



Δημοκρίτειο Πανεπιστήμιο
Θράκης
Democritus University of Thrace

ΔΗΜΟΚΡΙΤΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ
ΣΧΟΛΗ ΔΙΟΙΚΗΤΙΚΗΣ
ΕΠΙΣΤΗΜΗΣ ΚΑΙ
ΛΟΓΙΣΤΙΚΗΣ

ΤΜΗΜΑ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ ΚΑΙ
ΤΕΧΝΟΛΟΓΙΑΣ
ΠΜΣ «ΨΗΦΙΑΚΟ ΜΑΡΚΕΤΙΝΓΚ»



digital
marketing
MSc

ARM.P2

Δίκτυα Chord & Ανάλυση Δεδομένων χρησιμοποιώντας τη γλώσσα R

Λαυρεντίδου Βενέρα Αφροδίτη ΑΕΜ: 239

Ναζλίδου Μαρία ΑΕΜ: 225

Επιβλέποντες καθηγητές:

Επικ. Καθηγητής Στυλιανός Κρηνίδης

Καθ. Γεωργία Κούγκα

Περιεχόμενα

Εισαγωγή	3
Κεφάλαιο 1: Κατασκευή δικτύου CHORD	4
1.1 Υπολογισμός αριθμού κόμβων στο δίκτυο	4
1.2 Υπολογισμός αριθμού θέσεων (m) στο δίκτυο	4
1.3 Κατασκευή των Finger Tables.....	5
1.4 Ορισμός κλειδιών αναζήτησης.....	10
Κεφάλαιο 2: Ανάλυση δεδομένων αγορών και κανόνες συσχέτισης στο Dataset “Groceries”.....	12
2.1 Ανάλυση συνόλου δεδομένων “Groceries”	12
2.2 Εξερεύνηση του Dataset “Groceries” με τη συνάρτηση cat().....	13
2.3 Index Επιλεγμένων Προϊόντων	14
2.4 Γραφική παράσταση προϊόντων με συχνότητα εμφάνισης.....	15
2.5 Σπάνια προϊόντα στο dataset	16
2.6 Οπτική απεικόνιση της αραιότητας των δεδομένων	17
2.7 Εφαρμογή του Αλγορίθμου Apriori – Βήμα 1	18
2.8 Γραφική απεικόνιση συχνότητας αγοράς προϊόντων.....	20
2.9 Μοτίβο καταναλωτικής συμπεριφοράς	21
2.10 Συνδυασμοί Προϊόντων που Οδηγούν στην Αγορά Γάλακτος	22
2.11 Γραφική Απεικόνιση Συνδυασμών Προϊόντων που Συνοδεύουν την Αγορά Γάλακτος	24

Εισαγωγή

Σκοπός της εργασίας στο πρώτο μέρος, αποτελεί η υλοποίηση ενός κατακευματισμένου δικτύου CHORD, το οποίο στηρίζεται σε κυκλική δομή (ring) με θέσεις, όπου κάθε θέση έχει έναν αριθμό (0 έως 63 στην περίπτωση μας) και φιλοξενεί κόμβους.

Στο δεύτερο μέρος, σκοπός είναι η μελέτη των συναλλαγών του dataset “Groceries”, το οποίο περιλαμβάνει δεδομένα από αγορές πελατών. Μέσω της γλώσσας R και του πακέτου `arules`, εξετάζονται βασικά χαρακτηριστικά των δεδομένων, εντοπίζονται συχνές και σπάνιες αγορές, αναλύεται η αγοραστική συμπεριφορά των καταναλωτών και εφαρμόζεται ο αλγόριθμος Apriori για την εξαγωγή κανόνων συσχέτισης, με στόχο την κατανόηση των προτύπων κατανάλωσης.

Κεφάλαιο 1: Κατασκευή δικτύου CHORD

1.1 Υπολογισμός αριθμού κόμβων στο δίκτυο

Για τον εντοπισμό του αριθμού των κόμβων στο δίκτυο, θα αξιοποιηθεί ο εξής τύπος:

- $\text{Αριθμός κόμβων} = (\text{AM1} + \text{AM2}) \bmod 8 + 8$

Όπου ως AM1 και AM2 ορίζονται οι αριθμοί μητρώου μας, δηλαδή AM1 = 239 και AM2 = 225

Υπολογισμός:

Σύμφωνα με τον τύπο, πρώτα πραγματοποιείται πρόσθεση των δύο αριθμών, στην συνέχεια βρίσκεται το υπόλοιπο της διαίρεσης με το 8 (δηλαδή mod 8) και στο τέλος ακολουθεί πρόσθεση με το 8. Παρακάτω θα ακολουθήσει η διαδικασία του υπολογισμού:

- $239 + 225 = 464$
- $464 \bmod 8 = 0$
- $0 + 8 = 8$

Επομένως το δίκτυο θα έχει 8 κόμβους, όπως διαπιστώθηκε και στο excel.

1.2 Υπολογισμός αριθμού θέσεων (m) στο δίκτυο

Το δίκτυο CHORD, δουλεύει με θέσεις που είναι δυνάμεις του 2, δηλαδή ο αριθμός των θέσεων δίνεται από τον τύπο 2^m , όπου m ισούται με τον αριθμό bit που χρειάζεται για να καλύψει τον αριθμό των θέσεων .

