



Δίκτυα Chord & Ανάλυση Δεδομένων χρησιμοποιώντας τη γλώσσα R

Μαλλιαρού Σοφία (ΑΜ: 212)
Ραπτοπούλου Κωνσταντία (ΑΜ: 180)

Καβάλα 2024

Πίνακας περιεχομένων

Περιεχόμενα Πίνακα Γραφικών Παραστάσεων	4
Περιεχόμενα Πίνακα εικόνων.....	4
Περιεχόμενα Πίνακα Θερμικού χάρτη	4
ΘΕΜΑ 1ο.....	5
A.1. CHORD ΔΙΚΤΥΟ.....	5
1.1.1. Υπολογισμός κόμβων:.....	5
1.1.2. Κατανομή των κόμβων:.....	6
1.1.3. Δημιουργία Finger table για κάθε κόμβο:	6
1.1.4. Αναζητήσεις κόμβων:	7
ΘΕΜΑ 2ο.....	10
A.2. Πακέτο R arules, dataset 'Groceries'	10
2.1. Μελέτη των Δεδομένων: Ανάλυση και Καταγραφή Χαρακτηριστικών.....	10
2.1.1. Εισαγωγή	10
2.1.2. Τύπος δεδομένων	11
2.1.3. Αριθμός συναλλαγών	12
2.1.4. Κατηγορίες προϊόντων	12
2.1.5. Οπτικοποίηση δεδομένων.....	13
2.1.6. Σύνοψη των Αποτελεσμάτων	14
2.2. Απάντηση ερωτημάτων:.....	15
2.2.1 Να εντοπίσετε το index των προϊόντων citrus fruit, semi-finished bread, margarine, ready soups.....	15
2.2.2 Δημιουργία γραφικής παράστασης με τα προϊόντα που εμφανίζονται πιο συχνά	16
2.2.3 Δημιουργία γραφικής παράστασης με τα προϊόντα που αγοράζονται πιο σπάνια	18
2.2.4 Πυκνότητα ή Αραιότητα των Δεδομένων.....	20
2.2.5 Εφαρμογή αλγόριθμου Apriori.....	22
2.2.6 Εντοπισμός των top 5 συνόλων στοιχείων με το υψηλότερο επίπεδο confidence.	25
2.2.7 Δημιουργία Γραφικής Παράστασης για τη Σχέση μεταξύ Σπάνια και Συχνά Αγορασμένων Προϊόντων	27
2.2.8 Μοτίβο των αγορών των καταναλωτών.....	31
2.2.9 Συνδυασμοί που οδηγούν στην αγορά γάλακτος και η γραφική απεικόνιση αυτού	37
Ηλεκτρονικές πηγές	46

Περιεχόμενα Σχημάτων

Σχήμα 1 Απεικόνιση κόμβων	6
Σχήμα 2: Απεικόνιση διαδρομής ευρέσεως κλειδιού K212->20	8
Σχήμα 3: Απεικόνιση διαδρομής ευρέσεως κλειδιού K180->52	9

Περιεχόμενα Πινάκων

Πίνακας 1: Finger Table N0	6
Πίνακας 2: Finger Table N7	7
Πίνακας 3: Finger Table N15	7
Πίνακας 4: Finger Table N22	7
Πίνακας 5: Finger Table N30	7
Πίνακας 6: Finger Table N37	7
Πίνακας 7: Finger Table N45	7
Πίνακας 8: Finger Table N52	7

Περιεχόμενα Πίνακα Γραφικών Παραστάσεων

Γραφική παράσταση 1 Top 20 προϊόντα ανά συχνότητα.....	17
Γραφική παράσταση 2 Σπάνια αγορασμένα προϊόντα	20
Γραφική παράσταση 3 Γραφικής Παράσταση μοτίβο καταναλωτών με arulesViz	35
Γραφική παράσταση 4 Top 10 Κανόνες Συσχέτισης με το Υψηλότερο Lift	36
Γραφική παράσταση 5 Top Προϊόντα που Συνοδεύουν το Whole Milk.	42

Περιεχόμενα Πίνακα εικόνων

Εικόνα 1 Το index των προϊόντων citrus fruit, semi-finished bread, margarine, ready soups.....	16
Εικόνα 2 Προϊόντα που εμφανίζονται πιο συχνά.....	17
Εικόνα 3 Πυκνότητα ή Αραιότητα των Δεδομένων	21
Εικόνα 4 Αλγόριθμος Apriori.	23
Εικόνα 5 Εφαρμογή αλγόριθμου Apriori	25
Εικόνα 6 Εντοπισμός των top 5 συνόλων στοιχείων με το υψηλότερο επίπεδο confidence.	27
Εικόνα 7 Συχνότητα εμφάνισης των προϊόντων στο dataset	29
Εικόνα 8 Δημιουργία πίνακα συνάφειας.....	29
Εικόνα 9 Μετατροπή του Πίνακα Συνάφειας σε Dataframe	30
Εικόνα 10 Μοτίβο καταναλωτών.....	33
Εικόνα 11 Groceries	38
Εικόνα 12 Milk rules.....	39

Περιεχόμενα Πίνακα Θερμικού χάρτη

Θερμικός χάρτης 1 Σχέση μεταξύ Συχνά και Σπάνια Αγορασμένων Προϊόντων.....	31
----------------------------------------------------------------------------	----

ΘΕΜΑ 1ο

A.1. CHORD ΔΙΚΤΥΟ

1.1.1. Υπολογισμός κόμβων:

Σύμφωνα με τα δεδομένα του βοηθητικού αρχείου Excel , οι κόμβοι που προκύπτουν από τους αριθμούς μητρώων είναι οι εξής:

$$AM1 = 212$$

$$AM2 = 180$$

$$\text{Αριθμός κόμβων} = 8 \rightarrow$$

$$[(212+180)\bmod 8]+8=0+8= 8$$

$$\text{Βήμα} = 7$$

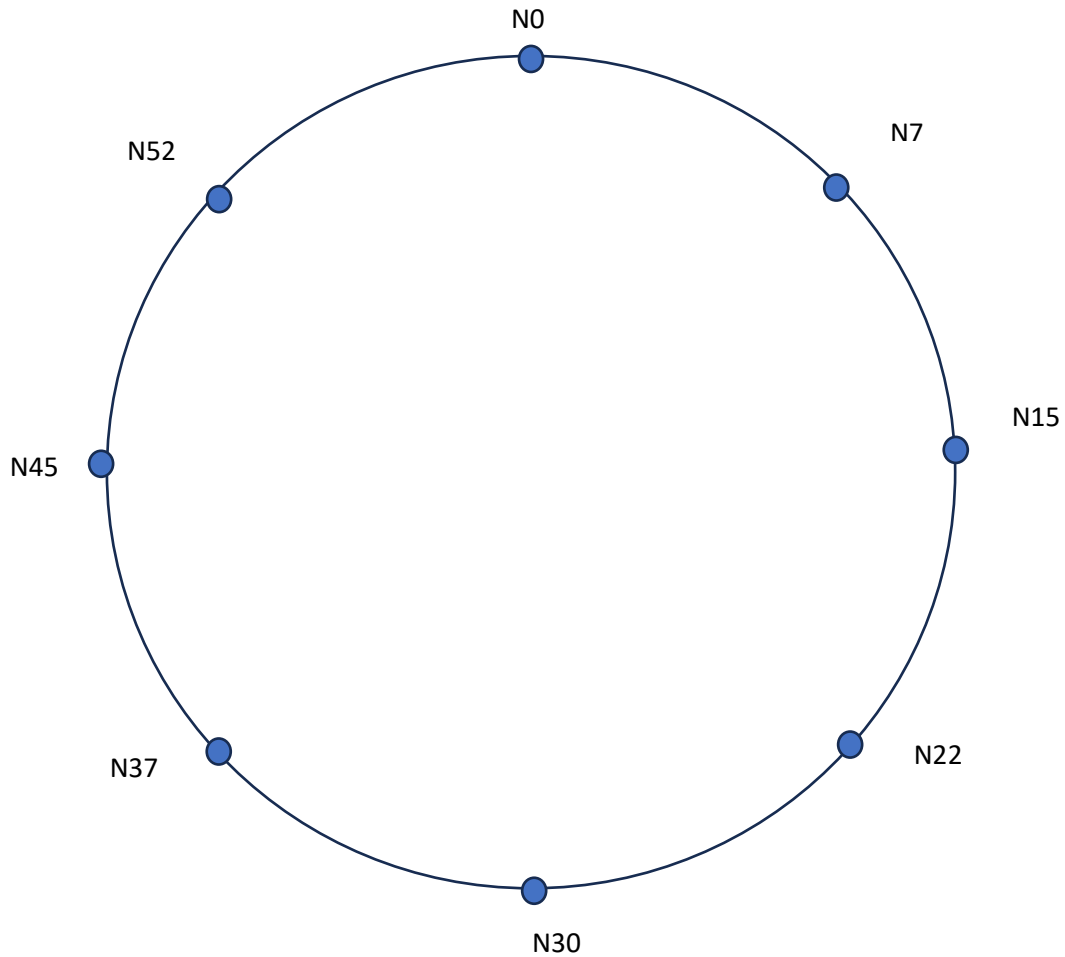
AM1	212
AM2	180
Αριθμός Κόμβων	8

Βήμα	7
------	---

	Θέση Κόμβων
1	0
2	7
3	15
4	22
5	30
6	37
7	45
8	52

1.1.2. Κατανομή των κόμβων:

Σχηματικά απεικονίζονται ως εξής:



Σχήμα 1 Απεικόνιση κόμβων

1.1.3. Δημιουργία Finger table για κάθε κόμβο:

N0			
$N0+2^0$	N0+1	N1 →	N7
$N0+2^1$	N0+2	N2 →	N7
$N0+2^2$	N0+4	N4 →	N7
$N0+2^3$	N0+8	N8 →	N15
$N0+2^4$	N0+16	N16 →	N22
$N0+2^5$	N0+32	N32 →	N37

Πίνακας 1: Finger Table N0

N7			
$N7+2^0$	N7+1	N8 →	N15
$N7+2^1$	N7+2	N9 →	N15
$N7+2^2$	N7+4	N11 →	N15
$N7+2^3$	N7+8	N15 →	N15
$N7+2^4$	N7+16	N23 →	N30
$N7+2^5$	N7+32	N39 →	N45

Πίνακας 2: Finger Table N7

N15			
$N15+2^0$	$N15+1$	$N16 \rightarrow$	N22
$N15+2^1$	$N15+2$	$N17 \rightarrow$	N22
$N15+2^2$	$N15+4$	$N19 \rightarrow$	N22
$N15+2^3$	$N15+8$	$N23 \rightarrow$	N30
$N15+2^4$	$N15+16$	$N31 \rightarrow$	N37
$N15+2^5$	$N15+32$	$N47 \rightarrow$	N52

Πίνακας 3: Finger Table N15

N22			
$N22+2^0$	$N22+1$	$N23 \rightarrow$	N30
$N22+2^1$	$N22+2$	$N24 \rightarrow$	N30
$N22+2^2$	$N22+4$	$N26 \rightarrow$	N30
$N22+2^3$	$N22+8$	$N30 \rightarrow$	N30
$N22+2^4$	$N22+16$	$N38 \rightarrow$	N45
$N22+2^5$	$N22+32$	$N54 \rightarrow$	N0

Πίνακας 4: Finger Table N22

N30			
$N30+2^0$	$N30+1$	$N31 \rightarrow$	N37
$N30+2^1$	$N30+2$	$N32 \rightarrow$	N37
$N30+2^2$	$N30+4$	$N34 \rightarrow$	N37
$N30+2^3$	$N30+8$	$N38 \rightarrow$	N45
$N30+2^4$	$N30+16$	$N46 \rightarrow$	N52
$N30+2^5$	$N30+32$	$N62 \rightarrow$	N0

Πίνακας 5: Finger Table N30

N37			
$N37+2^0$	$N37+1$	$N38 \rightarrow$	N45
$N37+2^1$	$N37+2$	$N39 \rightarrow$	N45
$N37+2^2$	$N37+4$	$N41 \rightarrow$	N45
$N37+2^3$	$N37+8$	$N45 \rightarrow$	N45
$N37+2^4$	$N37+16$	$N53 \rightarrow$	N0
$N37+2^5$	$N37+32$	$N69 \rightarrow$	N7

Πίνακας 6: Finger Table N37

N45			
$N45+2^0$	$N45+1$	$N46 \rightarrow$	N52
$N45+2^1$	$N45+2$	$N47 \rightarrow$	N52
$N45+2^2$	$N45+4$	$N49 \rightarrow$	N52
$N45+2^3$	$N45+8$	$N53 \rightarrow$	N0
$N45+2^4$	$N45+16$	$N61 \rightarrow$	N0
$N45+2^5$	$N45+32$	$N77 \rightarrow$	N15

Πίνακας 7: Finger Table N45

N52			
$N52+2^0$	$N52+1$	$N53 \rightarrow$	N0
$N52+2^1$	$N52+2$	$N54 \rightarrow$	N0
$N52+2^2$	$N52+4$	$N56 \rightarrow$	N0
$N52+2^3$	$N52+8$	$N60 \rightarrow$	N0
$N52+2^4$	$N52+16$	$N68 \rightarrow$	N7
$N52+2^5$	$N52+32$	$N84 \rightarrow$	N22

Πίνακας 8: Finger Table N52

1.1.4. Αναζητήσεις κόμβων:

Αναζήτηση δύο κλειδιών, η πρώτη αναζήτηση αφορά το κλειδί K₂₁₂ $[(212) \bmod 64]$ και η δεύτερη αφορά το κλειδί K₁₈₀ $[(180) \bmod 64]$.

A. Κλειδί K₂₁₂

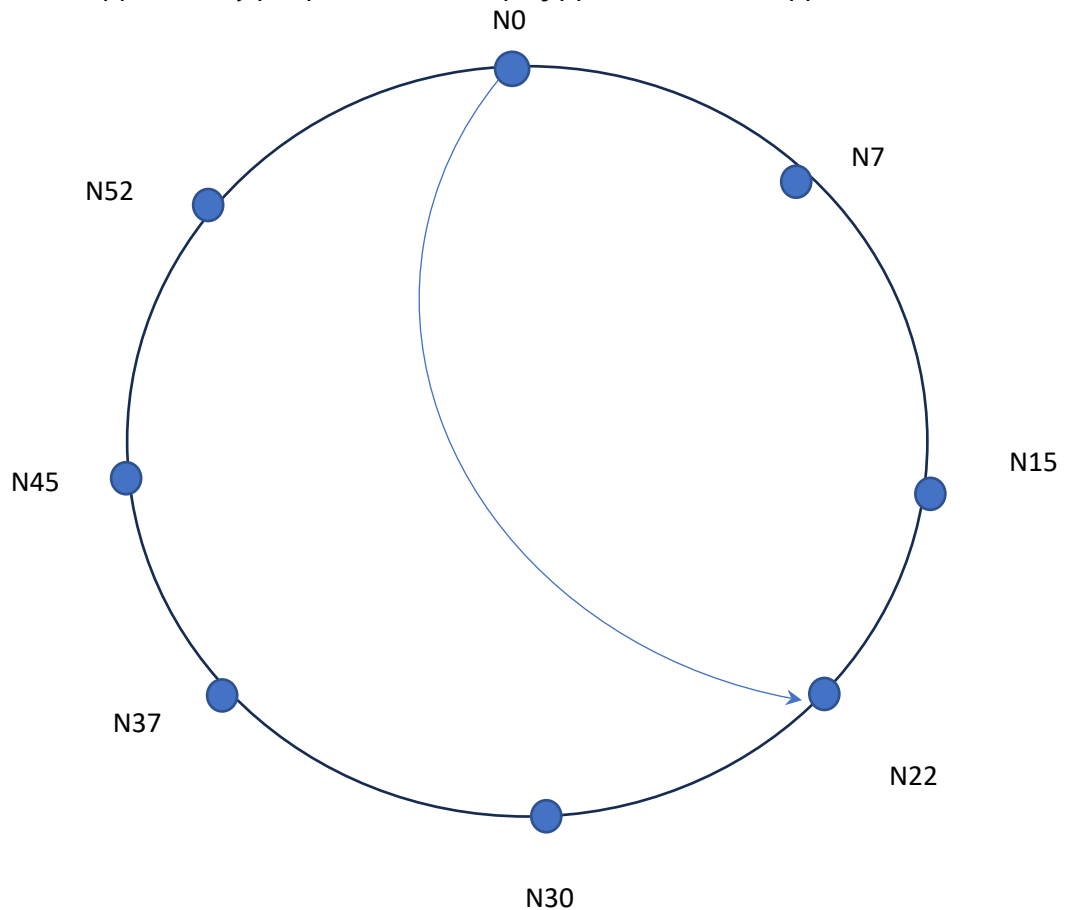
Για το κλειδί K₂₁₂, όπου

$$K_{212} [(212) \bmod 64] = 20$$

Ξεκινάμε την αναζήτηση από τον κόμβο N0.

1. Κόμβος N0:

- Κοιτάζουμε τον finger table του N0 και πάμε τα βήματα ένα ένα:
 Finger N1 → N7,
 Finger N2 → N7,
 Finger N4 → N7,
 Finger N8 → N15, και φτάνουμε στο
 Finger N16 → N22,
 Finger N32 → N37
 - Το κλειδί 20 βρίσκεται μετά τον κόμβο N15, αλλά επειδή δεν έχουμε ακριβώς τον κόμβο N20, ξέρουμε ότι το κλειδί μας βρίσκεται στον κόμβο N22.



Σχήμα 2: Απεικόνιση διαδρομής ευρέσεως κλειδιού K212->20

B. Κλειδί K180

Για το κλειδί K180, όπου

$$K180 [(180) \bmod 64] = 52$$

Ξεκινάμε την αναζήτηση από τον κόμβο N0.

1. Κόμβος N0:
 - Κοιτάζουμε τον finger table του N0:
 Finger N1 → N7,
 Finger N2 → N7,
 Finger N4 → N7,

Finger N8 \rightarrow N15,
Finger N16 \rightarrow N22,
Finger N32 \rightarrow N37

- Το κλειδί 52 βρίσκεται μετά το N32, οπότε κατευθυνόμαστε στον κόμβο N37.

