### -------------------------------------------------------------------------------

#Market basket analysis is used behind the scenes for the recommendation systems

#used in many brick-and-mortar and online retailers. The learned association rules

#indicate the combinations of items that are often purchased together. Knowledge

#of these patterns provides insight into new ways a grocery chain might optimize

#the inventory, advertise promotions,or organize the physical layout of the store.

#For instance, if shoppers frequently purchase coffee or orange juice with a breakfast

#pastry, it may be possible to increase profit by relocating pastries closer to coffee

#and juice.

#In this lab, we will perform a market basket analysis of transactional data

#from a grocery store.Our market basket analysis will utilize the purchase data

#collected from one month of operation at a real-world grocery store. The data contains

#9,835 transactions.

### ------------------------------------------------------------------------------

# 1. Import groceries.csv file

library(arules)

groceries <- read.transactions("groceries.csv", sep = ",")

# 2. Understanding of your data.

# Summary of dataset

summary(groceries)

# Inspect the first 5 transactions

inspect(groceries[1:5])

# How many transactions and items in this data?

#9835 transactions

#169 items

#also summary shows most frequent items.

#Also 2159 transactions include only 1 item.

#1643 transactions include 2 items. Transaction with the largest number of items includes 32 items.

# 3. Data exploration

# Examine the relative frequency of items in descending order

sort(itemFrequency(groceries, type="relative"), decreasing = TRUE)

#we use two function “sort” and “itemFrequency”. We want the items frequency to be sorted in descending order.

# What are the top 3 most frequent items?

# What are the support values for top 3 most frequent items?

# Examine the absolute frequency of items in descending order

# Plot the most frequent 8 items in the descending order of transaction frequency in percentage

itemFrequencyPlot(groceries, type="relative", topN = 8)

# 4. Use the apriori command to generate rules with minimal support = 0.01 and minimal confidence = 0.3 and max length = 2.

groceries\_rules <- apriori(groceries, parameter = list(support = 0.01, confidence = 0.3, maxlen=2))

summary(groceries\_rules)

# Display all rules sorted by confidence levels.

inspect(sort(groceries\_rules, by = "confidence"))

# Display top 5 rules

inspect(sort(groceries\_rules, by = "confidence")[1:5])

# What is the probability of buying whole milk when butter is purchased?

# What is the probability of buying butter and whole milk together?

# Interpret the first rule based on the values of the support, confidence, and lift.

# 5. Use the apriori command to generate rules with minimal support = 0.02 and minimal confidence = 0.4 and max length = 3.

# Display top 10 rules for Task 2 sorted by lift.

# Find and display rules containing "other vegetables"

# Find and display rules containing "other vegetables" on the left-hand side

# Interpret the first rule (containing "other vegetables" on the left-hand side) based on support, confidence, and lift values.

# Find and display rules containing "other vegetables" on the right-hand side

# 6. Use the apriori command to generate about 30 to 50 association rules. Set your own minimum support and confidence threshold levels.

# Remember if the thresholds are too low, you will get too many rules, or if you set them too high, you may not get any or enough rules.

# Inspect all of the rules in the descending lift values of the rules.

# Select an interesting rule and explain how it can benefit the grocery store.