

SUPERMARKET DATA VISUALISATION WITH "R"

supervisor

Prof. Svetlana Meißner

Presented By

Vikas Jivani (28403)

System Simulation, SS2021

- 1 Introduction
- 2 Task Overview
- 3 Item frequency plot
- 4 Generate the basket rule and paracode plot
- 5 Outcomes & Further business prediction
- 6 Code link & Fiels

Introduction

Vikas Jivani

Programing with R



- **Information Engineering and Computer Science**
- **Area of Interest:** Data analysis and machine learning
- **Project:** Data-analysis & visualization with R
- **Supermarket data visualization**
- **R plot functionality & further prediction**
- **Date:** 19 June 2021

Aim : Supermarket data visualizations and further prediction in “R”

- Task
- Create Transaction data set from CSV
 - Create an Item Frequency Plot
 - Generate rule and create the plot for most Frequent Item using paracord method
 - Generate rule and create the plot for most Frequent Item using paracord method

Create transaction data set from csv

Requirement & Installation

- R version 3.6 -> For arulesViz library installation
- Required library
 - library(arulesViz)
 - library(arules)
 - library(tidyverse)

```
> basket_list[1:2]
[[1]]
[1] lunch box          cereal to go box
3099 Levels: 'children's spring' cd 'christmas stories' cd 'guglhupfform' 3er '

[[2]]
[1] chocolate dessert mold      rolling pin set          ironing utensil stati
[6] shower shelf                tray with three herb seeds
3099 Levels: 'children's spring' cd 'christmas stories' cd 'guglhupfform' 3er '
```

Transaction data from csv

```
# Read CSV file from local computer and assign to the data variable
> data <- read.csv( file = "/home/vikas/Documents/academic/system-simulation/data.csv", sep = ";")

# Create Group from basket id for data set
> group_data <- group_by(data, basket_id)

# Make a list of item base on group set
> basket <- summarise(group_data, count = n(), basket_list = list(article_name))

# Filter data first 25000 basket list analysis
> basket_list <- basket$basket_list[1:25000]

# Convert list in to transaction
> retail_transaction <- as(basket_list, "transactions")
```

