

ΠΜΣ σε Διοίκηση, Αναλυτική και Πληροφοριακά
Συστήματα Επιχειρήσεων

MSc in Business Administration, Analytics and
Information Systems

Ακαδ. Έτος 2017-18

Διπλωματική Εργασία

**ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ ΣΥΜΠΕΡΙΦΟΡΑΣ
ΚΑΤΑΝΑΛΩΤΩΝ ΜΕΣΑ ΑΠΟ ΔΕΔΟΜΕΝΑ ΣΥΝΑΛΛΑΓΩΝ**

**RECOGNITION OF CONSUMER BEHAVIOR PATTERNS
THROUGH TRANSACTION DATA**

Διακούλη Μαριάννα

Επιβλέπων: Ευάγγελος Βασιλείου

MSc in Business Administration, Analytics and
Information Systems

ΠΜΣ σε Διοίκηση, Αναλυτική και Πληροφοριακά
Συστήματα Επιχειρήσεων

Ακαδ. Έτος 2017-18

Master Thesis

**RECOGNITION OF CONSUMER BEHAVIOR PATTERNS
THROUGH TRANSACTION DATA**

**ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ ΣΥΜΠΕΡΙΦΟΡΑΣ
ΚΑΤΑΝΑΛΩΤΩΝ ΜΕΣΑ ΑΠΟ ΔΕΔΟΜΕΝΑ ΣΥΝΑΛΛΑΓΩΝ**

Diakouli Marianna

Supervisor: Evangelos Vassiliou

Διπλωματική εργασία υποβληθείσα στο Τμήμα Οικονομικών Επιστημών του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών ως μέρους των απαιτήσεων για την απόκτηση Μεταπτυχιακού Διπλώματος Ειδίκευσης στην Διοίκηση, Αναλυτική και Πληροφοριακά Συστήματα και Επιχειρήσεων.

Η Μαριάννα Διακούλη του Ευσταθίου, βεβαιώνεται ότι η υποβληθείσα εργασία είναι προσωπική εκτός όπου γίνεται αναφορά σε εργασίες άλλων.

ΠΕΡΙΛΗΨΗ

Η Ανάλυση συσχέτισης (Association Analysis) και η μάθηση κανόνων συσχέτισης (Association Rule Learning), περιλαμβάνουν ένα σύνολο τεχνικών ανάλυσης που αποσκοπούν στην εύρεση συσχετίσεων μεταξύ συγκεκριμένων στοιχείων. Τα στοιχεία μπορεί να είναι είτε προϊόντα, είτε κατηγορίες προϊόντων σε καταστήματα λιανικής πώλησης, είτε άτομα που πραγματοποιούν αγορές σε ένα συγκεκριμένο ιστότοπο ή σε μη διαδικτυακό χώρο. Ακόμα τέτοιες τεχνικές μπορούν να λάβουν χώρα σε ιατρικά δεδομένα, περιβαλλοντολογικά δεδομένα και σε πολλά άλλα επιστημονικά πεδία. Από το σύνολο των τεχνικών ανάλυσης, η ανάλυση του καλαθιού αγοράς (Market Basket Analysis) είναι η τεχνική με την οποία θα ασχοληθούμε σε αυτή την εργασία.

Το αντικείμενο μελέτης της παρούσας εργασίας είναι η αναγνώριση προτύπων συμπεριφοράς καταναλωτών μέσα από πραγματικά δεδομένα συναλλαγών. Πιο συγκεκριμένα, η ανάλυσή μας αποτελείται από δύο μέρη. Αρχικά, ερευνούμε μέσα από τα καλάθια των καταναλωτών ενός συγκεκριμένου καταστήματος λιανικής πώλησης, για διάστημα δύο μηνών, εάν και σε πιο βαθμό υπάρχουν προϊόντα που αγοράζονται μαζί από τους καταναλωτές. Προσπαθούμε να διακρίνουμε τις καταναλωτικές ανάγκες βρίσκοντας συνδυασμούς προϊόντων και κατηγοριών προϊόντων που εμφανίζονται συχνά μαζί στα καλάθια των καταναλωτών. Εν συνεχεία, στο δεύτερο μέρος της ανάλυσης μας, χρησιμοποιούμε δεδομένα συναλλαγών από το ίδιο κατάστημα, για διάστημα δύο εβδομάδων όπου τη μία από τις δύο εβδομάδες κάποια προϊόντα βρίσκονται σε προσφορά. Με την ανάλυση αυτών των δεδομένων ελέγχουμε αν οι προωθητικές ενέργειες επηρεάζουν τη συμπεριφορά των καταναλωτών ως προς την επιλογή των προϊόντων που τοποθετούν στο καλάθι τους.

Για να αναλύσουμε τα καλάθια των καταναλωτών και να βρούμε πιθανές συσχετίσεις των προϊόντων που περιέχουν, χρησιμοποιούμε τα εργαλεία της ανάλυσης συσχέτισης. Προκειμένου να εξορύξουμε τις ανωτέρω συσχετίσεις χρησιμοποιούμε τον Αλγόριθμο Apriori, έναν αλγόριθμο συσχετίσεων, ο οποίος εντοπίζει αρχικά μεμονωμένα στοιχεία που εμφανίζονται συχνά στη βάση δεδομένων και τα επεκτείνει σε μεγαλύτερα σύνολα, εφόσον τα τελευταία έχουν και αυτά μεγάλη συχνότητα εμφάνισης.

Αξιολογώντας τα αποτελέσματα που εξόρυξε ο αλγόριθμος, καταλήγουμε σε αρκετά συμπεράσματα. Τα αποτελέσματα που προήλθαν από τα πραγματικά δεδομένα του καταστήματος δείχνουν ότι υπάρχουν συσχετίσεις τόσο μεταξύ των προϊόντων όσο και μεταξύ των κατηγοριών, εκ των οποίων οι περισσότερες είναι αναμενόμενες.

ABSTRACT

Association Analysis and Learning Association Rules include a set of analysis techniques designed to find correlations between items. Items can be either products or product categories in retail stores, or people who shop on a specific site or on a non-online site. Moreover, such techniques can take place in medical data, environmental data and in many other scientific fields. From all the analysis techniques, the market basket analysis is the technique with which we will deal with in this work.

The subject of this study is the recognition of consumer behavior patterns through real data transactions. More specifically, our analysis consists of two parts. Initially, we investigate through the consumer baskets of a specific store for a period of two months, if and to what extent there are products purchased at the same time by consumers. We try to discern consumer needs by finding combinations of products and product categories that often appear together in consumer baskets. Then, in the second part of our analysis, we used transaction data from the same store for a two-week period where several products on the second week are on offer. The analysis of these data determines whether the promotional actions affect the behavior of consumers in choosing the products they place together in their basket.

To analyze consumer baskets and find possible associations of the items they contain, we use the tools of association analysis. In order to extract the above associations, we use the Apriori Algorithm, an association algorithm, which initially identifies individual items most frequently displayed in the database and extends them to larger sets, since the latter also have a high incidence.

Evaluating the results that the algorithm exerted, we come up with several conclusions. The results from the actual store data show that there are associations between both products and categories, most of which are expected.

Ευχαριστίες

Στο σημείο αυτό να ευχαριστήσω τον καθηγητή μου κ. Ευάγγελο Βασίλειου για την εμπιστοσύνη που μου έδειξε αλλά και για την βοήθεια που μου προσέφερε στην επίλυση διαφόρων ζητημάτων κατά τη διάρκεια εκπόνησης αυτής της μεταπτυχιακής εργασίας. Επίσης θέλω να ευχαριστήσω τον κ. Δοδόπουλο για όλη την στήριξή του και για τα δεδομένα που μας παρείχε τα οποία αναλύσαμε στην συγκεκριμένη εργασία. Ακόμη, ευχαριστώ όλους τους φίλους που και αυτοί μέσω των συζητήσεων βοήθησαν όλον αυτόν τον καιρό. Τέλος το μεγαλύτερο ευχαριστώ το οφείλω στην οικογένειά μου που με την στήριξή τους ήταν δίπλα μου και κατάφερα να φέρω εις πέρας έναν απαιτητικό κύκλο σπουδών.

Αθήνα, Μάρτιος 2018

Μαριάννα Διακούλη

Πίνακας περιεχομένων

ΠΕΡΙΛΗΨΗ	i
ABSTRACT	ii
Ευχαριστίες.....	iii
Ευρετήριο Σχημάτων.....	III
Ευρετήριο Εικόνων	IV
Ευρετήριο Πινάκων.....	V
Ευρετήριο Ορολογιών.....	VI
ΚΕΦΑΛΑΙΟ 1	1
1.1 Εισαγωγή.....	1
1.2 Ανάλυση Συσχέτισης.....	2
1.2.1 Βασικές Έννοιες	2
ΚΕΦΑΛΑΙΟ 2	7
2.1 Ο Αλγόριθμος Apriori	7
2.2 Εναλλακτικές μέθοδοι για τη δημιουργία συχνών στοιχειοσυνόλων.....	19
2.2.1 Ο Αλγόριθμος Διαμέρισης	19
2.2.2 Ο Αλγόριθμος FP-Growth.....	19
2.2.3 Ο Αλγόριθμος Eclat.....	20
2.3 Αξιολόγηση Κανόνων Συσχέτισης.....	21
2.3.1 Lift.....	21
2.3.2 Conviction	23
2.3.3 Collective Strength	23
2.3.4 Weighted Association Rule Mining	24
2.3.5 Correlation Analysis.....	25
2.3.6 Effect of Skewed Support Distribution	26
2.3.7 All-confidence	27
ΚΕΦΑΛΑΙΟ 3	28
3.1 Πρώτο Μέρος Ανάλυσης.....	28

3.1.1 Περιγραφή Δεδομένων	28
3.1.2 Προετοιμασία Δεδομένων	34
3.1.3 Εξόρυξη Κανόνων Συσχέτισης Συχνών Στοιχειοσυνόλων σε Επίπεδο Κατηγοριών	38
3.2 Δεύτερο Μέρος Ανάλυσης	46
3.2.1 Περιγραφή των Δεδομένων της 1 ^{ης} Εβδομάδας	46
3.2.2 Περιγραφή των Δεδομένων της 2 ^{ης} Εβδομάδας	48
3.2.3 Προετοιμασία Δεδομένων 1 ^{ης} και 2 ^{ης} Εβδομάδας	54
3.2.4 Εξόρυξη Κανόνων Συσχέτισης για τα Δεδομένα της 1 ^{ης} Εβδομάδας	60
3.2.5 Εξόρυξη Κανόνων Συσχέτισης για τα Δεδομένα της 2 ^{ης} Εβδομάδας	65
ΣΥΜΠΕΡΑΣΜΑΤΑ.....	73
ΚΩΔΙΚΕΣ.....	75
ΑΝΑΦΟΡΕΣ.....	94

Ευρετήριο Σχημάτων

Σχήμα 1 Δικτυωτό Πλέγμα.....	9
Σχήμα 2 Υποσύνολα του Στοιχειοσυνόλου {ΓΜΨ}	9
Σχήμα 3 Υπερσύνολα του μη Συχνού Στοιχειοσυνόλου {ΑΒ}.....	10
Σχήμα 4 Κλειστά Συχνά Στοιχειοσύνολα.....	15
Σχήμα 5 Μέγιστα Συχνά Στοιχειοσύνολα	16
Σχήμα 6 Απόρριψη Κανόνων Συσχέτισης.....	18
Σχήμα 7 Κατανομή του Πλήθους των Προϊόντων ανά Συναλλαγή	30
Σχήμα 8 Κατανομή του Πλήθους των Κατηγοριών ανά Συναλλαγή	30
Σχήμα 9 Πλήθος Τμημάτων για τις 2- Πρώτες Συναλλαγές	31
Σχήμα 10 Κατανομή Προϊόντων ανά Κατηγορία.....	31
Σχήμα 11 Κατανομή Προϊόντων ανά Τμήμα	32
Σχήμα 12 Κατανομή Κατηγοριών ανα Τμήμα	32
Σχήμα 13 Δεκαπέντε Προϊόντα με τη Μεγαλύτερη Συχνότητα Εμφάνισης στις Συναλλαγές.....	33
Σχήμα 14 Δώδεκα Κατηγορίες με τη Μεγαλύτερη Συχνότητα Εμφάνισης	33
Σχήμα 15 Η Μεγαλύτερη Συχνότητα Εμφάνισης 20 Προϊόντων στις Συναλλαγές την Εβδομάδα 1	48
Σχήμα 16 Η Μεγαλύτερη Συχνότητα Εμφάνισης 20 Προϊόντων στις Συναλλαγές την Εβδομάδα 2	49
Σχήμα 17 Πωλήσεις Promo Προϊόντων την Εβδομάδα 1 και 2	50
Σχήμα 18 Πλήθος Προϊόντων ανά Καλάθι	51
Σχήμα 19 Αξία Καλαθιού Καταναλωτών για την Εβδομάδα 1 και 2.....	52
Σχήμα 20 Κατανομή Συναλλαγών ανά Πλήθος Προϊόντων για την Εβδομάδα 1.....	53
Σχήμα 21 Κατανομή Συναλλαγών ανά Πλήθος Προϊόντων για την Εβδομάδα 2.....	53

Ευρετήριο Εικόνων

Εικόνα 1 Εμπιστοσύνη των κανόνων που προκύπτουν από το {ΓΚΖ}.....	17
Εικόνα 2 Σύνολο Δεδομένων	29
Εικόνα 3 Επιλογή 9 Τμημάτων από τα 35.....	34
Εικόνα 4 Σύνολο δεδομένων, με την Ύπαρξη 9 Τμημάτων από τα 35.....	35
Εικόνα 5 Υποσύνολο Αρχικού Συνόλου Δεδομένων	36
Εικόνα 6 Αρχείο CSV	37
Εικόνα 7 Σύνοψη Αρχείου CSV.....	37
Εικόνα 8 Οι 10 Κατηγορίες με τη Μεγαλύτερη Συχνότητα Εμφάνισης	38
Εικόνα 9 Σύνοψη Κανόνων.....	39
Εικόνα 10 Κανόνες Συσχέτισης	40
Εικόνα 11 Συνέχεια Εικόνας 10	41
Εικόνα 12 Σύνοψη Κανόνων.....	44
Εικόνα 13 Εμφάνιση 10 Κανόνων	45
Εικόνα 14 Σύνολο Δεδομένων 1ης Εβδομάδας.....	47
Εικόνα 15 Σύνολο Δεδομένων 2ης Εβδομάδας.....	49
Εικόνα 16 Υποσύνολο Εβδομάδας 1.....	55
Εικόνα 17 Υποσύνολο Εβδομάδας 2.....	55
Εικόνα 18 Καλάθια Καταναλωτών Εβδομάδα 1	56
Εικόνα 19 Καλάθια Καταναλωτών Εβδομάδα 2.....	57
Εικόνα 20 Αρχείο CSV Εβδομάδα 1	57
Εικόνα 21 Αρχείο CSV Εβδομάδα 2	57
Εικόνα 22 Σύνοψη Εβδομάδας 1	58
Εικόνα 23 Σύνοψη Εβδομάδα 2	58
Εικόνα 24 Προϊόντα Μεγαλύτερης Συχνότητας Εμφάνισης Εβδομάδα 1	59
Εικόνα 25 Προϊόντα Μεγαλύτερης Συχνότητας Εμφάνισης Εβδομάδα 2	59
Εικόνα 26 Σύνοψη Κανόνων Εβδομάδα 1	60
Εικόνα 27 Κανόνες Συσχέτισης Εβδομάδα 1	61
Εικόνα 28 Συνέχεια Εικόνας 27	62
Εικόνα 29 Κανόνες με το Υψηλότερο Lift.....	64
Εικόνα 30 Σύνοψη Κανόνων Εβδομάδα 2	65
Εικόνα 31 Κανόνες Συσχέτισης Εβδομάδα 2.....	66
Εικόνα 32 Συνέχεια Εικόνας 31	67
Εικόνα 33 Συνέχεια Εικόνας 32	68

Ευρετήριο Πινάκων

Πίνακας 1 Παράδειγμα Συναλλαγών	1
Πίνακας 2 Παράδειγμα Βάσης Δεδομένων ενός Καταστήματος	5
Πίνακας 3 Παράδειγμα Αpriori	13
Πίνακας 4 Υποστήριξη Στοιχειοσυνόλων	17
Πίνακας 5 Περιγραφή Δεδομένων	29
Πίνακας 6 Περιγραφή Δεδομένων	35
Πίνακας 7 Περιγραφή Δεδομένων 1 ^{ης} Εβδομάδας	46
Πίνακας 8 Περιγραφή Δεδομένων 2 ^{ης} Εβδομάδας	48
Πίνακας 9 Προϊόντα που Εμφανίζονται και τις Δύο Εβδομάδες στο Δεξί Μέλος των Κανόνων .	70

Ευρετήριο Ορολογιών

Ελληνικός Όρος	Αγγλικός Όρος
A	
Αλφαβητική Σειρά	Lexicographic Order
Ανάλυση Συσχέτισης	Association Analysis
Αντιμονότονη Ιδιότητα	Anti-monotonicity
Από κάτω προς τα πάνω	Bottom- Up
B	
Βασίζεται στο πρόθεμα	Prefix- Based
Δ	
Δεδομένα συναλλαγών	Market Basket transactions
Δικτυωτό πλέγμα	Lattice Structure
E	
Εμπιστοσύνη	Confidence
Επακόλουθο	Consequent
K	
Καλάθι	Basket
Κανόνας Συσχέτισης	Association Rule
Κατώφλι Εμπιστοσύνης	Minconf
Κατώφλι Υποστήριξης	Minsup
Κλειστό συχνό στοιχειοσύνολο	Closed Frequent Itemset
Λ	
Λοξή προς τα αριστερά	Left Skewed
Λοξή προς τα δεξιά	Right Skewed
M	
Μέγιστο συχνό στοιχειοσύνολο	Maximal Frequent Itemset
Π	
Προς τα κάτω ιδιοκτησία κλεισίματος	Downward-closure property
Πρότερο	Antecedent
Πρώτα κατά βάθος	Depth-first
Πρώτα κατά πλάτος	Breadth-first
Σ	
Σταυροειδής υποστήριξη	Cross- Support
Στοιχείο	Item
Στοιχειοσύνολο	Itemset
Συχνό ή (μεγάλο) στοιχειοσύνολο	Frequent Itemset
Συχνότητα Εμφάνισης	Support Count
Υ	
Υπερσύνολο	Superset
Υποστήριξη	Support
Υποσύνολο	Subset
Υποψήφιοι	Candidates

ΚΕΦΑΛΑΙΟ 1

1.1 Εισαγωγή

Η παρούσα εργασία, στοχεύει στην αναγνώριση προτύπων συμπεριφοράς καταναλωτών μέσα από δεδομένα συναλλαγών. Το αντικείμενο μελέτης της εργασίας, αφορά στα καλάθια των καταναλωτών ενός συγκεκριμένου καταστήματος σουπερμάρκετ. Το καλάθι αγορών είναι τα προϊόντα που αγοράζει ένας καταναλωτής κατά την επίσκεψή του στο κατάστημα. Η επίσκεψη αυτή εισάγεται στη βάση δεδομένων του καταστήματος κατά την τιμολόγηση των προϊόντων από τις ταμειακές μηχανές. Επομένως, η επίσκεψη ενός καταναλωτή στο κατάστημα λέμε ότι έχει ως αποτέλεσμα την εισαγωγή μιας συναλλαγής μέσα στη βάση δεδομένων των πωλήσεων. Με τον τρόπο αυτό συσσωρεύονται τεράστιες ποσότητες τέτοιων δεδομένων στο κατάστημα και καταγράφονται όλες οι συναλλαγές σε συνεχόμενη βάση καθ' όλη τη διάρκεια λειτουργίας του καταστήματος.

