# Homework (HW09) Associated Ruled Mining

### General Instructions

For this homework you will upload 1 R file into blackboard.

Reminder:

* All HW must start with an Identification Block like this sample…

################################################

# IST 387/687, Standard Homework Heading

#

# Student name:

# Homework number:

# Date due:

#

# Attribution statement: (choose the statements that are true)

# 1. I did this work by myself, with help from the book and the professor

# 2. I did this work with help from the book and the professor and these Internet sources: <provide the urls>

# 3. I did this work with coaching from <Name of another student> but did not cut and paste any code

# Run these three functions to get a clean test of homework code

dev.off() # Clear the graph window

cat('\014') # Clear the console

rm(list=ls()) # Clear all user objects from the environment!!!

# Set working directory

# Change to the folder containing your homework data files

setwd("~/MyDesktop/ISTX87/Homework")

This homework builds on our efforts from the Prep Exercise and depends on a careful read of Chapter 17 of *An Introduction to Data Science*. As usual we’ll use the Prep Ex R file that you created as a starting point. We’ll revisit the “badboat” dataframe that you created in the PE. In this homework we will practice three of the four steps typically included in data mining, exploratory data analysis, model development, and interpretation of results. By the end of the homework assignment you should have a solid working knowledge of associated rules mining and the benefits of using this algorithm.

### HW09

**Step 1: Explore the Data Set.**

1. Re-run the *View()* command to examine the badboat dataframe; write a block comment describing what you see. Is this a sparse matrix?
2. Using the *table()* command, count the people in each category of the Survived variable. At a high-level, describe what you see in a comment.
3. Express the results of part B as percentages by sending the results of the *table()* command into the *prop.table()* command.
4. Following the same techniques used in the previous tasks, show the percentages for Class, Sex, and Age variables.
5. Show a contingency table of percentages for the Age and Sex variables together. Write a block comment of your observations.

**Step 2: Coercing the data frame into transactions.**

1. Coerce the badboat dataframe into a sparse transactions matrix using the code provided below. Comment what the components of the code tell RStudio to do.

**badboatX <- as(badboat,"transactions")**

1. Using the *inspect()*, *itemFrequncy()*, and *itemFrequencyPlot()* commands, explore the contents of “badboatX”. Write a comment about what each of the commands does.
2. Explore the spare matrix data object using the *View()* command and write a comment about your observations.
3. In a few sentences, write a block comment explaining the difference between “badboat” and “badboatX”.

**Step 3: Discovering patterns using associated rules mining.**

Before beginning this section, it is important to understand that support is the proportion of times that a particular set of items occurs relative to the whole dataset. Confidence is the proportion of times that the consequent occurs when the antecedent is present. In layman’s terms, confidence is the proportion of times that we observe the result given the preceding event has occurred.

1. Run the block of code below. Document, via comments, what each line of code does (the comments should be before each line of code below). Make sure that you are typing the code in RStudio, do not copy and paste.

**ruleset <- apriori(badboatX,   
 parameter=list(support=0.005,confidence=0.5),  
 appearance = list(default="lhs", rhs=("Survived=Yes")))**

1. Using the *inspect()* command, review the “ruleset”.
2. Experiment with the interactive ruleset interface by running the *inspectDT()* command.
3. In a block comment answer the following question. If you were to be onboard the titanic, what kind of person would you have wanted to be? Use the output of both *inspect()*  and *inspectDT()* commands to support your answer.

***You must submit all Homework to blackboard prior to the deadline specified for each assignment.***

Late HW assignments will not be accepted for credit.

*Materials provided by the instructors of this course are copyrighted 2018 by Jeffrey Saltz and Jeffrey Stanton, or by the respective instructor that produced them. These materials are provided to enrolled students for their exclusive use during the semester of instruction. Uploading instructor-produced materials to Internet sites (e.g., Coursehero) without permission is considered a breach of professional ethics and will be reported as a possible academic integrity violation.*