# ITM 6285 Data Mining Assignment

## Market Basket Analysis (expected time – 1~1.5 hours)

**Learning Objective:** This assignment is designed to help you practice the concepts of market basket analysis discussed in class, understand how to interpret the R outputs.

**Problem framework:**

REI, a national outdoor sport gear retailer, is trying to revamp the layout of its stores to make them more efficient and to help customers locate the desired merchandise with more ease. Moreover, REI is also trying to woo more customers in with bundled offers (multiple products grouped together for one price). For both undertakings, REI must understand what items customers tend to buy together. You have been recently hired by REI as a data analyst / statistician to help them with a market basket analysis.

Please download the **Market Basket.csv**. You will find REI's sales scanner data that traces 12 products over 100 transactions. Below is a list of the item names and their abbreviation.

| | | |
|---|---|---|
| Kayak - K | Ski Goggles - SKG | Swimsuit - SW |
| Tent - T | Sunglasses - SG | Snowboard - SNB |
| Flashlight - F | Backpack - BP | Sleeping Bag - SB |
| Portable Stove - PS | Sunscreen - SS | Running Shoes - R |

By the way, “Bag” means a purchase

**Task 1: Import the Data**

Please import the data set to R studio using read.csv.

mydata=read.csv("……/market basket.csv",header=TRUE)

If you try to find the path name of a file on a mac, right click (tap with 2 fingers) the file and then hit the option key, you will find an opinion “copy xxx as path name”

You can try File--Import Dataset--from CSV. But the data type read into R is not logical or categorical, which will prevent us from doing the analysis. However, if the data is all numerical, you may use this method.

**Task 2: Summary Statistics**

Please summarize the data using summary(the dataset name) and copy the R outputs to the answer sheet (you may use a txt file).

**Task 3: Default Apriori**

Do a classification rule mining using Apriori with default parameters. Important: don't inspect rules. It will take a lot of time. Answer how many rules you got.

Hint: the number of rules will be shown at the second last line, something like this:

writing ... [xxx rule(s)] done [0.02s].

**Task 4: Reduce the Number of Rules by Changing the Parameters**

Now limit the parameters to minlen=2,supp=0.05,conf=0.7, and limit our attention to the purchase of K (make K purchase as the **lhs**: appearance=list(lhs=c("K=Bag"),default="rhs"))

Inspect the rules and copy the results to your answer sheet.

**Task 5: Short Answer**

What the 3rd rule tells you about (confidence = 1)? What is rule 10? Is it meaningful??

**Task 6: Get Rid of the Rules with XXX=No on the rhs**

Now let's focus on the "Purchasing behavior", but not the "not-purchasing behavior" like the previous rule 10. Keep other parameters the same and use
appearance=list(lhs=c("K=Bag"),rhs=c("T=Bag","F=Bag","PS=Bag","SKG=Bag","SG=Bag","BP=Bag","SS=Bag","SW=Bag","SNB=Bag","SB=Bag","RS=Bag"),default="none")

Copy the R outputs to your answer sheet. Which rule has the highest lift?