2. Κόμβος N37:

- Κοιτάζουμε Finger table του N37:

Finger N38 \rightarrow N45,
Finger N39 \rightarrow N45,
Finger N41 \rightarrow N45,
Finger N45 \rightarrow N45,
Finger N53 \rightarrow N0,
Finger N69 \rightarrow N7

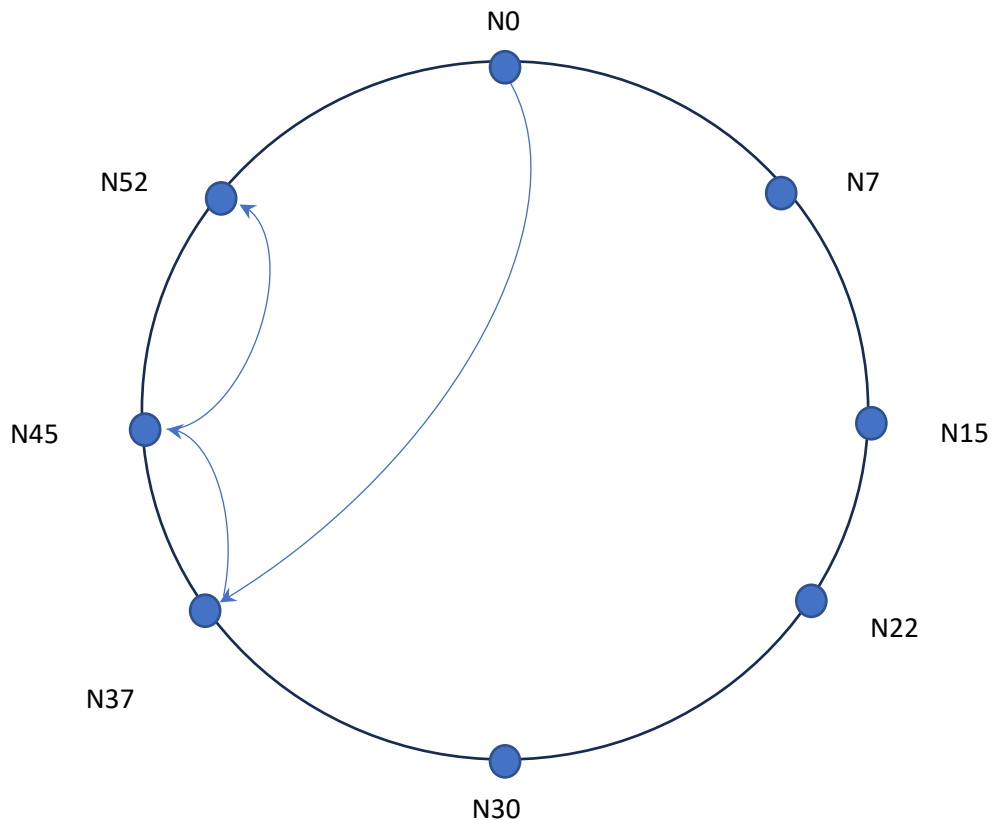
- Το κλειδί 52 βρίσκεται μετά τον N45 και πριν τον N53, οπότε κατευθυνόμαστε στον κόμβο N45.

3. Κόμβος N45:

- Κοιτάζουμε Finger table του N45:

Finger N46 \rightarrow N52,
Finger N47 \rightarrow N52,
Finger N49 \rightarrow N52,
Finger N53 \rightarrow N0,
Finger N61 \rightarrow N0,
Finger N77 \rightarrow N15

- Το κλειδί 52 βρίσκεται ακριβώς στον κόμβο N52, οπότε σταματάμε στον κόμβο N52



Σχήμα 3: Απεικόνιση διαδρομής ευρέσεως κλειδιού K180->52

ΘΕΜΑ 2ο

A.2. Πακέτο R arules, dataset 'Groceries'

2.1. Μελέτη των Δεδομένων: Ανάλυση και Καταγραφή Χαρακτηριστικών

2.1.1. Εισαγωγή

Η ανάλυση των δεδομένων αποτελεί ένα κρίσιμο βήμα για την κατανόηση και την αξιοποίηση ενός συνόλου δεδομένων. Στην παρούσα εργασία, θα μελετήσουμε το dataset 'Groceries', το οποίο είναι ενσωματωμένο στο πακέτο arules της γλώσσας R. Αρχικά, θα φορτώσουμε τα δεδομένα και θα εξετάσουμε τα βασικά χαρακτηριστικά τους.

*Αρχικά θα δίνονται οι εντολές για Run και μετά το βελάκι το αποτέλεσμα στην Console R 4.4.0

```
getwd()
setwd("C:/Users/mtsilogl/OneDrive/Εγγράφα/")
install.packages("arules")
library(arules)
install.packages("ggplot2")
library(ggplot2)
data("Groceries")
install.packages("Matrix")
library(Matrix) →
```

```
> getwd()
[1] "C:/Users/mtsilogl/OneDrive/Εγγράφα/"
> setwd("C:/Users/mtsilogl/OneDrive/Εγγράφα/Arm2")
> install.packages("arules")
Error in install.packages : Updating loaded packages

Restarting R session...

> install.packages("arules")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/mtsilogl/AppData/Local/R/win-library/4.4'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.4/arules_1.7-7.zip'
Content type 'application/zip' length 2614390 bytes (2.5 MB)
downloaded 2.5 MB

package 'arules' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:\Users\mtsilogl\AppData\Local\Temp\RtmpCAeACC\downloaded_packages
Loading required package: arules
Loading required package: Matrix
```

```
Attaching package: 'arules'
```

```
The following objects are masked from 'package:base':
```

```
abbreviate, write
```

```
> library(arules)
> data("Groceries")
> install.packages("Matrix")
Error in install.packages : Updating loaded packages
```

```
Restarting R session...
```

```
> install.packages("Matrix")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:
```

```
https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/mtsilogl/AppData/Local/R/win-library/4.4'
```

```
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.4/Matrix_1.7-0.zip'
Content type 'application/zip' length 4978220 bytes (4.7 MB)
downloaded 4.7 MB
```

```
package 'Matrix' successfully unpacked and MD5 sums checked
```

```
The downloaded binary packages are in
C:\Users\mtsilogl\AppData\Local\Temp\Rtmp6vKZ2d\downloaded_packages
```

```
Loading required package: arules
Loading required package: Matrix
```

```
Attaching package: 'arules'
```

```
The following objects are masked from 'package:base':
```

```
abbreviate, write
```

```
> library(Matrix)
>
```

Στην ανάλυση των δεδομένων του dataset 'Groceries', εξετάσαμε τα χαρακτηριστικά του dataset και καταγράψαμε τις βασικές πληροφορίες. Τα δεδομένα αφορούν συναλλαγές προϊόντων σε ένα παντοπωλείο και περιλαμβάνουν διάφορα προϊόντα κατηγοριοποιημένα ανά τύπο.

2.1.2. Τύπος δεδομένων

Ο τύπος των δεδομένων είναι θεμελιώδης για την κατανόηση της φύσης και της δομής τους. Στην προκειμένη περίπτωση, τα δεδομένα του dataset 'Groceries' είναι τύπου "transactions". Αυτό υποδεικνύει ότι τα δεδομένα καταγράφουν συναλλαγές που πραγματοποιήθηκαν σε ένα παντοπωλείο, περιλαμβάνοντας πληροφορίες σχετικά με τα προϊόντα που αγοράστηκαν σε κάθε συναλλαγή. Η χρήση του κατάλληλου τύπου δεδομένων

επιτρέπει την εφαρμογή εξειδικευμένων αλγορίθμων και μεθόδων ανάλυσης για τη μελέτη της αγοραστικής συμπεριφοράς.

```
cat("Τύπος των δεδομένων:", class(Groceries), "\n") →
```

```
> # Τύπος των Δεδομένων  
> cat("Τύπος των δεδομένων:", class(Groceries), "\n")  
Τύπος των δεδομένων: transactions  
>
```

2.1.3. Αριθμός συναλλαγών

Στη συνέχεια, υπολογίσαμε τον αριθμό των συναλλαγών που περιλαμβάνονται στο dataset. Ο αριθμός των συναλλαγών είναι σημαντικός για να κατανοήσουμε το μέγεθος του dataset και να αξιολογήσουμε την επάρκεια των δεδομένων για την ανάλυση. Η εντολή `length(Groceries)` επιστρέφει τον αριθμό των συναλλαγών, ο οποίος είναι 9835. Αυτό υποδεικνύει ότι το dataset περιλαμβάνει 9835 ξεχωριστές αγορές.

```
cat("Αριθμός συναλλαγών:", length(Groceries), "\n") →
```

```
> # Αριθμός Συναλλαγών  
> cat("Αριθμός συναλλαγών:", length(Groceries), "\n")  
Αριθμός συναλλαγών: 9835  
>
```

2.1.4. Κατηγορίες προϊόντων

Ένα άλλο σημαντικό χαρακτηριστικό των δεδομένων είναι ο αριθμός των κατηγοριών προϊόντων που περιλαμβάνονται. Στην ανάλυσή μας, εντοπίσαμε ότι υπάρχουν 169 διαφορετικές κατηγορίες προϊόντων. Αυτές οι κατηγορίες ποικίλουν από βασικά είδη διατροφής, όπως φρούτα και λαχανικά, μέχρι ειδικά προϊόντα, όπως καλλυντικά και προϊόντα καθαρισμού. Ο αριθμός των κατηγοριών των προϊόντων αλλά και οι ίδιες οι κατηγορίες είναι χρήσιμες πληροφορίες για την κατανόηση της ποικιλίας των προϊόντων και τη συχνότητα εμφάνισής τους στις συναλλαγές. Η εντολή `itemLabels(Groceries)` επιστρέφει όλες τις κατηγορίες προϊόντων, ενώ η εντολή `length(itemLabels(Groceries))` υπολογίζει τον αριθμό αυτών των κατηγοριών.

Οι κατηγορίες προϊόντων περιλαμβάνουν μια μεγάλη ποικιλία από τρόφιμα και άλλα αγαθά, όπως "frankfurter", "sausage", "citrus fruit", "tropical fruit", "whole milk", "yogurt", "frozen vegetables", "domestic eggs", "rolls/buns", "semi-finished bread", "margarine", "ready

soups", και πολλά άλλα. Αυτή η ποικιλία δείχνει την ευρεία γκάμα προϊόντων που είναι διαθέσιμα στους καταναλωτές και μπορούν να αγοραστούν σε ένα παντοπωλείο.

```
itemLabels <- itemLabels(Groceries)
cat("Αριθμός κατηγοριών προϊόντων:", length(itemLabels), "\n")
cat("Κατηγορίες προϊόντων:", paste(itemLabels, collapse = ", "),
"\n")
```

→

```
> itemLabels <- itemLabels(Groceries)
> cat("Αριθμός κατηγοριών προϊόντων:", length(itemLabels), "\n")
Αριθμός κατηγοριών προϊόντων: 169
> cat("Κατηγορίες προϊόντων:", paste(itemLabels, collapse = ", "), "\n")
Κατηγορίες προϊόντων: frankfurter, sausage, liver loaf, ham, meat,
finished products, organic sausage, chicken, turkey, pork, beef, hamburger
meat, fish, citrus fruit, tropical fruit, pip fruit, grapes, berries,
nuts/prunes, root vegetables, onions, herbs, other vegetables, packaged
fruit/vegetables, whole milk, butter, curd, dessert, butter milk, yogurt,
whipped/sour cream, beverages, UHT-milk, condensed milk, cream, soft
cheese, sliced cheese, hard cheese, cream cheese, processed cheese,
spread cheese, curd cheese, specialty cheese, mayonnaise, salad dressing,
tidbits, frozen vegetables, frozen fruits, frozen meals, frozen fish,
frozen chicken, ice cream, frozen dessert, frozen potato products,
domestic eggs, rolls/buns, white bread, brown bread, pastry, roll products
, semi-finished bread, zwieback, potato products, flour, salt, rice,
pasta, vinegar, oil, margarine, specialty fat, sugar, artif. sweetener,
honey, mustard, ketchup, spices, soups, ready soups, Instant food
products, sauces, cereals, organic products, baking powder, preservation
products, pudding powder, canned vegetables, canned fruit, pickled
vegetables, specialty vegetables, jam, sweet spreads, meat spreads, canned
fish, dog food, cat food, pet care, baby food, coffee, instant coffee,
tea, cocoa drinks, bottled water, soda, misc. beverages, fruit/vegetable
juice, syrup, bottled beer, canned beer, brandy, whisky, liquor, rum,
liqueur, liquor (appetizer), white wine, red/blush wine, prosecco,
sparkling wine, salty snack, popcorn, nut snack, snack products, long life
bakery product, waffles, cake bar, chewing gum, chocolate, cooking
chocolate, specialty chocolate, specialty bar, chocolate marshmallow,
candy, seasonal products, detergent, softener, decalcifier, dish cleaner,
abrasive cleaner, cleaner, toilet cleaner, bathroom cleaner, hair spray,
dental care, male cosmetics, make up remover, skin care, female sanitary
products, baby cosmetics, soap, rubbing alcohol, hygiene articles,
napkins, dishes, cookware, kitchen utensil, cling film/bags, kitchen
towels, house keeping products, candles, light bulbs, sound storage
medium, newspapers, photo/film, pot plants, flower soil/fertilizer, flower
(seeds), shopping bags, bags
```

2.1.5. Οπτικοποίηση δεδομένων

Για την οπτικοποίηση των δεδομένων μας, δημιουργήσαμε γραφικές παραστάσεις τις οποίες θα παρουσιάσουμε και θα αναλύσουμε στα επόμενα στάδια της εργασίας. Αυτές οι παραστάσεις, για παράδειγμα, μας επιτρέπουν να κατανοήσουμε καλύτερα τις συχνότητες εμφάνισης των προϊόντων και τις αγοραστικές τάσεις των καταναλωτών.

2.1.6. Σύνοψη των Αποτελεσμάτων

- ❖ **Τύπος δεδομένων:** transactions
- ❖ **Αριθμός συναλλαγών:** 9835
- ❖ **Αριθμός κατηγοριών προϊόντων:** 169
- ❖ **Κατηγορίες προϊόντων:** Περιλαμβάνονται όλες οι καταγεγραμμένες κατηγορίες προϊόντων, από τρόφιμα μέχρι προϊόντα καθαρισμού και καλλυντικά.

Αυτή η πρώτη ανάλυση των δεδομένων μας παρέχει τη βάση για περαιτέρω διερεύνηση και εξαγωγή χρήσιμων πληροφοριών σχετικά με τις αγοραστικές συνήθειες των καταναλωτών. Στα επόμενα βήματα, θα εμβαθύνουμε στην ανάλυση των δεδομένων, εντοπίζοντας συγκεκριμένα προϊόντα και εφαρμόζοντας αλγόριθμους για την εξαγωγή χρήσιμων κανόνων συσχέτισης.

2.2. Απάντηση ερωτημάτων:

2.2.1 Να εντοπίσετε το index των προϊόντων citrus fruit, semi-finished bread, margarine, ready soups

Για την περαιτέρω ανάλυση των δεδομένων μας, εντοπίσαμε τα συγκεκριμένα προϊόντα "citrus fruit", "semi-finished bread", "margarine", και "ready soups". Χρησιμοποιώντας τη συνάρτηση `which` της γλώσσας R, προσδιορίσαμε τους δείκτες αυτών των προϊόντων στο dataset. Οι δείκτες αυτοί είναι ουσιώδεις για την αναγνώριση των προϊόντων μέσα στο σύνολο δεδομένων και την εκτέλεση περαιτέρω αναλύσεων.

Εντοπισμός των δεικτών των συγκεκριμένων προϊόντων

```
citrus_fruit_index <- which(itemLabels(Groceries) == "citrus fruit")
```

```
semi_finished_bread_index <- which(itemLabels(Groceries) == "semi-finished bread")
```

```
margarine_index <- which(itemLabels(Groceries) == "margarine")
```

```
ready_soups_index <- which(itemLabels(Groceries) == "ready soups")
```

Εκτύπωση των δεικτών

```
cat("Index of citrus fruit:", citrus_fruit_index, "\n")
```

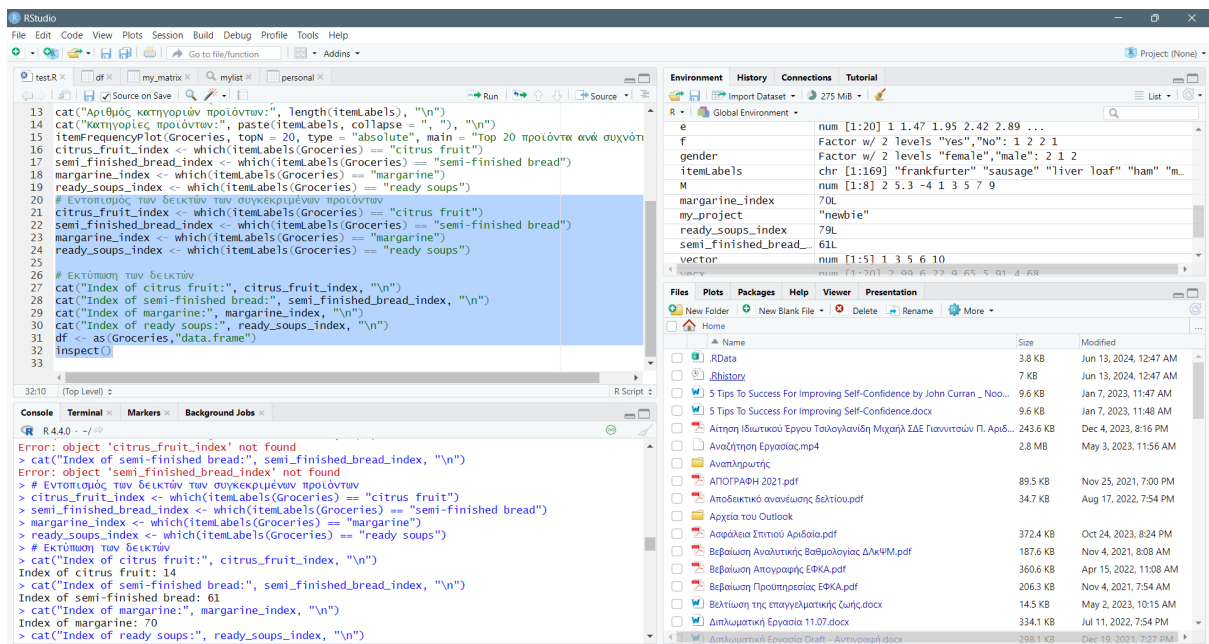
```
cat("Index of semi-finished bread:", semi_finished_bread_index, "\n")
```

```
cat("Index of margarine:", margarine_index, "\n")
```

```
cat("Index of ready soups:", ready_soups_index, "\n")
```

↓

```
> # Εντοπισμός των δεικτών των συγκεκριμένων προϊόντων
> citrus_fruit_index <- which(itemLabels(Groceries) == "citrus fruit")
> semi_finished_bread_index <- which(itemLabels(Groceries) == "semi-finished bread")
> margarine_index <- which(itemLabels(Groceries) == "margarine")
> ready_soups_index <- which(itemLabels(Groceries) == "ready soups")
> # Εκτύπωση των δεικτών
> cat("Index of citrus fruit:", citrus_fruit_index, "\n")
Index of citrus fruit: 14
> cat("Index of semi-finished bread:", semi_finished_bread_index, "\n")
Index of semi-finished bread: 61
> cat("Index of margarine:", margarine_index, "\n")
Index of margarine: 70
> cat("Index of ready soups:", ready_soups_index, "\n")
Index of ready soups: 79
>
```



Εικόνα 1 To index των προϊόντων citrus fruit, semi-finished bread, margarine, ready soups

Οι δείκτες των συγκεκριμένων προϊόντων είναι οι εξής:

- Index of citrus fruit: 14
- Index of semi-finished bread: 61
- Index of margarine: 70
- Index of ready soups: 79

Οι δείκτες αυτοί μας επιτρέπουν να εντοπίσουμε τα συγκεκριμένα προϊόντα στο dataset, διευκολύνοντας την ανάλυση της συχνότητας εμφάνισής τους καθώς και τη διερεύνηση της σχέσης τους με άλλα προϊόντα.