Στόχος της ανάλυσης του καλαθιού αγοράς, είναι η αναζήτηση προϊόντων, τα οποία συνήθως αγοράζονται μαζί από τους πελάτες του καταστήματος. Παράδειγμα καλαθιού αγοράς είναι ο Πίνακας 1, στον οποίο παρουσιάζονται 5 καλάθια συναλλαγών. Η κάθε σειρά αναφέρεται σε ένα καλάθι και εκπροσωπείται από έναν μοναδικό αριθμό συναλλαγής, ενώ η δεύτερη στήλη περιέχει τα προϊόντα που περιέχονται σε κάθε καλάθι.

Αρ.Συν	Προϊόντα
1	{Καφές, Γάλα, Μπύρες, Μακαρόνια, Πορτοκαλάδα}
2	{Καφές, Αυγά, Γάλα, Ψωμί}
3	{Ψωμί, Ρύζι, Ουίσκι, Γιαούρτι}
4	{Καφές, Αυγά, Γάλα, Ζάχαρη}
5	{Πατατάκια, Μπύρες, Ξηροί Καρποί, Καφές}

Πίνακας 1 Παράδειγμα Συναλλαγών

Αν οι πελάτες του καταστήματος τείνουν να αγοράζουν από κοινού κάποια προϊόντα, αυτό το γεγονός μπορεί να βοηθήσει τις προσπάθειες του τμήματος Μάρκετινγκ που παρακολουθεί τις πωλήσεις του εκάστοτε προϊόντος, στη διαχείριση των αποθεμάτων και σε εφαρμογές τακτικών. Τέτοιες εφαρμογές είναι και οι προωθητικές ενέργειες μέσω των οποίων ωφελείται και η επιχείρηση και ο πελάτης. Παραδείγματος χάριν, όταν ένας πελάτης αγοράζει καφέ με

ζάχαρη, και το τελευταίο βρίσκεται σε προσφορά, πιθανώς ενθαρρύνεται να ξοδέψει περισσότερα χρήματα στο καλάθι των αγορών του.

Η μεθοδολογία που αναγνωρίζει σχέσεις μεταξύ στοιχείων, καλείται **Association Analysis** και τα βασικά της εργαλεία είναι οι **κανόνες συσχέτισης**, το **support** και το **confidence**, όπου θα αναλυθούν στην επόμενη ενότητα. Τέλος, για την εξόρυξη σχέσεων μεταξύ στοιχείων, υπάρχει ένα φάσμα αλγορίθμων που διατίθενται. Στην παρούσα ανάλυση χρησιμοποιούμε τον Αλγόριθμο Apriori, έναν αλγόριθμο συσχέτισεων, την διαδικασία του οποίου θα αναλύσουμε διεξοδικά αργότερα.

Επίσης, εκτός από τα δεδομένα του καλάθιου αγοράς, η ανάλυση συσχέτισης εφαρμόζεται και σε άλλα πεδία εφαρμογής όπως η βιοπληροφορική, η ιατρική διάγνωση, η εξόρυξη ιστού και η ανάλυση επιστημονικών δεδομένων [1], [2], [3]. Στην ανάλυση των δεδομένων της γης, τα μοντέλα συσχέτισης μπορούν να αποκαλύψουν ενδιαφέρουσες συσχετίσεις μεταξύ των ωκεανών, της γης και των ατμοσφαιρικών διεργασιών. Τέτοιες πληροφορίες μπορούν να βοηθήσουν τους επιστήμονες να αναπτύξουν μια καλύτερη κατανόηση του τρόπου αλληλεπίδρασης των διαφόρων στοιχείων του συστήματος της γης. Παρόλο που οι τεχνικές που υπάρχουν είναι γενικά εφαρμόσιμες σε μια ευρύτερη ποικιλία συνόλων δεδομένων, στη παρούσα εργασία, η συζήτησή μας θα επικεντρωθεί κυρίως στα δεδομένα συναλλαγών αγοράς.

1.2 Ανάλυση Συσχέτισης

*Η μεθοδολογία η οποία μπορεί να αναγνωρίσει σχέσεις μεταξύ στοιχείων, ορίζεται ως Association Analysis[4]. Η εύρεση σχέσεων που μπορούν να ανακαλυφθούν μέσα από μια βάση δεδομένων, λέγονται κανόνες συσχέτισης οι οποίοι εισήχθησαν από τον Agrawal και τους συνεργάτες του το 1993 [5]. Κανόνας συσχέτισης ορίζεται ως το μοντέλο που καθορίζει το «πως» τα στοιχεία σχετίζονται το ένα με το άλλο και είναι της μορφής $A \rightarrow B$, όπου A καλείται πρότερο και B καλείται επακόλουθο [6]. Για να γίνει πιο κατανοητό τι ακριβώς ορίζουμε ως κανόνες συσχέτισης, πρέπει να ορίσουμε και τις έννοιες: **itemset**, **support** και **confidence**.*

1.2.1 Βασικές Έννοιες

Έστω $G = \{g_1, g_2, \dots, g_n\}$, το σύνολο όλων των προϊόντων του καταστήματος, και $\Sigma = \{\sigma_1, \sigma_2, \dots, \sigma_3\}$ το σύνολο όλων των συναλλαγών που πραγματοποιήθηκαν από καταναλωτές για μια συγκεκριμένη περίοδο στο συγκεκριμένο κατάστημα, έχουν καταγραφεί στη βάση δεδομένων του καταστήματος. Το μέγεθος της συναλλαγής καθορίζεται από το πλήθος των προϊόντων που περιέχονται στη συναλλαγή. Ένα σύνολο σ_n , είτε περιέχει στοιχεία είτε είναι κενό σύνολο, καλείται στοιχειοσύνολο. Όταν ένα στοιχειοσύνολο περιέχει πλήθος i στοιχείων καλείται i -στοιχειοσύνολο. Η συναλλαγή σ_n , λέγεται ότι περιέχει ένα στοιχειοσύνολο A , μόνο εάν το A

είναι υποσύνολο του σ_n . Κάθε συναλλαγή σ_n για $1 < n \leq N$, περιέχει τουλάχιστον ένα υποσύνολο του G . Ένας κανόνας συσχέτισης είναι ένα επαγωγικό συμπέρασμα της μορφής $A \rightarrow B$, όπου τα A και B δεν έχουν κανένα κοινό στοιχείο $A \cap B = \emptyset$, και ισχύει ότι $A \subseteq G$ και $B \subseteq G$. Παραδείγματος χάριν, ένας κανόνας θα μπορούσε να είναι της μορφής $\{\text{Καφές}, \text{Ζάχαρη}\} \rightarrow \{\text{Γάλα}\}$, που σημαίνει ότι ένα υποσύνολο των καταναλωτών που θα αγοράσουν καφέ και ζάχαρη, πιθανώς τείνουν να αγοράσουν και γάλα. Ένα από τα σημαντικότερα χαρακτηριστικά ενός στοιχειοσυνόλου είναι η συχνότητα εμφάνισης του δηλαδή σε πόσες συναλλαγές το συγκεκριμένο στοιχειοσύνολο εμφανίζεται. Από μαθηματικής απόψεως η υποστήριξη στοιχειοσυνόλου A εκφράζεται ως,

$$\sigma(A) = |\sigma_n | A \subseteq \sigma_n, \sigma_n \in \Sigma | \quad \text{Εξίσωση 2.1}$$

όπου το σύμβολο $|\cdot|$ υποδηλώνει το πλήθος των στοιχείων που βρίσκονται σε ένα σύνολο. Παραδείγματος χάριν, στον Πίνακα 1 του προηγούμενου κεφαλαίου, η συχνότητα εμφάνισης του στοιχειοσυνόλου $\{\text{Καφές}, \text{Γάλα}\}$ είναι τρία, καθώς το συγκεκριμένο σύνολο εμφανίζεται σε τρεις συναλλαγές.

Για κάθε κανόνα συσχέτισης ορίζουμε ένα μέτρο που ονομάζεται **υποστήριξη** και **εμπιστοσύνη**.

- Η Υποστήριξη δύο στοιχειοσυνόλων A και B που αποτελούν έναν κανόνα της μορφής $A \rightarrow B$ είναι το ποσοστό των συναλλαγών όπου τα A και B στοιχειοσύνολα εμφανίζεται μαζί σε όλες τις συναλλαγές, καθορίζοντας κατά πόσο ο κανόνας είναι εφαρμόσιμος σε ένα υφιστάμενο σύνολο δεδομένων [7].
- Η εμπιστοσύνη ορίζεται με βάση την υποστήριξη των στοιχειοσυνόλων A και $A \cup B$ και ερμηνεύεται ως το ποσοστό των φορών όπου τα στοιχειοσύνολα A και B εμφανίζονται μαζί στις συναλλαγές όπου υπάρχει το στοιχειοσύνολο A [8], [9]. Παρατηρούμε ότι και τα δύο μέτρα, αφορούν πιθανότητες, συνεπώς παίρνουν τιμές στο διάστημα $[0,1]$.

$$\text{Support} \quad \mathcal{S}(A \rightarrow B) = \frac{\text{supp}(A \cup B)}{N} \quad \text{Εξίσωση 2.2}$$

$$\text{Confidence} \quad \mathcal{C}(A \rightarrow B) = \frac{\text{supp}(A \cup B)}{\text{supp}(A)} \quad \text{Εξίσωση 2.3}$$

Σύμφωνα με όσα προαναφέραμε, στόχος μας είναι να βρούμε με αποδοτικό και γρήγορο τρόπο κανόνες συσχέτισης από ένα υφιστάμενο σύνολο δεδομένων. Εάν τα δεδομένα που χρησιμοποιούμε για την εύρεση κανόνων είναι πολλά, η εύρεση κανόνων συσχέτισης δεν είναι εύκολη διαδικασία εξαιτίας του όγκου των δεδομένων. Επομένως, η προσπάθεια να εξορύξουμε κανόνες από μεγάλες βάσεις δεδομένων είναι υπολογιστικά πολύ ακριβή. Φανταστείτε να κληθούμε να βρούμε πιθανές σχέσεις στοιχείων από μια βάση δεδομένων που περιέχει χιλιάδες στοιχεία. Κάτι τέτοιο θα απαιτούσε πολύ χρόνο και θα ήταν ιδιαίτερα

δύσκολο. Επίσης, κάποιοι από τους κανόνες που δημιουργούνται δεν είναι αληθείς. Πολλές φορές, η δημιουργία τους οφείλεται σε τυχαιότητα. Δεδομένου ενός συνόλου δεδομένων G που περιέχει g στοιχεία, μπορούν να δημιουργηθούν $2^g - 1$ πιθανά στοιχειοσύνολα μη συμπεριλαμβανομένου του κενού συνόλου [10]. Αυτό σημαίνει ότι για 6 μόλις στοιχεία, το πλήθος των δυνατών στοιχειοσυνόλων είναι 63. Σε ένα πραγματικό πρόβλημα, το 6 μπορεί να ανέρχεται σε χιλιάδες στοιχεία. Σε τέτοιες περιπτώσεις ο έλεγχος όλων των δυνατών στοιχειοσυνόλων είναι σχεδόν αδύνατος.

Επιπλέον, η εξίσωση που μας δίνει το πλήθος των κανόνων που μπορούν να δημιουργηθούν από g στοιχεία, είναι $R = 3^g - 2^{g+1} + 1$. Η διαφορά ανάμεσα στις δύο τελευταίες εξισώσεις, είναι ότι η πρώτη μας δίνει πιθανά στοιχειοσύνολα (συναλλαγές ή υποσύνολα συναλλαγών), ενώ η δεύτερη μας δίνει το πλήθος των συσχετίσεων που μπορούν να προκύψουν μεταξύ στοιχειοσυνόλων. Η απλούστερη εκδοχή εύρεσης κανόνων είναι να δημιουργηθούν όλοι οι δυνατοί κανόνες, στη συνέχεια να υπολογιστεί η υποστήριξη και η εμπιστοσύνη για κάθε έναν από αυτούς και τέλος να διατηρηθούν μόνο όσοι κανόνες έχουν για αυτά τα δύο μέτρα τιμές μεγαλύτερες ή ίσες από ένα προκαθορισμένο κατώφλι.

Το κατώφλι είναι τα κατώτατα όρια της υποστήριξης και της εμπιστοσύνης που καθορίζονται από τον χρήστη. Ένας τέτοιος τρόπος επίλυσης του προβλήματος είναι πρακτικά αδύνατος. Σύμφωνα με τον Πίνακα 2, ο αριθμός των προϊόντων ισούται με 6, συνεπώς δημιουργούνται 602 υποψήφιοι κανόνες. Πράγματι πρόκειται για μεγάλο αριθμό παρά το γεγονός ότι οι συναλλαγές είναι μόνο πέντε. Ο στόχος ενός αλγόριθμου εξόρυξης κανόνων συσχέτισης είναι να βρει τους ισχυρούς κανόνες σε μία μεγάλη βάση δεδομένων με τον πλέον αποτελεσματικό τρόπο. Ένας από τους τρόπους για να απορρίπτουμε αμέσως υποψήφιους κανόνες, κρατώντας μόνο τους ισχυρούς, είναι να θέσουμε κάποια κατώτατα όρια για την υποστήριξη και την εμπιστοσύνη και μπορεί να αναλυθεί σε δύο φάσεις:

1. Εύρεση των στοιχειοσυνόλων, που έχουν υποστήριξη μεγαλύτερη ή ίση με το κατώτατο όριο υποστήριξης. *Αυτά τα στοιχειοσύνολα καλούνται ως συχνά (ή μεγάλα) στοιχειοσύνολα. Το κατώτατο όριο υποστήριξης συμβολίζεται με **minsup**.* Ο υποψήφιος κανόνας που δεν ικανοποιεί αυτό το όριο, δηλαδή η υποστήριξή του είναι μικρότερη του κατώτατου ορίου της υποστήριξης, απορρίπτεται [11], [40], [41].
2. Εύρεση των κανόνων συσχέτισης, που έχουν εμπιστοσύνη μεγαλύτερη ή ίση με το κατώτατο όριο εμπιστοσύνης, κάνοντας χρήση των μεγάλων στοιχειοσυνόλων, που δημιουργήθηκαν στη προηγούμενη φάση. *Το κατώτατο όριο της εμπιστοσύνης συμβολίζεται με **minconf**.* Και σε αυτή την περίπτωση, ο υποψήφιος κανόνας που δεν ικανοποιεί αυτό το όριο, δηλαδή η εμπιστοσύνη του είναι μικρότερη του κατώτατου ορίου της εμπιστοσύνης, απορρίπτεται γιατί δεν θεωρείται ισχυρός [12].

Οι κανόνες που έχουν $support \geq minsup$ και $confidence \geq minconf$, ονομάζονται ΙΣΧΥΡΟΙ.

Για να κατανοήσουμε καλύτερα τις έννοιες που αναφέρθηκαν παραπάνω, δίνουμε ένα παράδειγμα με βάση τα δεδομένα του Πίνακα 2. Ας υποθέσουμε ότι αναπαριστούμε τις συναλλαγές ενός καταστήματος με δυαδικό τρόπο. Δηλαδή η ύπαρξη ενός προϊόντος σε μια συναλλαγή αναπαρίσταται με τιμή 1 και η απουσία του με τη τιμή 0. Κάθε σειρά στον Πίνακα 2 αναπαριστά μία συναλλαγή και κάθε στήλη απεικονίζει τα διαφορετικά προϊόντα. Αυτή η μορφή ονομάζεται όψη οριζόντιας βάσης δεδομένων [13]. Θεωρώντας τα προϊόντα μεταβλητές, θα λέμε ότι είναι ασύμμετρα δυαδικές μεταβλητές γιατί η παρουσία τους σε μια συναλλαγή είναι «χρησιμότερη» από την απουσία τους. Βεβαίως υπάρχουν και οι συμμετρικά δυαδικές μεταβλητές όπου η παρουσία τους είναι το ίδιο «χρήσιμη» με την απουσία τους, αλλά στην παρούσα εργασία θα ασχοληθούμε μόνο με τις μεταβλητές που παρουσιάζονται σε συναλλαγές.

Αρ.Συν	Γάλα	Καφές	Μακαρόνια	Μπύρες	Πορτοκαλάδα	Ψωμί
1	1	1	0	0	0	0
2	0	1	1	1	0	1
3	1	0	1	0	1	1
4	1	1	1	0	0	1
5	1	1	1	0	1	0

Πίνακας 2 Παράδειγμα Βάσης Δεδομένων ενός Καταστήματος

Για παράδειγμα, έστω η συναλλαγή $\sigma_1 = \{\text{Γάλα}, \text{Καφές}\}$ και το στοιχειοσύνολο $A = \{\text{Καφές}\}$. Το στοιχειοσύνολο A είναι υποσύνολο της συναλλαγής σ_1 , καθώς περιέχει προϊόντα από τη σ_1 . Η τέταρτη συναλλαγή $\sigma_4 = \{\text{Γάλα}, \text{Καφές}, \text{Μακαρόνια}, \text{Ψωμί}\}$ περιέχει το στοιχειοσύνολο $\Gamma = \{\text{Μακαρόνια}, \text{Ψωμί}\}$ αλλά όχι το $B = \{\text{Μπύρες}, \text{Ψωμί}\}$. Επομένως το στοιχειοσύνολο Γ και όχι το B μπορεί να θεωρηθεί υποσύνολο της τέταρτης συναλλαγής. Ας θεωρήσουμε τον κανόνα $\{\text{Γάλα}, \text{Μακαρόνια}\} \rightarrow \{\text{Ψωμί}\}$. Από τη στιγμή που η συχνότητα εμφάνισης του $\{\text{Γάλα}, \text{Μακαρόνια}, \text{Ψωμί}\}$ είναι ίση με δύο και όλες οι συναλλαγές είναι πέντε, η υποστήριξη του κανόνα είναι $2/5 = 0.4$, ενώ η εμπιστοσύνη προκύπτει ως ο λόγος της συχνότητας εμφάνισης του στοιχειοσυνόλου $\{\text{Γάλα}, \text{Μακαρόνια}, \text{Ψωμί}\}$ με την συχνότητα εμφάνισης του στοιχειοσυνόλου $\{\text{Γάλα}, \text{Μακαρόνια}\}$.

Η συχνότητα εμφάνισης του $\{\text{Γάλα}, \text{Μακαρόνια}\}$ ισούται με 3, επομένως η εμπιστοσύνη του κανόνα είναι ίση με $2/3 = 0,67$.

