

## 1. Importing the file

- ✓ Firstly, all the environment variables are removed from the environment.
- ✓ All the required packages and libraries are installed and loaded which consists of tidyverse, arules and arulesviz.
- ✓ tidyverse package is used here for data manipulation, and tidying.
- ✓ arules and arulesviz package are used for association mining and visualizing association respectively.
- ✓ Then the csv file is loaded, checked and analyzed.

## 2. Data Conversion

- ✓ The basket data is converted as factor to input as parameters to the split function.
- ✓ After the split is done, the data frame is converted to transactions to ease the association rule mining.

## 3. Data Visualization

- ✓ The absolute item frequency of the top 30 items in the transactions is shown in the figure at the right.
- ✓ The kitchen gadgets and dishes are highest bought item and household helpers is 30th frequently bought item.
- ✓ So, to increasing the sales of cutlery items, retailer can put it near kitchen gadgets.
- ✓ The summary(retail\_transactions) function gave the top five items as Kitchen gadgets, dishes, lighting, bathroom accessories and bags with the frequencies 20250, 16898, 15312, 14375, and 12680 respectively.

## 4. Association Rule Visualization - I

- ✓ The scatter plot illustrates the 85 rules produced by apriori algorithm at support=0.001 and confidence=0.1
- ✓ The darker the color of the circle, higher the value of the lift.
- ✓ According to fig 4 of the slide, the higher the correlation between the itemset (lift value), lower is the support.

## 5. Association Rule Visualization - II

- ✓ The scatter plot illustrates the 3 rules produced by apriori algorithm at support=0.002 and confidence=0.2.
- ✓ It shows only 3 rules at the minimum threshold support of 0.002 and min. threshold confidence of 0.2.
- ✓ There is only one rule with high lift according to the fig 4 in the slide.

## 6. Annex

```
#removing the environment variables
ls()
rm(list = ls())
ls()

#setting the working directory
getwd()
setwd("F:/R-Data-Mining/ Assignment 2")
getwd()

#loading the package for data manipulation, tidy and visualization
library(tidyverse)

#installing and loading package for a ssociation mining and association visualization
install.packages("arules")
library(arules)
install.packages("arulesViz")
library(arulesViz)

#importing data
marketdata <- read.csv2('exercise_WS2020_data.csv')
marketdata
str(marketdata)

#converting basket as a factor
marketdata$article_group <- factor(marketdata$article_group)
marketdata$article_group

#converting baskets into different kinds of objects
shopping_basket <- split(marketdata$article_group, marketdata$basket_id)
shopping_basket
retail_transactions <- as(shopping_basket, "transactions")
retail_transactions
#summarizing the data
summary(retail_transactions)

#items frequency plot
itemFrequencyPlot(retail_transactions, topN=30, type='absolute')

#using association rules
rule1 <- apriori(retail_transactions, parameter = list(supp=0.001, conf=0.1))
#sorting rule1 generated by apriori algorithm in decreasing confidence value
rule1 <- sort(rule1, by='confidence', decreasing = TRUE)
summary(rule1)
#inspecting a ssociation rules
inspect(rule1)
plot(rule1)

#using association rule with different support and confidence
rule2 <- apriori(retail_transactions, parameter = list(supp=0.002, conf=0.2))
inspect(rule2)
plot(rule2, measure = c("support", "lift"), shading = "confidence")
```