Στην εργασία δίνεται δίκτυο με 64 θέσεις , επομένως πρέπει να βρεθεί ποιο m δίνει 64: $2^6 = 64$

Απο τα παραπάνω είναι προκύπτει ότι:

- $m = 6$

- Το δίκτυο έχει θέσεις από 0 έως 63

B3 $\text{=MOD}(B1+B2;8)+8$										
	A	B	C	D	E	F	G	H	I	J
1	AM1	239				Θέση Κόμβων				
2	AM2	225			1	0				
3	Αριθμός Κόμβων	8			2	7				
4					3	15				
5	Βήμα	7			4	22				
6					5	30				
7					6	37				
8					7	45				
9					8	52				
10					9	60				
11					10	67				
12					11	75				
13					12	82				
14					13	90				

Εικόνα: 1

Οι 8 κόμβοι τοποθετούνται ομοιόμορφα στις 64 θέσεις.

Στο excel (Εικόνα:1) φαίνεται πως η τοποθέτηση γίνεται στις θέσεις:

0, 7, 15, 22, 30, 37, 45, 52

Δηλαδή σχεδόν κάθε 7 θέσεις, λόγω $64/8 \approx 8$ θέσεων απόσταση μεταξύ τους, αλλά το βήμα στο excel είναι 7.

1.3 Κατασκευή των Finger Tables

Ο Finger Table, είναι ένας μικρός χάρτης που έχει κάθε κόμβος στο CHORD δίκτυο, ο οποίος τον βοηθά να πηγαίνει γρήγορα σε άλλους κόμβους, χωρίς να χρειάζεται να ψάχνει σε όλο τον κύκλο.

Έτσι, ο κάθε κόμβος διατηρεί έναν πίνακα με m γραμμές (δηλαδή 6 γραμμές για $m=6$) , δηλαδή έναν finger table.

Ο τύπος για κάθε γραμμή είναι:

$$\text{start}(i) = (37 + 2^{i-1}) \bmod 64$$

Παρακάτω θα αναλύσουμε ένα παράδειγμα υπολογισμού για τον κόμβο N0.

i	2^{i-1}	$0+2^{i-1}$	Mod64	Start(i)
---	-----------	-------------	-------	----------

1	1	$0+1=1$	1	1
2	2	$0+2=2$	2	2
3	4	$0+4=4$	4	4
4	8	$0+8=8$	8	8
5	16	$0+16=16$	16	16
6	32	$0+32=32$	32	32

Στην συνέχεια, πρέπει για κάθε $start(i)$ να βρεθεί ο πρώτος κόμβος (successor) που υπάρχει μετά από το $start(i)$ (ή στην ίδια θέση αν υπάρχει).

Όπως ήδη αναφέρθηκε, οι κόμβοι μας βρίσκονται στις θέσεις 0, 7, 15, 22, 30, 37, 45, 52, επομένως οι πρώτοι κόμβοι που υπάρχουν για τον κόμβο $N0$, μετά το $start(i)$, είναι οι εξείς:

Start(i)	Finger[i] (successor)
1	Ο επόμενος κόμβος ≥ 1 είναι 7
2	Ο επόμενος κόμβος ≥ 2 είναι 7
4	Ο επόμενος κόμβος ≥ 4 είναι 7
8	Ο επόμενος κόμβος ≥ 8 είναι 15
16	Ο επόμενος κόμβος ≥ 16 είναι 22
32	Ο επόμενος κόμβος ≥ 32 είναι 37

Ομοίως πράττουμε και με τους υπόλοιπους κόμβους

Finger Table για κόμβο N7

i	Start (i)	Finger[i] (successor)
1	8	15
2	9	15
3	11	15
4	15	15
5	23	30
6	39	45

Finger Table για κόμβο N15

i	Start (i)	Finger[i] (successor)
1	16	22
2	17	22
3	19	22
4	23	30
5	31	37
6	47	52

Finger Table για κόμβο N22

i	Start (i)	Finger[i] (successor)
1	23	30
2	24	30

3	26	30
4	30	30
5	38	45
6	54	0

Finger Table για τον Κόμβο N30

i	start(i)	Finger[i] (successor)
1	31	37
2	32	37
3	34	37
4	38	45
5	46	52
6	62	N0

Finger Table για τον Κόμβο N37

i	start(i)	Finger[i] (successor)
1	38	45
2	39	45
3	41	45
4	45	45
5	53	0
6	5	7

Finger Table για τον Κόμβο N45

i	start(i)	Finger[i] (successor)
1	46	52
2	47	52

3	49	52
4	53	0
5	61	0
6	13	15

Finger Table για τον Κόμβο N52

i	start(i)	Finger[i] (successor)
1	53	0
2	54	0
3	56	0
4	60	0
5	4	7
6	20	22

1.4 Ορισμός κλειδιών αναζήτησης

Στο πρωτόκολλο Chord, κάθε κλειδί τοποθετείται στον κυκλικό χώρο τιμών (identifier space). Στην παρούσα εργασία, το δίκτυο αποτελείται από 64 θέσεις (0 έως 63).

Ο στόχος της αναζήτησης είναι να εντοπιστεί, για κάθε κλειδί, ο $\text{successor}(k)$ δηλαδή ο πρώτος κόμβος του κύκλου με τιμή μεγαλύτερη ή ίση από το κλειδί. Η διαδικασία πραγματοποιείται ξεκινώντας από έναν αρχικό κόμβο και ακολουθώντας τις εγγραφές του finger table μέχρι να φτάσουμε στον υπεύθυνο κόμβο.