Έστω ότι θέτουμε $\text{minsup} = 0.2$ και $\text{minconf} = 0.5$ για να εξορύξουμε ισχυρούς κανόνες και συχνά στοιχειοσύνολα. Η υποστήριξη του κανόνα $\{\text{Μπύρες}, \text{Ψωμί}\} \rightarrow \{\text{Μακαρόνια}\}$, εξαρτάται μόνο από την υποστήριξη του στοιχειοσυνόλων $\{\text{Μακαρόνια}, \text{Μπύρες}, \text{Ψωμί}\}$. Επομένως, οι παρακάτω κανόνες που προκύπτουν από το στοιχειοσύνολο $\{\text{Μακαρόνια}, \text{Μπύρες}, \text{Ψωμί}\}$, έχουν πανομοιότυπη υποστήριξη, γιατί περιέχουν προϊόντα από το ίδιο στοιχειοσύνολο.

$$\begin{aligned} \{\text{Μπύρες}, \text{Ψωμί}\} &\rightarrow \{\text{Μακαρόνια}\}, & \{\text{Μακαρόνια}, \text{Ψωμί}\} &\rightarrow \{\text{Μπύρες}\}, \\ \{\text{Μακαρόνια}, \text{Μπύρες}\} &\rightarrow \{\text{Ψωμί}\}, & \{\text{Ψωμί}\} &\rightarrow \{\text{Μακαρόνια}, \text{Μπύρες}\}, \\ \{\text{Μπύρες}\} &\rightarrow \{\text{Μακαρόνια}, \text{Ψωμί}\}, & \{\text{Μακαρόνια}\} &\rightarrow \{\text{Μπύρες}, \text{Ψωμί}\} \end{aligned}$$

Η υποστήριξη του στοιχειοσυνόλου $\{\text{Μακαρόνια}, \text{Μπύρες}, \text{Ψωμί}\}$ είναι $\frac{2}{5} = 0.4 > \text{minsup}$, καθώς το στοιχειοσύνολο εμφανίζεται σε δύο από τις πέντε συναλλαγές. Από τη στιγμή που το στοιχειοσύνολο ικανοποιεί το όριο του minsup , όλοι οι υποψήφιοι κανόνες που μπορούν να προέλθουν από τα στοιχεία αυτού του στοιχειοσυνόλου, έχουν την ίδια υποστήριξη. Αντίστοιχα, η υποστήριξη του στοιχειοσυνόλου $\{\text{Γάλα}, \text{Καφές}, \text{Ψωμί}\}$ είναι $\frac{1}{5} = 0.2 < \text{minsup}$, γιατί εμφανίζεται μόνο σε μία συναλλαγή από τις πέντε.

$$\begin{aligned} \{\text{Γάλα}, \text{Καφές}\} &\rightarrow \{\text{Ψωμί}\}, & \{\text{Καφές}\} &\rightarrow \{\text{Γάλα}, \text{Ψωμί}\}, \\ \{\text{Γάλα}\} &\rightarrow \{\text{Καφές}, \text{Ψωμί}\}, & \{\text{Γάλα}, \text{Ψωμί}\} &\rightarrow \{\text{Καφές}\}, \\ \{\text{Ψωμί}\} &\rightarrow \{\text{Γάλα}, \text{Καφές}\}, & \{\text{Καφές}, \text{Ψωμί}\} &\rightarrow \{\text{Γάλα}\} \end{aligned}$$

Επομένως, το στοιχειοσύνολο $\{\text{Γάλα}, \text{Καφές}, \text{Ψωμί}\}$ απορρίπτεται και μαζί του και όλοι οι υποψήφιοι κανόνες που δημιουργήθηκαν αφού έχουν την ίδια υποστήριξη με το στοιχειοσύνολο $\{\text{Γάλα}, \text{Καφές}, \text{Ψωμί}\}$, χωρίς να χρειάζεται να υπολογίσουμε την εμπιστοσύνη τους.

Η συνολική πολυπλοκότητα ενός αλγορίθμου εξόρυξης κανόνων συσχέτισης κυριαρχείται από το πρώτο βήμα, που είναι και το πιο “βαρύ” υπολογιστικά όπως προαναφέραμε. Μετά την εύρεση των μεγάλων στοιχειοσυνόλων, οι αντίστοιχοι κανόνες συσχέτισης μπορούν να βρεθούν πιο εύκολα με τη βοήθεια αλγορίθμων εξόρυξης κανόνων.

Στη συνέχεια παρουσιάζουμε έναν αλγόριθμο, ο οποίος έλυσε το πρόβλημα της αποτελεσματικής μέτρησης των μεγάλων στοιχειοσυνόλων. Ο αλγόριθμος αυτός είναι γνωστός ως Αλγόριθμος Apriori και κάνει χρήση της ακόλουθης ιδιότητας των στοιχειοσυνόλων: *Εάν $A \subseteq B \Rightarrow \text{supp}(A) \geq \text{supp}(B)$, δηλαδή, εάν ένα στοιχειοσύνολο A είναι υποσύνολο ενός στοιχειοσυνόλου B , τότε η υποστήριξη του στοιχειοσυνόλου A είναι μεγαλύτερη ή τουλάχιστον ίση με αυτή του στοιχειοσυνόλου B .*

ΚΕΦΑΛΑΙΟ 2

2.1 Ο Αλγόριθμος Apriori

Η εύρεση συχνών στοιχειοσυνόλων είναι ένα ενδιαφέρον πρόβλημα. Ο πρώτος σχετικός αλγόριθμος που προτάθηκε ονομάζεται Apriori (Agrawal & Srikant, 1994) [14]. Το όνομα του οφείλεται στο γεγονός ότι χρησιμοποιεί προηγούμενη (prior) γνώση σχετικά με τη συχνότητα i -Στοιχειοσυνόλων, για να βρει συχνά $(i + 1)$ -Στοιχειοσύνολα.

Η προηγούμενη γνώση που χρησιμοποιείται αφορά στην αντιμονότονη ιδιότητα της υποστήριξης. Η ιδιότητα αυτή, (γνωστή ως support-based pruning), βασίζεται στην υποστήριξη των στοιχειοσυνόλων, και αναφέρει ότι η υποστήριξη ενός στοιχειοσυνόλου ποτέ δεν ξεπερνά την υποστήριξη των υποσυνόλων του. Αυτό το χαρακτηριστικό είναι ένα **μη μονότονο χαρακτηριστικό** [15]. Η ιδιότητα αυτή ορίζει ότι η υποστήριξη ενός στοιχειοσυνόλου είναι ίση ή μικρότερη από την υποστήριξη κάθε δυνατού υποσυνόλου του. Η αντιμονότονη ιδιότητα της υποστήριξης μαθηματικά ορίζεται με τη Σχέση 2.1

$$\forall A, B : (A \subseteq B) \Rightarrow \text{supp}(A) \geq \text{supp}(B)$$

Σχέση 2.1

Η αντιμονότονη ιδιότητα της υποστήριξης, καλείται επίσης και ιδιότητα συχνών στοιχειοσυνόλων καθώς οποιοδήποτε υποσύνολο ενός συχνού στοιχειοσυνόλου πρέπει να είναι συχνό. Τα συχνά στοιχειοσύνολα, ονομάζονται επίσης και κλειστά προς τα κάτω (closure property) [16], επειδή εάν ένα στοιχειοσύνολο ικανοποιεί τις απαιτήσεις της ελάχιστης υποστήριξης, το ίδιο συμβαίνει και για όλα τα υποσύνολά του. Αντίστροφα, εάν ένα στοιχειοσύνολο δεν ικανοποιεί τα όρια της ελάχιστης υποστήριξης, (minimum support), τότε κάθε μη κενό υπερσύνολο αυτού του στοιχειοσυνόλων δεν ικανοποιεί την ελάχιστη υποστήριξη επίσης. Εάν γνωρίζουμε ότι ένα στοιχειοσύνολο δεν είναι συχνό, δεν χρειάζεται να δημιουργήσουμε κανένα υπερσύνολο του, σαν υποψήφιο επειδή και αυτό αποκλείεται να είναι συχνό.

Χρησιμοποιούμε το δικτυωτό πλέγμα [17] που φαίνεται στο Σχήμα 1 για να παρουσιάσουμε την έννοια αυτής της σημαντικής ιδιότητας. Σε αυτή τη περίπτωση υπάρχουν πέντε προϊόντα «Αυγά, Βούτυρο, Γάλα, Μπύρες, Ψωμί» και χάριν ευκολίας συμβολίζουμε τα Αυγά με «Α», το βούτυρο με «Β», το Γάλα με «Γ», τις Μπύρες με «Μ», και το Ψωμί με «Ψ». Οι γραμμές στο πλέγμα αναπαριστούν τη σχέση των υποσυνόλων, και έτσι η ιδιότητα του συχνού

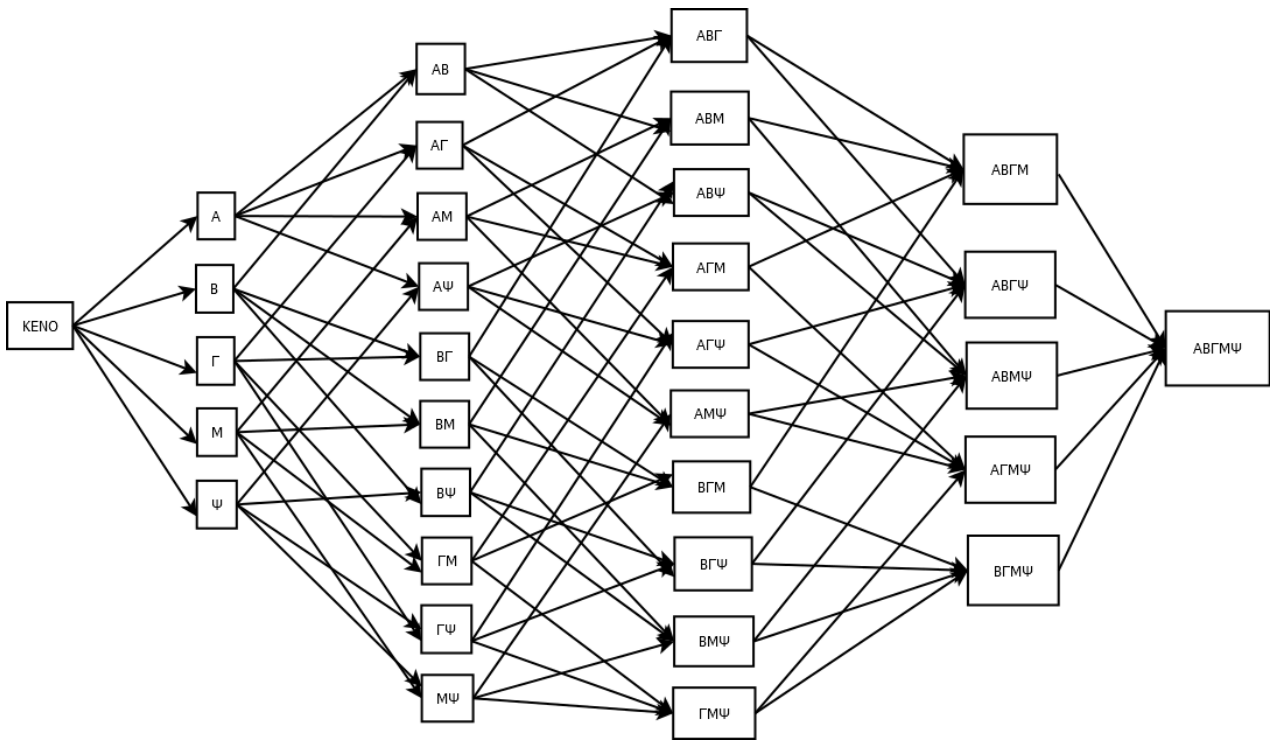
στοιχειοσυνόλου καθορίζει ότι οποιοδήποτε σύνολο στο μονοπάτι επάνω από ένα στοιχειοσύνολο πρέπει να είναι συχνό εάν το αρχικό στοιχειοσύνολο είναι συχνό.

Στο Σχήμα 2, τα μη κενά υποσύνολα του $\{\Gamma, M, \Psi\}$ φαίνεται να είναι τα $\{\Gamma\}$, $\{M\}$, $\{\Psi\}$, $\{\Gamma, M\}$, $\{M, \Psi\}$ και $\{\Gamma, \Psi\}$. Εάν το $\{\Gamma M \Psi\}$ είναι συχνό, το ίδιο ισχύει και για όλα τα υποσύνολά του. Αντίθετα εάν ένα σύνολο δεν είναι συχνό, το ίδιο συμβαίνει και με τα υπερσύνολά του. (Σχήμα 3).

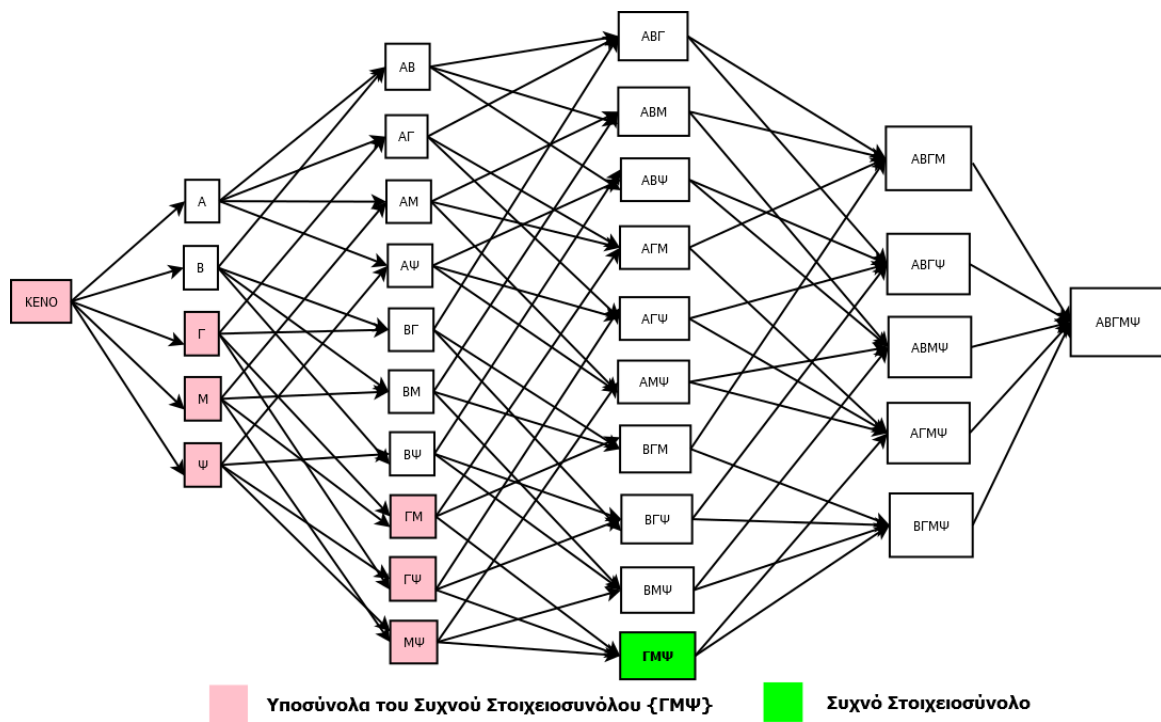
Η βασική ιδέα του αλγόριθμου Apriori, είναι η δημιουργία υποψηφίων στοιχειοσυνόλων ενός συγκεκριμένου μεγέθους και στη συνέχεια η σάρωση της βάσης δεδομένων για να μετρήσει και να δει αν αυτά είναι συχνά. Ο Apriori είναι αναδρομικός αλγόριθμος και δουλεύει σε επίπεδα (levelwise) όπου ο αριθμός του κάθε επιπέδου αντιστοιχεί στο πλήθος των στοιχείων ενός στοιχειοσυνόλου που εξετάζεται. Ξεκινάει από το πρώτο επίπεδο και αναδρομικά προχωράει σε επόμενα επίπεδα μέχρι το επίπεδο εκείνο όπου για πρώτη φορά δεν θα εμφανιστούν μεγάλα στοιχειοσύνολα.

Ο αλγόριθμος Apriori σε κάθε επίπεδο κάνει χρήση δύο στοιχειοσυνόλων, C_i και L_i , για την αποθήκευση των υποψηφίων C_i (Candidates) και των μεγάλων στοιχειοσυνόλων L_i (Large). Τα στοιχειοσύνολα C_i είναι αυτά που βρίσκονται στο επίπεδο i και δεν έχει βρεθεί ακόμα η υποστήριξή τους, ενώ τα στοιχειοσύνολα L_i είναι αυτά για τα οποία έχει βρεθεί η υποστήριξή τους και θα χρησιμοποιηθούν ως στοιχειοσύνολα C_{i+1} στο επίπεδο $i + 1$. Κατά τη διάρκεια του i - οστού περάσματος, καταμετρούνται τα υποψήφια στοιχειοσύνολα C_i μεγέθους i και δημιουργείται ένα σύνολο με εκείνα τα στοιχεία όπου η συχνότητα εμφάνισης τους ισούται ή υπερβαίνει το καθορισμένο κατώφλι (minsup). Ένα στοιχειοσύνολο θεωρείται ως υποψήφιο μόνο όταν όλα τα υποσύνολα του είναι επίσης συχνά.

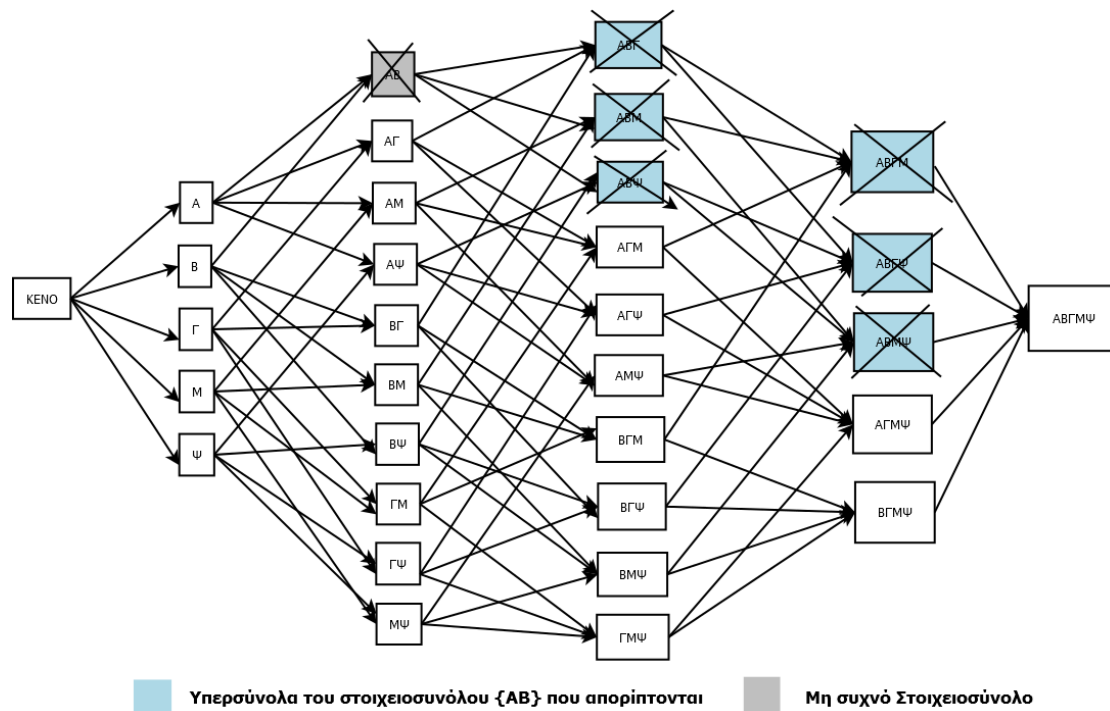
Μία σημαντική σημείωση είναι ότι πρέπει σε κάθε επίπεδο τα στοιχεία των συχνών στοιχειοσυνόλων να ταξινομούνται με αλφαβητική σειρά. Αυτό συμβαίνει γιατί όταν θέλουμε να συνενώσουμε τα συχνά στοιχειοσύνολα από το επίπεδο i για το επίπεδο $i + 1$ πρέπει να υπάρχουν $i - 1$ κοινά στοιχεία στις πρώτες θέσεις των στοιχειοσυνόλων. Για τη δημιουργία υποψηφίων στοιχειοσυνόλων μεγέθους $i + 1$, γίνονται συνενώσεις συχνών στοιχειοσυνόλων ($L_i * L_i$) που βρίσκονται στο προηγούμενο επίπεδο. Το C_{i+1} ελέγχεται για ύπαρξη μη συχνών στοιχειοσυνόλων σύμφωνα με την αντιμονότονη ιδιότητα της υποστήριξης. Όσα $(i + 1)$ - στοιχειοσύνολα βρεθούν να περιέχουν ένα τουλάχιστον μη συχνό i - στοιχειοσύνολο διαγράφονται από το C_{i+1} . Με την ολοκλήρωση του τελευταίου βήματος όπου κρατάμε τα στοιχειοσύνολα που καλύπτουν το κριτήριό μας, έχει βρεθεί το σύνολο των συχνών $(i+1)$ - στοιχειοσυνόλων $L_{(i+1)}$ [18].