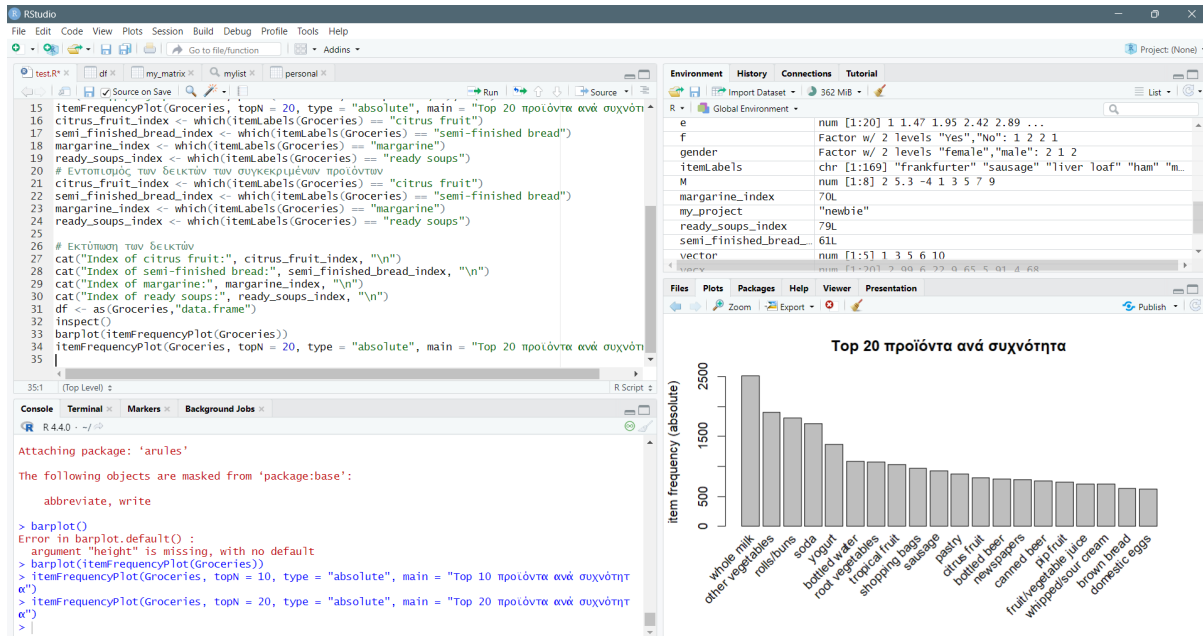
2.2.2 Δημιουργία γραφικής παράστασης με τα προϊόντα που εμφανίζονται πιο συχνά

Για την καλύτερη κατανόηση στις αγοραστικές συνήθειες των καταναλωτών, δημιουργήσαμε μια γραφική παράσταση που απεικονίζει τα προϊόντα που εμφανίζονται πιο συχνά στο dataset. Η γραφική αυτή παράσταση μας βοηθά να εντοπίσουμε τα προϊόντα που αγοράζονται συχνότερα και να κατανοήσουμε τις προτιμήσεις των καταναλωτών.

Η παρακάτω γραφική παράσταση (Γραφική παράσταση 1) δείχνει τα είκοσι (20) προϊόντα τα οποία εμφανίζονται πιο συχνά στο dataset.

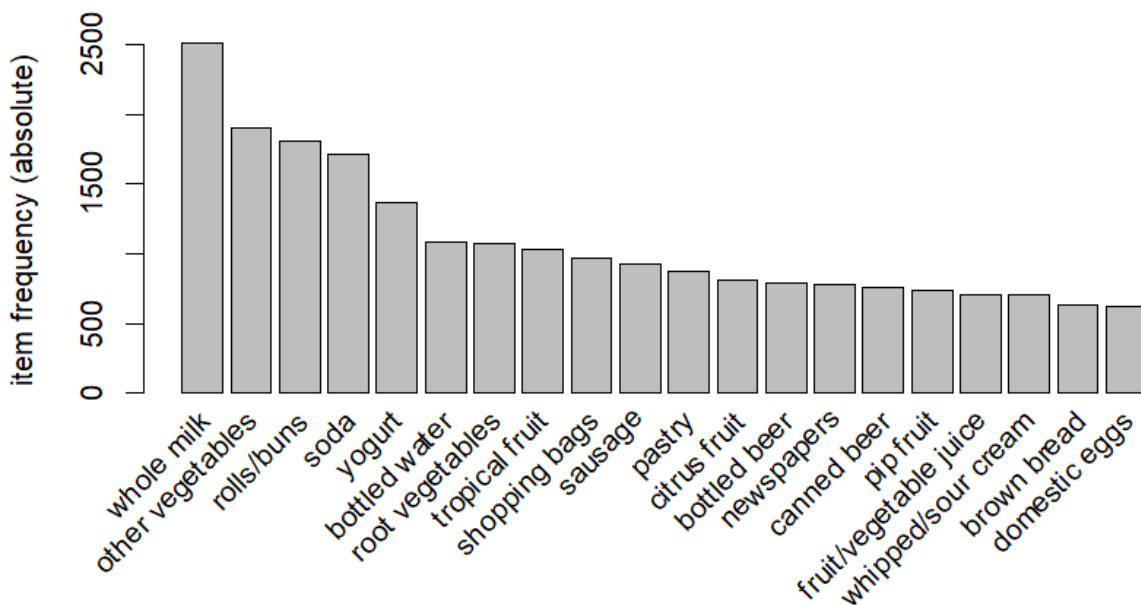
itemFrequencyPlot(Groceries, topN = 20, type = "absolute", main = "Top 20 προϊόντα ανά συχνότητα") →

> itemFrequencyPlot(Groceries, topN = 20, type = "absolute", main = "Top 20 προϊόντα ανά συχνότητα")
>



Εικόνα 2 Προϊόντα που εμφανίζονται πιο συχνά

Top 20 προϊόντα ανά συχνότητα



Γραφική παράσταση 1 Top 20 προϊόντα ανά συχνότητα

Από την ανάλυση της γραφικής παράστασης, παρατηρούμε ότι τα πιο συχνά εμφανιζόμενα προϊόντα είναι τα εξής:

1. Whole milk
2. Other vegetables
3. Rolls/buns
4. Soda
5. Yogurt
6. Bottled water
7. Root vegetables
8. Tropical fruit
9. Shopping bags
10. Sausage
11. Pastry
12. Citrus fruit
13. Bottled beer
14. Newspapers
15. Canned beer
16. Pip fruit
17. Fruit/vegetable juice
18. Whipped/sour cream
19. Brown bread
20. Domestic eggs

Η αναγνώριση των πιο συχνά αγορασμένων προϊόντων είναι σημαντική για την κατανόηση των προτιμήσεων των καταναλωτών και μπορεί να χρησιμοποιηθεί για την ανάπτυξη στρατηγικών μάρκετινγκ και τη βελτίωση της τοποθέτησης των προϊόντων στα καταστήματα.

2.2.3 Δημιουργία γραφικής παράστασης με τα προϊόντα που αγοράζονται πιο σπάνια

Στο σημείο αυτό θα εντοπίσουμε τα προϊόντα που αγοράζονται σπάνια. Υπολογίσαμε την συχνότητα που εμφανίζονται τα προϊόντα και φιλτράραμε αυτά με συχνότητα κάτω από 0,001, ώστε να δώσουμε μια τιμή αναφοράς. Τα σπάνια προϊόντα είναι σημαντικά για να

εντοπίσουμε τα λιγότερο δημοφιλή είδη και να κατανοήσουμε καλύτερα τη συμπεριφορά των καταναλωτών.

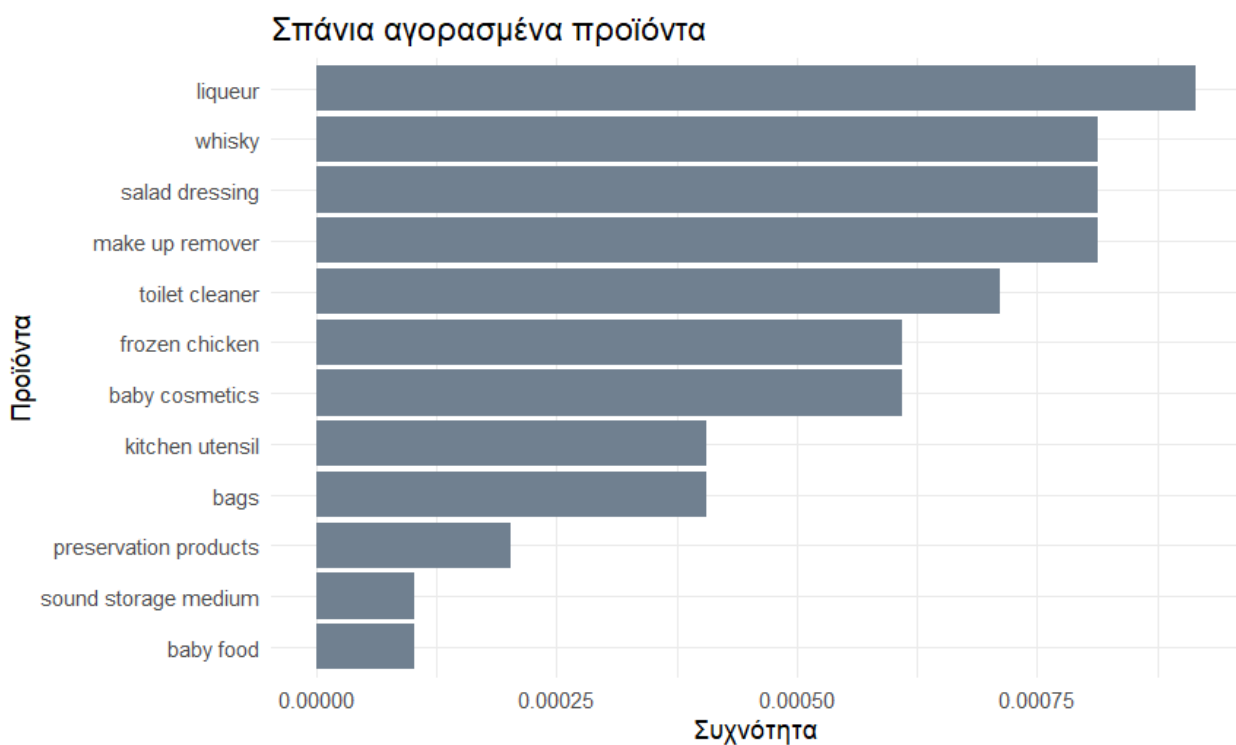
```
item_freq <- itemFrequency(Groceries)
rare_items <- item_freq[item_freq < 0.001]
rare_items_df <- data.frame(
  item = names(rare_items),
  frequency = rare_items
)

ggplot(rare_items_df, aes(x = reorder(item, frequency), y = frequency)) +
  geom_bar(stat = "identity", fill="#708090") +
  coord_flip() +
  labs(title = "Σπάνια αγορασμένα προϊόντα", x = "Προϊόντα", y = "Συχνότητα") +
  theme_minimal()
```

→

```
> item_freq <- itemFrequency(Groceries)
> rare_items <- item_freq[item_freq < 0.001]
> rare_items_df <- data.frame(
+   item = names(rare_items),
+   frequency = rare_items
+ )
> ggplot(rare_items_df, aes(x = reorder(item, frequency), y = frequency))
+   geom_bar(stat = "identity", fill="#708090") +
+   coord_flip() +
+   labs(title = "Σπάνια αγορασμένα προϊόντα", x = "Προϊόντα", y = "Συχν
ότητα") +
+   theme_minimal()
>
```

Η παρακάτω γραφική παράσταση δείχνει τα προϊόντα που αγοράζονται σπάνια, δηλαδή με συχνότητα εμφάνισης κάτω από 0.001.



Γραφική παράσταση 2 Σπάνια αγορασμένα προϊόντα

Από την ανάλυση της γραφικής παράστασης, παρατηρούμε ότι τα προϊόντα που αγοράζονται σπάνια είναι κυρίως εξειδικευμένα ή μη βασικά είδη, τα οποία μπορεί να ενδιαφέρουν μικρότερες ομάδες καταναλωτών.

2.2.4 Πυκνότητα ή Αραιότητα των Δεδομένων

Η πυκνότητα των δεδομένων υπολογίζεται ως ο λόγος του συνολικού αριθμού των στοιχείων (δηλαδή του αριθμού των προϊόντων που περιλαμβάνονται σε όλες τις συναλλαγές) προς το γινόμενο του αριθμού των συναλλαγών και του αριθμού των κατηγοριών προϊόντων. Μια χαμηλή τιμή πυκνότητας υποδηλώνει ότι οι περισσότερες συναλλαγές περιλαμβάνουν μόνο ένα μικρό ποσοστό των διαθέσιμων προϊόντων, δηλαδή τα δεδομένα είναι αραιά.

Η αραιότητα των δεδομένων υπολογίζεται ως το συμπλήρωμα της πυκνότητας. Δηλαδή, είναι το ποσοστό των δυνατών συνδυασμών προϊόντων που δεν εμφανίζονται στις συναλλαγές. Υπολογίζεται αφαιρώντας την πυκνότητα από το 1. Η έννοια της αραιότητας (sparsity) αναφέρεται στο πόσο αραιά είναι τα δεδομένα, δηλαδή πόσο λίγες είναι οι μη μηδενικές τιμές σε σύγκριση με το συνολικό αριθμό των τιμών. Στο πλαίσιο των συναλλαγών, η αραιότητα δείχνει πόσο λίγες συναλλαγές περιέχουν ένα συγκεκριμένο προϊόν σε σύγκριση με το σύνολο των συναλλαγών.

Οπότε, με την παρακάτω εντολή θα έχουμε μια ένδειξη για το πόσο πυκνά ή αραιά είναι τα δεδομένα μας.

```

density <- sum(size(Groceries)) / (length(Groceries) * length(itemLabels(Groceries)))

cat("Πυκνότητα των δεδομένων:", density, "\n")

# Υπολογισμός της αραιότητας των δεδομένων

sparsity <- 1 - density

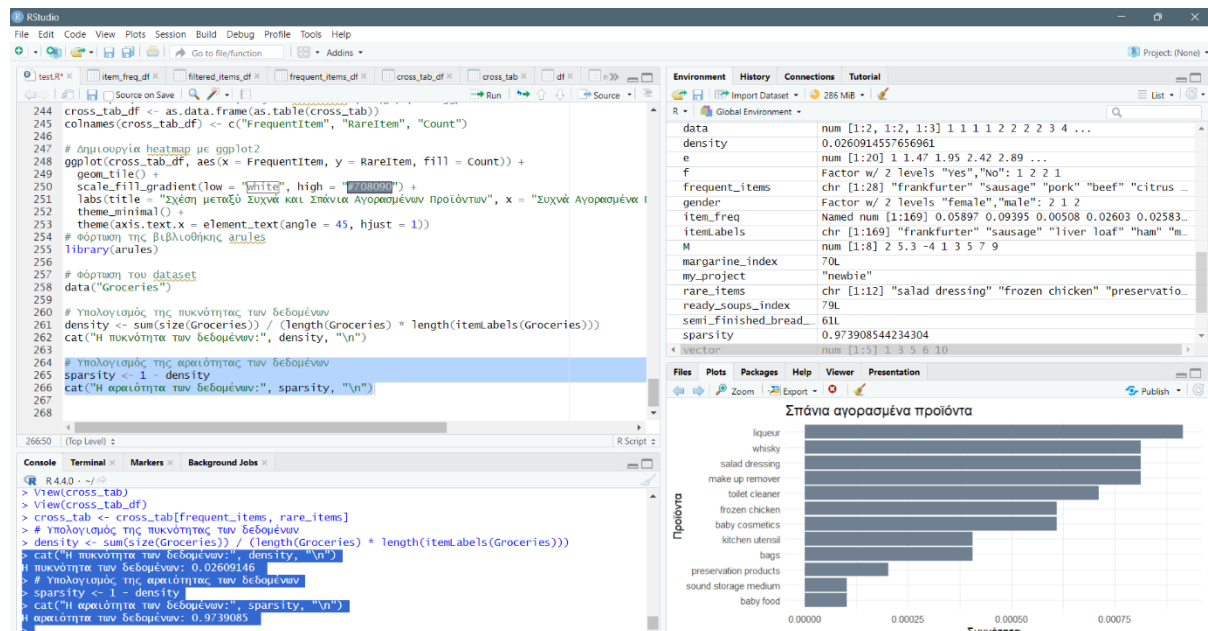
cat("Η αραιότητα των δεδομένων:", sparsity, "\n") ** →

> density <- sum(size(Groceries)) / (length(Groceries) * length(itemLabels(Groceries)))
> cat("Η πυκνότητα των δεδομένων:", density, "\n")
Η πυκνότητα των δεδομένων: 0.02609146
> # Υπολογισμός της αραιότητας των δεδομένων
> sparsity <- 1 - density
> cat("Η αραιότητα των δεδομένων:", sparsity, "\n") **
Η αραιότητα των δεδομένων: 0.9739085

>

** Μπορεί να γίνει και με την εντολή
> cat("Η αραιότητα των δεδομένων:", 1 - density, "\n")

```



Εικόνα 3 Πυκνότητα ή Αραιότητα των Δεδομένων

Στην περίπτωση μας, η πυκνότητα είναι περίπου 2,6% ενώ η αραιότητα 97,4%.

Το γεγονός ότι η πυκνότητα είναι μόλις 2.6% δείχνει ότι η περισσότερες συναλλαγές περιλαμβάνουν λίγα προϊόντα. Αυτό μπορεί να υποδεικνύει ότι τα προϊόντα που αγοράζονται συνήθως μαζί είναι περιορισμένα σε αριθμό και ότι οι καταναλωτές τείνουν να αγοράζουν συγκεκριμένους συνδυασμούς προϊόντων.

Αντίθετα, η υψηλή αραιότητα (97.4%) δείχνει ότι οι περισσότερες πιθανές συναλλαγές (δηλαδή οι συνδυασμοί προϊόντων) δεν πραγματοποιούνται. Αυτό σημαίνει ότι υπάρχουν πολλοί συνδυασμοί προϊόντων που δεν αγοράζονται ποτέ μαζί, ή που αγοράζονται πολύ σπάνια.

Η υψηλή αραιότητα μπορεί να παρουσιάζει ευκαιρίες για ανάπτυξη στρατηγικών μάρκετινγκ που προωθούν συγκεκριμένους συνδυασμούς προϊόντων. Για παράδειγμα, τα προϊόντα που δεν αγοράζονται συχνά μαζί μπορούν να προωθηθούν ως συνδυαστικά πακέτα.

Αυτό είναι σημαντικό για την περαιτέρω ανάλυση και την εφαρμογή αλγορίθμων όπως ο Apriori, οι οποίοι λειτουργούν πιο αποτελεσματικά σε αραιά δεδομένα.

2.2.5 Εφαρμογή αλγόριθμου Apriori.

