Assignment 7

Due: 11/9

Note: Show all your work.

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

TID	Items
100	2, 3, 4, 5, 6, 8
200	1, 2, 3, 5, 6
300	1, 4, 5, 7, 8
400	2, 3, 4, 5, 6
500	1, 2, 3, 4, 5, 7
600	1, 3, 8

(1) Mine all frequent itemsets using the Apriori algorithm that we discussed in the class. Show all candidate itemsets and frequent itemsets. You should follow the step by step process that we discussed in the class (i.e., C1 → L1 → C2 → L2 → …). You don't need to show the pruning steps. Minimum support = 50% (or 3 or more transactions). To save your time, L1 is given below:

L1:

Itemset	1	2	3	4	5	6	8
Count	4	4	5	4	5	3	3

(2) Sort all frequent 4-itemsets by their item number. Then, select the first frequent 4-itemset from 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 (20 points). Consider the following contingency table.

	C (buys coffee = Yes)	$\overline{C}$ (buys coffee = No)
T (buys tea = Yes)	287	195
$\overline{T}$ (buys tea = No)	45	143

(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 5% significance level and determine whether they are correlated or not.

Problem 3 (20 points). This problem is about mining frequent itemsets and strong rules from a grocery data using R. Use *hw7.csv* data for this problem. It is a slightly modified version of the *groceries.csv* data that is included in R.

(1). Mine all frequent itemsets with the minimum support = 0.005 and include the screenshot of the summary in your submission.
(2). What is the total number of frequent itemsets? List the top five frequent itemsets and their supports in descending order of support.
(3). What is the total number of frequent 1-itemsets (L1)? List the top five frequent 1-itemsets and their supports in descending order of support.
(4). What is the total number of frequent 2-itemsets (L2)? List the top five frequent 2-itemsets and their supports in descending order of support.
(5). What is the total number of frequent 3-itemsets (L3)? Show the frequent 3-itemset and its support that has the highest support.
(6). What is the total number of frequent 4-itemsets (L4)? Show the frequent 4-itemset and its support that has the highest support.

(7). Mine all rules with minimum support = 0.005 and minimum confidence = 0.15 and include the screenshot of the summary in your submission.
(8). Sort rules in descending order of confidences and show the top five rules and their confidences.
(9). Sort rules in descending order of supports and show the top five rules and their supports.
(10). Show all rules that include *coffee* (in either side of a rule) along with their supports and confidences.

Submission:

Submit all solutions in a single Word or PDF document and upload it to Blackboard. Use *LastName_FirstName_hw7*.docx or *LastName_FirstName_hw7*.pdf as the file name. You must also submit your R code file. Name it *hw7.R*.