Σχήμα 1 Δικτυωτό Πλέγμα



Σχήμα 2 Υποσύνολα του Στοιχειοσυνόλου {ΓΜΨ}



Σχήμα 3 Υπερσύνολα του μη Συχνού Στοιχειοσυνόλου {ΑΒ}

Ο αλγόριθμος Arriogi, περιλαμβάνει μια επαναλαμβανόμενη διαδικασία και κατά τη διάρκεια των επαναλήψεων πραγματοποιούνται δύο βήματα:

- Δημιουργούνται i - στοιχειοσύνολα από τα συχνά $(i - 1)$ -στοιχειοσύνολα. Αν L_{i-1} είναι το σύνολο των $(i - 1)$ στοιχειοσυνόλων, τότε δημιουργείται ένα σύνολο C_i υποψηφίων i -στοιχειοσυνόλων. Αυτή η διαδικασία επαναλαμβάνεται έως ότου να μην μπορούν να δημιουργηθούν άλλα συχνά στοιχειοσύνολα. Θεωρούμε ότι τα στοιχεία των στοιχειοσυνόλων είναι διατεταγμένα με αλφαβητική σειρά. Δηλαδή ακολουθούν την αρχή της λεξικογραφικής σειράς [19]. Παραδείγματος χάριν, {Γάλα, Καφές, Ζάχαρη}. Το στοιχειοσύνολο {Γάλα, Καφές, Ζάχαρη} μπορεί να συνδεθεί με το στοιχειοσύνολο {Γάλα, Καφές, Ψωμί} γιατί τα δύο πρώτα στοιχεία (Γάλα και Καφές) είναι κοινά. Το αποτέλεσμα είναι το σύνολο {Γάλα, Καφές, Ζάχαρη, Ψωμί}. Αντιθέτως τα σύνολα {Γάλα, Καφές, Ζάχαρη} και {Γάλα, Ζάχαρη, Καρότα} δεν μπορούν να συνδεθούν γιατί τα δύο πρώτα τους στοιχεία δεν είναι κοινά Γάλα, Καφές \neq Γάλα, Ζάχαρη.
- Αφού εκτελεστεί το πρώτο στάδιο και βρεθούν όλα τα συχνά στοιχειοσύνολα, ο αλγόριθμος προχωρά στην εξόρυξη κανόνων συσχέτισης χρησιμοποιώντας τα συχνά στοιχειοσύνολα που βρήκε στο προηγούμενο βήμα. Οι κανόνες που δημιουργούνται είναι εξασφαλισμένο ότι ικανοποιούν την κατώτατη τιμή της υποστήριξης αφού έχουν δημιουργηθεί από συχνά στοιχειοσύνολα και πρέπει να ικανοποιούν και το κατώτατο όριο της εμπιστοσύνης. Μόνο σε αυτή τη περίπτωση, οι κανόνες που δημιουργούνται

θεωρούνται ισχυροί. Σε περίπτωση που οι υποψήφιοι κανόνες που δημιουργούνται δεν ικανοποιούν το κατώτατο όριο της εμπιστοσύνης, απορρίπτονται [20].

Για την καλύτερη κατανόηση του αλγορίθμου Αργιορί παραθέτουμε το παράδειγμα του Πίνακα 3.

Η βάση δεδομένων μας περιέχει συνολικά εννέα συναλλαγές (Τμήμα Α). Για κάθε συναλλαγή παρατίθενται τα προϊόντα που περιέχει. Ορίζουμε ένα κατώφλι υποστήριξης ίσο με 0,4. Η βάση δεδομένων σαρώνεται και υπολογίζεται η συχνότητα εμφάνισης κάθε προϊόντος (Τμήμα Β). Θεωρούμε το σύνολο C_1 των προϊόντων ως 1-itemset στο επίπεδο 1. Παρατηρούμε ότι η υποστήριξη κάθε προϊόντος, είναι ίση ή μεγαλύτερη από 0,4. Αυτό σημαίνει ότι όλα τα 1-itemset είναι συχνά και κανένα δεν πρέπει να διαγραφεί. Σε αυτή τη περίπτωση $C_1 = L_1$.

Στη συνέχεια, το L_1 συνενώνεται με τον εαυτό του και προκύπτει το 2-itemset C_2 (Τμήμα Γ). Πραγματοποιείται καταμέτρηση των εμφανίσεων των προϊόντων του C_2 , στη βάση δεδομένων, και παρατηρούμε ότι τρία στοιχειοσύνολα, το {Γάλα, Καρότα}, το {Ζάχαρη, Καρότα} και το {Καρότα, Ψωμί}, έχουν υποστήριξη μικρότερη από 0,4 (Τμήμα Γ).

Τα τρία αυτά στοιχειοσύνολα απομακρύνονται από το C_2 και προκύπτει το σύνολο των συχνών 2-itemset L_2 (Τμήμα Δ). Στη συνέχεια, το L_2 συνενώνεται με τον εαυτό του. Το {Γάλα, Καφές} συνδυάζεται με το {Γάλα, Ζάχαρη} και το {Γάλα, Ψωμί}. Επίσης, το {Γάλα, Ζάχαρη} συνδυάζεται με το {Γάλα, Ψωμί}. Με τον ίδιο τρόπο συνδυάζονται και τα μέλη του L_2 των οποίων το πρώτο στοιχείο είναι το {Καφές}. Το {Ζάχαρη, Ψωμί} δεν μπορεί να συνδυαστεί με κανένα μέλος του L_2 . Το αποτέλεσμα της συνένωσης είναι το σύνολο C_3 και απεικονίζεται στο τμήμα Ε.

Πραγματοποιείται έλεγχος στο C_3 για να βρεθούν μέλη που έχουν μη συχνά υποσύνολα. Δύο μέλη του C_3 έχουν μη συχνά υποσύνολα. Ειδικότερα, το {Καφές, Ζάχαρη, Καρότα} έχει υποσύνολο το {Ζάχαρη, Καρότα} το οποίο είναι μη συχνό καθώς δεν ανήκει στο L_2 . Επίσης, το {Καφές, Καρότα, Ψωμί} έχει υποσύνολο το {Καρότα, Ψωμί} το οποίο είναι και αυτό μη συχνό. Από την αντιμονότονη ιδιότητα της υποστήριξης συμπεραίνουμε ότι και τα {Καφές, Ζάχαρη, Καρότα}, {Καφές, Καρότα, Ψωμί} είναι μη συχνά. Για τον λόγο αυτό, τα δύο στοιχειοσύνολα απομακρύνονται από το C_3 , και για τα υπόλοιπα στοιχειοσύνολα του C_3 , πραγματοποιείται καταμέτρηση στη βάση δεδομένων (Τμήμα Ε). Τα αποτελέσματα παρουσιάζονται στο τμήμα ΣΤ.

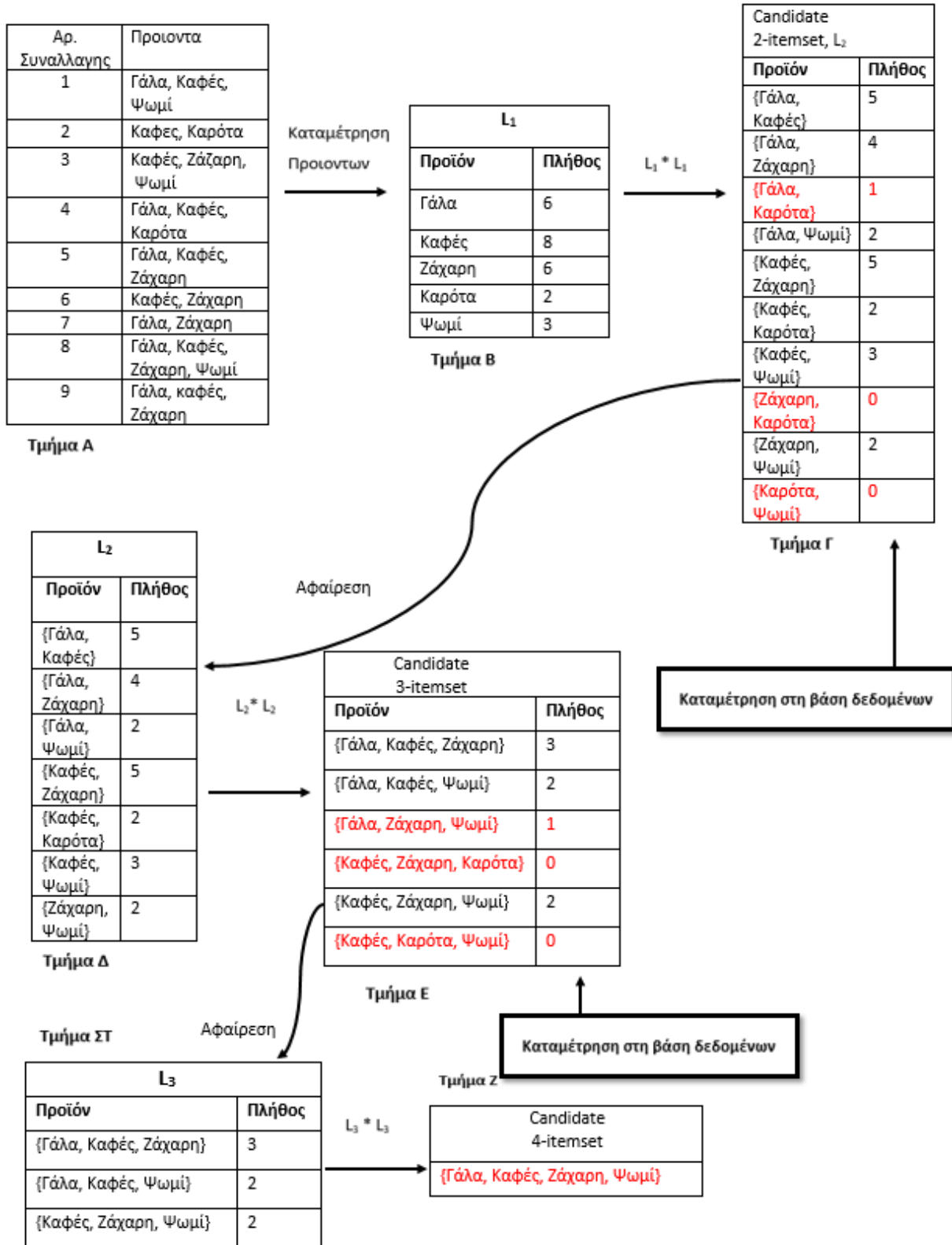
Παρατηρούμε ότι ένα μέλος του C_3 , το {Γάλα, Ζάχαρη, Ψωμί}, έχει υποστήριξη ίση με 0,2, δηλαδή εμφανίζεται σε 1 συναλλαγή. Το στοιχειοσύνολο αυτό είναι μη συχνό και απομακρύνεται από το C_3 . Το αποτέλεσμα είναι το σύνολο συχνών 3-itemset L_3 , το οποίο παρουσιάζεται στο τμήμα ΣΤ.

Στη συνέχεια, το L_3 συνενώνεται με τον εαυτό του. Τα μέλη του που μπορούν να συνδυαστούν είναι το $\{\mathbf{Γάλα, Καφές, Ζάχαρη}\}$ και το $\{\mathbf{Γάλα, Καφές, Ψωμί}\}$, γιατί αυτά τα δύο στοιχειοσύνολα έχουν τα δυο πρώτα μέλη τους κοινά.

Το αποτέλεσμα της συνένωσης είναι το C_4 , το οποίο απεικονίζεται στο τμήμα Z. Το C_4 έχει ένα μόνο μέλος, το $\{\mathbf{Γάλα, Καφές, Ζάχαρη, Ψωμί}\}$. Παρατηρούμε ότι το $\{\mathbf{Γάλα, Καφές, Ζάχαρη, Ψωμί}\}$ έχει υποσύνολο το $\{\mathbf{Γάλα, Ζάχαρη, Ψωμί}\}$ το οποίο είναι μη συχνό, οπότε συμπεραίνουμε ότι και το $\{\mathbf{Γάλα, Καφές, Ζάχαρη, Ψωμί}\}$ είναι μη συχνό.

Με την ολοκλήρωση της εκτέλεσης του αλγορίθμου έχουμε εντοπίσει τα συχνά στοιχειοσύνολα, τα οποία είναι τα μέλη του L_2 και του L_3 (Τμήμα Δ και Τμήμα ΣΤ).

Όπως ήδη ορίσαμε, συχνά καλούνται όλα τα στοιχειοσύνολα που προκύπτουν από μία βάση δεδομένων, όταν η υποστήριξή τους είναι μεγαλύτερη ή ίση από το προκαθορισμένο κατώφλι. Ένα υποσύνολο των συχνών στοιχειοσυνόλων είναι τα κλειστά στοιχειοσύνολα των οποίων ο ορισμός δίδεται παρακάτω.



Πίνακας 3 Παράδειγμα Apriori

Ορισμός: Κλειστά λέμε τα στοιχειοσύνολα, τα οποία είναι συχνά και δεν υπάρχει κανένα υπερσύνολο με την ίδια υποστήριξη [21], [22].

Όπως βλέπουμε, το Σχήμα 4 δείχνει 5 συναλλαγές. Τα διαφορετικά προϊόντα που βρέθηκαν στα καλάθια είναι τα Αυγά, το Βούτυρο, το Γάλα, οι Μπύρες και το Ψωμί. Χάριν ευκολίας, αναπαριστούμε τα Αυγά με «Α», το Βούτυρο με «Β», το Γάλα με «Γ», τις Μπύρες με «Μ» και το Ψωμί με «Ψ». Θέτουμε κατώφλι υποστήριξης ίσο με 0.4. Σε κάθε επίπεδο l του δέντρου, δημιουργούνται στοιχειοσύνολα μεγέθους l και οι αριθμοί που αναγράφονται πάνω από κάθε στοιχειοσύνολο δηλώνουν τον αριθμό της συναλλαγής που υπάρχει το κάθε στοιχειοσύνολο.

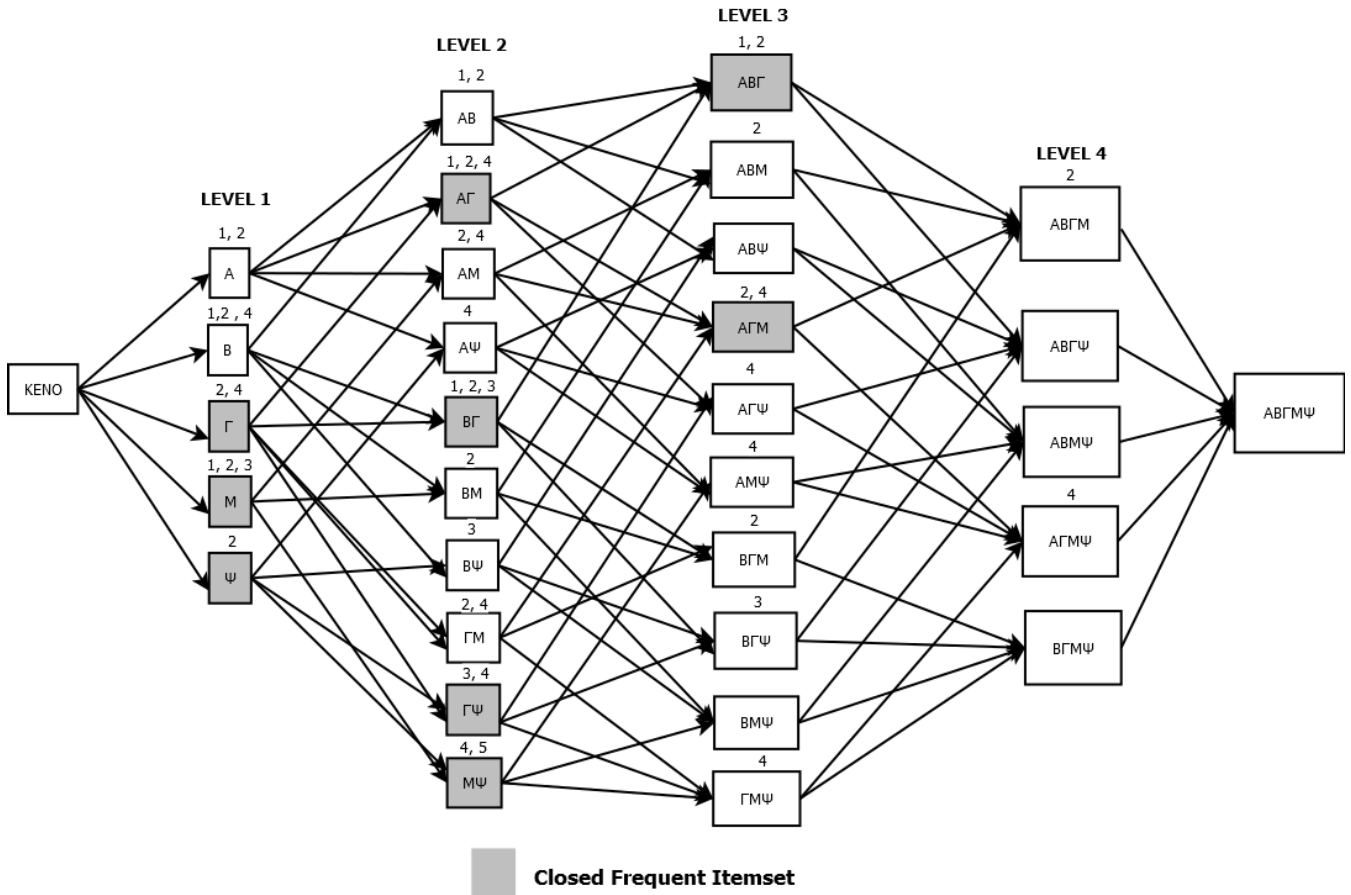
Για κάθε επίπεδο στο δέντρο εκτελούμε την ίδια διαδικασία όπως παραπάνω, καταγράφοντας τον αριθμό της συναλλαγής του κάθε υπερσυνόλου που δημιουργήθηκε από το προηγούμενο επίπεδο. Όπως βλέπουμε, το προϊόν «Γάλα» είναι συχνό γιατί ικανοποιεί το κατώτατο όριο υποστήριξης που έχουμε θέσει και εμφανίζεται στις συναλλαγές 1, 2, 3 και 4. Το στοιχειοσύνολο {Βούτυρο, Γάλα}, παρόλο που ικανοποιεί το προκαθορισμένο κατώφλι υποστήριξης, δεν έχει την ίδια υποστήριξη με το προϊόν «Γάλα» καθώς εμφανίζεται στις συναλλαγές 1, 2 και 3. Επομένως το 1-στοιχειοσύνολο {Γάλα}, μπορεί να θεωρηθεί κλειστό συχνό στοιχειοσύνολο. Ομοίως κανένα υπερσύνολο του στοιχειοσύνολο {Βούτυρο, Γάλα} δεν έχει την ίδια υποστήριξη με αυτό, επομένως μπορεί να θεωρηθεί κλειστό συχνό στοιχειοσύνολο.