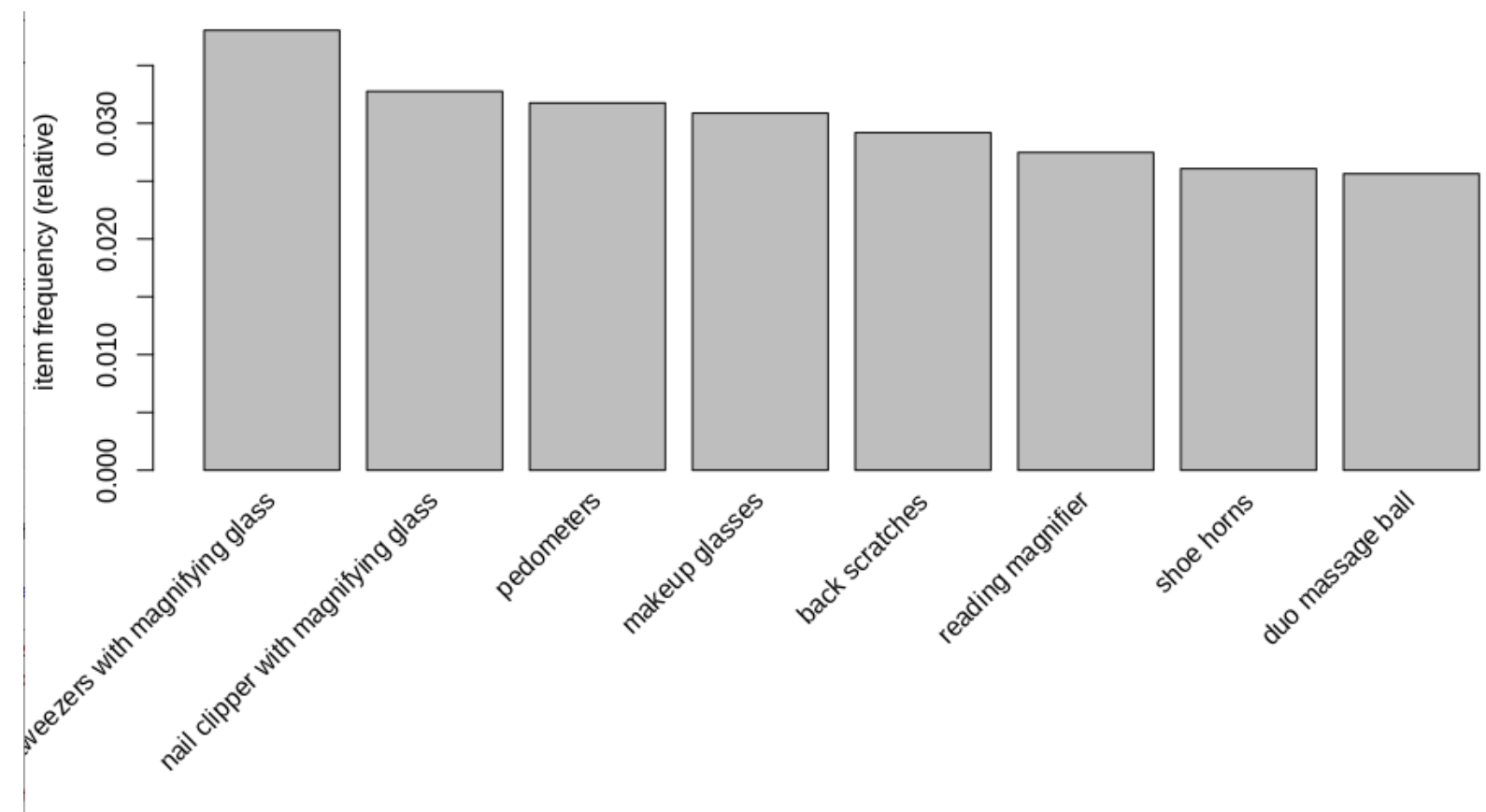
Item frequency plot overview

Transaction data overview

- 25000 transactions (rows) > (filter from 214,530 row)
- 2319 items (columns)

Top-8 most frequent item in shopping basket

- support of item “Tweezers with magnifying glass” is more than 0.03
- more then 3% chance of “Tweezers with magnifying glass” in shopping basket
- Identify E-question :
 - support = $\sigma\{\text{item in set}\} / \{\text{total Transaction}\}$
- Plot generation code
 - > itemFrequencyPlot(retail_transaction, topN = 8)



Generate rule & plot for a most frequent item using paracord method

- Generate rule with support 0.0002 and confidence 0.66 minimum item is 2 and max 3 from create 54 rule

```
> basket_rule <- apriori(retail_transaction, parameter = list(support = 0.0002, conf = 0.66, minlen = 2, maxlen = 3))
```

#filter rule apply on tweezer with magnifying glass

```
> tweezer_rule <- subset(basket_rule, subset = rhs %in% "tweezers with magnifying glass")
```

```
> tweezer_rule_head <- head(sort(tweezer_rule, by = "lift"), 4)
```

```
> inspect(tweezer_rule_head)
```

