STAT 515 Applied Statistics and Visualization for Analytics (Fall 2016)

Assignment 5

Due Date: 12/04/2016 Sunday 11:59 PM

1. Groceries Transaction Study (50pts)

Groceries Transaction Data Study

For this problem, you'll be working with the **Groceries.csv** file. This data file reflects 43,366 items purchased across 9,835 transactions at a grocery store. Each row represents a transaction/product pair. The first value is the transaction ID and the second value is the product name (such as citrus fruit, yogurt, coffee, etc.). The manager of the store wants to better understand his customer's buying habits so that he can modify the store layout and run more effective promotions.

Please answer the following questions by conducting association / market basket analysis.

a. Generate a bar-chart plot for the top 5 items with highest frequency and report the corresponding items?

b. Generate a set of association rules by setting the minimum support threshold value at 0.005, and the minimum confidence threshold value at 0.2. Report the total number of associations rules obtained and the top 5 rules with highest lift values. What is the association rule with the highest lift value? How do you interpret the lift value (or, what insights can you derive from this specific rule)?

c. Explore the association rules with soda. Using the same minimum threshold values of support and confidence, generate a subset of rules with only one item as the antecedent (lhs) item and only " soda" as the consequent (rhs) item. Report the top 5 rules with highest lift values.

Hint for R: you will need to add "maxlen=2" in the parameter argument setting, and the appearance argument settings in the aprori function. For example.

```
soda_rules = apriori(Groceries, parameter = list(support = 0.005,confidence = 0.2, maxlen = 2), appearance = list(default="lhs",rhs=" soda"))
```

(Note that there is a space in " soda" due to the format in the original dataset.)

d. Your store manager comes to you and suggests that they should place the soda and fruit/vegetable juice together because, as beverages, they "go together." Is that supported by the data and association rules? Report the statistic values you used to reach this conclusion (using the same minimum threshold values of support and confidence).
(Hint: look for rules { fruit/vegetable juice} => { soda} and { soda} => { fruit/vegetable juice}.)

e. You decide to start a promotion that will help cross-sell butter. Which product(s) have the highest predictive power in determining who will purchase butter? Report the statistic values you used to reach this conclusion (using the same minimum threshold values of support and confidence).

Submission:

1. Prepare a pdf (or MS word) file with answers to above questions, brief explanations for the analysis you perform. Please provide your R codes at the end of the submitted file as appendix.
2. Name the file as "LastName, FirstName-HW5.pdf" (for example, "Ji,Ran-HW5.pdf") and submit it on blackboard.