Ένα υποσύνολο των συχνών κλειστών στοιχειοσυνόλων είναι τα μέγιστα στοιχειοσύνολα.

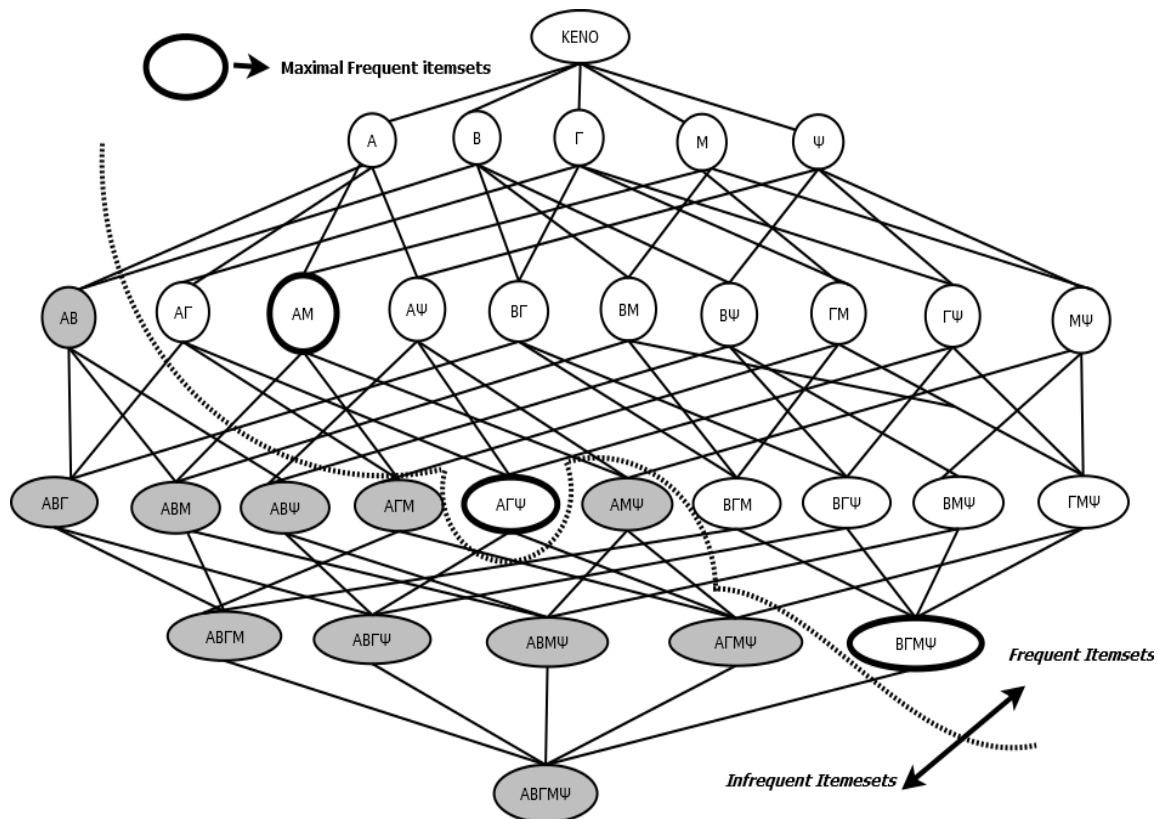
Ορισμός: Μέγιστο στοιχειοσύνολο ορίζεται αυτό που είναι συχνό, και κανένα από τα αμέσως επόμενα υπερσύνολα του δεν είναι συχνά[23].

Αντίθετα, όταν ένα στοιχειοσύνολο δεν είναι μέγιστο, τότε όλα τα αμέσως επόμενα υπερσύνολά του είναι συχνά. Κάτι τέτοιο μπορεί να αναπαρασταθεί σε ένα δικτυωτό πλέγμα (Σχήμα 5), όπου διαχωρίζουμε τα στοιχειοσύνολα που είναι συχνά (frequent itemset), δεξιά της διακεκομμένης γραμμής και τα στοιχειοσύνολα που δεν είναι συχνά (infrequent itemset), αριστερά της διακεκομμένης. Όπως βλέπουμε, τα στοιχειοσύνολα {ΑΜ}, {ΑΓΨ} και {ΒΓΜΨ} είναι μέγιστα συχνά στοιχειοσύνολα γιατί τα υπερσύνολα αυτών των συνόλων δεν είναι συχνά (infrequent itemset) δηλαδή δεν ικανοποιούν το κατώτατο όριο της υποστήριξης που έχουμε θέσει.

Αρ. Συν	Προϊόντα
1	Αυγά, Βούτυρο, Γάλα
2	Αυγά, Βούτυρο, Γάλα, Μπύρες
3	Βούτυρο, Γάλα, Ψωμί
4	Αυγά, Γάλα, Μπύρες, Ψωμί
5	Μπύρες, Ψωμί



Σχήμα 4 Κλειστά Συχνά Στοιχειοσύνολα



Σχήμα 5 Μέγιστα Συχνά Στοιχειοσύνολα

Με την ολοκλήρωση του πρώτου σταδίου έχουν βρεθεί τα συχνά στοιχειοσύνολα της βάσης δεδομένων. Η υποστήριξη αυτών των στοιχειοσυνόλων είναι μεγαλύτερη από την τιμή κατωφλιού υποστήριξης. Τα συχνά στοιχειοσύνολα χρησιμοποιούνται για την εύρεση των κανόνων συσχέτισης. Ο αλγόριθμος δημιουργίας κανόνων είναι ο ακόλουθος:

Κάθε εμφανιζόμενο i -στοιχειοσύνολο, μπορεί να παραγάγει $2^i - 2$ κανόνες συσχέτισης.

Για κάθε συχνό στοιχειοσύνολο X, Y , δημιουργείται ο κανόνας της μορφής $X \rightarrow Y$. Κάθε κανόνας που δημιουργήθηκε στο προηγούμενο βήμα ελέγχεται ως προς την τιμή της εμπιστοσύνης. Αν η εμπιστοσύνη είναι μεγαλύτερη από την τιμή κατωφλιού εμπιστοσύνης, τότε ο κανόνας θεωρείται ισχυρός. Μετά την εκτέλεση του αλγορίθμου έχουν εντοπιστεί όλοι οι ισχυροί κανόνες, οι οποίοι έχουν υποστήριξη και εμπιστοσύνη μεγαλύτερη από τις ελάχιστες καθορισμένες τιμές. Αξίζει να σημειώσουμε ότι οι κανόνες δημιουργήθηκαν από συχνά στοιχειοσύνολα, οπότε είναι εξασφαλισμένο ότι η υποστήριξη τους ισούται ή υπερβαίνει την τιμή κατωφλιού υποστήριξης.

Η εμπιστοσύνη ενός κανόνα της μορφής $X \rightarrow Y$ υπολογίζεται σύμφωνα με την Εξίσωση 2.4.

$$\text{conf}(X \rightarrow Y) = \frac{\text{supp_count}(XY)}{\text{supp_count}(X)}$$

Εξίσωση 2.4

Υπενθυμίζουμε ότι το $\{X\}$ και το $\{Y\}$ είναι συχνά στοιχειοσύνολα, οπότε η συχνότητα εμφάνισης τους είναι ήδη γνωστή από το πρώτο στάδιο, και η υποστήριξη του κανόνα

$X \rightarrow Y$ μπορεί εύκολα να υπολογιστεί.

Στην Εικόνα 1, παρουσιάζεται η δημιουργία κανόνων για το συχνό στοιχειοσύνολο {Γάλα, Καφές, Ζάχαρη} του προηγούμενου παραδείγματος. Συμβολίζουμε για ευκολία αναπαράστασης, το Γάλα με (Γ), τον Καφέ με (Κ) και τη Ζάχαρη με (Ζ). Ο Πίνακας 4, δείχνει την συχνότητα εμφάνισης κάθε υποσυνόλου του συχνού στοιχειοσυνόλου {Γάλα, Καφές, Ζάχαρη}.

Στοιχειοσύνολα	Sup_count
Γ	6
Κ	8
Ζ	6
Γ,Κ	5
Γ,Ζ	4
Κ,Ζ	5
Γ, Κ, Ζ	3

Πίνακας 4 Υποστήριξη Στοιχειοσυνόλων

Κανόνας	Εμπιστοσύνη	
$\Gamma \rightarrow \text{Κ, Ζ}$	$\text{sup_count}(\Gamma, \text{Κ, Ζ})/\text{sup_count}(\Gamma)$	0,5
$\text{Κ} \rightarrow \Gamma, \text{Ζ}$	$\text{sup_count}(\Gamma, \text{Κ, Ζ})/\text{sup_count}(\text{Κ})$	0,38
$\text{Ζ} \rightarrow \text{Κ, Γ}$	$\text{sup_count}(\Gamma, \text{Κ, Ζ})/\text{sup_count}(\text{Ζ})$	0,5
$\Gamma, \text{Κ} \rightarrow \text{Ζ}$	$\text{sup_count}(\Gamma, \text{Κ, Ζ})/\text{sup_count}(\Gamma, \text{Κ})$	0,6
$\Gamma, \text{Ζ} \rightarrow \text{Κ}$	$\text{sup_count}(\Gamma, \text{Κ, Ζ})/\text{sup_count}(\Gamma, \text{Ζ})$	0,75
$\text{Κ, Ζ} \rightarrow \Gamma$	$\text{sup_count}(\Gamma, \text{Κ, Ζ})/\text{sup_count}(\text{Κ, Ζ})$	0,6

Εικόνα 1 Εμπιστοσύνη των κανόνων που προκύπτουν από το {ΓΚΖ}

Από το συχνό στοιχειοσύνολο $\{\Gamma, \text{Κ, Ζ}\}$ δημιουργούνται οι κανόνες, $(\Gamma \rightarrow \text{Κ, Ζ})$, $(\text{Κ} \rightarrow \Gamma, \text{Ζ})$, $(\text{Ζ} \rightarrow \text{Κ, Γ})$, $(\Gamma, \text{Κ} \rightarrow \text{Ζ})$, $(\Gamma, \text{Ζ} \rightarrow \text{Κ})$ και $(\text{Κ, Ζ} \rightarrow \Gamma)$. Οι συχνότητες εμφάνισης των στοιχειοσυνόλων που συμμετέχουν στους κανόνες είναι γνωστές, οπότε υπολογίζονται εύκολα οι τιμές εμπιστοσύνης των κανόνων. Αν έχουμε ορίσει ως κατώφλι εμπιστοσύνης την τιμή 0,6% (minconf), τότε μόνο οι κανόνες $(\Gamma, \text{Κ} \rightarrow \text{Ζ})$, $(\Gamma, \text{Ζ} \rightarrow \text{Κ})$ και $(\text{Κ, Ζ} \rightarrow \Gamma)$ είναι ισχυροί.

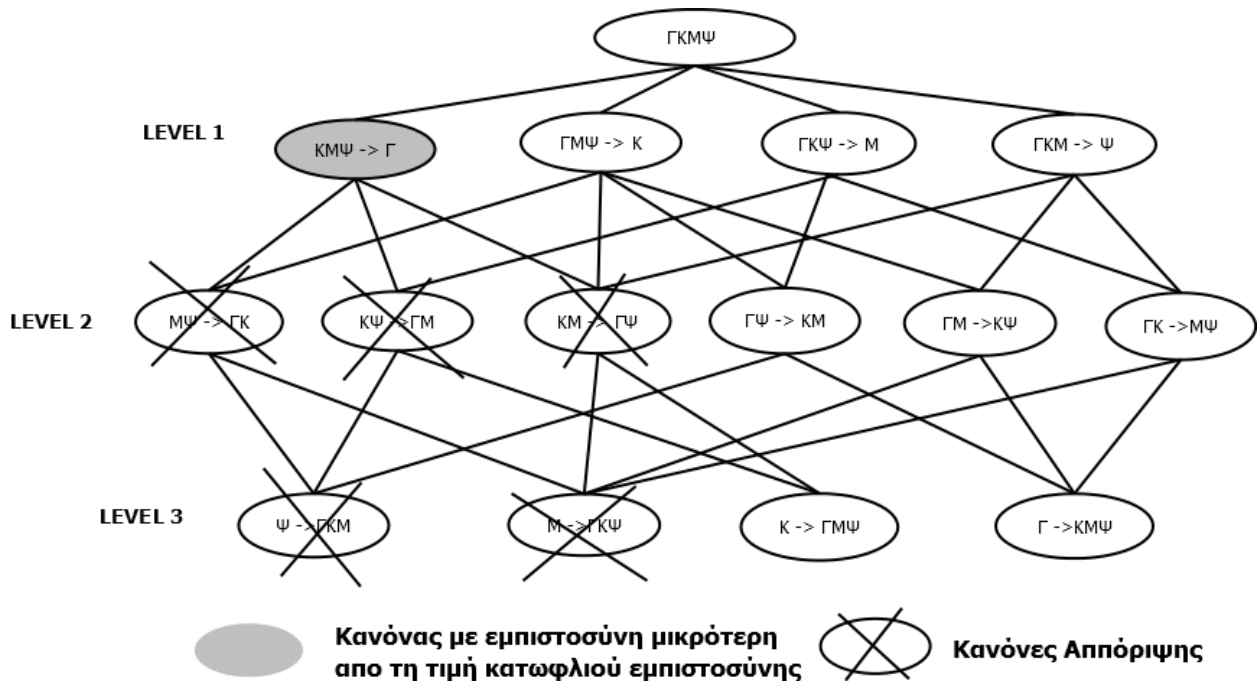
Δεδομένου ότι από ένα στοιχειοσύνολο μεγέθους i δημιουργούνται συνολικά $2^i - 2$ κανόνες, σημαίνει ότι για ένα στοιχειοσύνολο με 100 μέλη δημιουργούνται περίπου 10^{30} υποψήφιοι κανόνες. Το πλήθος των υποψήφιων κανόνων μπορεί να είναι πολύ μεγάλο. Για τον λόγο αυτό είναι αναγκαίος ο περιορισμός του.

Όπως φαίνεται από την Εξίσωση 2.4, για κανόνες της μορφής $X \rightarrow Y$ που δημιουργούνται από ένα στοιχειοσύνολο X, Y , η εμπιστοσύνη επηρεάζεται από τη συχνότητα εμφάνισης του X , καθώς η συχνότητα εμφάνισης του X, Y είναι η ίδια για όλους τους κανόνες. Η συχνότητα εμφάνισης του X βρίσκεται στον παρονομαστή της Εξίσωσης 2.4 οπότε και είναι αντιστρόφως ανάλογη με την εμπιστοσύνη.

Από την αντιμονότονη ιδιότητα της υποστήριξης γνωρίζουμε ότι εάν $X_1 \subseteq X_2$ τότε ισχύει ότι $\text{sup_count}(X_1) \geq \text{sup_count}(X_2)$ και κατά συνέπεια θα ισχύει ότι $\text{conf}(X_2 \rightarrow X_1) \leq \text{conf}(X_1 \rightarrow X_2)$. Σε ελεύθερη διατύπωση αυτό σημαίνει ότι ένας κανόνας της μορφής $X \rightarrow Y$, έχει εμπιστοσύνη ίση ή μεγαλύτερη από κάθε κανόνα που στο αριστερό του μέλος έχει ένα

υποσύνολο του X . Συνεπώς το πρακτικό συμπέρασμα που προκύπτει είναι ότι, εάν η εμπιστοσύνη ενός κανόνα $X \rightarrow Y$ είναι μικρότερη από την τιμή κατωφλιού εμπιστοσύνης, τότε και όλοι οι κανόνες που στο αριστερό τους μέλος έχουν ένα υποσύνολο του X θα είναι και αυτοί μη ισχυροί, συνεπώς απορρίπτονται. Αυτό γίνεται κατανοητό μέσω του Σχήματος 6.

Έστω ότι η εμπιστοσύνη του κανόνα $\{KM\psi\} \rightarrow \{\Gamma\}$, είναι μικρότερη από τη τιμή κατωφλιού εμπιστοσύνης. Οι κανόνες που έχουν διαγραφεί, έχουν στο αριστερό τους μέλος υποσύνολο του X . Επομένως είναι και αυτοί μη ισχυροί και απορρίπτονται.



Σχήμα 6 Απόρριψη Κανόνων Συσχέτισης

Παρόλο που ο Αλγόριθμος Apriori, είναι ένας από τους πιο γνωστούς και δημοφιλείς αλγορίθμους στην Εξόρυξη Δεδομένων, η απόδοση του ίσως υστερεί σημαντικά σε πυκνά δεδομένα καθώς το μέγεθος των συναλλαγών αυξάνει. Όσο περισσότερες συναλλαγές υπάρχουν στη βάση δεδομένων, τόσο περισσότερο χρόνο χρειάζεται ο αλγόριθμος για να αναπαράγει υποψήφια στοιχειοσύνολα. Ο μέσος όρος μεγέθους των συναλλαγών σε πυκνά δεδομένα είναι ανάλογος με το μέγιστο μέγεθος των συχνών στοιχειοσυνόλων. Κάτι τέτοιο επηρεάζει τη πολυπλοκότητα του αλγορίθμου αφού πραγματοποιεί περισσότερες επαναλήψεις.

Επιπλέον, όσο περισσότερα είναι τα συχνά στοιχειοσύνολα που αναπαράγονται, τόσο μεγαλύτερο είναι το κόστος υπολογισμού των κανόνων συσχέτισης που δημιουργούνται. Όσο ο αριθμός των στοιχείων που αποθηκεύονται στην βάση δεδομένων αυξάνεται, ο αλγόριθμος χρειάζεται περισσότερο χώρο για να αποθηκεύσει την υποστήριξη τους. Υπάρχουν αρκετές

εναλλακτικές μέθοδοι που έχουν αναπτυχθεί για να ξεπεράσουν αυτούς τους περιορισμούς και μπορούν να βελτιώσουν την αποτελεσματικότητα του Αλγορίθμου Apriori.