Αρχικά, για τον καθορισμό των δύο κλειδιών που θα αναζητηθούν στο δίκτυο CHORD, χρησιμοποιούνται οι αριθμοί μητρώου:

- $AM1 = 239$
- $AM2 = 225$

Χρησιμοποιώντας τον τελεστή modulo 64, που ορίζει το μέγεθος του δακτυλίου ($2^6 = 64$), έχουμε:

1. $Kx = 239 \bmod 64 = 47$
2. $Ky = 225 \bmod 64 = 33$

1. Αναζήτηση για $Kx = 47$

Το κλειδί 47 αναζητείται ξεκινώντας από τον κόμβο N0. Στόχος είναι να βρούμε τον πρώτο κόμβο που είναι μεγαλύτερος ή ίσος από το 47, δηλαδή τον $\text{successor}(47)$. Με βάση τη λίστα των κόμβων: 0, 7, 15, 22, 30, 37, 45, 52, αυτός είναι ο κόμβος N52.

Έπειτα για να φτάσουμε εκεί, ακολουθούμε τα βήματα του πρωτοκόλλου Chord:

- Από τον N0, κοιτάζουμε το finger table και βλέπουμε ότι το $\text{finger}[6]$ οδηγεί στον N37, που είναι ο κοντινότερος κόμβος πριν το 47.

- Από τον N37, επιλέγεται το `finger[4]`, που οδηγεί στον N45, επίσης πριν από το 47.
- Από τον N45, πηγαίνει στον N52, που είναι ο πρώτος κόμβος ≥ 47 .

Επομένως, η διαδρομή αναζήτησης είναι: **N0** $\rightarrow$ **N37** $\rightarrow$ **N45** $\rightarrow$ **N52**
Και τελικά, ο N52 είναι υπεύθυνος για το κλειδί 47.

2. Αναζήτηση για Ky = 33

Η διαδικασία είναι αντίστοιχη. Η αναζήτηση για το κλειδί 33 ξεκινά και πάλι από τον κόμβο N0. Στόχος είναι να βρούμε τον πρώτο κόμβο στον κύκλο που είναι μεγαλύτερος ή ίσος από το 33. Από τη λίστα των κόμβων: 0, 7, 15, 22, 30, 37, 45, 52, βλέπουμε ότι ο πρώτος κόμβος ≥ 33 είναι ο N37.

Κοιτώντας το `finger table` του N0, βλέπουμε ότι το `finger[6]` δείχνει απευθείας στον N37. Επομένως, η αναζήτηση ολοκληρώνεται σε ένα μόνο βήμα, χωρίς να χρειάζονται ενδιάμεσοι κόμβοι.

Η διαδρομή αναζήτησης είναι: **N0** $\rightarrow$ **N37**

Και τελικά, ο N37 είναι υπεύθυνος για το κλειδί 33.

Κεφάλαιο 2: Ανάλυση δεδομένων αγορών και κανόνες συσχέτισης στο Dataset “Groceries”

2.1 Ανάλυση συνόλου δεδομένων “Groceries”

Σκοπός της παρούσας εργασίας είναι η ανάλυση του συνόλου δεδομένων “Groceries”, στο οποίο περιλαμβάνονται οι συναλλαγές πελατών από ένα σουπερμάρκετ. Κάθε συναλλαγή περιέχει τα προϊόντα που αγόρασε ένας καταναλωτής, δηλαδή το καλάθι αγορών του.

Για την υλοποίηση της παρούσας εργασίας χρησιμοποιήθηκε η γλώσσα προγραμματισμού R, η οποία εξειδικεύεται στην στατιστική ανάλυση και στην επεξεργασία δεδομένων. Η εγκατάσταση πραγματοποιήθηκε από τον επίσημο ιστότοπο του R: <https://cran.r-project.org>

Αφού εγκαταστάθηκε η γλώσσα R, στη συνέχεια εγκαταστάθηκε και το γραφικό περιβάλλον RStudio. Η εγκατάσταση του RStudio έγινε και αυτή από την επίσημη ιστοσελίδα: <https://posit.co/download/rstudio-desktop/>

Μετά την εγκατάσταση, δημιουργήθηκε ένα νέο R script όπου φορτώθηκαν τα απαραίτητα πακέτα όπως το `arules`, και πραγματοποιήθηκε η ανάλυση του dataset Groceries. Οι κώδικες για την υλοποίηση των παραπάνω διαδικασιών δίνονται παρακάτω:

Κώδικες:

```
•library(arules)  
  
•data("Groceries")
```

2.2 Εξερεύνηση του Dataset “Groceries” με τη συνάρτηση cat()

Μετά την εισαγωγή της βιβλιοθήκης `arules` και του dataset `Groceries`, χρησιμοποιήθηκε η συνάρτηση `cat()` για να εμφανιστούν τα βασικά στατιστικά στοιχεία του συνόλου δεδομένων, με έναν πιο φιλικό και κατανοητό τρόπο.

Πρωτού αξιοποιηθεί όμως η συνάρτηση `cat()`, εντοπίστηκε με τον κώδικα `class(Groceries)`, ο τύπος δεδομένων του dataset.

```
> class(Groceries)
[1] "transactions"
attr(,"package")
[1] "arules"
```

Εικόνα 2

Το αποτέλεσμα της εκτέλεσης του κώδικα φαίνεται στην Εικόνα 2, δηλαδή το dataset είναι τύπου `transactions`, με άλλα λόγια η κάθε γραμμή είναι ένα καλάθι (μία συναλλαγή) με προϊόντα.