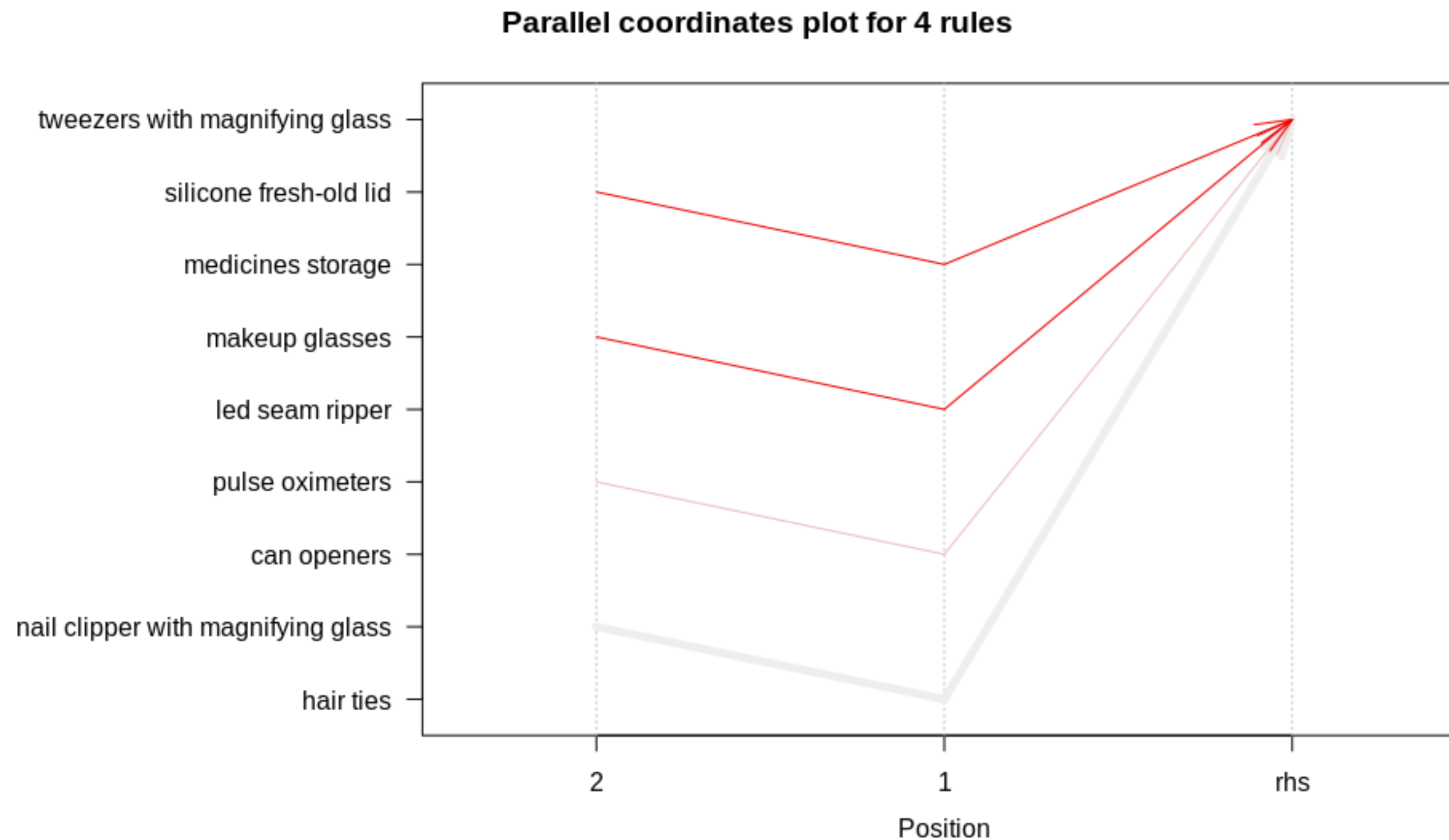
```
> inspect(tweezer_rule_head)
```

lhs	rhs	support	confidence	coverage	lift	count
[1] {led seam ripper,makeup glasses}	=> {tweezers with magnifying glass}	0.00020	0.8333333	0.00024	21.90676	5
[2] {medicines storage,silicone fresh-old lid}	=> {tweezers with magnifying glass}	0.00020	0.8333333	0.00024	21.90676	5
[3] {can openers,pulse oximeters}	=> {tweezers with magnifying glass}	0.00020	0.7142857	0.00028	18.77723	5
[4] {hair ties,nail clipper with magnifying glass}	=> {tweezers with magnifying glass}	0.00024	0.6666667	0.00036	17.52541	6

```
> |
```