Στην επόμενη ενότητα θα παρουσιάσουμε τον Αλγόριθμο διαμέρισης, τον Αλγόριθμο FP-Growth και τον Αλγόριθμο Eclat.

2.2 Εναλλακτικές μέθοδοι για τη δημιουργία συχνών στοιχειοσυνόλων.

2.2.1 Ο Αλγόριθμος Διαμέρισης

Ο αλγόριθμος διαμέρισης ξεκινάει την διαδικασία αναπαραγωγής υποψηφίων στοιχειοσυνόλων με τον διαχωρισμό της βάσης δεδομένων Σ σε διαμερίσεις, $(\Sigma_1, \Sigma_2, \dots, \Sigma_n)$.

Η βασική ιδέα είναι ότι κάθε συχνό στοιχειοσύνολο πρέπει να είναι συχνό σε μία τουλάχιστον διαμέριση. Στη συνέχεια εφαρμόζεται ο Αλγόριθμος Apriori σε κάθε διαμέριση Σ_i και προκύπτουν τα αντίστοιχα συχνά στοιχειοσύνολα (L_1, L_2, \dots, L_i) . Τα υποψήφια συχνά στοιχειοσύνολα είναι $C = L_1 \cup L_2 \cup \dots \cup L_i$.

Τέλος, γίνεται μια τελική καταμέτρηση της υποστήριξης των υποψηφίων συχνών στοιχειοσυνόλων C στη βάση Σ , για να βρεθούν τα τελικά συχνά στοιχειοσύνολα L . Τα βασικά πλεονεκτήματα αυτού του αλγορίθμου, είναι ότι προσαρμόζεται εύκολα στη διαθέσιμη κύρια μνήμη, μπορεί εύκολα να υπολογίζει ταυτόχρονα τα υποψήφια συχνά στοιχειοσύνολα και τέλος ο μέγιστος αριθμός σαρωμάτων είναι ίσος με δύο. Το μοναδικό του μειονέκτημα, είναι ότι στο δεύτερο σάρωμα πιθανά παράγονται πολλά υποψήφια στοιχειοσύνολα [24].

2.2.2 Ο Αλγόριθμος FP-Growth

Ένας εναλλακτικός αλγόριθμος είναι ο FP-Growth ο οποίος ανακαλύπτει στοιχειοσύνολα με διαφορετική προσέγγιση. Δεν αναπαράγει κατ' επανάληψη στοιχειοσύνολα όπως ο Apriori. Η αναπαραγωγή συνόλων αντικειμένων, γίνεται μέσω της αποκρυπτογράφησης των στοιχειοσυνόλων που περιέχονται σε κάθε συναλλαγή, χρησιμοποιώντας μία συμπαγή δομή απεικόνισης, μέσω μιας ειδικής δενδρικής δομής, γνωστής ως (frequent-pattern) FP-tree και το εξερευνά με bottom up αναζήτηση [25]. Με αυτόν τον τρόπο εξάγονται συχνά στοιχειοσύνολα, κατευθείαν από το δέντρο.

Με απλά λόγια, ο αλγόριθμος αρχικά σαρώνει τη βάση δεδομένων για πρώτη φορά, βρίσκει τα συχνά στοιχειοσύνολα μήκους 1 (1-itemset) και δημιουργεί μια λίστα, ταξινομημένη κατά φθίνουσα σειρά υποστήριξης. Στη συνέχεια, με βάση αυτή τη λίστα ταξινομεί τα στοιχεία κάθε συναλλαγής, ενώ τα μη συχνά στοιχεία τα διαγράφει.

Τέλος κάνει ένα δεύτερο σάρωμα και δημιουργεί μια δενδρική δομή FP-tree, τοποθετώντας κάθε ταξινομημένη (κατά συχνότητα στοιχείων) συναλλαγή στη δομή αυτή. Αρχικά, η δενδρική δομή έχει έναν κενό κόμβο, ο οποίος είναι η ρίζα του δένδρου και συμβολίζεται ως «{}». Έπειτα, σαρώνονται με τη σειρά οι ταξινομημένες συναλλαγές και προστίθενται νέοι κόμβοι όποτε αυτό είναι απαραίτητο. Οι κόμβοι περιέχουν τα στοιχεία των συναλλαγών και τον τρέχοντα αριθμό εμφάνισής τους στις συναλλαγές, οι οποίες έχουν ήδη προσπελαστεί. Αν υπάρχει κοινό πρόθεμα μεταξύ δύο ταξινομημένων συναλλαγών, π.χ. **FCAHL**, **FCABH**, τότε δεν εισάγονται νέοι κόμβοι για το κοινό πρόθεμα, αλλά αυξάνεται ο μετρητής κάθε κόμβου, που συμμετέχει στο κοινό πρόθεμα κατά ένα.

Στην περίπτωση που μια δοσοληψία είναι υποσύνολο κάποιας προηγούμενης δοσοληψίας, τότε δεν προστίθεται κανένας νέος κόμβος. Αφού δημιουργηθεί η δενδρική δομή (FP-tree) η εύρεση των συχνών στοιχειοσυνόλων γίνεται με προσπέλαση του δένδρου από κάτω προς τα πάνω, δηλαδή από τα φύλλα προς τη ρίζα.

Για την εύρεση των συχνών στοιχειοσυνόλων εντοπίζονται τα προθεματικά υποδέντρα που καταλήγουν σε κάθε απλό στοιχειοσύνολο (στοιχείο), χρησιμοποιώντας τους δείκτες μεταξύ των ίδιων στοιχείων που υπάρχουν στη δενδρική δομή. Κάθε υποδέντρο προσπελαύνεται αναδρομικά για την εξαγωγή όλων των συχνών στοιχειοσυνόλων. Για παράδειγμα, έστω ότι έχουμε τα στοιχεία ABCF και το προθεματικό υποδέντρο για το στοιχείο B. Η προσπέλαση θα γίνει αναδρομικά, αναζητώντας στοιχειοσύνολα που τελειώνουν σε B, έπειτα σε FB, CB, AB, έπειτα σε FCB, FAB, CAB κοκ. Παράλληλα, γίνεται η καταμέτρηση υποστήριξης για κάθε στοιχειοσύνολο.

Στο τέλος της εκτέλεσης του αλγορίθμου, γίνεται συνένωση των συνόλων για κάθε προθεματικό υποδέντρο και προκύπτει το σύνολο όλων των συχνών στοιχειοσυνόλων μαζί με τις αντίστοιχες υποστηρίξεις τους [26], [39].

2.2.3 Ο Αλγόριθμος Eclat

Ο συγκεκριμένος αλγόριθμος χρησιμοποιεί ισοδύναμες κλάσεις, δηλαδή κλάσεις που βασίζονται στο πρόθεμα, σε συνδυασμό με bottom up αναζήτηση μέσω της δενδρικής δομής FT-tree. Αυτό σημαίνει ότι όταν δύο στοιχειοσύνολα έχουν το ίδιο πρόθεμα ανήκουν στην ίδια κλάση και η αναζήτηση από κάτω προς τα πάνω βασίζεται σε μια αναδρομική αποσύνθεση κάθε κόμβου σε μικρότερες ομάδες. Το πλέγμα της ισοδύναμης κλάσης μπορεί να διασχίζεται πρώτα κατά βάθος με το να ξεκινάει ο αλγόριθμος από τον πρώτο κόμβο, μετρώντας την υποστήριξή του στοιχείου και ελέγχοντας αν είναι συχνό.

Κατόπιν αυτού, συνεχίζει στους επόμενους κόμβους που περιέχουν το στοιχείο του πρώτου κόμβου μέχρι να φτάσει στον κόμβο όπου το στοιχειοσύνολο δεν θα είναι συχνό. Αμέσως μετά, ξεκινάει με τον επόμενο κόμβο που δεν περιέχει το στοιχείο του πρώτου κόμβου ακολουθώντας την ίδια διαδικασία. Αυτή η προσέγγιση συνήθως χρησιμοποιείται για να βρεθούν τα μέγιστα συχνά στοιχειοσύνολα. Το πλέγμα της ισοδύναμης κλάσης μπορεί να διασχίζεται και πρώτα κατά πλάτος με το να βρίσκει ο αλγόριθμος όλα τα συχνά στοιχειοσύνολα μήκους 1, μετά τα στοιχειοσύνολα μήκους 2 κ.ο.κ, μέχρι να μην αναπαραχθεί κανένα συχνό στοιχειοσύνολο [23], [27].

2.3 Αξιολόγηση Κανόνων Συσχέτισης

Όπως ήδη έχει αναφερθεί, η ανάλυση κανόνων συσχέτισης αποδίδει κανόνες οι οποίοι ικανοποιούν το κριτήριο ελάχιστης υποστήριξης και της ελάχιστης εμπιστοσύνης και επομένως είναι ισχυροί. Αυτό όμως δεν σημαίνει ότι οι κανόνες είναι ενδιαφέροντες ή ότι συνιστούν θετικές συσχετίσεις μεταξύ του αριστερού και του δεξιού μέλους του κανόνα. Η ανάλυση κανόνων συσχέτισης αποδίδει πολλούς κανόνες, που αναδεικνύουν διαφόρων ειδών πληροφορίες. Παρόλο που η αξιολόγηση των κανόνων διαφέρει από αναλυτή σε αναλυτή, μπορούν να εφαρμοστούν κάποια αντικειμενικά κριτήρια, τα οποία θα περιορίσουν τον αριθμό των κανόνων και θα καταδείξουν πρόσθετα στοιχεία σχετικά με το είδος της συσχέτισης μεταξύ του αριστερού και του δεξιού μέλους των κανόνων. Στην παρακάτω ενότητα παρατίθενται μερικά από τα μέτρα αξιολόγησης των κανόνων.

2.3.1 Lift

Ένα αντικειμενικό κριτήριο αποτίμησης των κανόνων συσχέτισης είναι το στατιστικό μέτρο Lift. Το Lift είναι ένα ποσοτικό μέτρο, το οποίο δείχνει πόσο καλύτερη ή χειρότερη είναι η επίδοση ενός κανόνα σε σχέση με έναν κανόνα τυχαίας επιλογής. Αποδίδει δηλαδή τον βαθμό και το είδος συσχέτισης μεταξύ των μέλους του κανόνα καθώς αντίθετα με τα μέτρα της υποστήριξης και της εμπιστοσύνης, λαμβάνει υπόψιν του την υποστήριξη του δεξιού μέλους του κανόνα. Από τη Στατιστική γνωρίζουμε ότι αν δύο γεγονότα X και Y είναι μεταξύ τους ανεξάρτητα, τότε η πιθανότητα ταυτόχρονης εμφάνισης τους ισούται με το γινόμενο των πιθανοτήτων της εμφάνισης του κάθε γεγονότος ξεχωριστά.

Δύο γεγονότα είναι θετικά συσχετισμένα, αν η εμφάνιση του ενός ενισχύει την πιθανότητα εμφάνισης του άλλου και τότε η πιθανότητα ταυτόχρονης εμφάνισης τους είναι μεγαλύτερη από το γινόμενο των πιθανοτήτων εμφάνισης του κάθε γεγονότος ξεχωριστά. Αντιστρόφως, αν δύο γεγονότα είναι αρνητικά συσχετισμένα, τότε η εμφάνιση του ενός δεν αποτελεί κίνητρο

για την εμφάνιση του άλλου και η πιθανότητα ταυτόχρονης εμφάνισης τους είναι μικρότερη από το γινόμενο των πιθανοτήτων εμφάνισης του καθενός ξεχωριστά[28].

Εάν θεωρήσουμε λοιπόν, έναν κανόνα της μορφής $A \rightarrow B$ και η πιθανότητα εμφάνισης του κανόνα είναι ίση με το γινόμενο των πιθανοτήτων εμφάνισης του αριστερού και δεξιού μέλους ($P(A) * P(B)$), τότε τα δύο σκέλη είναι μεταξύ τους ανεξάρτητα. Ένας κανόνας είναι πραγματικά ενδιαφέρον όταν η πιθανότητα εμφάνισης του διαφέρει από το γινόμενο $P(A) * P(B)$ (Piatetsky-Shapiro, 1991). Το μέτρο Lift χρησιμοποιείται για την αποτίμηση του βαθμού ενδιαφέροντος ενός κανόνα. Το Lift ορίζεται από την Εξίσωση 2.5.

$$\mathbf{Lift} = \frac{\mathbf{conf}(A \rightarrow B)}{\mathbf{supp}(B)} \quad \text{Εξίσωση 2.5}$$

- Αν κάποιος κανόνας έχει $Lift = 1$, η πιθανότητα εμφάνισης του προηγούμενου και του επακόλουθου είναι ανεξάρτητη η μία από την άλλη. Συνεπώς δεν υπάρχει συσχέτιση μεταξύ του A και του B και ο κανόνας που δημιουργείται δεν είναι αξιολογήσιμος.
- Εάν κάποιος κανόνας έχει $Lift > 1$, υπάρχει θετική σχέση μεταξύ του A και του B , και η εμφάνιση του A , ενισχύει την πιθανότητα εμφάνισης του B .
- Εάν κάποιος κανόνας έχει $Lif < 1$, τότε τα A και B είναι αρνητικά συσχετισμένα και η εμφάνιση του A μειώνει τη πιθανότητα εμφάνισης του B .

Η σημασία του μέτρου Lift, καταδεικνύεται καλύτερα με το ακόλουθο παράδειγμα.

Ας υποθέσουμε τον κανόνα $\{\text{Καφές}\} \rightarrow \{\text{Τσάι}\}$, δηλαδή κάποιος που θα αγοράσει καφέ, πιθανά θα αγοράσει και τσάι.

Έστω ότι έχουμε 100 καταναλωτές και από τους καταναλωτές που πίνουν καφέ, το 30% πίνει και τσάι. Έστω τέλος ότι μόνο το 10% από τους 100 καταναλωτές πίνουν τσάι. Αφού η τιμή του Lift είναι ίση με $Lift = \frac{\mathbf{conf}\{\text{Καφές}\} \rightarrow \{\text{Τσάι}\}}{\mathbf{supp}(\text{Τσάι})} = \frac{30\%}{10\%} = 3$, σημαίνει ότι στον υποπληθυσμό των καταναλωτών που πίνουν καφέ, το ποσοστό των καταναλωτών που πίνουν τσάι, είναι 3 φορές μεγαλύτερο από ότι είναι στο συνολικό πληθυσμό (10%).

Το Lift είναι ένα χρήσιμο μέτρο για την κατανόηση της συσχέτισης μεταξύ του αριστερού και του δεξιού μέλους του κανόνα. Ουσιαστικά δείχνει τον παράγοντα αύξησης ή μείωσης της πιθανότητας πώλησης του προϊόντος B , με δεδομένο ότι έχει πωληθεί το προϊόν A . Κανόνες των οποίων η τιμή Lift είναι μεγαλύτερη από 1 είναι πραγματικά ενδιαφέροντες, καθώς καταδεικνύουν περιπτώσεις προϊόντων που η συχνότητα ταυτόχρονης πώλησης τους υπερβαίνει το τυχαίο.

2.3.2 Conviction

Το μέτρο conviction αναπτύχθηκε ως εναλλακτική λύση του μέτρου εμπιστοσύνης, καθώς το τελευταίο δεν αποτυπώνει επαρκώς την κατεύθυνση των συσχετίσεων. Το μέτρο conviction συγκρίνει την πιθανότητα εμφάνισής του αριστερού μέλους του κανόνα χωρίς την ύπαρξη του δεξιού μέλους μόνο αν και τα δύο μέλη του κανόνα είναι ανεξάρτητα από την πραγματική συχνότητα εμφάνισης του αριστερού μέλους χωρίς το δεξί. Από αυτή την άποψη αυτό το μέτρο είναι όμοιο με το μέτρο Lift. Παρόλα αυτά, σε αντίθεση με το lift, είναι ένα προσανατολισμένο μέτρο δεδομένου ότι χρησιμοποιεί τις πληροφορίες από την απουσία του επακόλουθου [28], [29]. Ο τύπος του μέτρου παρουσιάζεται μέσω της Εξίσωσης 2.6.

$$\mathit{conv}(A \Rightarrow B) = \frac{1 - \mathit{sup}(B)}{1 - \mathit{conf}(A \Rightarrow B)} = \frac{P(A) * P(\bar{B})}{P(A \cap \bar{B})} \quad \text{Εξίσωση 2.6}$$

Όπου \bar{B} , είναι το γεγονός ότι το B δεν εμφανίζεται στις συναλλαγές.

2.3.3 Collective Strength

Προκειμένου να αποφευχθούν οι κανόνες που είναι παραπλανητικοί ή ψευδείς επειδή η εμφάνισή τους οφείλεται σε τυχαιότητα, ως λύση προτείνεται το μέτρο της συλλογικής δύναμης [30]. Το επίπεδο ενδιαφέροντος ενός κανόνα είναι ο λόγος της πραγματικής ισχύος προς την αναμενόμενη δύναμη που βασίζεται στην υπόθεση της στατιστικής ανεξαρτησίας (Piatetsky – Shapiro, 1991).

Οι περισσότερες μέθοδοι χρησιμοποιούν το μέτρο συλλογικής δύναμης ως εργαλείο απόρριψης, ώστε να καταργούνται οι κανόνες που δεν έχουν κάποιο ενδιαφέρον. Ωστόσο, για όσο διάστημα η υποστήριξη εξακολουθεί να είναι ο πρωταρχικός καθοριστικός παράγοντας στην αρχική αναπαραγωγή στοιχειοσυνόλων, είτε ο χρήστης πρέπει να ρυθμίσει την αρχική υποστήριξη της παράμετρου αρκετά χαμηλή ώστε να μην χάσει κανέναν ενδιαφέροντα κανόνα είτε να ρισκάρει χάνοντας κάποιους σημαντικούς κανόνες. Στην πρώτη περίπτωση, η υπολογιστική αποτελεσματικότητα μπορεί να είναι ένα πρόβλημα, ενώ η δεύτερη περίπτωση έχει το πρόβλημα να μην είναι σε θέση να διατηρήσει κανόνες που μπορεί να είναι ενδιαφέροντες από την άποψη ενός χρήστη.

Οι κανόνες αντιστοίχισης έχουν συζητηθεί από τον Brin κ.α το 1997[31]. Η ισχύς υπολογισμού ενός κανόνα είναι ένας αριθμός μεταξύ 0 και 1. Ο αριθμός 1 υποδεικνύει ότι η ισχύς του κανόνα είναι ακριβώς όπως θα ήταν υπό την υπόθεση της στατιστικής ανεξαρτησίας. Μια τιμή μεγαλύτερη από 1 υποδηλώνει παρουσία μεγαλύτερη από την αναμενόμενη παρουσία του

κανόνα. Αυτό το μέτρο είναι προτιμότερο από το μέτρο της εμπιστοσύνης, διότι ασχολείται με μέτρα μεγαλύτερα από τα αναμενόμενα για την εξεύρεση κανόνων συσχέτισης.

Η διαφορετική έννοια που ονομάζεται “συλλογική δύναμη” επιθυμεί να βρει ενδιαφέροντες κανόνες συσχέτισης χρησιμοποιώντας μεγαλύτερες τιμές από τις αναμενόμενες [32].