Στην συνέχεια με την αξιοποίηση της συνάρτησης `cat()` εκτελέστηκε ο εξής κώδικας:

```
cat("Αριθμός καλαθιών είναι:", length(Groceries), "\n")
```

Ο οποίος μας εμφάνισε ότι το dataset περιέχει 9835 συναλλαγές, δηλαδή τόσοι πελάτες πραγματοποίησαν αγορές.

Με παρόμοιο τρόπο εκτελέστηκαν και οι εξείς κώδικες :

```
1. cat("Αριθμός διαφορετικών προϊόντων:", length(itemLabels(Groceries)),
"\n")
```

♦ Οδήγησε στο αποτέλεσμα :

Αριθμός διαφορετικών προϊόντων: 169

Δηλαδή το dataset περιλαμβάνει 169 μοναδικά είδη προς πώληση

2. itemLabels(Groceries)

♦ Εμφάνισε αναλυτικά όλα τα προϊόντα, από τρόφιμα όπως “whole milk”, “butter”, μέχρι προϊόντα προσωπικής υγιεινής όπως “soap”, “napkins”.

3. `cat("Ελάχιστος αριθμός προϊόντων σε καλάθι:", min(size(Groceries)), "\n")`

♦ Επέστρεψε ότι ο ελάχιστος αριθμός προϊόντων στο καλάθι είναι 1 ,επομένως υπάρχουν πελάτες που αγόρασαν μόνο 1 προϊόν

4. `cat("Μέγιστος αριθμός προϊόντων σε καλάθι:", max(size(Groceries)), "\n")`

♦ Επέστρεψε ότι ο μέγιστος αριθμός προϊόντων στο καλάθι είναι 32, δηλαδή το μεγαλύτερο καλάθι περιείχε 32 διαφορετικά προϊόντα.

5. `cat("Μέσος αριθμός προϊόντων ανά καλάθι:", mean(size(Groceries)), "\n")`

♦ Επέστρεψε ότι ο μέσος αριθμός ανά καλάθι είναι 4.409456, που σημαίνει ότι κατά μέσο όρο, οι καταναλωτές αγοράζουν περίπου 4 προϊόντα ανά συναλλαγή.

2.3 Index Επιλεγμένων Προϊόντων

Στο dataset Groceries, κάθε προϊόν καταλαμβάνει μια συγκεκριμένη θέση στη λίστα με τα διαθέσιμα προϊόντα (`itemLabels(Groceries)`), και αυτή η θέση ονομάζεται δείκτης (`index`).

Με χρήση της συνάρτησης `which(itemLabels(Groceries) == "όνομα προϊόντος")`, εντοπίστηκε η θέση (`index`) τεσσάρων επιλεγμένων προϊόντων στη λίστα.

```
> cat("Index του citrus fruit:", which(itemLabels(Groceries) == "citrus fruit"), "\n")
Index του citrus fruit: 14
> cat("Index του semi-finished bread:", which(itemLabels(Groceries) == "semi-finished bread"),
"\n")
Index του semi-finished bread: 61
> cat("Index του margarine:", which(itemLabels(Groceries) == "margarine"), "\n")
Index του margarine: 70
> cat("Index του ready soups:", which(itemLabels(Groceries) == "ready soups"), "\n")
Index του ready soups: 79
> |
```

Εικόνα 3

Τα αποτελέσματα (Εικόνα 3) έχουν ως εξής:

- Το προϊόν citrus fruit έχει δείκτη: 14
- Το προϊόν semi-finished bread έχει δείκτη: 61
- Το προϊόν margarine έχει δείκτη: 70
- Το προϊόν ready soups έχει δείκτη: 79

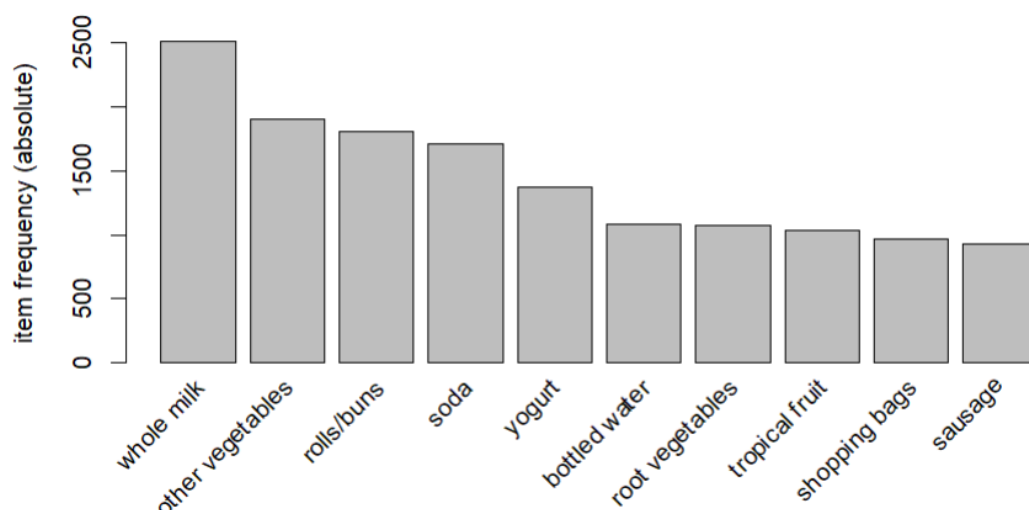
Αυτό σημαίνει ότι τα συγκεκριμένα προϊόντα βρίσκονται στην αντίστοιχη σειρά στη λίστα `itemLabels`, η οποία, όπως αναφέρθηκε προηγουμένως, περιλαμβάνει 169 διαφορετικά προϊόντα.