Ο αλγόριθμος Apriori είναι μια σειρά βημάτων που χρησιμοποιούνται για να βρεθεί το πιο συχνό σύνολο στοιχείων σε μια βάση δεδομένων. Αυτή η τεχνική εξόρυξης δεδομένων εφαρμόζει επαναληπτικά τα βήματα της ένωσης και της απαλοιφής, μέχρι να βρεθεί το πιο συχνό σύνολο στοιχείων. Ένα ελάχιστο όριο υποστήριξης είτε δίνεται ως μέρος του προβλήματος είτε καθορίζεται από τον χρήστη.

Για να εξαγάγουμε κανόνες συσχέτισης από το dataset Groceries, θα εφαρμόσουμε τον αλγόριθμο Apriori με τα ακόλουθα κριτήρια:

- Ελάχιστο support: 0.006
- Ελάχιστο length: 2
- Ελάχιστο confidence: 0.25

Αυτή η προσέγγιση μας επιτρέπει να ανακαλύψουμε σημαντικές συσχετίσεις μεταξύ των προϊόντων που αγοράζονται συχνά μαζί.

Φόρτωση της βιβλιοθήκης arules

library(arules)

Εφαρμογή του αλγορίθμου Apriori

rules <- apriori(Groceries, parameter = list(support = 0.006, confidence = 0.25, minlen = 2))



```
> # Φόρτωση της βιβλιοθήκης arules
```

```
> library(arules)
```

```
> # Εφαρμογή του αλγορίθμου Apriori
```

```
> rules <- apriori(Groceries, parameter = list(support = 0.006, confidence = 0.25, minlen = 2))
```

```
Apriori
```

```
Parameter specification:
```

confidence	minval	smax	arem	aval	originalSupport	maxtime	support	minlen	maxlen	target	ext
0.25	0.1	1	none	FALSE	TRUE	5	0.006	2	10	rules	TRUE

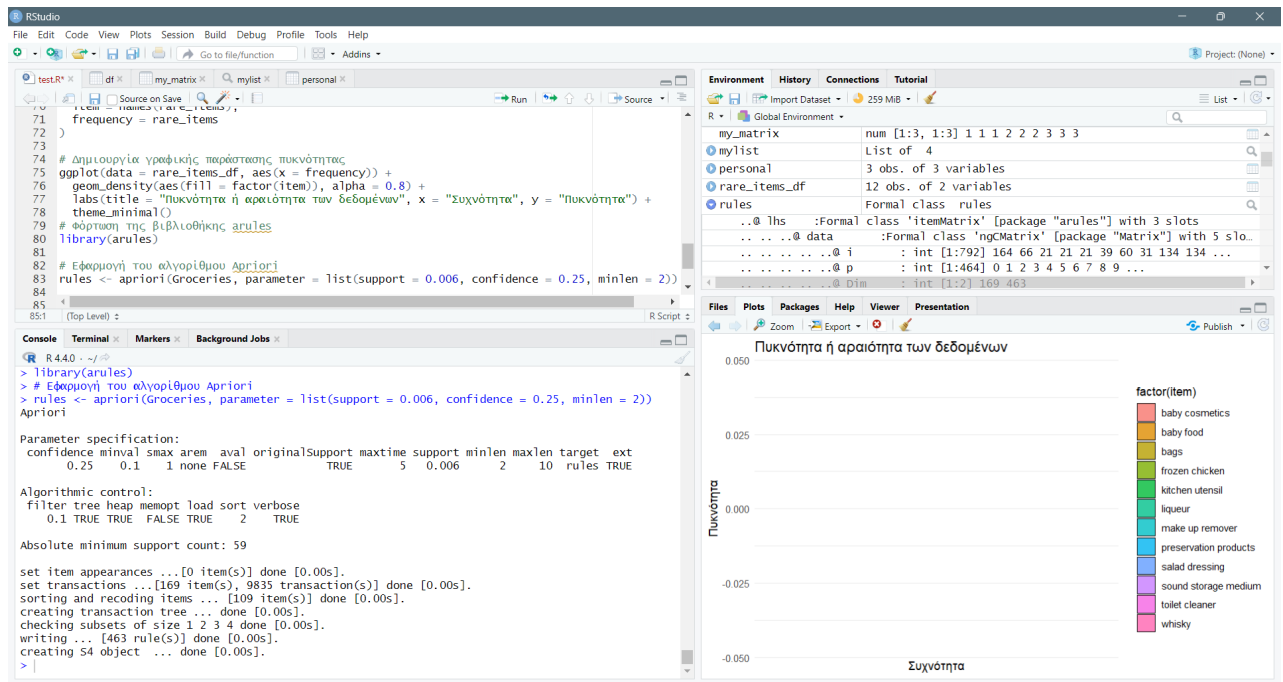
```
Algorithmic control:
```

filter	tree	heap	memopt	load	sort	verbose
0.1	TRUE	TRUE	FALSE	TRUE	2	TRUE

Absolute minimum support count: 59

```
set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
sorting and recoding items ... [109 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 done [0.00s].
writing ... [463 rule(s)] done [0.00s].
creating S4 object ... done [0.00s].
```

>



Εικόνα 4 Αλγόριθμος Apriori.

Τα αποτελέσματα που προκύπτουν, εμφανίζονται στην παρακάτω λίστα με την εντολή:

Εμφάνιση των κανόνων που προέκυψαν

inspect(rules)

→

			support	confidence	coverage	lift	count
[1]	{pot plants}	=> {whole milk}	0.006914082	0.4000000	0.01728521	1.5654596	68
[2]	{pasta}	=> {whole milk}	0.006100661	0.4054054	0.01504830	1.5866145	60
[3]	{herbs}	=> {root vegetables}	0.007015760	0.4312500	0.01626843	3.9564774	69
[4]	{herbs}	=> {other vegetables}	0.00727504	0.4750000	0.01626843	2.4548739	76
[5]	{herbs}	=> {whole milk}	0.00727504	0.4750000	0.01626843	1.8589833	76
[6]	{processed cheese}	=> {whole milk}	0.007015760	0.4233129	0.01657346	1.6566981	69
[7]	{semi-finished bread}	=> {whole milk}	0.007117438	0.4022989	0.01769192	1.5744565	70
[8]	{beverages}	=> {whole milk}	0.006812405	0.2617188	0.02602949	1.0242753	67
[9]	{detergent}	=> {other vegetables}	0.006405694	0.3333333	0.01921708	1.7227185	63
[10]	{detergent}	=> {whole milk}	0.008947636	0.4656085	0.01921708	1.8222281	88
[11]	{pickled vegetables}	=> {other vegetables}	0.006405694	0.3579545	0.01789527	1.8499648	63
[12]	{pickled vegetables}	=> {whole milk}	0.007117438	0.3972723	0.01789527	1.5565650	70
[13]	{baking powder}	=> {other vegetables}	0.007320793	0.4137931	0.01769192	2.1385471	72
[14]	{baking powder}	=> {whole milk}	0.009252669	0.5229885	0.01769192	2.0467935	91
[15]	{flour}	=> {other vegetables}	0.006304016	0.3625731	0.01738688	1.8738342	62
[16]	{flour}	=> {whole milk}	0.008439248	0.4853801	0.01738688	1.8996074	83
[17]	{soft cheese}	=> {other vegetables}	0.007117438	0.4166667	0.01708185	2.1533981	70
[18]	{soft cheese}	=> {whole milk}	0.007524148	0.4404762	0.01708185	1.7238692	74
[19]	{specialty bar}	=> {soda}	0.007219115	0.2639405	0.02735130	1.5136181	71
[20]	{misc. beverages}	=> {soda}	0.007320793	0.2580645	0.02836807	1.4799210	72
[21]	{grapes}	=> {tropical fruit}	0.006100661	0.2727273	0.02236909	2.5991015	60
[22]	{grapes}	=> {other vegetables}	0.009049314	0.4045455	0.02236909	2.0907538	89
[23]	{cat food}	=> {yogurt}	0.006202339	0.2663755	0.02328419	1.9094778	61
[24]	{cat food}	=> {other vegetables}	0.006507372	0.2794760	0.02328419	1.4443753	64
[25]	{cat food}	=> {whole milk}	0.008845958	0.3799127	0.02328419	1.4868448	87
[26]	{specialty chocolate}	=> {whole milk}	0.008032537	0.2642140	0.03040163	1.0340410	79
[27]	{meat}	=> {rolls/buns}	0.006914082	0.2677165	0.02582613	1.4554959	68
[28]	{meat}	=> {other vegetables}	0.009964413	0.3858268	0.02582613	1.9940128	98
[29]	{meat}	=> {whole milk}	0.009964413	0.3858268	0.02582613	1.5099906	98
[30]	{frozen meals}	=> {other vegetables}	0.007524148	0.2652330	0.02836807	1.3707653	74
[31]	{frozen meals}	=> {whole milk}	0.009862735	0.3476703	0.02836807	1.3606593	97
[32]	{hard cheese}	=> {yogurt}	0.006405694	0.2614108	0.02450432	1.8738886	63
[33]	{hard cheese}	=> {other vegetables}	0.009456024	0.3858921	0.02450432	1.9943505	93
[34]	{hard cheese}	=> {whole milk}	0.010066090	0.4107884	0.02450432	1.6076815	99
[35]	{whole milk}	=> {whole milk}	0.008540035	0.3084845	0.02706136	2.1806104	84

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Source

Console Terminal Markers Background Jobs

R 4.4.0 ~ /

[33]	{hard cheese}	=>	{yogurt}	0.006405694	0.2614108	0.02450432	1.8738886	63
[34]	{hard cheese}	=>	{other vegetables}	0.009456024	0.3858921	0.02450432	1.9943505	93
[35]	{hard cheese}	=>	{whole milk}	0.010066090	0.4107884	0.02450432	1.6076815	99
[36]	{butter milk}	=>	{yogurt}	0.008540925	0.3054545	0.02796136	2.1896104	84
[37]	{butter milk}	=>	{rolls/buns}	0.007625826	0.2727273	0.02796136	1.4827378	75
[38]	{butter milk}	=>	{other vegetables}	0.010371124	0.3709091	0.02796136	1.9169159	102
[39]	{butter milk}	=>	{whole milk}	0.011591256	0.4145455	0.02796136	1.6223854	114
[40]	{candy}	=>	{soda}	0.008642603	0.2891156	0.02989324	1.6579897	85
[41]	{candy}	=>	{whole milk}	0.008235892	0.2755102	0.02989324	1.0782502	81
[42]	{ham}	=>	{yogurt}	0.006710727	0.2578125	0.02602949	1.8480947	66
[43]	{ham}	=>	{rolls/buns}	0.006914082	0.2656250	0.02602949	1.4441249	68
[44]	{ham}	=>	{other vegetables}	0.009150991	0.3515625	0.02602949	1.8169297	90
[45]	{ham}	=>	{whole milk}	0.011489578	0.4414062	0.02602949	1.7275091	113
[46]	{sliced cheese}	=>	{sausage}	0.007015760	0.2863071	0.02450432	3.0474349	69
[47]	{sliced cheese}	=>	{yogurt}	0.008032537	0.3278008	0.02450432	2.3497968	79
[48]	{sliced cheese}	=>	{rolls/buns}	0.007625826	0.3112033	0.02450432	1.6919208	75
[49]	{sliced cheese}	=>	{other vegetables}	0.009049314	0.3692946	0.02450432	1.9085720	89
[50]	{sliced cheese}	=>	{whole milk}	0.010777834	0.4398340	0.02450432	1.7213560	106
[51]	{oil}	=>	{root vegetables}	0.007015760	0.2500000	0.02806304	2.2936101	69
[52]	{oil}	=>	{other vegetables}	0.009964413	0.3550725	0.02806304	1.8350697	98
[53]	{oil}	=>	{whole milk}	0.011286223	0.4021739	0.02806304	1.5739675	111
[54]	{onions}	=>	{root vegetables}	0.009456024	0.3049180	0.03101169	2.7974523	93
[55]	{onions}	=>	{other vegetables}	0.014234875	0.4590164	0.03101169	2.3722681	140
[56]	{onions}	=>	{whole milk}	0.012099644	0.3901639	0.03101169	1.5269647	119
[57]	{berries}	=>	{whipped/sour cream}	0.009049314	0.2721713	0.03324860	3.7968855	89
[58]	{berries}	=>	{yogurt}	0.010574479	0.3180428	0.03324860	2.2798477	104
[59]	{berries}	=>	{other vegetables}	0.010269446	0.3088685	0.03324860	1.5962805	101
[60]	{berries}	=>	{whole milk}	0.011794611	0.3547401	0.03324860	1.3883281	116
[61]	{hamburger meat}	=>	{rolls/buns}	0.008642603	0.2599388	0.03324860	1.4132109	85
[62]	{hamburger meat}	=>	{other vegetables}	0.013828165	0.4159021	0.03324860	2.1494470	136
[63]	{hamburger meat}	=>	{whole milk}	0.014743264	0.4434251	0.03324860	1.7354101	145
[64]	{hygiene articles}	=>	{other vegetables}	0.009557702	0.2901235	0.03294357	1.4994032	94
[65]	{hygiene articles}	=>	{whole milk}	0.012811388	0.3888889	0.03294357	1.5219746	126
[66]	{salty snack}	=>	{other vegetables}	0.010777834	0.2849462	0.03782410	1.4726465	106
[67]	{salty snack}	=>	{whole milk}	0.011184545	0.2956989	0.03782410	1.1572618	110
[68]	{sugar}	=>	{other vegetables}	0.010777834	0.3183183	0.03385867	1.6451186	106
[69]	{sugar}	=>	{whole milk}	0.015048297	0.4444444	0.03385867	1.7393996	148
[70]	{waffles}	=>	{other vegetables}	0.010066090	0.2619048	0.03843416	1.3535645	99
[71]	{waffles}	=>	{whole milk}	0.012709710	0.3306878	0.03843416	1.2941961	125
[72]	{long life bakery product}	=>	{other vegetables}	0.010676157	0.2853261	0.03741739	1.4746096	105
[73]	{long life bakery product}	=>	{whole milk}	0.013523132	0.3614130	0.03741739	1.4144438	133

Source

Console Terminal Markers Background Jobs

R 4.4.0 ~ /

[71]	{waffles}	=>	{whole milk}	0.012709710	0.3306878	0.03843416	1.2941961	125
[72]	{long life bakery product}	=>	{other vegetables}	0.010676157	0.2853261	0.03741739	1.4746096	105
[73]	{long life bakery product}	=>	{whole milk}	0.013523132	0.3614130	0.03741739	1.4144438	133
[74]	{dessert}	=>	{soda}	0.009862735	0.2657534	0.03711235	1.5240145	97
[75]	{dessert}	=>	{yogurt}	0.009862735	0.2657534	0.03711235	1.9050182	97
[76]	{dessert}	=>	{other vegetables}	0.011591256	0.3123288	0.03711235	1.6141636	114
[77]	{dessert}	=>	{whole milk}	0.013726487	0.3698630	0.03711235	1.4475140	135
[78]	{cream cheese }	=>	{yogurt}	0.012404677	0.3128205	0.03965430	2.2424123	122
[79]	{cream cheese }	=>	{rolls/buns}	0.009964413	0.2512821	0.03965430	1.3661465	98
[80]	{cream cheese }	=>	{other vegetables}	0.013726487	0.3461538	0.03965430	1.7889769	135
[81]	{cream cheese }	=>	{whole milk}	0.016471784	0.4153846	0.03965430	1.6256696	162
[82]	{chicken}	=>	{root vegetables}	0.010879512	0.2535545	0.04290798	2.3262206	107
[83]	{chicken}	=>	{other vegetables}	0.017895272	0.4170616	0.04290798	2.1554393	176
[84]	{chicken}	=>	{whole milk}	0.017590239	0.4099526	0.04290798	1.6044106	173
[85]	{white bread}	=>	{other vegetables}	0.013726487	0.3260870	0.04209456	1.6852681	135
[86]	{white bread}	=>	{whole milk}	0.017081851	0.4057971	0.04209456	1.5881474	168
[87]	{chocolate}	=>	{soda}	0.013523132	0.2725410	0.04961871	1.5629391	133
[88]	{chocolate}	=>	{other vegetables}	0.012709710	0.2561475	0.04961871	1.3238103	125
[89]	{chocolate}	=>	{whole milk}	0.016675140	0.3360656	0.04961871	1.3152427	164
[90]	{coffee}	=>	{whole milk}	0.018708693	0.3222417	0.05805796	1.2611408	184
[91]	{frozen vegetables}	=>	{yogurt}	0.012404677	0.2579281	0.04809354	1.8489235	122
[92]	{frozen vegetables}	=>	{other vegetables}	0.017793594	0.3699789	0.04809354	1.9121083	175
[93]	{frozen vegetables}	=>	{whole milk}	0.020437214	0.4249471	0.04809354	1.6630940	201
[94]	{beef}	=>	{root vegetables}	0.017386884	0.3313953	0.05246568	3.0403668	171
[95]	{beef}	=>	{rolls/buns}	0.013624809	0.2596899	0.05246568	1.4118576	134
[96]	{beef}	=>	{other vegetables}	0.019725470	0.3759690	0.05246568	1.9430662	194
[97]	{beef}	=>	{whole milk}	0.021250635	0.4050388	0.05246568	1.5851795	209
[98]	{curd}	=>	{yogurt}	0.017285206	0.3244275	0.05327911	2.3256154	170
[99]	{curd}	=>	{other vegetables}	0.017183528	0.3225191	0.05327911	1.6668288	169
[100]	{curd}	=>	{whole milk}	0.026131164	0.4904580	0.05327911	1.9194805	257
[101]	{napkins}	=>	{other vegetables}	0.014438231	0.2757282	0.05236401	1.4250060	142
[102]	{napkins}	=>	{whole milk}	0.019725470	0.3766990	0.05236401	1.4742678	194
[103]	{pork}	=>	{other vegetables}	0.021657346	0.3756614	0.05765125	1.9414764	213
[104]	{pork}	=>	{whole milk}	0.022165735	0.3844797	0.05765125	1.5047187	218
[105]	{frankfurter}	=>	{rolls/buns}	0.019217082	0.3258621	0.05897306	1.7716161	189
[106]	{frankfurter}	=>	{other vegetables}	0.016471784	0.2793103	0.05897306	1.4435193	162
[107]	{frankfurter}	=>	{whole milk}	0.020538892	0.3482759	0.05897306	1.3630295	202
[108]	{bottled beer}	=>	{whole milk}	0.020437214	0.2537879	0.08052872	0.9932367	201
[109]	{brown bread}	=>	{other vegetables}	0.018708693	0.2884013	0.06487036	1.4905025	184
[110]	{brown bread}	=>	{whole milk}	0.025216065	0.3887147	0.06487036	1.5212930	248
[111]	{brown bread}	=>	{rolls/buns}	0.014743264	0.2517361	0.05856634	1.3686151	145