Plot of most frequent item with method paracoord

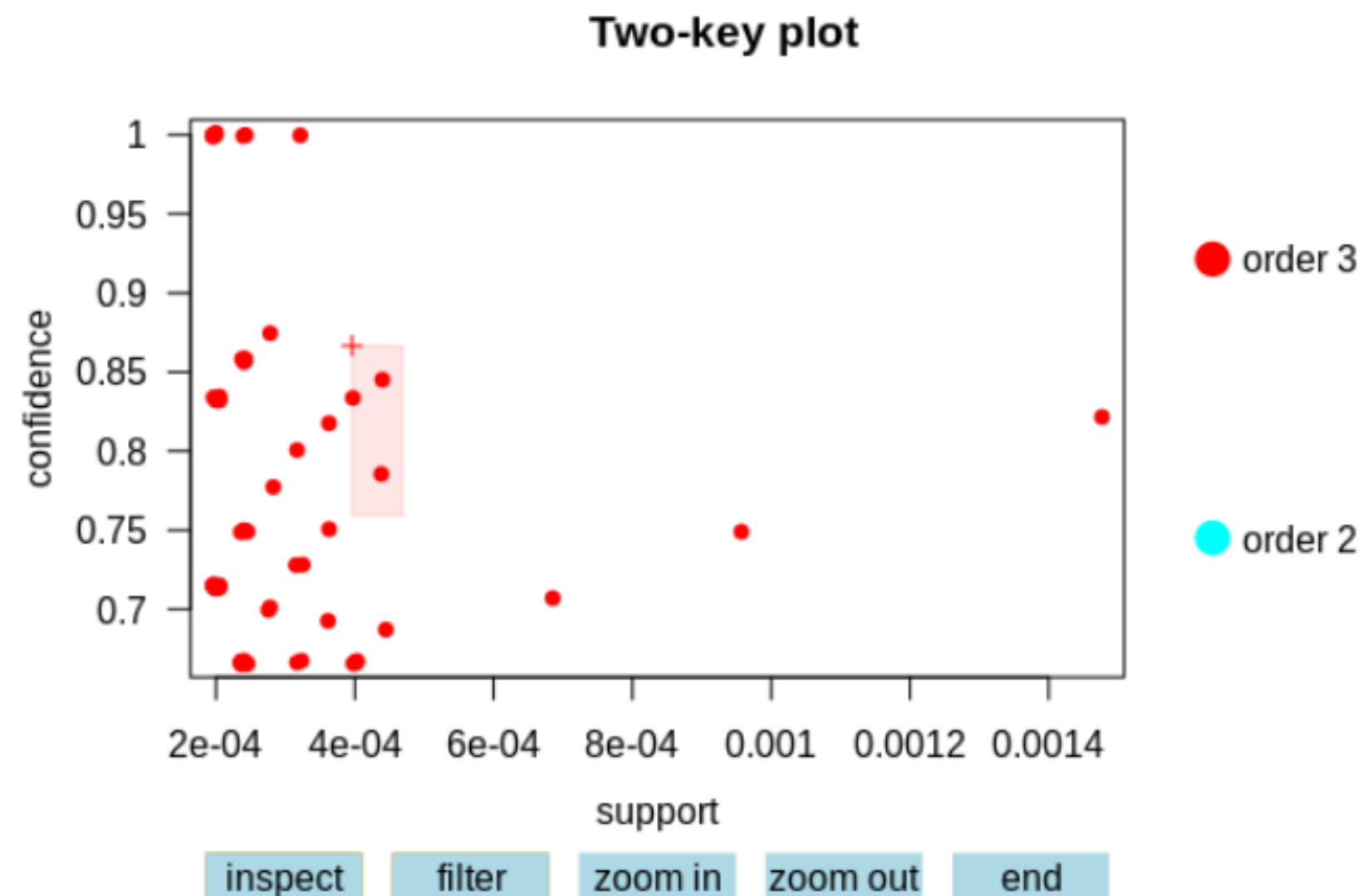
● `> plot(tweezer_rule_head,method="paracoord")`



Generate rule for entire basket and create plot with two-key plot method

> plot(basket_rule,method = "two-key plot",interactive = TRUE)

NOTE : Selected data shows the inspect of item list from “Two-key plot” graph



Selected data of region

Number of rules selected: 3

lhs	rhs	support	conf
[1] {crystal rocks ms bracelet,ms ear}	=> {ms bracelet}	0.00040	0.83
[2] {bath buckets,toilet paper holder black}	=> {hooks}	0.00044	0.84
[3] {ms chain,ms ring crystal gray}	=> {ms ring}	0.00044	0.78

Outcomes & Further Prediction

Outcomes

- The item “tweezer with a magnifying glass” and “nail clipper with magnifying glass” both the most frequent items from the same category so these two product keep gathering in shop. For easy to customer.

Further Prediction

- Also, identify which category of the item should keep together
- Which item should close to selling due to very low frequency of buy.
- How to arrange section of item with daily usage

Project and Code Link

The project code push on GitHub repository link has been given below

Code Link

- **GitHub:** <https://github.com/vji-axelor/simulation>

Related Fiels

- **data.csv:** Supermarket transaction data records
- **documents.pdf:** The brief description about code and project
- **system-simulation.r :** R code with comments
- **presentation.pdf :** Presentation with pdf format
- **Simulation.pptx :** Presentation with power-point format

Thank You...!



M-IE_1.01 System Simulation, SS2021