Η συλλογική ισχύς στοιχειοσυνόλου, είναι ένας αριθμός μεταξύ 0 και ∞ . Η τιμή 0 υποδεικνύει τέλεια αρνητική συσχέτιση, ενώ η τιμή ∞ δείχνει τέλεια θετική συσχέτιση και 1 εάν τα στοιχεία συνυπάρχουν όπως αναμένεται με ανεξαρτησία.

Ένα στοιχειοσύνολο A λέγεται ότι παραβιάζει μια συναλλαγή, εάν κάποια από τα στοιχεία είναι παρόντα στη συναλλαγή ενώ κάποια άλλα δεν είναι. Έτσι, η έννοια της παραβίασης δηλώνει πόσες φορές ένας πελάτης μπορεί να αγοράσει τουλάχιστον μερικά από τα στοιχεία ενός στοιχειοσυνόλου χωρίς πιθανά να αγοράζει τα υπόλοιπα στοιχεία. Το ποσοστό παραβίασης ενός στοιχειοσυνόλου A υποδηλώνεται με $\nu(A)$ και είναι ο λόγος των συναλλαγών που περιέχουν κάποια από τα στοιχεία ενός στοιχειοσυνόλου αλλά όχι όλα.

Η συλλογική δύναμη στοιχειοσυνόλου A , ορίζεται ως,

$$C(A) = \frac{1-\nu(A)}{1-E[\nu(A)]} * \frac{E[\nu(A)]}{\nu(A)} \quad \text{Εξίσωση 2.7}$$

Θεωρούμε ότι η παραβίαση στοιχειοσυνόλου σε μία συναλλαγή ως ένα κακό γεγονός από την άποψη ότι προσπαθούμε να δημιουργήσουμε υψηλή συσχέτιση μεταξύ των αντίστοιχων στοιχείων. Για αυτό τον λόγο το $\nu(A)$ αναφέρεται στα κακά γεγονότα, ενώ το $1 - E[\nu(A)]$, στα καλά. Το πρόβλημα είναι ότι για τα στοιχεία με μεσαίες έως χαμηλές πιθανότητες οι παρατηρήσεις των αναμενόμενων τιμών του ποσοστού παραβίασης κυριαρχούν από το ποσοστό των συναλλαγών που δεν περιέχουν κανένα από τα στοιχεία του A . Για τέτοια στοιχειοσύνολα, η συλλογική δύναμη παράγει τιμές κοντά σε ένα, ακόμα και αν ένα στοιχειοσύνολο εμφανίζεται πολλές φορές από ότι θα αναμενόταν να εμφανίζονταν μαζί όλα τα στοιχειοσύνολα.

2.3.4 Weighted Association Rule Mining

Αυτή η προσέγγιση γενικεύει την παραδοσιακή εξόρυξη κανόνων συσχέτισης, αναθέτοντας βάρη στα αντικείμενα βάση της σπουδαιότητάς τους. Το βάρος ενός στοιχείου είναι ένας μη αρνητικός πραγματικός αριθμός που δείχνει τη σημασία κάθε στοιχείου. Ένα ζεύγος (x, w) ονομάζεται σταθμισμένο στοιχείο όπου το $x \in G$ (όλα τα στοιχεία) και $w \in W$ είναι το βάρος που σχετίζεται με το x . Μια συναλλαγή είναι ένα σύνολο σταθμισμένων στοιχείων, το καθένα από τα οποία μπορεί να εμφανίζεται σε πολλαπλές συναλλαγές με διαφορετικά βάρη.

Ο Ramkumar κ.α, το 1998 [33] εισήγαγε την έννοια της σταθμισμένης υποστήριξης των κανόνων συσχέτισης με βάση το κόστος που αφιερώθηκε τόσο στα στοιχεία όσο και στις συναλλαγές. Ένας αλγόριθμος που ονομάζεται WIS προτάθηκε για να αντλήσει τους κανόνες που έχουν σταθμισμένη υποστήριξη μεγαλύτερη από ένα δεδομένο όριο. Ο Cai και οι συνεργάτες του το 1998, [34] καθόρισαν την έννοια της σταθμισμένης υποστήριξης με παρόμοιο τρόπο, αλλά έλαβαν υπόψη μόνο τα βάρη των στοιχείων και όχι των συναλλαγών. Ο ορισμός έσπασε την ιδιότητα κλεισίματος (closure property) προς τα κάτω. Ως αποτέλεσμα, ο προτεινόμενος αλγόριθμος εξόρυξης έγινε πιο περίπλοκος και χρονοβόρος.

Οι Ταο κ.α [35] προέβλεψαν έναν άλλο ορισμό για τη διατήρηση της "σταθμισμένης περιουσιακής κατάστασης κλεισίματος". Συμπερασματικά, η μεθοδολογία WARM είναι για να αποδίδουμε βάρη σε στοιχεία, να εφεύρουμε νέα μέτρα (σταθμισμένη υποστήριξη) βάσει αυτών των βαρών και να αναπτύξουμε τους αντίστοιχους αλγόριθμους εξόρυξης.

Η σταθμισμένη υποστήριξη δύο στοιχειοσυνόλων A,B, όπου $A \cap B = \emptyset$, που ανήκουν στο σύνολο συναλλαγών T, ορίζεται ως ο λόγος του βάρους των συναλλαγών που περιέχουν τόσο το A όσο και το B, σε σχέση με το βάρος όλων των συναλλαγών.

$$wsp(AB) = \frac{\sum_{k=1}^{|WST| \& A \cup B \subseteq tk} weight(t_k)}{\sum_{k=1}^{|WST|} weight(t_k)} \quad \text{Εξίσωση 2.8}$$

2.3.5 Correlation Analysis

Πρόκειται για μια τεχνική που βασίζεται σε στατιστική ανάλυση, και εξετάζει τις σχέσεις ανάμεσα σε ζεύγη μεταβλητών π.χ. ($A \rightarrow B$). Για συνεχείς μεταβλητές, η συσχέτιση καθορίζεται από τον συντελεστή συσχέτισης pearson[37]. Για δυαδικές μεταβλητές, η συσχέτιση μπορεί να μετρηθεί κάνοντας χρήση του ϕ - συντελεστή[36].

$$\phi = \frac{f_{11}f_{00} - f_{01}f_{10}}{\sqrt{f_{1+}f_{+1}f_{0+}f_{+0}}} \quad \text{Εξίσωση 2.9}$$

Κάθε f_{ij} υποδηλώνει τη συχνότητα της κάθε μεταβλητής στις συναλλαγές.

f_{11} , δείχνει πόσες φορές το A και το B, εμφανίζονται μαζί στην ίδια συναλλαγή

f_{01} , δείχνει τον αριθμό των συναλλαγών που υπάρχει το B αλλά όχι το A

f_{10} , δείχνει τον αριθμό των συναλλαγών που υπάρχει το A αλλά όχι το B

f_{1+} Η υποστήριξη του A

$f + 1$ Η υποστήριξη του B

$f0 +$ Η υποστήριξη όταν το A δεν εμφανίζεται

$f + 0$ Η υποστήριξη όταν το B δεν εμφανίζεται

Η τιμή της συσχέτισης κυμαίνεται από -1 (τέλεια αρνητική συσχέτιση) έως +1 (τέλεια θετική συσχέτιση). Ο συντελεστής ϕ είναι ανάλογος του συντελεστή συσχέτισης pearson για συνεχείς μεταβλητές. Για μεταβλητές που είναι στατιστικά ανεξάρτητες ο συντελεστής $\phi = 0$.

2.3.6 Effect of Skewed Support Distribution

Τα χαρακτηριστικά των δεδομένων που εισάγονται σε αλγόριθμους συσχέτισης, επηρεάζουν την συμπεριφορά τους. Στην περίπτωση που εξετάζουμε, επικεντρωνόμαστε σε σύνολα δεδομένων με αλλοιωμένες κατανομές υποστήριξης, όπου τα περισσότερα στοιχεία έχουν σχετικά χαμηλές προς μεσαίες συχνότητες εμφάνισης. Αυτό εξαρτάται από το όριο της υποστήριξης που θέτουμε. Με όριο υποστήριξης πολύ υψηλό, υπάρχει πιθανότητα να χάσουμε πολλά ενδιαφέροντα μοντέλα (κανόνες) που περιέχουν χαμηλή υποστήριξη. Αυτή η χαμηλή υποστήριξη μπορεί να οφείλεται στο γεγονός ότι τα αντικείμενα είναι υψηλής χρηματικής αξίας, συνεπώς δεν επιλέγονται με τόσο μεγάλη συχνότητα από τους αγοραστές (κοσμήματα).

Από την άλλη μεριά θέτοντας όριο υποστήριξης πολύ χαμηλό, είναι δύσκολο να βρούμε κανόνες συσχέτισης καθώς οι απαιτήσεις υπολογισμού και χώρου αποθήκευσης τόσων πολλών κανόνων συσχέτισης αυξάνεται σημαντικά όταν το όριο υποστήριξης είναι χαμηλό, γιατί εξάγονται πολλοί κανόνες. Επίσης υπάρχει πιθανότητα να εξαχθούν ψευδείς κανόνες που συσχετίζουν συχνά στοιχειοσύνολα με μη συχνά στοιχειοσύνολα. Αυτοί οι κανόνες λέγονται cross-support patterns, γιατί φαίνεται να είναι αναληθείς εφόσον η συσχέτιση τους είναι αδύναμη και η υποστήριξή τους είναι μικρότερη από αυτή που έχει καθοριστεί από τον χρήστη.

Για αυτό τον λόγο εισάγουμε την ιδέα cross-support property, η οποία είναι χρήσιμη για την εξάλειψη των υποψηφίων μοντέλων που έχουν στοιχεία με πολύ διαφορετική υποστήριξη. Το μέτρο **h-confidence** έχει ένα τέτοιο χαρακτηριστικό και αναπτύσσει έναν αποδοτικό αλγόριθμο που ονομάζεται hyperclique miner που χρησιμοποιεί τόσο τις σταυροειδείς υποστηρίξεις όσο και τις αντιμονοτονικές ιδιότητες του μέτρου εμπιστοσύνης h- confidence [38].

$$r(X) = \frac{\min[s(i1),s(i2),\dots,s(ik)]}{\max[s(i1),s(i2),\dots,s(ik)]} \quad \text{Εξίσωση 2.10}$$

2.3.7 All-confidence

Το μέτρο ***all – confidence*** ορίζει ότι οι κανόνες που μπορούν να αναπαραχθούν από ένα στοιχειοσύνολο X έχουν τουλάχιστον εμπιστοσύνη από το μέτρο ***all – confidence***(X)[15].

$$\mathbf{all - confidence} = \frac{[\mathbf{sup}(X)]}{\mathbf{max}_{x \in X}(\mathbf{sup}(x \in X))} \quad \text{Εξίσωση 2.11}$$

Όπου ***sup***(X), είναι η υποστήριξη ενός στοιχειοσυνόλου και ***max*** $_{x \in X}(\mathbf{sup}(x \in X))$ είναι η υποστήριξη ενός στοιχείου με την υψηλότερη υποστήριξη μέσα στο στοιχειοσύνολο X .

ΚΕΦΑΛΑΙΟ 3

Στο παρόν κεφάλαιο, θα εφαρμόσουμε σε πραγματικά δεδομένα τη μέθοδο που αναλύσαμε στο Κεφάλαιο 1 και με τη βοήθεια του αλγόριθμου Apriori θα προσπαθήσουμε να αναγνωρίσουμε τα πρότυπα συμπεριφοράς των καταναλωτών. Το πρώτο μέρος της ανάλυσης, προβαίνουμε στη δημιουργία συσχετίσεων, σε επίπεδο κατηγοριών προϊόντων (high-level). Όταν δύο ή περισσότερες κατηγορίες προϊόντων συσχετίζονται, σημαίνει ότι και τα προϊόντα που ανήκουν σε αυτές τις κατηγορίες αγοράζονται μαζί από τους καταναλωτές. Η ανάλυση δεν θα πραγματοποιηθεί σε επίπεδο προϊόντων, γιατί όπως θα εξηγήσουμε παρακάτω οι συσχετίσεις που προκύπτουν θα αναφέρονται σε πολύ μικρά ποσοστά για να είναι μετρήσιμα και έτσι η ανάλυσή μας δεν αποκτά μεγάλο επιχειρησιακό ενδιαφέρον.

Στο δεύτερο μέρος της ανάλυσης, θα προβούμε στην αναγνώριση προτύπων συμπεριφοράς καταναλωτών μέσα από δεδομένα συναλλαγών για δύο διαφορετικές εβδομάδες όπου στη μία από τις δύο εβδομάδες κάποια προϊόντα είναι σε προσφορά. Με αυτό το τρόπο, θα προσπαθήσουμε να διακρίνουμε τυχόν αλλαγές στο καλάθι ενός καταναλωτή την εβδομάδα που κάποια προϊόντα ήταν σε προσφορά, σε σχέση με την εβδομάδα όπου τα ίδια προϊόντα, δεν ήταν σε προσφορά. Η προωθητική ενέργεια διαφημίζεται είτε σε έντυπη μορφή μέσω φυλλαδίου, είτε μέσω προβολής της στην τηλεόραση. Ουσιαστικά θα προβούμε στη δημιουργία συσχετίσεων σε επίπεδο προϊόντων (low-level), ξεχωριστά για κάθε εβδομάδα. Την πρώτη εβδομάδα θα εξετάσουμε τα προϊόντα που δεν είναι σε προσφορά, αλλά είναι σε προσφορά την δεύτερη, με ποια προϊόντα αγοράζονται μαζί, από τους καταναλωτές. Τη δεύτερη εβδομάδα θα εξετάσουμε πάλι τα παραπάνω προϊόντα που είναι σε προσφορά αυτή τη φορά, με ποια προϊόντα αγοράζονται μαζί. Όλη η ανάλυση θα πραγματοποιηθεί με τη γλώσσα προγραμματισμού ανοιχτού κώδικα R, μέσω του περιβάλλοντος RStudio.

3.1 Πρώτο Μέρος Ανάλυσης

3.1.1 Περιγραφή Δεδομένων

Για την εκπόνηση της ανωτέρω ανάλυσης, χρησιμοποιήσαμε δεδομένα συναλλαγών καταναλωτή για τη περίοδο 1/10/2017- 30/11/2017, τα οποία μας χορηγήθηκαν από το κατάστημα του Αιγάλεω, του σουπερμάρκετ «ΑΒ Βασιλόπουλος». Πιο συγκεκριμένα, το αρχείο που επεξεργαστήκαμε (αρχείο Excel), περιέχει, τον μοναδικό αριθμό συναλλαγής (Transaction ID) ο οποίος εκφράζει τον μοναδικό αριθμό απόδειξης, τον κωδικό του κάθε

προϊόντος (Product ID), την περιγραφή του κάθε προϊόντος (Description), το τμήμα στο οποίο ανήκει κάθε προϊόν (Department) και τέλος την κατηγορία στην οποία ανήκει (Category).

Στην Εικόνα 2 παρουσιάζονται οι 18 πρώτες εγγραφές του αρχείου που επεξεργαστήκαμε ενώ στον Πίνακα 5, βλέπουμε τη δομή του αρχείου μας.

Columns	Labels	Count
1 st Column	TransID	76.810
2 nd Column	ProductID	14.674
3 rd Column	Desc	14.674
4 th Column	Dept	35
5 th Column	Categ	198
Rows of the Dataset	720.374	

Πίνακας 5 Περιγραφή Δεδομένων

	transID	ProductID	Desc	Dept	Categ
1	1	40713397700	ΝΟΥΝΟΥ LIGHT 400GR	ΠΑΝΤΟΠΩΛΕΙΟ	ΓΑΛΑ Μ.ΔΙΑΡΚΕΙΑΣ
2	1	40714506000	FITNESS CRUNCHY CARAMEL 6X23.5G	ΠΑΝΤΟΠΩΛΕΙΟ	ΔΗΜΗΤΡΙΑΚΑ-MUSLI
3	1	40722396500	ΜΠΟΥΤΑΚΙΑ ΜΙΜΙΚΟΣ 590ΓΡ+120ΓΡ ΔΩΡΟ	ΚΡΕΟΠΩΛΕΙΟ	ΝΩΠΑ ΠΟΥΛΕΡΙΚΑ
4	1	40723214900	AMICO ΓΥΝ.ΜΠΛΟΥΖΑ RIB	ΕΝΔΥΣΗ	ΕΝΔΥΣΗ ΓΥΝΑΙΚΕΙΑ
5	1	90020010100	ΑΧΛΑΔΙΑ ΚΡΥΣΤΑΛΙΑ ΕΛΛΗΝΙΚΑ	ΟΠΩΡΟΠΩΛΕΙΟ	ΦΡΟΥΤΑ
6	1	90020221400	ΠΟΡΤΟΚΑΛΙΑ ΒΑΛΕΝ.ΧΥΜΟΥ ΕΛΛΗΝ.ΣΥΣΚ.	ΟΠΩΡΟΠΩΛΕΙΟ	ΦΡΟΥΤΑ
7	1	90025680000	ΑΒ ΕΛΛ ΓΗ ΧΥΛΟΠΙΤΑΚΙ ΚΟΡΙΝΘΟΥ 500Γ	ΠΑΝΤΟΠΩΛΕΙΟ	ΖΥΜΑΡΙΚΑ
8	1	90040100100	ΤΟΜΑΤΕΣ ΧΥΜΑ ΕΛΛΗΝΙΚΕΣ	ΟΠΩΡΟΠΩΛΕΙΟ	ΑΓΓΟΥΡΙΑ - ΤΟΜΑΤΕΣ
9	1	90052170000	ΑΒ ΠΕΡΑΣΤΗ ΤΟΜΑΤΑ 500GR	ΠΑΝΤΟΠΩΛΕΙΟ	ΤΟΜΑΤΟΕΙΔΗ
10	1	90110590301	ΙΟΝ ΣΟΚΟΦΡΕΤΑ 4X38GR 0.20Ε	ΖΑΧΑΡΩΔΗ	ΣΟΚΟΛΑΤΟΕΙΔΗ
11	1	90117650000	FITNESS ΜΑΥΡΗ ΣΟΚΟΛΑΤΑ 375ΓΡ	ΠΑΝΤΟΠΩΛΕΙΟ	ΔΗΜΗΤΡΙΑΚΑ-MUSLI
12	1	90163610100	LIFTON ΠΥΡΑ.ΑΦΕΨ.ΜΗΛΟ.ΚΑΝ.20ΦΧ2.2Γ	ΠΑΝΤΟΠΩΛΕΙΟ	ΡΟΦΗΜΑΤΑ
13	1	93000242100	ΑΒ ΣΑΚ. ΤΡΟΦΙΜΩΝ ΨΥΓ. ΜΕΖΑΙΕΣ 50Τ	ΟΙΚΙΑΚΑ - ΠΑΝΤΟΠΩΛΕΙΟ	ΕΙΔΗ ΣΥΝΤΗΡΗΣΗΣ ΤΡΟΦΙΜΩΝ
14	1	93010123300	ΑΒ ΣΑΚ.ΑΠΟΡ.ΚΟΡΔΟΝΙ ΑΡΩΜ.52Χ75 10Τ	ΟΙΚΙΑΚΑ - ΠΑΝΤΟΠΩΛΕΙΟ	ΣΑΚΟΥΛΕΣ ΑΠΟΡΡΙΜΜΑΤΩΝ
15	2	40715106500	BREF WC POWER ACTIVE OCEAN 50GR	ΑΠΟΡΡΥΠΑΝΤΙΚΑ - ΚΑΘΑΡΙΣΤΙΚΑ	ΚΑΘΑΡΙΣΤΙΚΑ ΣΠΙΤΙΟΥ
16	2	40723401000	ΦΥΛΛΟ ΠΑΡΑΔΟΣΙΑΚΟ ΚΑΝΑΚΙ ΚΤΨ 650G	ΚΑΤΕΨΥΓΜΕΝΑ	ΚΑΤΕΨΥΓΜΕΝΑ ΠΑΡΑΣΚΕΥΑΣΜ.ΠΡΟΙΟΝΤΑ
17	2	40729352400	BREF WC BLUE AKTIV FLORAL 50GR	ΑΠΟΡΡΥΠΑΝΤΙΚΑ - ΚΑΘΑΡΙΣΤΙΚΑ	ΚΑΘΑΡΙΣΤΙΚΑ ΣΠΙΤΙΟΥ
18	2	90570560200	ΤΥΡΙ ΦΕΤΑ ΤΡΙΠΟΛΕΩΣ ΒΑΡΕΛΙ	ΤΥΡΙΑ	ΦΕΤΕΣ