Source						
Console Terminal Markers Background Jobs						
R 4.4.0 ~ /						
[88]	{chocolate}	=> {other vegetables}	0.012709710	0.2561475	0.04961871	1.3238103 125
[89]	{chocolate}	=> {whole milk}	0.016675140	0.3360656	0.04961871	1.3152427 164
[90]	{coffee}	=> {whole milk}	0.018708693	0.3222417	0.05805796	1.2611408 184
[91]	{frozen vegetables}	=> {yogurt}	0.012404677	0.2579281	0.04809354	1.8489235 122
[92]	{frozen vegetables}	=> {other vegetables}	0.017793594	0.3699789	0.04809354	1.9121083 175
[93]	{frozen vegetables}	=> {whole milk}	0.020437214	0.4249471	0.04809354	1.6630940 201
[94]	{beef}	=> {root vegetables}	0.017386884	0.3313953	0.05246568	3.0403668 171
[95]	{beef}	=> {rolls/buns}	0.013624809	0.2596899	0.05246568	1.4118576 134
[96]	{beef}	=> {other vegetables}	0.019725470	0.3759690	0.05246568	1.9430662 194
[97]	{beef}	=> {whole milk}	0.021250635	0.4050388	0.05246568	1.5851795 209
[98]	{curd}	=> {yogurt}	0.017285206	0.3244275	0.05327911	2.3256154 170
[99]	{curd}	=> {other vegetables}	0.017183528	0.3225191	0.05327911	1.6668288 169
[100]	{curd}	=> {whole milk}	0.026131164	0.4904580	0.05327911	1.9194805 257
[101]	{napkins}	=> {other vegetables}	0.014438231	0.2757282	0.05236401	1.4250060 142
[102]	{napkins}	=> {whole milk}	0.019725470	0.3766990	0.05236401	1.4742678 194
[103]	{pork}	=> {other vegetables}	0.021657346	0.3756614	0.05765125	1.9414764 213
[104]	{pork}	=> {whole milk}	0.022165735	0.3844797	0.05765125	1.5047187 218
[105]	{frankfurter}	=> {rolls/buns}	0.019217082	0.3258621	0.05897306	1.7716161 189
[106]	{frankfurter}	=> {other vegetables}	0.016471784	0.2793103	0.05897306	1.4435193 162
[107]	{frankfurter}	=> {whole milk}	0.020538892	0.3482759	0.05897306	1.3630295 202
[108]	{bottled beer}	=> {whole milk}	0.020437214	0.2537879	0.08052872	0.9932367 201
[109]	{brown bread}	=> {other vegetables}	0.018708693	0.2884013	0.06487036	1.4905025 184
[110]	{brown bread}	=> {whole milk}	0.025216065	0.3887147	0.06487036	1.5212930 248
[111]	{margarine}	=> {rolls/buns}	0.014743264	0.2517361	0.05856634	1.3686151 145
[112]	{margarine}	=> {other vegetables}	0.019725470	0.3368056	0.05856634	1.7406635 194
[113]	{margarine}	=> {whole milk}	0.024199288	0.4131944	0.05856634	1.6170980 238
[114]	{butter}	=> {yogurt}	0.014641586	0.2642202	0.05541434	1.8940273 144
[115]	{butter}	=> {other vegetables}	0.020030503	0.3614679	0.05541434	1.8681223 197
[116]	{butter}	=> {whole milk}	0.027554652	0.4972477	0.05541434	1.9460530 271
[117]	{newspapers}	=> {whole milk}	0.027351296	0.3426752	0.07981698	1.3411103 269
[118]	{domestic eggs}	=> {other vegetables}	0.022267412	0.3509615	0.06344687	1.8138238 219
[119]	{domestic eggs}	=> {whole milk}	0.029994916	0.4727564	0.06344687	1.8502027 295
[120]	{fruit/vegetable juice}	=> {soda}	0.018403660	0.2545710	0.07229283	1.4598869 181
[121]	{fruit/vegetable juice}	=> {yogurt}	0.018708693	0.2587904	0.07229283	1.8551049 184
[122]	{fruit/vegetable juice}	=> {other vegetables}	0.021047280	0.2911392	0.07229283	1.5046529 207
[123]	{fruit/vegetable juice}	=> {whole milk}	0.026639553	0.3684951	0.07229283	1.4421604 262
[124]	{whipped/sour cream}	=> {yogurt}	0.020742247	0.2893617	0.07168277	2.0742510 204
[125]	{whipped/sour cream}	=> {other vegetables}	0.028876462	0.4028369	0.07168277	2.0819237 284
[reached 'max' / getOption("max.print") -- omitted 338 rows]						
>						

Εικόνα 5 Εφαρμογή αλγόριθμου Apriori

Ερμηνεία Αποτελεσμάτων

Support: Η σχετική συχνότητα των συναλλαγών που περιέχουν και τα δύο στοιχεία.

Confidence: Η πιθανότητα να περιέχεται το rhs στοιχείο στη συναλλαγή, δεδομένου ότι περιέχεται το lhs στοιχείο.

Lift: Ένας δείκτης της ισχύος του κανόνα. Αν το lift είναι μεγαλύτερο από 1, ο κανόνας είναι πιο ισχυρός από την τυχαία συσχέτιση.

2.2.6 Εντοπισμός των top 5 συνόλων στοιχείων με το υψηλότερο επίπεδο confidence.

Και προχωράμε αμέσως με την παρακάτω εντολή, για τον εντοπισμό των top 5 σύνολα στοιχείων που έχουν το υψηλότερο επίπεδο confidence.

Εντοπισμός των top 5 συνόλων στοιχείων με το υψηλότερο confidence

```
top_5_rules <- head(sort(rules, by = "confidence"), 5)
```

```
inspect(top_5_rules)
```

→

```
> # Εντοπισμός των top 5 συνόλων στοιχείων με το υψηλότερο confidence
> top_5_rules <- head(sort(rules, by = "confidence"), 5)
> inspect(top_5_rules)
```

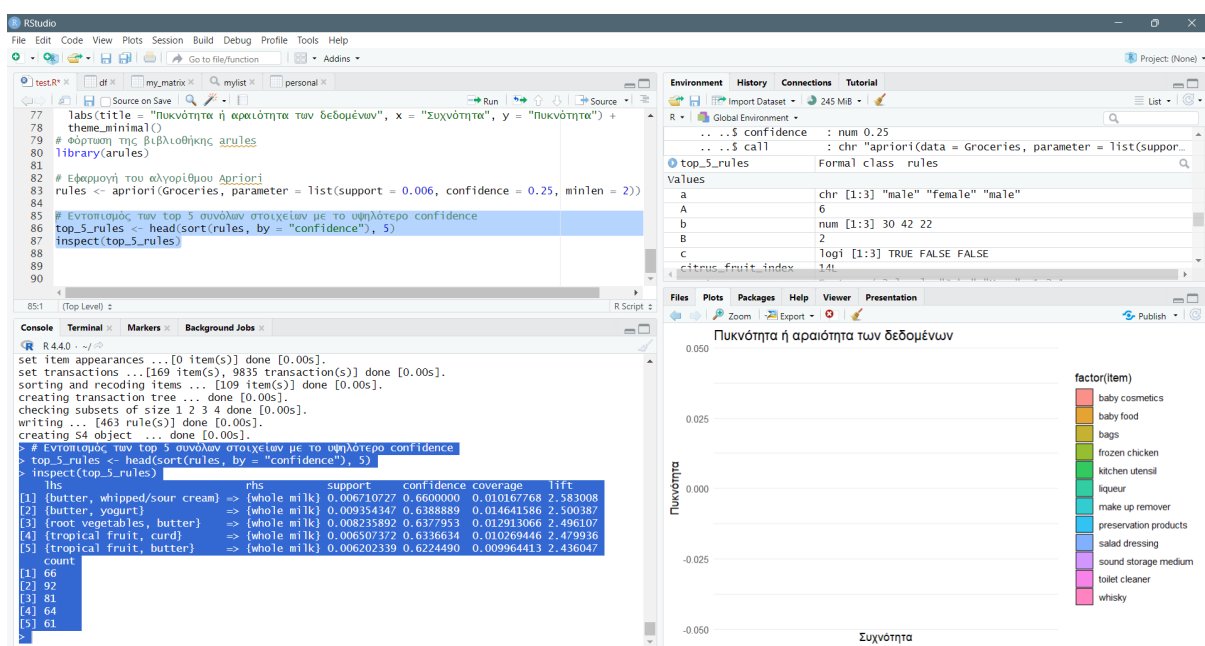
	lhs	rhs	support	confidence	coverage	lift
[1]	{butter, whipped/sour cream}	=> {whole milk}	0.006710727	0.6600000	0.010167768	2.583008
[2]	{butter, yogurt}	=> {whole milk}	0.009354347	0.6388889	0.014641586	2.500387
[3]	{root vegetables, butter}	=> {whole milk}	0.008235892	0.6377953	0.012913066	2.496107
[4]	{tropical fruit, curd}	=> {whole milk}	0.006507372	0.6336634	0.010269446	2.479936
[5]	{tropical fruit, butter}	=> {whole milk}	0.006202339	0.6224490	0.009964413	2.436047

```
count
[1] 66
[2] 92
[3] 81
[4] 64
[5] 61
```

>

Επεξήγηση των στηλών:

- Lhs: Το αριστερό μέρος του κανόνα (προϊόντα που αγοράστηκαν μαζί),
- Rhs: Το δεξί μέρος του κανόνα (προϊόντα που συσχετίζονται),
- Support: Η υποστήριξη του κανόνα, δηλαδή το ποσοστό των συναλλαγών στις οποίες εμφανίζεται ο κανόνας,
- Confidence: Η εμπιστοσύνη του κανόνα, δηλαδή το ποσοστό των συναλλαγών στις οποίες τα προϊόντα του αριστερού μέρους συνυπάρχουν με τα προϊόντα του δεξιού μέρους.
- Coverage: Η κάλυψη του κανόνα, δηλαδή το ποσοστό των συναλλαγών στις οποίες εμφανίζεται το αριστερό μέρος του κανόνα.
- Lift: Ο συντελεστής ανύψωσης του κανόνα, δηλαδή πόσο περισσότερο πιθανό είναι να αγοραστεί το δεξί μέρος του κανόνα όταν αγοράζεται το αριστερό μέρος σε σχέση με το αν ήταν ανεξάρτητα.
- count: Ο αριθμός των συναλλαγών στις οποίες εμφανίζεται ο κανόνας.



Ανάλυση αποτελεσμάτων:

Οι top 5 κανόνες με το υψηλότερο confidence είναι:

1. {butter, whipped/sour cream} => {whole milk}: Όταν οι πελάτες αγοράζουν butter και whipped/sour cream, υπάρχει πιθανότητα 66% να αγοράσουν επίσης whole milk. Ο κανόνας αυτός έχει lift 2.583, που σημαίνει ότι η αγορά whole milk είναι 2.583 φορές πιο πιθανή όταν αγοράζονται butter και whipped/sour cream μαζί.
2. {butter, yogurt} => {whole milk}: Όταν οι πελάτες αγοράζουν butter και yogurt, υπάρχει πιθανότητα 63.9% να αγοράσουν επίσης whole milk. Ο κανόνας αυτός έχει lift 2.500.
3. {root vegetables, butter} => {whole milk}: Όταν οι πελάτες αγοράζουν root vegetables, και butter, υπάρχει πιθανότητα 63.8% να αγοράσουν επίσης whole milk. Ο κανόνας αυτός έχει lift 2.496.
4. {tropical fruit, curd} => {whole milk}: Όταν οι πελάτες αγοράζουν tropical fruit και curd, υπάρχει πιθανότητα 63.4% να αγοράσουν επίσης whole milk. Ο κανόνας αυτός έχει lift 2.480.
5. {tropical fruit, butter} => {whole milk}: Όταν οι πελάτες αγοράζουν tropical fruit και butter, υπάρχει πιθανότητα 62.2% να αγοράσουν επίσης whole milk. Ο κανόνας αυτός έχει lift 2.436.

2.2.7 Δημιουργία Γραφικής Παράστασης για τη Σχέση μεταξύ Σπάνια και Συχνά Αγορασμένων Προϊόντων

Για να δημιουργήσουμε μια γραφική παράσταση που απεικονίζει τη σχέση μεταξύ των σπάνια και συχνά αγορασμένων προϊόντων, θα συνδυάσουμε τα δεδομένα συχνότητας εμφάνισης προϊόντων και θα δημιουργήσουμε ένα heatmap για να δείξουμε τη συνάφεια μεταξύ τους.

Ακολουθούμε τα παρακάτω βήματα:

Υπολογισμός συχνότητας εμφάνισης των προϊόντων

```
item_freq <- itemFrequency(Groceries)
```

Φιλτράρισμα των συχνά και σπάνια αγορασμένων προϊόντων

```
frequent_items <- names(item_freq[item_freq >= 0.05])
```

```
rare_items <- names(item_freq[item_freq < 0.001])
```

```

# Δημιουργία πίνακα συνάφειας
cross_tab <- crossTable(Groceries, measure = "count")
cross_tab <- cross_tab[frequent_items, rare_items]
# Μετατροπή του πίνακα συνάφειας σε dataframe για χρήση στο ggplot2
cross_tab_df <- as.data.frame(as.table(cross_tab))
colnames(cross_tab_df) <- c("FrequentItem", "RareItem", "Count")
# Δημιουργία heatmap με ggplot2
ggplot(cross_tab_df, aes(x = FrequentItem, y = RareItem, fill = Count)) +
  geom_tile() +
  scale_fill_gradient(low = "white", high = "steelblue") +
  labs(title = "Σχέση μεταξύ Συχνά και Σπάνια Αγορασμένων Προϊόντων", x = "Συχνά
Αγορασμένα Προϊόντα", y = "Σπάνια Αγορασμένα Προϊόντα") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) →

```

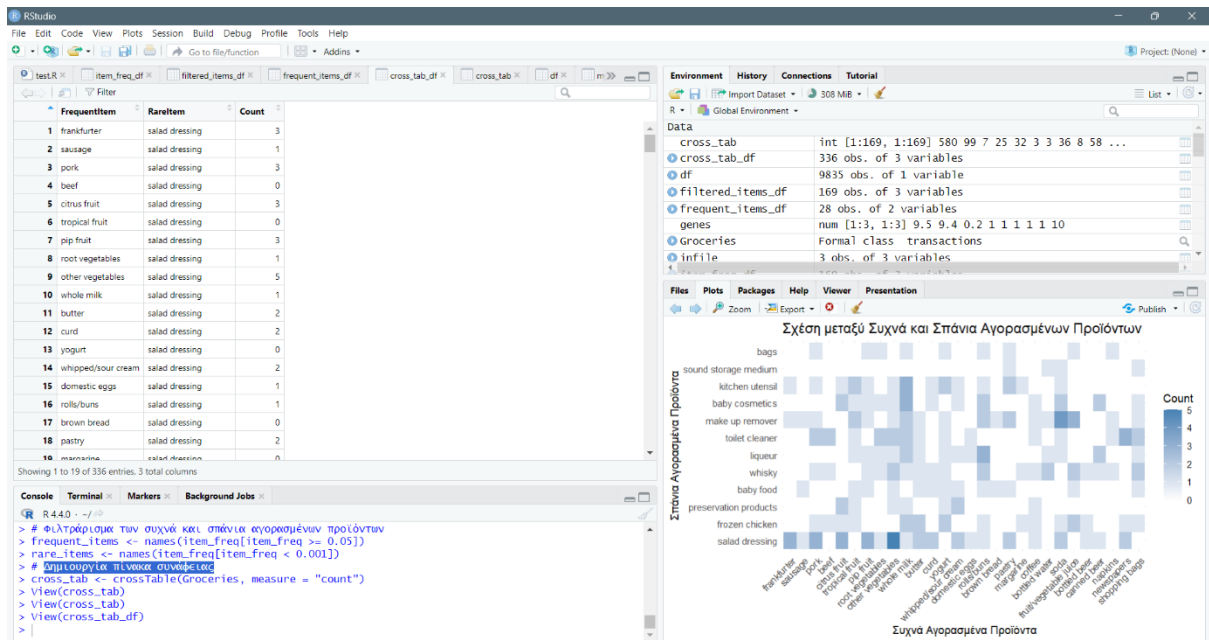
```

> # Υπολογισμός συχνότητας εμφάνισης των προϊόντων
> item_freq <- itemFrequency(Groceries)
> # Φιλτράρισμα των συχνά και σπάνια αγορασμένων προϊόντων
> frequent_items <- names(item_freq[item_freq >= 0.05])
> rare_items <- names(item_freq[item_freq < 0.001])
> # Δημιουργία πίνακα συνάφειας
> cross_tab <- crossTable(Groceries, measure = "count")
> cross_tab <- cross_tab[frequent_items, rare_items]
> view(cross_tab)
> cross_tab <- cross_tab[frequent_items, rare_items]
> # Μετατροπή του πίνακα συνάφειας σε dataframe για χρήση στο ggplot2
> cross_tab_df <- as.data.frame(as.table(cross_tab))
> view(df)
> view(cross_tab_df)
> colnames(cross_tab_df) <- c("FrequentItem", "RareItem", "Count")
> # Δημιουργία heatmap με ggplot2
> ggplot(cross_tab_df, aes(x = FrequentItem, y = RareItem, fill = Count))
+   geom_tile() +
+   scale_fill_gradient(low = "white", high = "steelblue") +
+   labs(title = "Σχέση μεταξύ Συχνά και Σπάνια Αγορασμένων Προϊόντων", x
= "Συχνά Αγορασμένα Προϊόντα", y = "Σπάνια Αγορασμένα Προϊόντα") +
+   theme_minimal() +
+   theme(axis.text.x = element_text(angle = 45, hjust = 1))
>

```

Αναλυτικά βήματα παρακάτω:

Μετατρέπουμε τον πίνακα συνάφειας σε μορφή dataframe για να το χρησιμοποιήσουμε με το ggplot2.

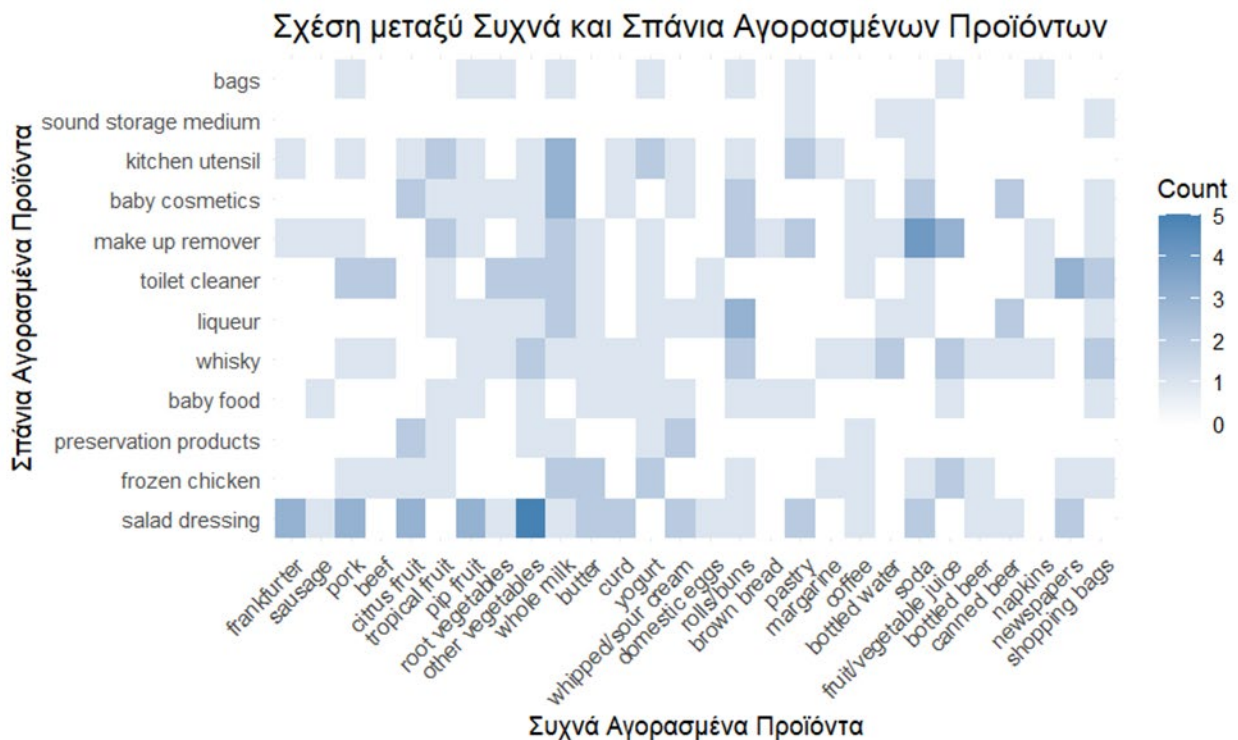


Εικόνα 9 Μετατροπή του Πίνακα Συνάφειας σε Dataframe.