2.4 Γραφική παράσταση προϊόντων με συχνότητα εμφάνισης

Για την αναγνώριση των πιο συχνά αγορασμένων προϊόντων στο dataset `Groceries`, χρησιμοποιήθηκε η συνάρτηση `itemFrequencyPlot()`. Η συνάρτηση εμφανίζει τα προϊόντα με τη μεγαλύτερη συχνότητα εμφάνισης (δηλαδή, πόσες φορές εμφανίζονται σε όλες τις συναλλαγές).

Κώδικας :

```
itemFrequencyPlot(Groceries, topN=10, type="absolute")
```



Εικόνα 4

Στο γράφημα (Εικόνα 4) που προκύπτει, παρατηρούμε ότι τα προϊόντα με τη μεγαλύτερη συχνότητα εμφάνισης είναι:

- whole milk

- other vegetables
- rolls/buns
- soda
- yogurt
- bottled water
- root vegetables
- tropical fruit
- shopping bags
- sausage

Τα παραπάνω προϊόντα είναι κάποια βασικά είδη διατροφής και καθημερινής χρήσης, γεγονός που δικαιολογεί την υψηλή τους συχνότητα.

2.5 Σπάνια προϊόντα στο dataset

Για τον εντοπισμό των προϊόντων που εμφανίζονται πιο σπάνια στο dataset, χρησιμοποιήθηκε η συνάρτηση `itemFrequency()` σε συνδυασμό με `sort()`, ώστε να εμφανιστούν τα είδη με τη μικρότερη απόλυτη συχνότητα.

Κώδικας :

```
sort(itemFrequency(Groceries, type = "absolute"), decreasing = FALSE)[1:10]
```

```
> sort(itemFrequency(Groceries, type = "absolute"), decreasing = FALSE)[1:10]
      baby food      sound storage medium preservation products      kitchen utensil
           1              1              2              4
      bags      frozen chicken      baby cosmetics      toilet cleaner
           4              6              6              7
      salad dressing              whisky
           8              8
```

Εικόνα 5

Με βάση τα αποτελέσματα της Εικόνας 5, τα προϊόντα που εμφανίζονται με μικρότερη συχνότητα είναι τα εξής:

- baby food 1
- sound storage 1
- preservation products 2
- kitchen utensil 4
- bags 4
- frozen chicken 6
- baby cosmetics 6
- toilet cleaner 7
- salad dressing 8
- whisky 8

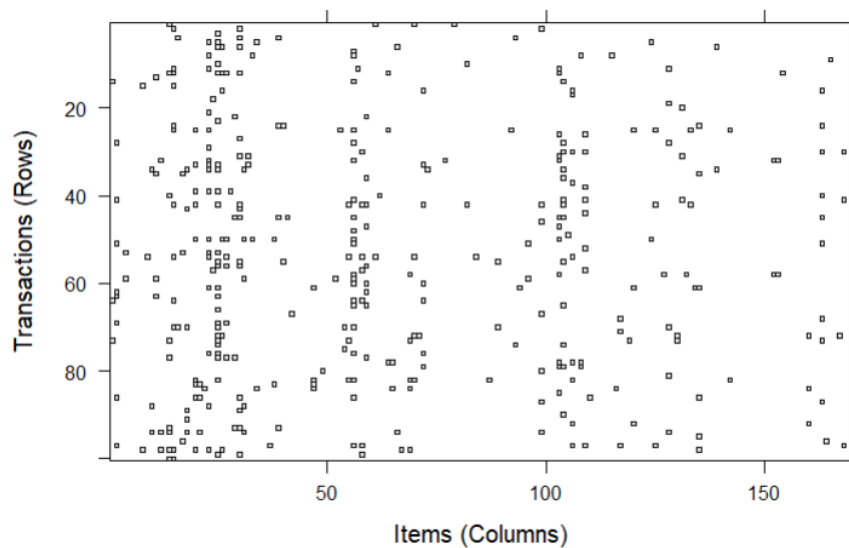
Τα παραπάνω προϊόντα εμφανίζονται μόλις μία με 8 φορές στο σύνολο των 9.835 συναλλαγών, και θεωρούνται σπάνια.

2.6 Οπτική απεικόνιση της αραιότητας των δεδομένων

Για την καλύτερη κατανόηση της δομής του dataset “Groceries”, δημιουργήθηκε ένα γράφημα το οποίο δείχνει ποια προϊόντα περιέχονται σε ποια καλάθια αγορών.

Χρησιμοποιήθηκε η εντολή:

```
image(Groceries[1:100])
```



Εικόνα 6

Με αυτή την εντολή δημιουργείται ένα εικονικό γράφημα (Εικόνα 6) που απεικονίζει τις πρώτες 100 συναλλαγές (καλάθια). Παρατηρώντας το γράφημα που απεικονίζει τις συναλλαγές σε σχέση με τα προϊόντα, διαπιστώνεται ότι το μεγαλύτερο μέρος του είναι κενό, με λίγες μόνο θέσεις να έχουν τιμή. Αυτό δείχνει ότι το dataset είναι αραιό, δηλαδή σε κάθε συναλλαγή αγοράζονται λίγα προϊόντα σε σχέση με τα διαθέσιμα είδη.

