Apriori Algorithm

```
#install.packages("plyr")
library(plyr)
#install.packages("arules")
library(arules)
#install.packages("arulesViz")
#install.packages("ggplot2")
library(ggplot2)
library(arulesViz)
```

 # Read the data

```
df_groceries <- read.csv("Groceries_dataset.csv")
```

 # Data cleaning and manipulations using R

#First make sure that the Member numbers are of numeric data type and then

#sort the dataframe based on the Member_number.

```
df_sorted <- df_groceries[order(df_groceries$Member_number),]
View(df_sorted)
df_sorted$Member_number <- as.numeric(df_sorted$Member_number)
```

#convert the dataframe into transactions format such that all

#the items bought at the same time in one row.

#ie; convert the dataframe into basket format based on the Member_number

#and Date of transaction

#ddply: Split data frame, apply function, and return results in a data frame.

```
df_itemList <- ddply(df_groceries,c("Member_number","Date"),
                     function(df1)paste(df1$itemDescription,collapse = ","))

View(df_itemList)
```

#Once we have the transactions, we no longer need the date and member numbers in

#our analysis.Delete those columns.

```
df_itemList$Member_number <- NULL  # drop (delete) columns
df_itemList$Date <- NULL

View(df_itemList)
```

#Rename column headers for ease of use

```
colnames(df_itemList) <- c("ItemList")

View(df_itemList)
```

#Write dataframe to a csv file using write.csv()

```
write.csv(df_itemList,"new_Grocery_ItemList1.csv", row.names = TRUE)
```

#Find the association rules

#Run algorithm on Grocery_ItemList.csv to find relationships among the items

#Using the read.transactions() functions, we can read the file ItemList.csv and convert it to a transaction format

```
txn = read.transactions(file="Grocery_ItemList1.csv", rm.duplicates= TRUE,
format="basket",sep=",",cols=1);

txn
```

#run the apriori algorithm on the transactions by specifying minimum values for

#support and confidence.

```
basket_rules <- apriori(txn,parameter = list(sup = 0.01, conf = 0.01));

print(basket_rules)
#inspect() function prints the internal representation of an R object or the result of an expression.
inspect(basket_rules)

plot(basket_rules)
```

#Graph to display top 5 items

```
itemFrequencyPlot(txn, topN = 5)
```

****

Example 2:

```
# Loading Libraries
library(arules)
library(arulesViz)
library(RColorBrewer)

# import dataset
data("Groceries")

#  apriori() function
rules <- apriori(Groceries,parameter = list(supp = 0.01, conf = 0.2))

# using inspect() function
inspect(rules[1:10])

#  itemFrequencyPlot() function
itemFrequencyPlot(Groceries, topN = 10)
```

****