5. Δημιουργία Heatmap με ggplot2:

Δημιουργούμε ένα heatmap, καθώς με τα ραβδογράμματα ή τα ιστογράμματα δεν απεικονίζονταν σωστά ο συνδυασμός των δύο, για να απεικονίσουμε τη σχέση μεταξύ των συχνά και των σπάνια αγορασμένων προϊόντων.

Η παρακάτω γραφική παράσταση δείχνει τη σχέση μεταξύ των συχνά και των σπάνια αγορασμένων προϊόντων στο dataset Groceries. Το heatmap απεικονίζει την συνάφεια μεταξύ αυτών των προϊόντων, με το χρώμα να υποδηλώνει τη συχνότητα συνύπαρξης των προϊόντων στις συναλλαγές.



Θερμικός χάρτης 1 Σχέση μεταξύ Συχνά και Σπάνια Αγορασμένων Προϊόντων

2.2.8 Μοτίβο των αγορών των καταναλωτών

Για να κατανοήσουμε το μοτίβο των αγορών του καταναλωτή, μπορούμε να αναλύσουμε τους κανόνες συσχέτισης που εξάγονται από τον αλγόριθμο Apriori και να δούμε ποια προϊόντα αγοράζονται συχνά μαζί.

Εφαρμογή του αλγορίθμου Apriori για την εξαγωγή κανόνων συσχέτισης

```
rules <- apriori(Groceries, parameter = list(support = 0.01, confidence = 0.2, minlen = 2))
```

Εντοπισμός των top 10 κανόνων με το υψηλότερο lift

```
top_rules <- head(sort(rules, by = "lift"), 10)
```

```
inspect(top_rules) →
```

```
Console Terminal Markers Background Jobs
R 4.4.0 ~ /
> library(arules)
> # Φόρτωση του dataset
> data("Groceries")
> # Εφαρμογή του αλγορίθμου Apriori για την εξαγωγή κανόνων συσχέτισης
> rules <- apriori(Groceries, parameter = list(support = 0.01, confidence = 0.2, minlen = 2))
Apriori

Parameter specification:
confidence minval smax arem aval originalSupport maxtime support minlen maxlen target ext
0.2 0.1 1 none FALSE TRUE 5 0.01 2 10 rules TRUE

Algorithmic control:
filter tree heap memopt load sort verbose
0.1 TRUE TRUE FALSE TRUE 2 TRUE

Absolute minimum support count: 98

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
sorting and recoding items ... [88 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 done [0.00s].
writing ... [231 rule(s)] done [0.00s].
creating S4 object ... done [0.00s].
> View(top_5_rules)
> # Εντοπισμός των top 10 κανόνων με το υψηλότερο lift
> top_rules <- head(sort(rules, by = "lift"), 10)
> View(top_rules)
> View(rules)
> inspect(top_rules)

lhs rhs support confidence
[1] {citrus fruit, other vegetables} => {root vegetables} 0.01037112 0.3591549
[2] {other vegetables, yogurt} => {whipped/sour cream} 0.01016777 0.2341920
[3] {tropical fruit, other vegetables} => {root vegetables} 0.01230300 0.3427762
[4] {beef} => {root vegetables} 0.01738688 0.3313953
[5] {citrus fruit, root vegetables} => {other vegetables} 0.01037112 0.5862069
[6] {tropical fruit, root vegetables} => {other vegetables} 0.01230300 0.5845411
[7] {other vegetables, whole milk} => {root vegetables} 0.02318251 0.3097826
[8] {whole milk, curd} => {yogurt} 0.01006609 0.3852140
[9] {other vegetables, yogurt} => {root vegetables} 0.01291307 0.2974239
[10] {other vegetables, yogurt} => {tropical fruit} 0.01230300 0.2833724

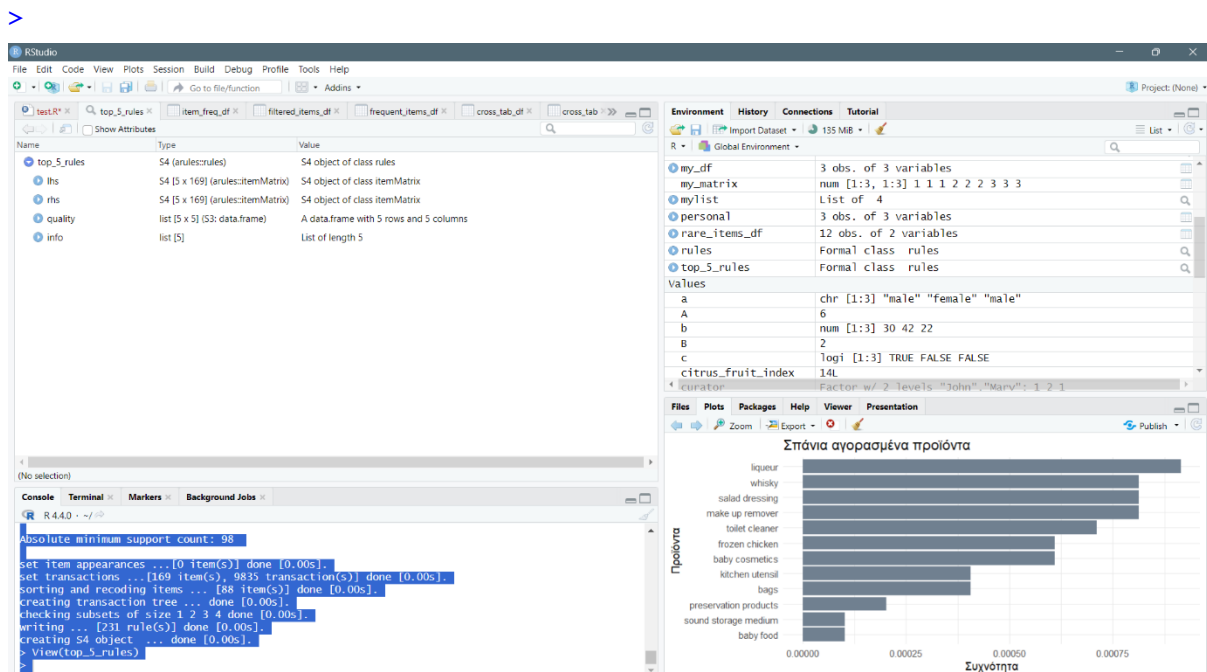
Console Terminal Markers Background Jobs
R 4.4.0 ~ /
> filter tree heap memopt load sort verbose
0.1 TRUE TRUE FALSE TRUE 2 TRUE

Absolute minimum support count: 98

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
sorting and recoding items ... [88 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 done [0.00s].
writing ... [231 rule(s)] done [0.00s].
creating S4 object ... done [0.00s].
> View(top_5_rules)
> # Εντοπισμός των top 10 κανόνων με το υψηλότερο lift
> top_rules <- head(sort(rules, by = "lift"), 10)
> View(top_rules)
> View(rules)
> inspect(top_rules)

lhs rhs support confidence
[1] {citrus fruit, other vegetables} => {root vegetables} 0.01037112 0.3591549
[2] {other vegetables, yogurt} => {whipped/sour cream} 0.01016777 0.2341920
[3] {tropical fruit, other vegetables} => {root vegetables} 0.01230300 0.3427762
[4] {beef} => {root vegetables} 0.01738688 0.3313953
[5] {citrus fruit, root vegetables} => {other vegetables} 0.01037112 0.5862069
[6] {tropical fruit, root vegetables} => {other vegetables} 0.01230300 0.5845411
[7] {other vegetables, whole milk} => {root vegetables} 0.02318251 0.3097826
[8] {whole milk, curd} => {yogurt} 0.01006609 0.3852140
[9] {other vegetables, yogurt} => {root vegetables} 0.01291307 0.2974239
[10] {other vegetables, yogurt} => {tropical fruit} 0.01230300 0.2833724

coverage lift count
[1] 0.02887646 3.295045 102
[2] 0.04341637 3.267062 100
[3] 0.03589222 3.144780 121
[4] 0.05246568 3.040367 171
[5] 0.01769192 3.029608 102
[6] 0.02104728 3.020999 121
[7] 0.07483477 2.842082 228
[8] 0.02613116 2.761356 99
[9] 0.04341637 2.728698 127
[10] 0.04341637 2.700550 121
> |
```

Εικόνα 10 Μοτίβο καταναλωτών

Ταξινομούμε τους κανόνες βάσει του lift και επιλέγουμε τους δέκα πρώτους κανόνες με το υψηλότερο lift. Χρησιμοποιούμε τη συνάρτηση inspect για να δούμε τα αποτελέσματα.

Το lift είναι ένα μέτρο της ισχύος ενός κανόνα συσχέτισης. Υποδεικνύει πόσο περισσότερο πιθανό είναι τα προϊόντα στον δεξιό μέρος του κανόνα να αγοραστούν μαζί με τα προϊόντα στον αριστερό μέρος του κανόνα σε σύγκριση με το να αγοραστούν τυχαία.

Από την ανάλυση των κανόνων συσχέτισης, μπορούμε να εξάγουμε πληροφορίες για τα αγοραστικά μοτίβα των καταναλωτών. Οι κανόνες με το υψηλότερο lift μας δείχνουν ποια προϊόντα αγοράζονται συχνά μαζί και υποδεικνύουν ισχυρές συσχετίσεις μεταξύ των προϊόντων.

Για να κατανοήσουμε καλύτερα τα αγοραστικά μοτίβα, μπορούμε να δημιουργήσουμε μια γραφική παράσταση των κανόνων συσχέτισης χρησιμοποιώντας το πακέτο arulesViz και αυτό γιατί το ggplot2 δεν έχει ενσωματωμένες συναρτήσεις για την απεικόνιση των κανόνων συσχέτισης απευθείας. Παρόλα αυτά έγινε και μία απεικόνιση με ggplot2.

Εγκατάσταση του πακέτου arulesViz

```
install.packages("arulesViz")
```

```
library(arulesViz)
```

Δημιουργία γραφικής παράστασης των κανόνων με το υψηλότερο lift

```
plot(top_rules, method = "graph", control = list(type = "items"))
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins

test.R top_rules top_5_rules rules item_freq_df filtered_items_df frequent_items_df
Source on Save Run Source
261 density <- sum(size(Groceries))/ (length(Groceries) * length(itemLabels(Groceries)))
262 cat("Η πυκνότητα των δεδομένων:", density, "\n")
263
264
265 # φόρτωση της βιβλιοθήκης arules
266 library(arules)
267
268 # φόρτωση του dataset
269 data("Groceries")
270
271 (Top Level) R Script

Console Terminal Markers Background Jobs
R 4.4.0 - ~/
L/ 0.0/4034/ 2.042002 220
[8] 0.02613116 2.761356 99
[9] 0.04341637 2.728698 127
[10] 0.04341637 2.700550 121
> # Εγκατάσταση του πακέτου arulesviz (αν δεν είναι ήδη εγκατεστημένο)
> install.packages("arulesviz")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and
install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/mtsilogl/AppData/Local/R/win-library/4.4'
(as 'lib' is unspecified)
also installing the dependencies 'iterators', 'permute', 'zoo', 'Rcpp', 'tweenr', 'polyclip', 'Rcpp
Eigen', 'gridExtra', 'RcppArmadillo', 'later', 'ca', 'foreach', 'gclus', 'qap', 'registry', 'TSP',
'vegan', 'lme4', 'ggforce', 'ggrepel', 'viridis', 'tidygraph', 'graphlayouts', 'htmlwidgets', 'ht
tput', 'crosstalk', 'promises', 'lazyeval', 'seriation', 'vcd', 'igraph', 'scatterplot3d', 'gggrap
h', 'DT', 'plotly', 'visNetwork'

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.4/iterators_1.0.14.zip'
Content type 'application/zip' length 353197 bytes (344 KB)
downloaded 344 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.4/permute_0.9-7.zip'
Content type 'application/zip' length 232415 bytes (226 KB)
downloaded 226 KB

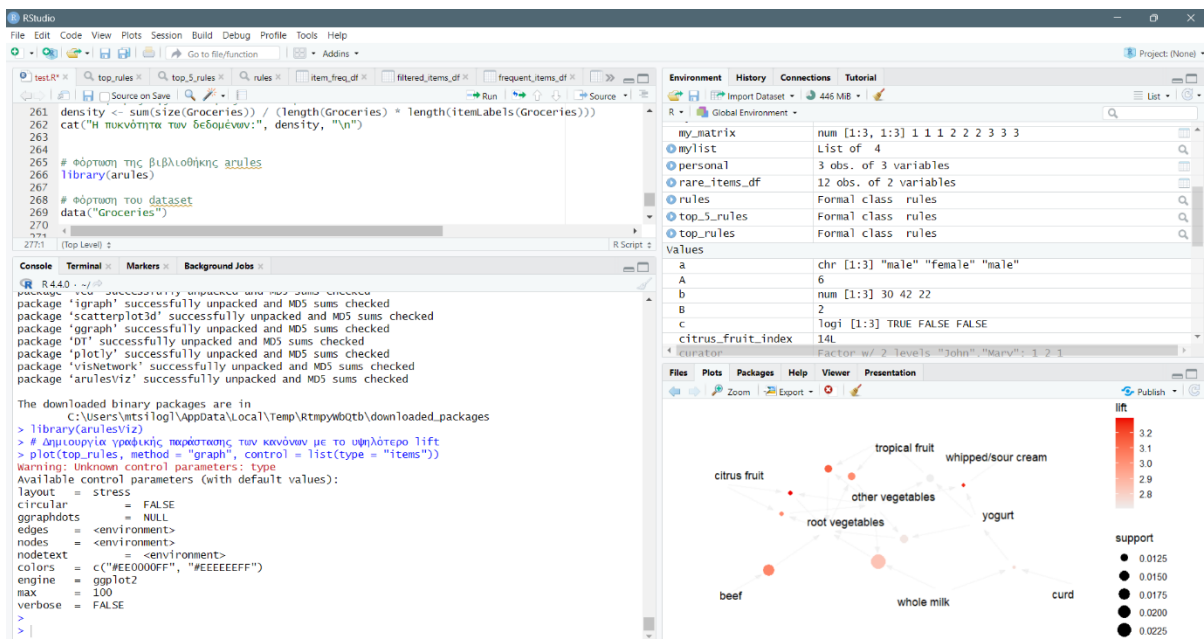
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.4/zoo_1.8-12.zip'
Content type 'application/zip' length 1027879 bytes (1003 KB)
downloaded 1003 KB
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins

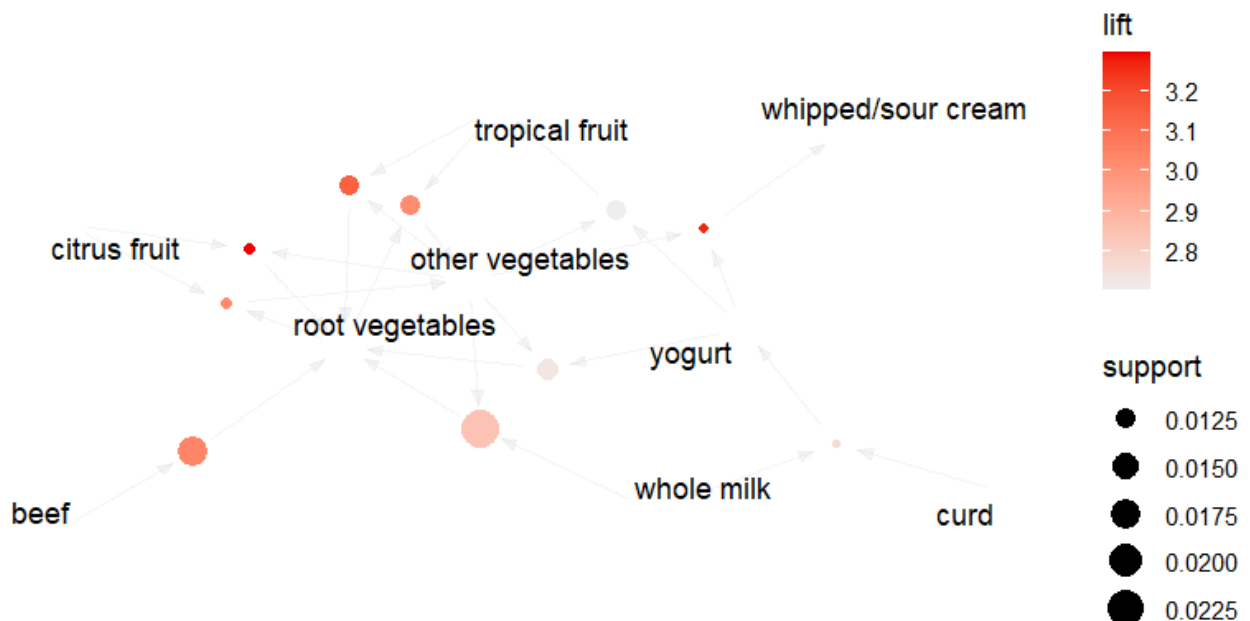
test.R top_rules top_5_rules rules item_freq_df filtered_items_df frequent_items_df
Source on Save Run Source
261 density <- sum(size(Groceries))/ (length(Groceries) * length(itemLabels(Groceries)))
262 cat("Η πυκνότητα των δεδομένων:", density, "\n")
263
264
265 # φόρτωση της βιβλιοθήκης arules
266 library(arules)
267
268 # φόρτωση του dataset
269 data("Groceries")
270
271 (Top Level) R Script

Console Terminal Markers Background Jobs
R 4.4.0 - ~/
Content type 'application/zip' length 1877794 bytes (1.8 MB)
downloaded 1.8 MB

package 'iterators' successfully unpacked and MD5 sums checked
package 'permute' successfully unpacked and MD5 sums checked
package 'zoo' successfully unpacked and MD5 sums checked
package 'Rcpp' successfully unpacked and MD5 sums checked
package 'tweenr' successfully unpacked and MD5 sums checked
package 'polyclip' successfully unpacked and MD5 sums checked
package 'RcppEigen' successfully unpacked and MD5 sums checked
package 'gridExtra' successfully unpacked and MD5 sums checked
package 'RcppArmadillo' successfully unpacked and MD5 sums checked
package 'later' successfully unpacked and MD5 sums checked
package 'ca' successfully unpacked and MD5 sums checked
package 'foreach' successfully unpacked and MD5 sums checked
package 'gclus' successfully unpacked and MD5 sums checked
package 'qap' successfully unpacked and MD5 sums checked
package 'registry' successfully unpacked and MD5 sums checked
package 'TSP' successfully unpacked and MD5 sums checked
package 'vegan' successfully unpacked and MD5 sums checked
package 'lme4' successfully unpacked and MD5 sums checked
package 'ggforce' successfully unpacked and MD5 sums checked
package 'ggrepel' successfully unpacked and MD5 sums checked
package 'viridis' successfully unpacked and MD5 sums checked
package 'tidygraph' successfully unpacked and MD5 sums checked
package 'graphlayouts' successfully unpacked and MD5 sums checked
package 'htmlwidgets' successfully unpacked and MD5 sums checked
package 'httpuv' successfully unpacked and MD5 sums checked
```



Από την ανάλυση των αποτελεσμάτων, παρατηρούμε τα αγοραστικά μοτίβα των καταναλωτών. Οι κανόνες με το υψηλότερο lift δείχνουν ποια προϊόντα αγοράζονται συχνά μαζί, βοηθώντας μας να κατανοήσουμε τις συνήθειες και τις προτιμήσεις των καταναλωτών. Η γραφική παράσταση των κανόνων μας παρέχει μια οπτική αναπαράσταση αυτών των συσχετίσεων, διευκολύνοντας την κατανόηση των αγοραστικών μοτίβων.