2.7 Εφαρμογή του Αλγορίθμου Apriori – Βήμα 1

Για την εξαγωγή κανόνων συσχέτισης στο dataset “Groceries”, χρησιμοποιήθηκε ο αλγόριθμος Apriori, ο οποίος εφαρμόζεται όταν θέλουμε να εντοπίσουμε ποια προϊόντα αγοράζονται συχνά μαζί.

Η εντολή που χρησιμοποιήθηκε ήταν η εξής:

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

Επεξήγηση παραμέτρων της εντολής :

•support = 0.006 → Ο κανόνας πρέπει να εμφανίζεται τουλάχιστον στο 0.6% των συναλλαγών (δηλαδή σε περίπου 59 καλάθια).

•confidence = 0.25 → Ο κανόνας πρέπει να έχει τουλάχιστον 25% πιθανότητα να ισχύει.

•minlen = 2 → Κάθε κανόνας πρέπει να περιλαμβάνει τουλάχιστον 2 προϊόντα, για να είναι ουσιαστικός

Η εντολή δημιούργησε ένα αντικείμενο με το όνομα rules, το οποίο περιέχει όλους τους συχνούς κανόνες που πληρούν τα παραπάνω κριτήρια.

Στην συνέχεια, η εντολή που χρησιμοποιήθηκε για την ταξινόμηση των κανόνων ως προς το confidence ήταν:

```
top5_conf <- head(sort(rules, by = "confidence", decreasing = TRUE), 5)
```

```
inspect(top5_conf)
```

και το αποτέλεσμα είναι:

Συνδυασμός Προϊόντων (lhs)	Προϊόν που Προκύπτει (rhs)	Support	Confidence	Lift	Πλήθος (count)
butter, whipped/sour cream	whole milk	0.0067	66.00%	2.58	66
butter, yogurt	whole milk	0.0093	63.89%	2.50	92
root vegetables, butter	whole milk	0.0082	63.78%	2.49	81
tropical fruit, curd	whole milk	0.0065	63.37%	2.48	64
tropical fruit, butter	whole milk	0.0062	62.24%	2.43	61

Προκύπτει, λοιπόν, από τους κανόνες συσχέτισης που παρήχθησαν μέσω του αλγορίθμου Apriori ότι το πλήρες γάλα αποτελεί το πιο συχνό "αποτέλεσμα" (rhs), δηλαδή αγοράζεται μαζί με άλλα προϊόντα. Για παράδειγμα, όταν κάποιος αγοράζει βούτυρο και κρέμα γάλακτος, υπάρχει 66% πιθανότητα να αγοράσει και γάλα. Το ίδιο ισχύει και για άλλους συνδυασμούς όπως βούτυρο με γιαούρτι ή τροπικά φρούτα με curd. Οι συνδυασμοί αυτοί έχουν υψηλό confidence (πάνω

από 62%) και υψηλό lift, γεγονός που δείχνει ότι δεν είναι τυχαίες συσχετίσεις αλλά αντιπροσωπεύουν πραγματικές καταναλωτικές συνήθειες.

2.8 Γραφική απεικόνιση συχνότητας αγοράς προϊόντων

Για τη σύγκριση των συχνών και των σπάνιων προϊόντων, δημιουργήθηκε μια γραφική παράσταση τύπου ραβδογράμματος (barplot), η οποία παρουσιάζει τα 5 προϊόντα με τη μεγαλύτερη συχνότητα αγορών και τα 5 προϊόντα με τη μικρότερη συχνότητα αγορών

Κώδικας:

```
# Συχνότητα προϊόντων (σε απόλυτους αριθμούς)
freq <- itemFrequency(Groceries, type = "absolute")

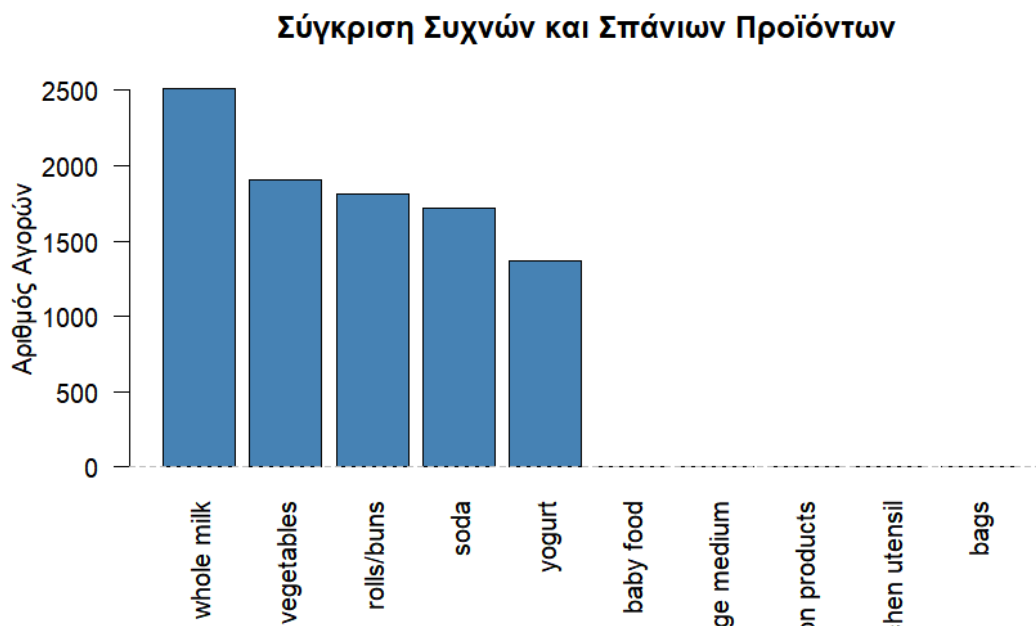
# Top 5 πιο συχνά προϊόντα
top5 <- sort(freq, decreasing = TRUE)[1:5]

# Top 5 πιο σπάνια προϊόντα
bottom5 <- sort(freq, decreasing = FALSE)[1:5]

# Συνδυασμός και γραφική απεικόνιση
combined <- c(top5, bottom5)

# Γράφημα: σύγκριση συχνών vs σπάνιων
barplot(combined,
        col = c(rep("steelblue", 5), rep("tomato", 5)),
        las = 2,
        main = "Σύγκριση Συχνών και Σπάνιων Προϊόντων",
        ylab = "Αριθμός Αγορών")
abline(h = 1, col = "gray", lty = 2) # Προσθέτει μια γραμμή για τις πολύ χαμηλές τιμές
```

