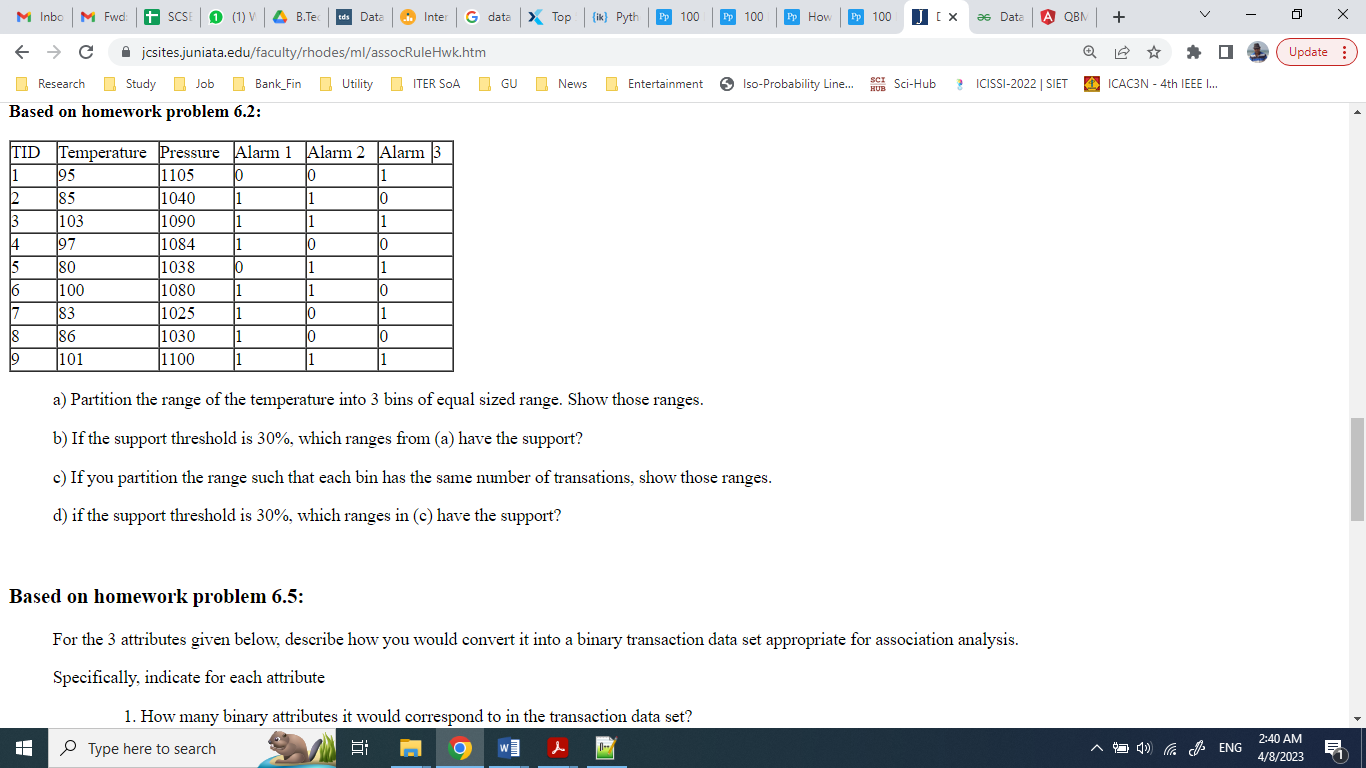
1. List all candidate 4-itemsets obtained by a candidate generation procedure using the *Fk-1× Fk-1* merging strategy.
2. List all candidate 4-itemsets that survive the candidate pruning step of the *Apriori* algorithm.

**Q3.** Consider the following dataset:



1. Show a binarized version of the data set.
2. What is the maximum width of each transaction now that it is binarized?
3. Assuming that support threshold is 30%, how many candidate and frequent itemsets will begenerated?
4. Create a data set that contains only the following asymmetric binary attributes: (Weather = Bad,Driver’s condition = Alcohol-impaired, Traffic violation = Yes, Seat Belt = No, Crash Severity = Major). For Traffic violation, only None has a value of 0. The rest of the attribute values are assigned to 1.
5. Assuming that support threshold is 30%, how many candidate and frequent itemsets will begenerated?

**Q4.** Consider the following dataset:



1. Partition the range of the temperature into 3 bins of equal sized range. Show those ranges.
2. If the support threshold is 30%, which ranges from (a) have the support?
3. If you partition the range such that each bin has the same number of transations, show thoseranges.
4. if the support threshold is 30%, which ranges in (c) have the support?