Assignment 8

Due: 4/3

Note: Show all your work.

Problem 1 (20 points). Consider the following transactional database.

TID	Items
100	2, 3, 4, 5, 7
200	1, 2, 5, 6
300	1, 4, 5, 7, 8
400	1, 2, 3, 5, 6, 7
500	2, 3, 5, 6, 7, 8

(1) Mine all frequent itemsets using Apriori. Show all candidate itemsets and frequent itemsets. You should follow the process described in the book and lecture (i.e., C1 → L1 → C2 → L2 → …). Minimum support = 60% (or 3 or more transactions). To save your time, L1 and L2 are given below:

L1:

Itemset	1	2	3	5	6	7
Count	3	4	3	5	3	4

L2:

Itemset	{1,5}	{2,3}	{2,5}	{2,6}	{2,7}	{3,5}	{3,7}	{5,6}	{5,7}
Count	3	3	4	3	3	3	3	3	4

(2) Sort all frequent 4-itemsets by their item number. Then, select the first frequent 4-itemset form the sorted list of frequent 4-itemsets and mine all strong rules from this itemset that have the format {W, X} => {Y, Z}, where W, X, Y, and Z are individual items. Assume that minimum confidence = 80%.

Problem 2 (10 points). Consider the following transactional database for sequential pattern mining.

CID	Day	Items
1	1	B, D, H
	14	A, C, E
	24	B, C, H
	31	D, F, G
2	4	A, B, G, H
	9	B, D, E, G
	14	C, E, H
3	1	B, G, H
	24	A, C, D, E
	51	A, B, G, H
4	2	B, G
	12	A, B, E, H
	25	B, C, D, E. G

Determine the supports of the following sequences:

<{A}, {D}>, <{B, D}, {G}>, <{A}, {D, E}>

Problem 3 (20 points). Consider the following contingency table.

|	C (buys coffee = Yes)	$\overline{C}$ (buys coffee = No)
T (buys tea = Yes)	238	197
$\overline{T}$ (buys tea = No)	82	85

(1). Compute the *lift*, *all-confidence*, *cosine*, *Kulczynski* and *imbalance ratio* measure, and determine whether buying coffee and buying tea are positively correlated, negatively correlated, or not correlated.
(2). Perform the chi-square test with 95% significance level and determine whether they are correlated or not.

Problem 4 (20 points). This problem has two parts. Problem 2-1 uses Oracle and Problem 2-2 uses JMP Pro. Choose one of the two.

Problem 2-1 (Oracle) Follow the instructions in *oracle-association-rule-assignment.pdf* file. The submission requirements are indicated with **"Required."**

Problem 2-1 (JMP Pro) You will perform association analysis using JMP Pro. There is a secion in *Predictive and Specialized Modeling.pdf* documentation that shows how to do association analysis. You may want to read this section before starting the assignment. For this assignment, follow the instructions given below.

1. Start JMP Pro
2. Select Help > Sample Data Library and open Grocery Purchases.jmp.
3. Select Analyze > Screening > Association Analysis.
4. Select Product and click Item.
5. Select Customer ID and click ID.
6. Set the following parameters
   Minimum Support: 0.15
   Minimum Confidence 0.35
   Lift: 1.2
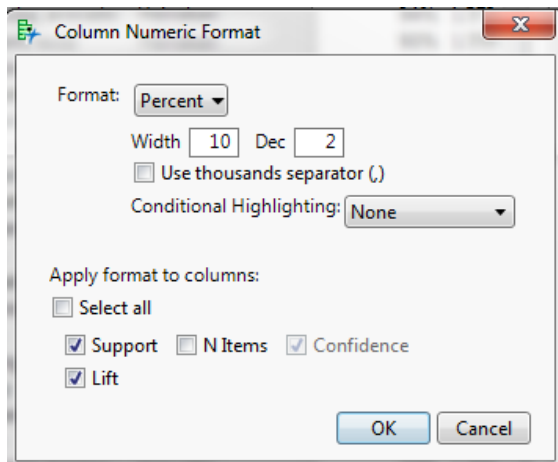   Maximum Antecedents: 3
   Maximum Rule Size: 10
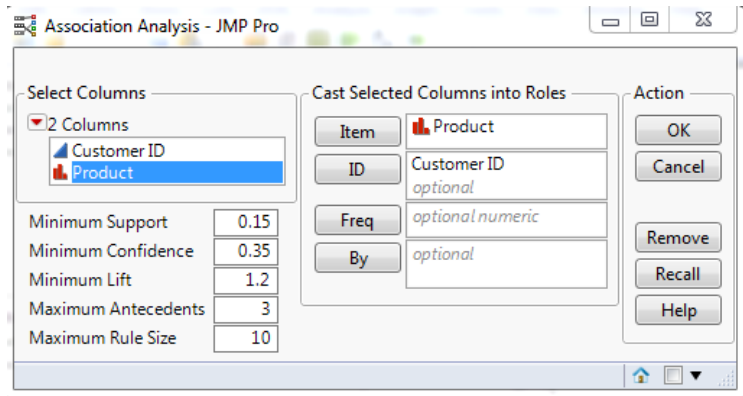
7. Click OK.
   Rules report will appear. Capture a top part of the report, including about 20 rules, and include this screenshot in your report.

8. Right-click Confidence > select Format Column
   Change Dec from 0 to 2
   Select Support, Confidence, and Lift

   Click OK. Rules report will now show 2 digits below the decimal point for those three measures. Capture a top part of the report, including about 20 rules, and include this screenshot in your report.

9. Find the 3-itemset {crackers, Heineken, soda} under Frequent Item Sets.
10. Manually mine all rules from this 3-itemset and calculate the confidences of these rules using the method discussed in the class. Show, in your report, all rules and their confidences. You need to show how you calculated the confidences.
11. If the minimum confidence is 70%, which ones are strong rules?