Γραφική παράσταση 3 Γραφικής Παράσταση μοτίβο καταναλωτών με `arulesViz`

Παρακάτω έχουμε και την γραφική παράσταση με χρήση ggplot2:

Δημιουργία γραφικής παράστασης με ggplot2

```
ggplot(rules_df, aes(x = reorder(lhs, lift), y = lift, fill = rhs)) +
```

```
  geom_bar(stat = "identity") +
```

```
  coord_flip() +
```

```
  labs(title = "Top 10 Κανόνες Συσχέτισης με το Υψηλότερο Lift", x = "Συνδυασμοί Προϊόντων (lhs)", y = "Lift", fill = "Προϊόν (rhs)") +
```

```
  theme_minimal() →
```

```
> # Δημιουργία γραφικής παράστασης με ggplot2
```

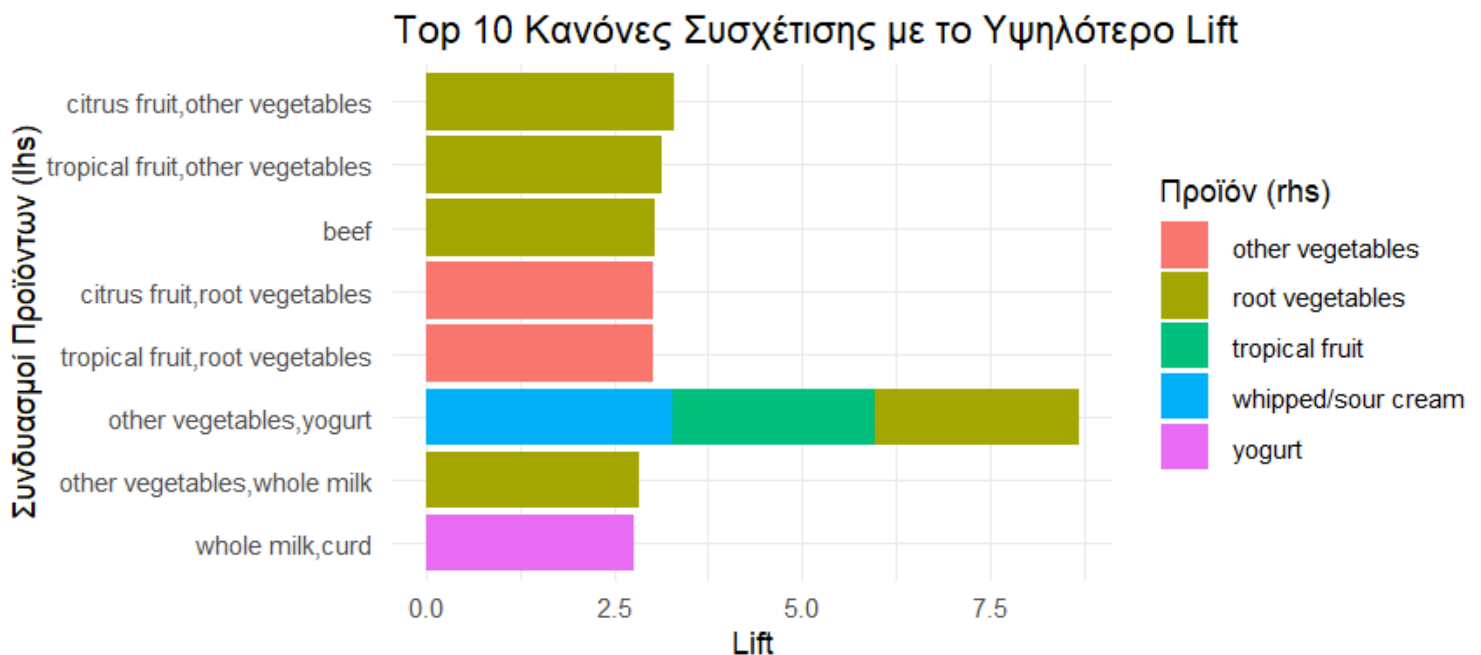
```
> ggplot(rules_df, aes(x = reorder(lhs, lift), y = lift, fill = rhs)) +
```

```
+   geom_bar(stat = "identity") +
```

```
+   coord_flip() +
```

```
+   labs(title = "Top 10 Κανόνες Συσχέτισης με το Υψηλότερο Lift", x = "Συνδυασμοί Προϊόντων (lhs)", y = "Lift", fill = "Προϊόν (rhs)") +
```

```
+   theme_minimal()
```



Γραφική παράσταση 4 Top 10 Κανόνες Συσχέτισης με το Υψηλότερο Lift

Με βάση τα δεδομένα που έχουμε αναλύσει μέχρι τώρα από το dataset "Groceries", το μοτίβο των αγορών των καταναλωτών μπορεί να περιγραφεί ως εξής:

Συχνά Αγοραζόμενα Προϊόντα

Τα προϊόντα που αγοράζονται πιο συχνά περιλαμβάνουν το whole milk, το other vegetables, το rolls/buns και το soda. Αυτά τα προϊόντα έχουν υψηλό support στα δεδομένα μας, υποδεικνύοντας ότι είναι συνηθισμένες αγορές για τους καταναλωτές. Οι καταναλωτές τείνουν να επιλέγουν αυτά τα προϊόντα στις καθημερινές τους αγορές, γεγονός που υποδηλώνει ότι είναι βασικά είδη που χρησιμοποιούνται συχνά.

Σπάνια Αγοραζόμενα Προϊόντα

Από την άλλη πλευρά, υπάρχουν και προϊόντα που αγοράζονται λιγότερο συχνά, όπως το προϊόν baby food, το sound storage medium, το preservation products και το bags. Αυτά τα προϊόντα έχουν χαμηλότερο support και εμφανίζονται σπανιότερα στις αγορές. Αυτό μπορεί να υποδηλώνει ότι αυτά τα προϊόντα είτε είναι ειδικά είδη που δεν χρειάζονται συχνά, είτε δεν είναι δημοφιλή μεταξύ των καταναλωτών.

Συνδυασμοί Προϊόντων

Οι καταναλωτές συνήθως κάνουν αγορές που περιλαμβάνουν πολλά προϊόντα μαζί. Για παράδειγμα, συχνά βλέπουμε συνδυασμούς όπως "butter" και "whipped/sour cream" με "whole milk", "butter" και "yogurt" με "whole milk", "root vegetables" και "butter" με "whole milk" και άλλα. Αυτοί οι συνδυασμοί απεικονίζονται στους κανόνες που ανακαλύψαμε με τον αλγόριθμο Apriori, όπου τα σύνολα προϊόντων με υψηλό confidence δείχνουν τις συνήθειες αγορές των καταναλωτών.

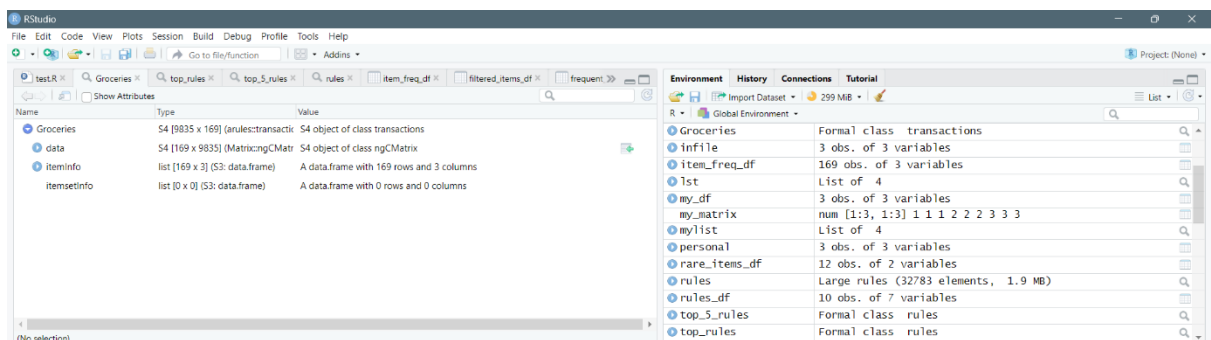
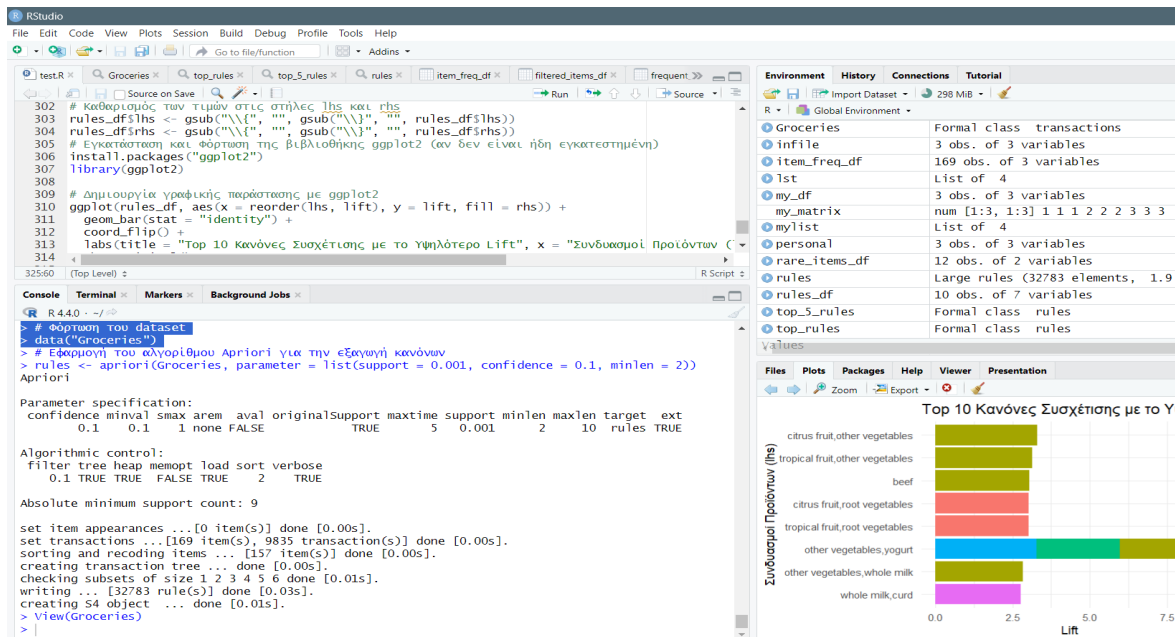
Από την ανάλυση των δεδομένων μπορούμε να διαπιστώσουμε ότι οι καταναλωτές συχνά επιλέγουν προϊόντα που είναι κοινά και δημοφιλή, μερικά από τα οποία συνήθως αγοράζονται μαζί. Υπάρχουν, επίσης, προϊόντα που αγοράζονται σπανιότερα, υποδηλώνοντας εξειδικευμένες ή λιγότερο συνηθισμένες αγορές. Αυτά τα μοτίβα αγορών μπορούν να βοηθήσουν στην κατανόηση των προτιμήσεων των καταναλωτών και στην ανάπτυξη στρατηγικών μάρκετινγκ και διαχείρισης αποθεμάτων.

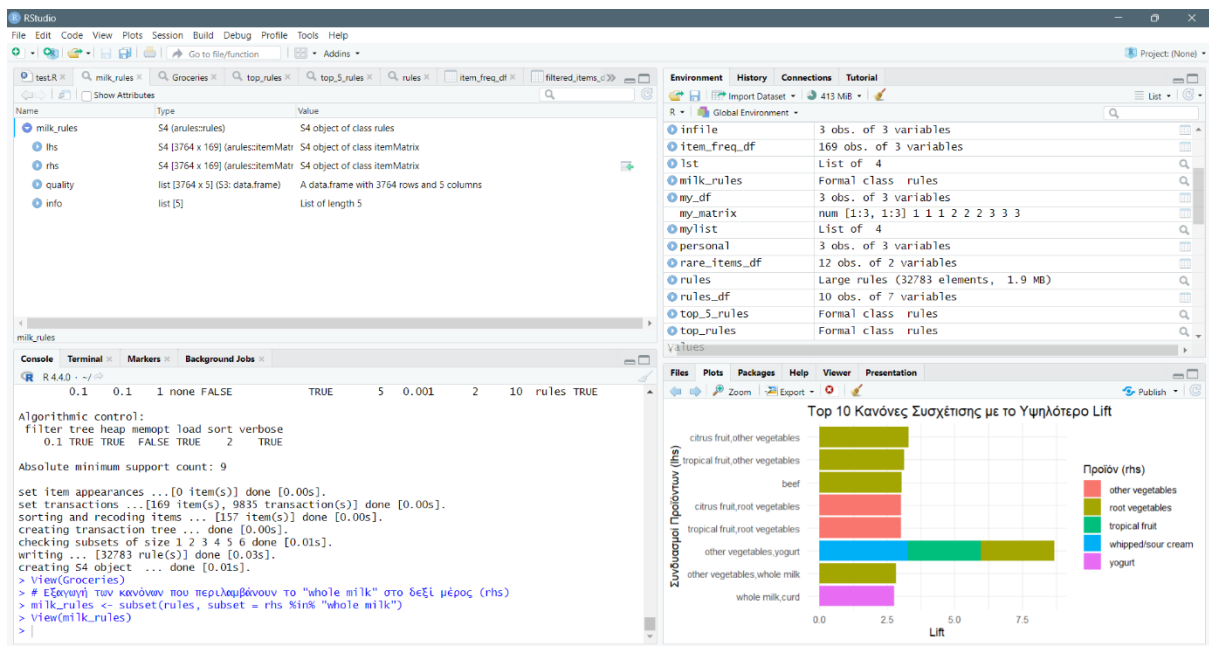
2.2.9 Συνδυασμοί που οδηγούν στην αγορά γάλακτος και η γραφική απεικόνιση αυτού

Σε αυτό το βήμα, θα εξαγάγουμε τους συνδυασμούς προϊόντων που οδηγούν στην αγορά γάλακτος χρησιμοποιώντας τον αλγόριθμο Apriori και θα δημιουργήσουμε μια γραφική απεικόνιση αυτών των συνδυασμών.

```
# Εφαρμογή του αλγορίθμου Apriori για την εξαγωγή κανόνων
rules <- apriori(Groceries, parameter = list(support = 0.001, confidence = 0.1, minlen = 2))

# Εξαγωγή των κανόνων που περιλαμβάνουν το "whole milk" στο δεξί μέρος (rhs)
milk_rules <- subset(rules, subset = rhs %in% "whole milk")
```





Εικόνα 12 Milk rules

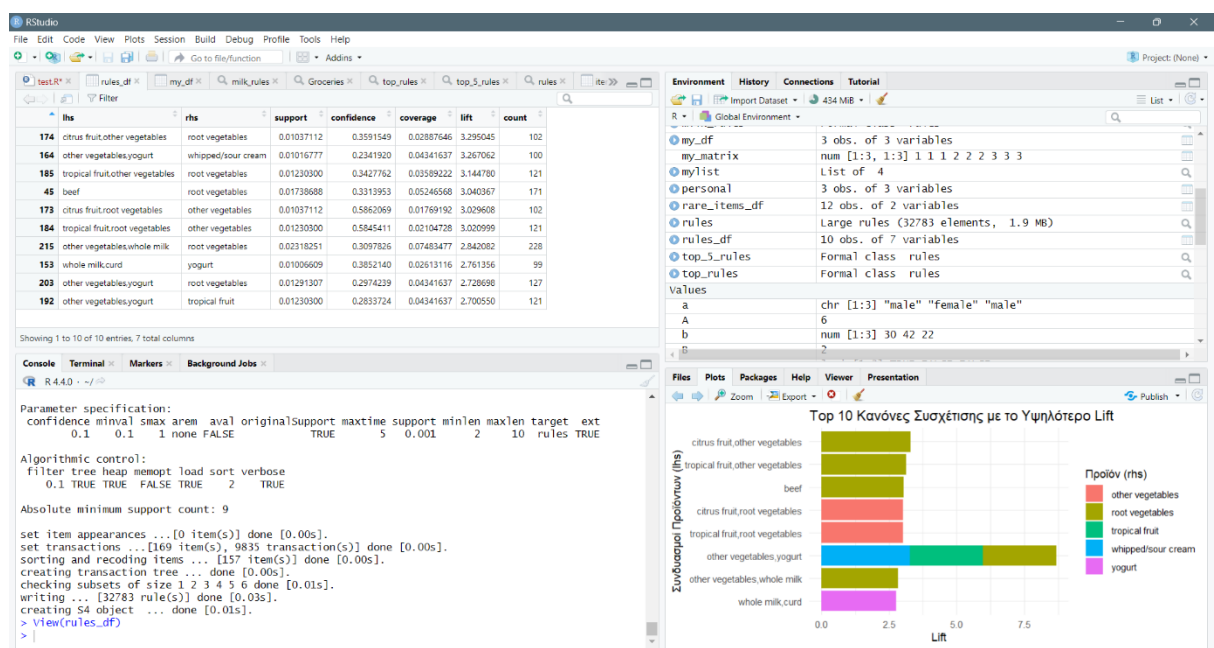
Στη συνέχεια, θα υπολογίσουμε τη συχνότητα εμφάνισης των προϊόντων στο αριστερό μέρος (lhs) των κανόνων που περιλαμβάνουν το "whole milk".