Εικόνα 7



Εικόνα 8

Στο γράφημα παρατηρείται μια ξεκάθαρη αντίθεση ανάμεσα στα συχνά και τα σπάνια προϊόντα. Όπως φαίνεται στο διάγραμμα, προϊόντα όπως το γάλα, τα λαχανικά και τα ψωμιά έχουν πολύ υψηλό αριθμό αγορών, σε αντίθεση με προϊόντα όπως το βρεφικό φαγητό, τα μαγειρικά σκεύη και οι σακούλες, τα οποία εμφανίζονται ελάχιστα. Προστέθηκε διακεκομμένη γραμμή στο ύψος 1 για να γίνει πιο εύκολα αντιληπτή η διαφορά. Το φαινόμενο αυτό αντικατοπτρίζει αυτό που συμβαίνει και στην πραγματική αγορά: κάποια πράγματα είναι απαραίτητα για όλους και τα αγοράζουμε συχνά, ενώ άλλα είναι για πιο συγκεκριμένες περιπτώσεις και τα αγοράζουμε σπάνια ή και καθόλου.

2.9 Μοτίβο καταναλωτικής συμπεριφοράς

Από την ανάλυση του dataset Groceries προκύπτει ότι οι καταναλωτές ακολουθούν ένα σταθερό μοτίβο αγορών, το οποίο βασίζεται κυρίως σε βασικά και καθημερινά είδη. Τα προϊόντα που εμφανίζονται με τη μεγαλύτερη συχνότητα, όπως το γάλα, τα λαχανικά, τα ψωμάκια, το γιαούρτι και τα αναψυκτικά, δείχνουν ότι οι περισσότεροι πελάτες αγοράζουν προϊόντα που καλύπτουν καθημερινές διατροφικές ανάγκες.

Επίσης, η εφαρμογή του αλγορίθμου Apriori δείχνει την ύπαρξη ισχυρών συσχετίσεων μεταξύ προϊόντων. Πιο συγκεκριμένα, εντοπίστηκαν επαναλαμβανόμενοι συνδυασμοί προϊόντων όπως:

- butter + whipped/sour cream → whole milk
- butter + yogurt → whole milk
- root vegetables + butter → whole milk

Οι καταναλωτές δεν επιλέγουν τυχαία τα προϊόντα που αγοράζουν, αλλά αγοράζουν προϊόντα που σχετίζονται μεταξύ τους για παράδειγμα, υλικά για ένα γεύμα ή βασικά είδη για το σπίτι.

Τέλος, από την ανάλυση των καλαθιών προέκυψε ότι σε κάθε συναλλαγή αγοράζονται κατά μέσο όρο περίπου 4 προϊόντα. Το στοιχείο αυτό υποδηλώνει ότι οι καταναλωτές πραγματοποιούν μικρές και συχνές αγορές, πιθανότατα για να καλύψουν άμεσες ή καθημερινές ανάγκες.

2.10 Συνδυασμοί Προϊόντων που Οδηγούν στην Αγορά Γάλακτος

Για την εξαγωγή των κανόνων συσχέτισης, εφαρμόστηκε ο αλγόριθμος Apriori στο dataset "Groceries" με παραμέτρους:

- Support ≥ 0.006
- Confidence ≥ 0.25
- Minlen = 2

Στη συνέχεια, φιλτραρίστηκαν μόνο οι κανόνες των οποίων το αποτέλεσμα (rhs) είναι το προϊόν "whole milk", με σκοπό να εντοπιστούν οι πιο σημαντικοί συνδυασμοί που οδηγούν στην αγορά του.

Ο κώδικας που χρησιμοποιήθηκε είναι: `milk_rules <- subset(rules, rhs %in% "whole milk")`

