**IST687 – HW 9 - Finding Patterns with arules**

**Part A: Explore Data Set**

1. Load the dataset: hotelSurveyBarriot.json (similar to HW8, but a different dataset)
2. Name the dataframe hotelSurvey

**Part B: Explore Data Set**

1. Ensure hotelSurvey is a dataframe, and look at the structure via the *str*() command
2. Map each numeric attribute to a category – Since we want to create rules, we should convert the attributes that have a numeric range into buckets (ex. low or high)

*Hint: For Survey attributes that range from 0 to 10 one can use the following:*

*vBuckets <- replicate(length(vec), "Average")*

*vBuckets[vec > 7] <- "High"*

*vBuckets[vec < 7] <- "Low"  
  
For other attributes, you can use the following code:*

*q <- quantile(vec, c(0.4, 0.6))*

*vBuckets <- replicate(length(vec), "Average")*

*vBuckets[vec <= q[1]] <- "Low"*

*vBuckets[vec > q[2]] <- "High"*

*🡪 make sure to document the code you use!!!*

1. Count the people in each category of for the age and friendliness attributes

*Hint: Use the table( ) command.*

1. Express the results of problem 3 as percentages by sending the results of the *table( )* command into the *prop.table( )* command
2. Show a “contingency table” of percentages for the age and the overall satisfaction variables together. Write a block comment about what you see.

**Part C: Coerce the data frame into transactions**

1. Install and library two packages: arules and arulesViz.
2. Coerce the hotelSurvey data frame into a sparse transactions matrix using:  
   hotelSurveyX *<- as(*hotelSurvey*,"transactions")*
3. Use the *inspect( )*, *itemFrequency( ),* and *itemFrequencyPlot( )* commands to explore the contents of hotelSurveyX.

**Part C: Use arules to discover patterns**Support is the proportion of times that a particular set of items occurs relative to the whole dataset. Confidence is proportion of times that the consequent occurs when the antecedent is present. See the review on the next page.

1. Run the apriori command to try and predict happy customers (as defined by their overall satisfaction being high – above 7).

1. Once you have a reasonable number of rules, use inspect( ) to view the ruleset.
2. If you had to provide two rules to the hotel owner (in terms of what helps drive high overall customer satisfaction, what would those two rules be? Use a block comment to explain your answer.