Εφαρμογή του αλγορίθμου Apriori για την εξαγωγή κανόνων

rules <- apriori(Groceries, parameter = list(support = 0.001, confidence = 0.1, minlen = 2))

Εξαγωγή των κανόνων που περιλαμβάνουν το "whole milk" στο δεξι μέρος (rhs)

milk_rules <- subset(rules, subset = rhs %in% "whole milk")



Στη συνέχεια, θα υπολογίσουμε τη συχνότητα εμφάνισης των προϊόντων στο αριστερό μέρος (lhs) των κανόνων που περιλαμβάνουν το "whole milk".

Υπολογισμός της συχνότητας των προϊόντων στο αριστερό μέρος (lhs)

```
lhs_items <- lhs(milk_rules)
```

```
lhs_freq <- itemFrequency(lhs_items)
```

Μετατροπή της συχνότητας σε dataframe

```
lhs_freq_df <- data.frame(item = names(lhs_freq), freq = lhs_freq)
```

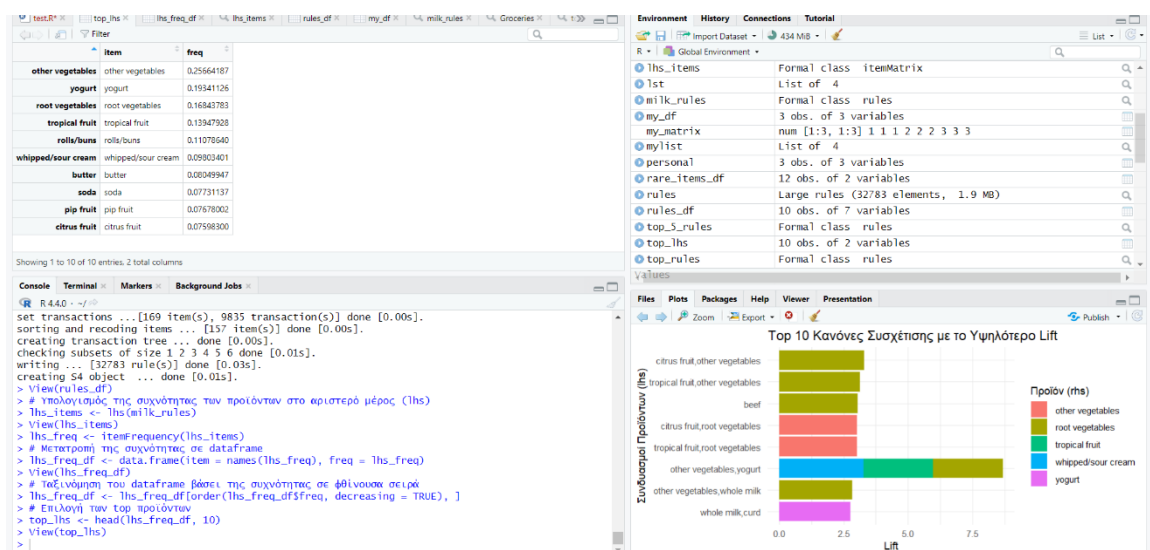
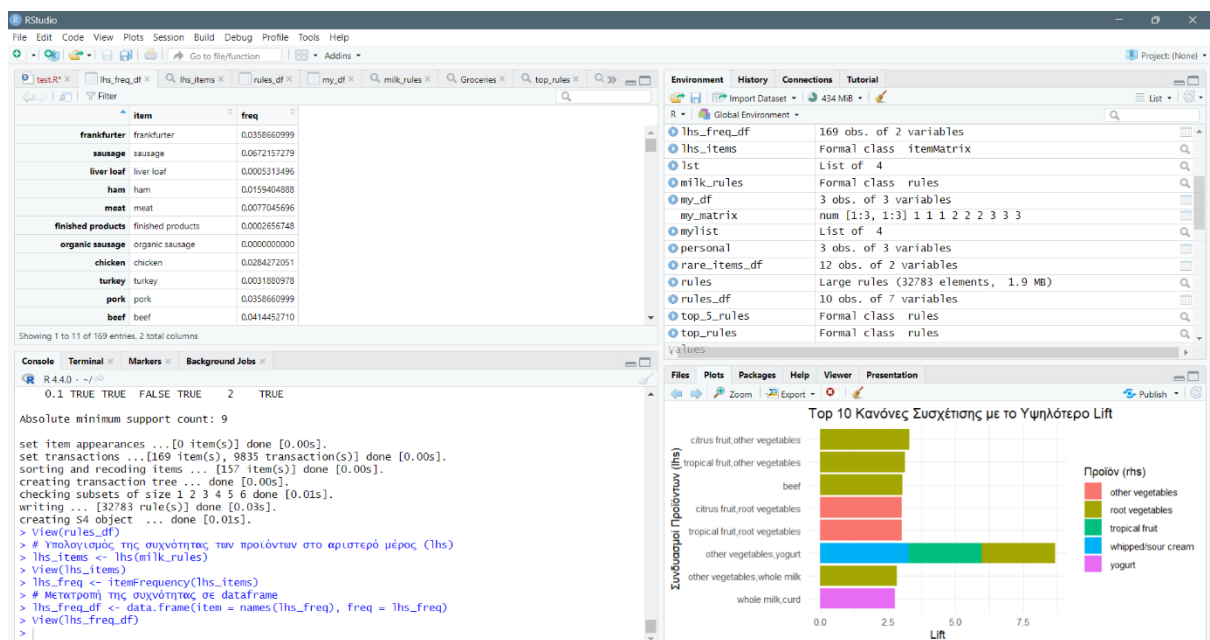
Ταξινόμηση του dataframe βάσει της συχνότητας σε φθίνουσα σειρά

```
lhs_freq_df <- lhs_freq_df[order(lhs_freq_df$freq, decreasing = TRUE),]
```

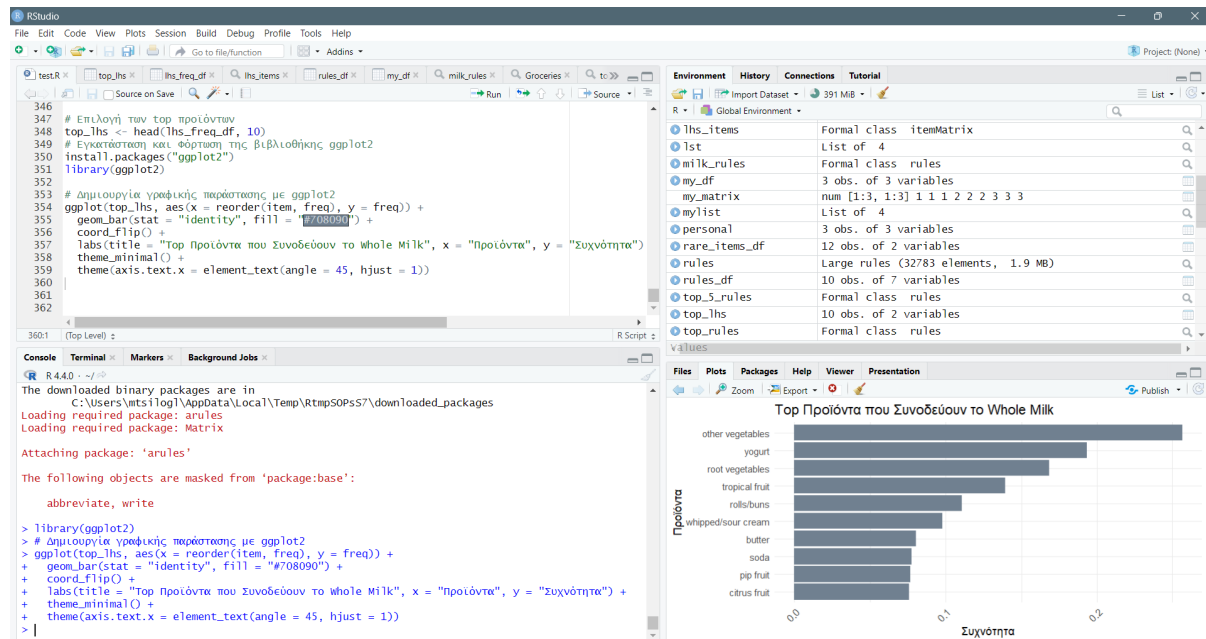
Επιλογή των top προϊόντων

```
top_lhs <- head(lhs_freq_df, 10)
```

→



Τέλος, θα δημιουργήσουμε μια γραφική παράσταση με το ggplot2 για να απεικονίσουμε τη συχνότητα των προϊόντων που συνοδεύουν το "whole milk".



Δημιουργία γραφικής παράστασης με ggplot2

ggplot(top_lhs, aes(x = reorder(item, freq), y = freq)) +

geom_bar(stat = "identity", fill = "#708090") +

coord_flip() +

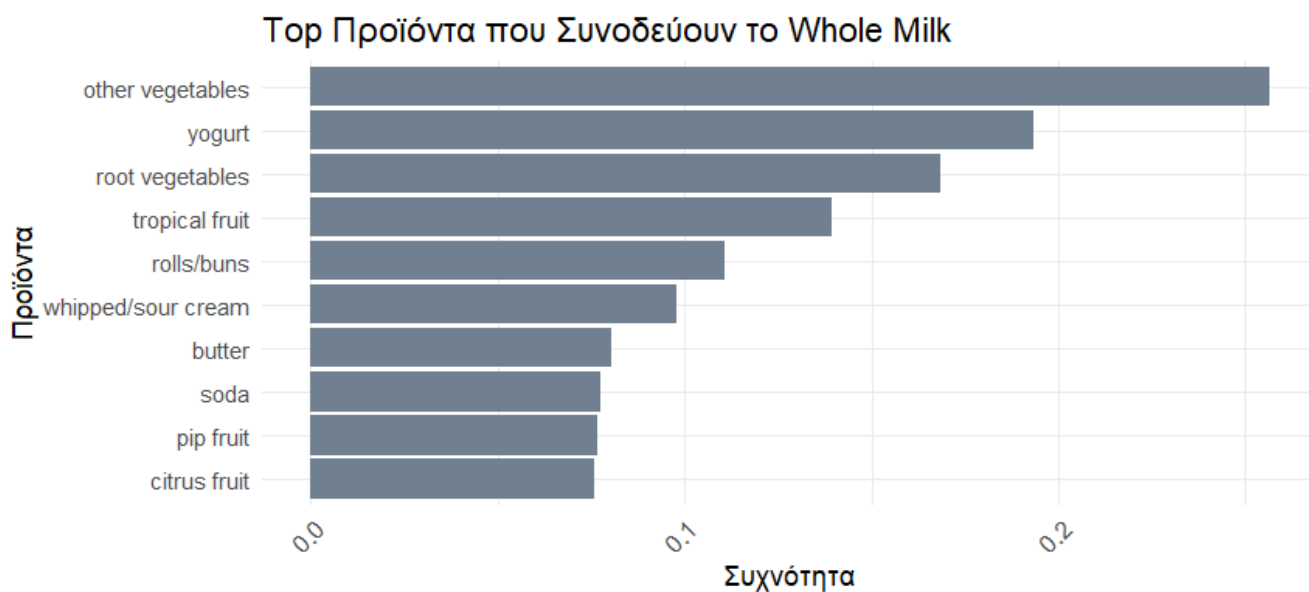
labs(title = "Top Προϊόντα που Συνοδεύουν το Whole Milk", x = "Προϊόντα", y = "Συχνότητα") +

theme_minimal() +

theme(axis.text.x = element_text(angle = 45, hjust = 1))



Το γράφημα που προκύπτει από τον παραπάνω κώδικα απεικονίζει τις συχνότητες των αριστερών πλευρών (LHS) των κανόνων που έχουν ως δεξιά πλευρά το "whole milk". Κάθε μπάρα στο γράφημα αντιστοιχεί σε ένα συγκεκριμένο προϊόν (από τις αριστερές πλευρές των κανόνων), και το ύψος της μπάρας δείχνει τη συχνότητα εμφάνισης αυτού του συνδυασμού στο σύνολο των συναλλαγών.



Γραφική παράσταση 5 Τορ Προϊόντα που Συνοδεύουν το Whole Milk.

Συγκεκριμένα, το γράφημα βοηθά να κατανοήσουμε τις συχνότητες με τις οποίες οι καταναλωτές αγοράζουν κάποια προϊόντα ταυτόχρονα με το γάλα. Αν ένας συνδυασμός προϊόντων έχει υψηλή μπάρα, αυτό σημαίνει ότι αυτός ο συνδυασμός αγοράζεται συχνά μαζί με το γάλα.

Έτσι παρατηρούμε ότι τα 10 κορυφαία μεμονωμένα προϊόντα που αγοράζονται μαζί με το γάλα είναι:

1. Other vegetables
2. Yogurt
3. Root vegetables
4. Tropical fruit
5. Rolls/Buns
6. Whipped/sour cream
7. Butter
8. Soda
9. Pip fruit
10. Citrus fruit

Τα αποτελέσματα που παρουσιάζονται στον πίνακα δείχνουν τη συχνότητα εμφάνισης συγκεκριμένων προϊόντων στο αριστερό μέρος (LHS) των κανόνων συσχέτισης που έχουν ως δεξιά πλευρά (RHS) το "whole milk". Αυτές οι συχνότητες αντικατοπτρίζουν πόσο συχνά τα συγκεκριμένα προϊόντα αγοράζονται μαζί με το γάλα.

File	Edit	Code	View	Plots	Session	Build	Debug	Profile	Tools	Help
Go to file/function										
Arm2_180_212.R	mylist	milk_rules	lhs_items							
Filter										
	item	freq								
other vegetables	other vegetables	0.25664187								
yogurt	yogurt	0.19341126								
root vegetables	root vegetables	0.16843783								
tropical fruit	tropical fruit	0.13947928								
rolls/buns	rolls/buns	0.11078640								
whipped/sour cream	whipped/sour cream	0.09803401								
butter	butter	0.08049947								
soda	soda	0.07731137								
pip fruit	pip fruit	0.07678002								
citrus fruit	citrus fruit	0.07598300								

❖ Αναλυτική Εξέταση των Προϊόντων:

➤ **Other vegetables**

Συχνότητα: 25.66%

Ανάλυση: Τα other vegetables είναι το προϊόν που εμφανίζεται πιο συχνά στους κανόνες που περιλαμβάνουν το γάλα. Αυτό υποδηλώνει ότι οι καταναλωτές συχνά αγοράζουν γάλα μαζί με άλλα λαχανικά, πιθανώς για να καλύψουν τις διατροφικές τους ανάγκες με προϊόντα υψηλής διατροφικής αξίας.

➤ **Yogurt**

Συχνότητα: 19.34%

Ανάλυση: Το yogurt είναι το δεύτερο πιο συχνά εμφανιζόμενο προϊόν. Η συσχέτιση αυτή μπορεί να οφείλεται στο γεγονός ότι τόσο το γάλα όσο και το γιαούρτι είναι γαλακτοκομικά προϊόντα, και οι καταναλωτές που αγοράζουν το ένα είναι πιθανό να αγοράζουν και το άλλο.

➤ **Root vegetables**

Συχνότητα: 16.84%

Ανάλυση: Τα root vegetables έχουν επίσης υψηλή συχνότητα.

➤ **Tropical fruit**

Συχνότητα: 13.95%

Ανάλυση: Τα tropical fruit εμφανίζονται συχνά μαζί με το γάλα, κάτι που μπορεί να υποδηλώνει ότι οι καταναλωτές τα επιλέγουν για να συνδυάσουν τη φρεσκάδα των φρούτων με την πλούσια διατροφική αξία του γάλακτος.

➤ ***Rolls/Buns***

Συχνότητα: 11.08%

Ανάλυση: Τα rolls/buns είναι επίσης ένα κοινό συνοδευτικό του γάλακτος. Αυτό μπορεί να σχετίζεται με το γεγονός ότι αποτελούν μια δημοφιλή επιλογή για πρωινό ή σνακ.

➤ ***Whipped/sour cream***

Συχνότητα: 9.80%

Ανάλυση: Τα whipped/sour cream είναι άλλο ένα προϊόν που συχνά συνοδεύεται από γάλα, δείχνοντας ότι οι καταναλωτές μπορεί να προτιμούν να αγοράζουν μαζί γαλακτοκομικά προϊόντα για τη μαγειρική ή το ψήσιμο.

➤ ***Butter***

Συχνότητα: 8.05%

Ανάλυση: Το butter έχει επίσης υψηλή συχνότητα εμφάνισης μαζί με το γάλα, υποδηλώνοντας ότι οι καταναλωτές συνδυάζουν συχνά αυτά τα δύο προϊόντα στις αγορές τους, πιθανόν για μαγειρική ή άλλες καθημερινές ανάγκες.

➤ ***Soda***

Συχνότητα: 7.73%

Ανάλυση: Τα soda εμφανίζονται μαζί με το γάλα σε ένα αξιοσημείωτο ποσοστό αγορών, δείχνοντας ότι οι καταναλωτές μπορεί να τα αγοράζουν για να καλύψουν διαφορετικές διατροφικές ανάγκες ή προτιμήσεις στην ίδια επίσκεψη στο κατάστημα.

➤ ***Pip fruit***

Συχνότητα: 7.68%

Ανάλυση: Τα pip fruit αγοράζονται επίσης συχνά με το γάλα, ενισχύοντας την ιδέα ότι οι καταναλωτές προτιμούν να συνδυάζουν φρούτα με γαλακτοκομικά προϊόντα.

➤ ***Citrus fruit***

Συχνότητα: 7.60%

Ανάλυση: Τα citrus fruit είναι το δέκατο πιο συχνά αγοραζόμενο προϊόν μαζί με το γάλα.

Τα αποτελέσματα υποδεικνύουν σαφώς τα προϊόντα που οι καταναλωτές συνήθως αγοράζουν μαζί με το γάλα. Τα "other vegetables" και το "yogurt" είναι τα πιο συχνά συνοδευτικά προϊόντα, ενώ και root vegetables και whipped/sour cream εμφανίζονται επίσης συχνά. Αυτή η ανάλυση μας βοηθά να κατανοήσουμε τις διατροφικές προτιμήσεις των καταναλωτών και να προσαρμόσουμε καλύτερα τις στρατηγικές μάρκετινγκ και διαχείρισης αποθεμάτων για να καλύψουμε αυτές τις ανάγκες.

Ηλεκτρονικές πηγές

1. Software Testing Help, <https://www.softwaretestinghelp.com/apriori-algorithm/> , [πρόσβαση 14/06/2024].
2. Apriori - Association Rule Induction / Frequent Item Set Mining, <https://borgelt.net/apriori.html> , [πρόσβαση 14/06/2024].
3. ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics, <https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf> , [πρόσβαση 10/06/2024].
4. RPubS by RStudio, <https://rpubs.com/anguyenbus/associationrulecasestudy> , [πρόσβαση 13/06/2024].
5. piknotita_apantisi_fyllou_s.pdf
https://blogs.sch.gr/wnikdim/files/2016/09/piknotita_apantisi_fyllou_s.pdf?x71199 , [πρόσβαση 12/06/2024].
6. Εισαγωγή στην R, <https://cran.r-project.org/doc/contrib/mainfokianoscharalambous.pdf> , [πρόσβαση 11/06/2024].