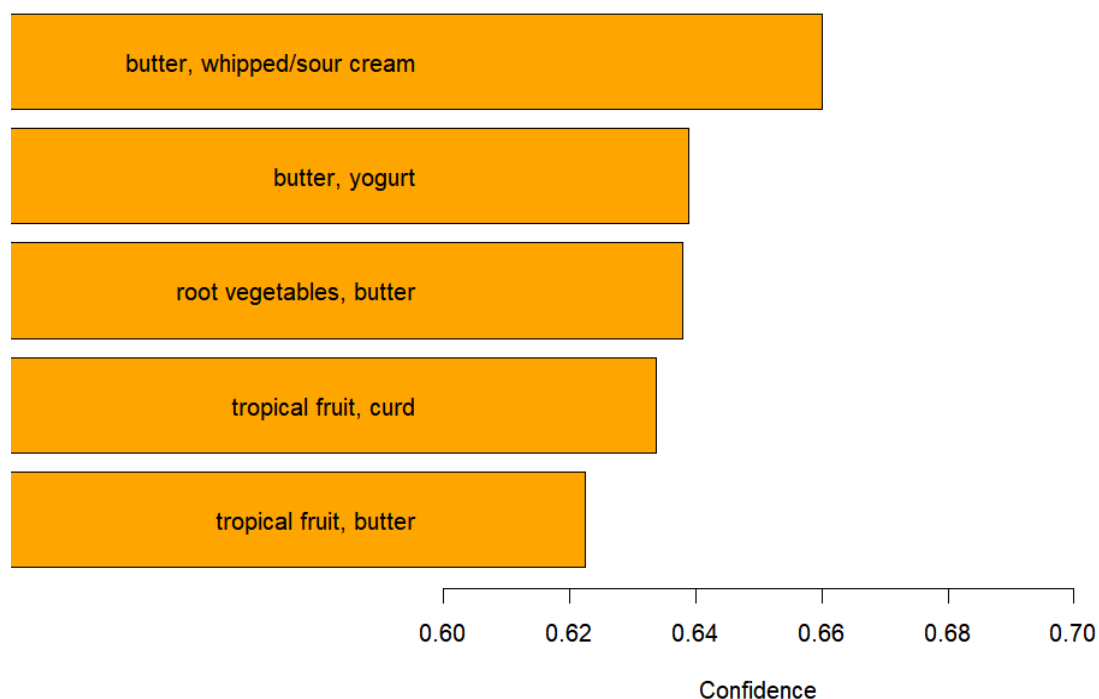
`inspect(head(sort(milk_rules, by = "confidence", decreasing = TRUE), 5))`

Συνδυασμός Προϊόντων (lhs)	Προϊόν που Προκύπτει (rhs)	Confidence
butter, whipped/sour cream	whole milk	66.00%
butter, yogurt	whole milk	63.89%
root vegetables, butter	whole milk	63.78%
tropical fruit, curd	whole milk	63.36%
tropical fruit, butter	whole milk	62.24%

Παρατηρείται πως το γάλα εμφανίζεται σταθερά ως βασικό αγαθό στις αγορές των καταναλωτών, υποδηλώνοντας τη σημασία του στην καθημερινή κατανάλωση. Όταν αγοράζονται προϊόντα όπως βούτυρο, γιαούρτι, φρούτα ή λαχανικά, είναι πολύ πιθανό να προστεθεί και γάλα στο καλάθι. Αυτό δείχνει μια τάση για ομαδικές αγορές συμπληρωματικών ειδών, κυρίως σχετικών με το πρωινό, όπως γαλακτοκομικά και φρούτα. Το μοτίβο αυτό μπορεί να χρησιμοποιηθεί στη στρατηγική τοποθέτηση προϊόντων ή στη δημιουργία προσφορών.

2.11 Γραφική Απεικόνιση Συνδυασμών Προϊόντων που Συνοδεύουν την Αγορά Γάλακτος

Τop 5 Συνδυασμοί που Οδηγούν στην Αγορά Γάλακτος



Εικόνα 9

Για να εντοπιστούν οι συνδυασμοί προϊόντων που αυξάνουν σημαντικά την πιθανότητα να αγοραστεί γάλα χρησιμοποιήθηκε πάλι ο αλγόριθμος Apriori. Από το σύνολο των κανόνων που προέκυψαν, επιλέχθηκαν μόνο εκείνοι στους οποίους το αποτέλεσμα είναι το ίδιο το γάλα, ενώ στη συνέχεια ταξινομήθηκαν κατά φθίνουσα σειρά ως προς την τιμή του confidence, δηλαδή την πιθανότητα να αγοραστεί γάλα όταν υπάρχουν συγκεκριμένα προϊόντα στο καλάθι.

Το παραπάνω γράφημα παρουσιάζει τους 5 ισχυρότερους συνδυασμούς που οδηγούν στην αγορά γάλακτος. Ανάμεσα στους πιο ισχυρούς συνδυασμούς περιλαμβάνονται ζεύγη προϊόντων όπως: butter με whipped/sour cream, butter με yogurt, root vegetables με butter, tropical fruit με curd, και tropical fruit με butter. Οι τιμές του confidence για αυτούς τους συνδυασμούς κυμαίνονται από 62% έως 66%, πράγμα που σημαίνει ότι όταν αυτά τα προϊόντα υπάρχουν σε μια αγορά, είναι εξαιρετικά πιθανό να προστεθεί και γάλα.

Τέλος, τα παραπάνω μπορούν να αξιοποιηθούν εμπορικά με αποτελεσματικούς τρόπους. Για παράδειγμα, μπορούν να δημιουργηθούν προσφορές (όπως γιαούρτι + βούτυρο + γάλα), να τοποθετούν τα προϊόντα κοντά μεταξύ τους στο ράφι για να ενισχυθεί η πιθανότητα συνδυαστικής αγοράς ή να εφαρμοστούν προγράμματα επιβράβευσης που βασίζονται σε αυτούς τους συνδυασμούς. Με αυτόν τον τρόπο, θα ενισχυθεί η εμπειρία του καταναλωτή και ταυτόχρονα θα αυξηθούν οι πωλήσεις.