Showing 1 to 18 of 720,374 entries

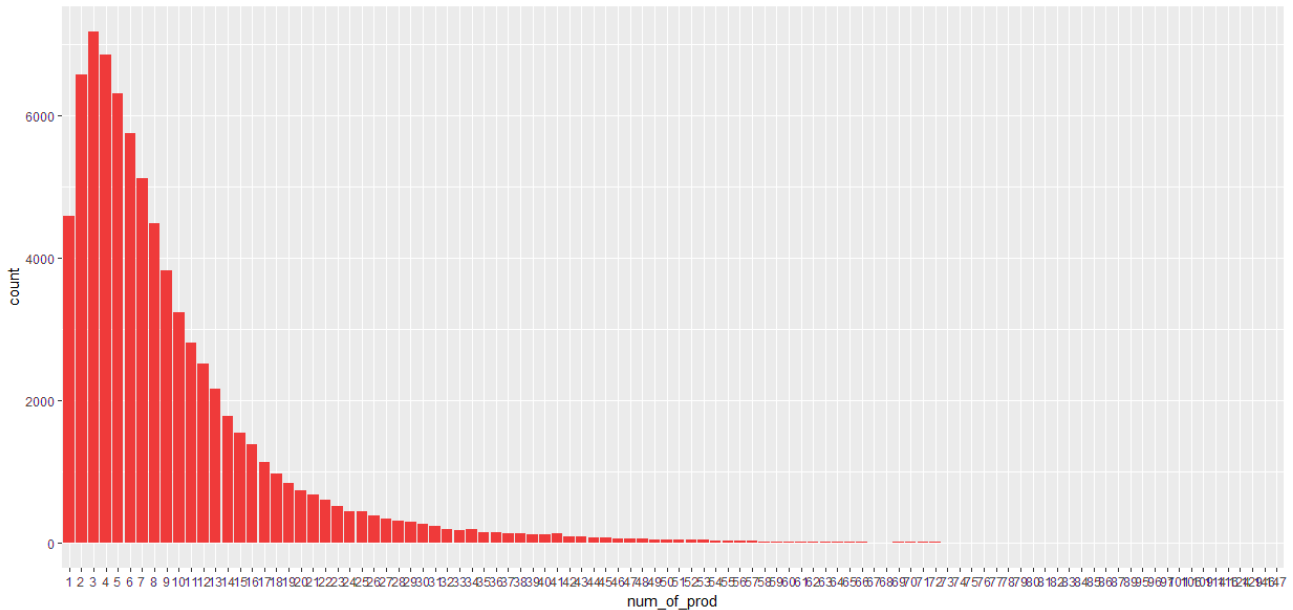
Εικόνα 2 Σύνολο Δεδομένων

Σύμφωνα με τον Πίνακα 5 η βάση δεδομένων, αποτελείται από 76.810 συναλλαγές για το διάστημα των 2 μηνών. Κάθε προϊόν αντιστοιχεί σε έναν κωδικό και μία περιγραφή.

Όσον αφορά στη περιγραφή κάθε διαφορετικού προϊόντος που βρέθηκε στα καλάθια των καταναλωτών, το πλήθος τους είναι 14.678, όσο και το πλήθος των διαφορετικών κωδικών.

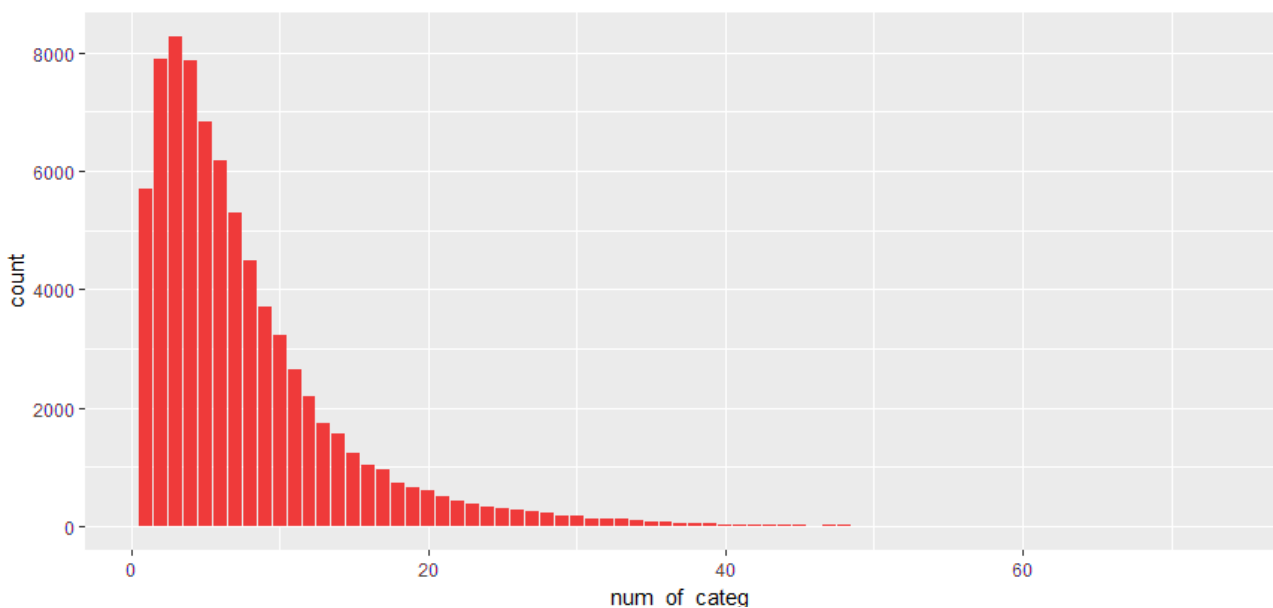
Τέλος, υπάρχουν 35 διαφορετικά τμήματα και 198 διαφορετικές κατηγορίες, όπου μπορεί να ανήκει το κάθε προϊόν. Στη παρούσα φάση, θα παρουσιάσουμε τα χαρακτηριστικά των δεδομένων μέσω γραφικής απεικόνισης, ώστε να τα κατανοήσουμε καλύτερα.

Παρακάτω στο Σχήμα 7, παρουσιάζεται η κατανομή του πλήθους των προϊόντων ανά συναλλαγή. Πιο συγκεκριμένα, ένα προϊόν βρίσκεται σε πάνω από 4.000 συναλλαγές, 2 προϊόντα υπάρχουν σε πάνω από 6.000 συναλλαγές, 147 προϊόντα βρίσκονται σε μία συναλλαγή κοκ.



Σχήμα 7 Κατανομή του Πλήθους των Προϊόντων ανά Συναλλαγή

Όσον αφορά στη κατανομή του πλήθους των κατηγοριών ανά συναλλαγή, μας τη δείχνει το Σχήμα 8.



Σχήμα 8 Κατανομή του Πλήθους των Κατηγοριών ανά Συναλλαγή

Όπως παρατηρούμε, περισσότερες από 5.000 συναλλαγές περιέχουν 1 κατηγορία, περίπου 8.000 συναλλαγές περιέχουν 2 κατηγορίες κοκ. Όπως παραπάνω από ένα προϊόντα μπορούν να ανήκουν στην ίδια κατηγορία, ένα ή περισσότερα προϊόντα μπορεί να ανήκουν σε ένα τμήμα. Το Σχήμα 9 δείχνει πόσα τμήματα έχει η κάθε συναλλαγή για τις είκοσι πρώτες συναλλαγές.



Σχήμα 9 Πλήθος Τμημάτων για τις 2- Πρώτες Συναλλαγές

Παρατηρούμε, ότι η πρώτη συναλλαγή περιλαμβάνει 6 διαφορετικά τμήματα, παρόλο που περιλαμβάνει 12 κατηγορίες και 14 προϊόντα. Αντίθετα, η τέταρτη συναλλαγή, αποτελείται από δύο προϊόντα, που ανήκουν το καθένα σε δύο διαφορετικές κατηγορίες και σε δύο διαφορετικά τμήματα. Το σύνολο δεδομένων μας, αποτελείται από 14.674 διαφορετικά προϊόντα, και 198 κατηγορίες. Στο Σχήμα 10 που ακολουθεί, βλέπουμε την κατανομή των προϊόντων ανά κατηγορία..



Σχήμα 10 Κατανομή Προϊόντων ανά Κατηγορία

Η κατηγορία «ΦΡΟΝΤΙΔΑ ΣΩΜΑΤΟΣ», περιλαμβάνει 920 προϊόντα, η κατηγορία «ΦΡΟΝΤΙΔΑ ΜΑΛΛΙΩΝ» περιλαμβάνει 609 προϊόντα, η κατηγορία «ΠΡΟΣΩΠΙΚΗ ΥΓΙΕΙΝΗ» περιλαμβάνει 532 προϊόντα, η κατηγορία «ΑΠΟΡΡΥΠΑΝΤΙΚΑ ΡΟΥΧΩΝ», περιλαμβάνει 408 προϊόντα, κοκ.

Η κατανομή των προϊόντων ανά τμήμα, παρουσιάζεται στο Σχήμα 11.

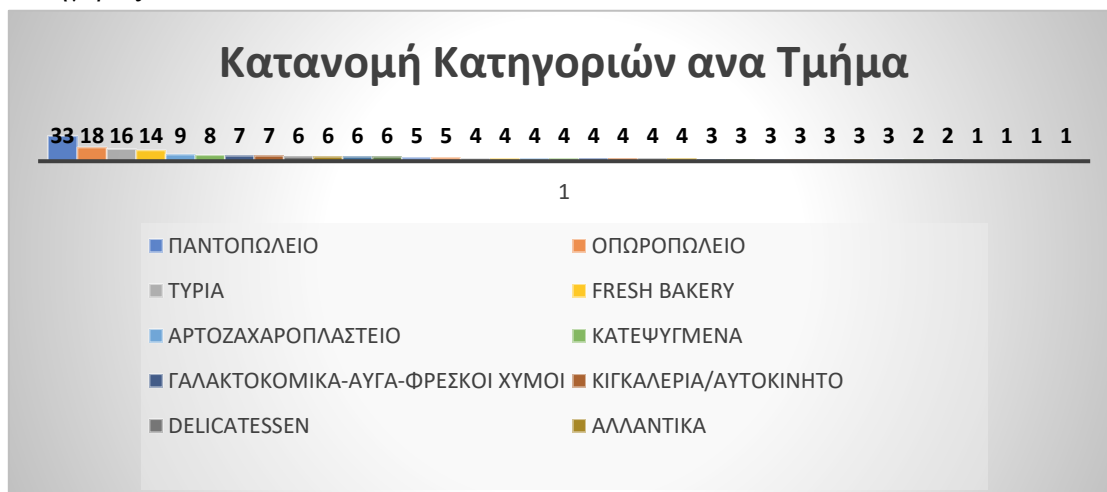


Σχήμα 11 Κατανομή Προϊόντων ανά Τμήμα

Το τμήμα που περιλαμβάνει τα περισσότερα προϊόντα είναι το «ΠΑΝΤΟΠΩΛΕΙΟ». Ακολουθεί το τμήμα «ΥΓΙΕΙΝΗ-ΟΜΟΡΦΙΑ» και το τμήμα «ΑΠΟΡΡΥΠΑΝΤΙΚΑ-ΚΑΘΑΡΙΣΤΙΚΑ».

Η κατανομή των κατηγοριών ανά τμήμα, παρουσιάζεται στο Σχήμα 12.

Οι περισσότερες κατηγορίες, ανήκουν στο τμήμα «ΠΑΝΤΟΠΩΛΕΙΟ». Ακολουθούν οι κατηγορίες, «ΟΠΩΡΟΠΩΛΕΙΟ», «ΤΥΡΙΑ» και «FRESH BAKERY».



Σχήμα 12 Κατανομή Κατηγοριών ανα Τμήμα

Το Σχήμα 13, δείχνει 15 από τα 14.674 προϊόντα του καταστήματος, τα οποία έχουν τη μεγαλύτερη συχνότητα εμφάνισης στα καλάθια των καταναλωτών.



Σχήμα 13 Δεκαπέντε Προϊόντα με τη Μεγαλύτερη Συχνότητα Εμφάνισης στις Συναλλαγές

Όπως φαίνεται οι μπανάνες έχουν τη μεγαλύτερη συχνότητα εμφάνισης στα καλάθια των καταναλωτών, και ακολουθούν οι πατάτες και τα κρεμμύδια. Αντίστοιχα, δώδεκα από τις κατηγορίες με τη μεγαλύτερη συχνότητα εμφάνισης, παρουσιάζονται μέσω του Σχήματος 14.



Σχήμα 14 Δώδεκα Κατηγορίες με τη Μεγαλύτερη Συχνότητα Εμφάνισης

Η κατηγορία με τη μεγαλύτερη συχνότητα Εμφάνισης, είναι η «ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ».

Κατόπιν παρουσίασης κάποιων βασικών χαρακτηριστικών των δεδομένων, θα προχωρήσουμε σε μια προ επεξεργασία των δεδομένων μας για να μπορούμε να εξασφαλίσουμε την εισαγωγή των δεδομένων αυτών στον αλγόριθμο Apriori ώστε να εξαχθούν κανόνες συσχέτισης τους οποίους στη συνέχεια θα αξιολογήσουμε.

3.1.2 Προετοιμασία Δεδομένων

Κατόπιν εύρεσης της συχνότητας εμφάνισης κάθε διαφορετικού προϊόντος στις συναλλαγές, αθροίσαμε τις συχνότητες εμφάνισης όλων των διαφορετικών προϊόντων και τις διαιρέσαμε με το πλήθος τους. Βρήκαμε ότι κατά μέσο όρο κάθε προϊόν εμφανίζεται σε 49,09186 συναλλαγές από τις 76.810. Αυτό σημαίνει ότι η συχνότητα εμφάνισης κάθε προϊόντος σε μία συναλλαγή κατά μέσο όρο είναι πολύ μικρή της τάξεως $6 \cdot 10^{-4}$.

Ακολουθώντας την ίδια διαδικασία και για τις κατηγορίες των προϊόντων βρήκαμε ότι κατά μέσο όρο κάθε κατηγορία, εμφανίζεται σε 3.638,253 συναλλαγές από τις 76.810 συνολικά. Δηλαδή η συχνότητα εμφάνισης κάθε κατηγορίας σε μία συναλλαγή είναι περίπου 4,73%.

Σύμφωνα με όσα αναφέραμε, εάν χρησιμοποιήσουμε όλο το σύνολο δεδομένων που έχουμε στη διάθεσή μας προσπαθώντας να ανακαλύψουμε συσχετίσεις μεταξύ των προϊόντων, θα αναφερόμαστε σε πολύ μικρά ποσοστά για να είναι μετρήσιμα. Με άλλα λόγια, οι συσχετίσεις που θα προκύψουν, θα αναφέρονται στο 0,6 τις χιλίους των καταναλωτών.

Για να αποκτήσει η ανάλυση μας μεγαλύτερο επιχειρησιακό ενδιαφέρον, θα επικεντρωθούμε σε κάποια τμήματα προϊόντων, και εν συνεχεία με τη βοήθεια του αλγόριθμου Apriori, θα αναζητήσουμε συσχετίσεις σε επίπεδο κατηγοριών. Τα τμήματα των προϊόντων που επιλέξαμε να χρησιμοποιήσουμε είναι τα ακόλουθα εννέα από τα τριανταπέντε (Εικόνα 3).

▲	Dept
1	ΠΑΝΤΟΠΩΛΕΙΟ
2	ΚΡΕΟΠΩΛΕΙΟ
3	ΑΠΟΡΡΥΠΑΝΤΙΚΑ - ΚΑΘΑΡΙΣΤΙΚΑ
4	ΤΥΡΙΑ
5	ΓΑΛΑΚΤΟΚΟΜΙΚΑ-ΑΥΓΑ-ΦΡΕΣΚΟΙ ΧΥΜΟΙ
6	LIQUIDE GROCERY FOOD
7	ΥΓΙΕΙΝΗ - ΟΜΟΡΦΙΑ
8	ΑΛΛΑΝΤΙΚΑ
9	ΚΑΒΑ

Εικόνα 3 Επιλογή 9 Τμημάτων από τα 35

Επομένως μετά την επιλογή μόνο αυτών των τμημάτων, το σύνολο δεδομένων που θα εισάγουμε στον αλγόριθμο, φαίνεται στην Εικόνα 4.

Columns	Labels	Count
1st Column	TransID	72.543
2st Column	Productid	9.988
3rd Column	Desc	9.988
4th Column	Dept	9
5th Column	Categ	83
Rows of the Dataset	477.310	

Εικόνα 4 Σύνολο δεδομένων, με την Ύπαρξη 9 Τμημάτων από τα 35

Σε αυτό το υποσύνολο του αρχικού συνόλου δεδομένων, κάθε προϊόν κατά μέσο όρο υπάρχει σε 47,78835 συναλλαγές από τις εναπομένουσες 72.543. Καθώς, η συχνότητα εμφάνισης κάθε προϊόντος σε κάθε συναλλαγή κατά μέσο όρο είναι πολύ μικρή, τάξεως 10^{-3} , αναμένουμε ότι και οι κανόνες συσχέτισης που θα εξαχθούν θα έχουν πολύ μικρή υποστήριξη. Γι' αυτό τον λόγο, θα εξορύξουμε κανόνες συσχέτισης με βάση τις κατηγορίες των προϊόντων καθώς κατά μέσο όρο κάθε κατηγορία εμφανίζεται στο 8% των συναλλαγών και όχι με βάση τις περιγραφές των προϊόντων. Κατά μέσο όρο κάθε κατηγορία υπάρχει σε 5.750,823 συναλλαγές από τις 72.543. Η δομή των δεδομένων δίδεται στην Εικόνα 4 και ο Πίνακας 6 παρουσιάζει ενδεικτικά τις πρώτες 10 εγγραφές του συνόλου δεδομένων που θα χρησιμοποιήσουμε τελικά στην ανάλυσή μας.

	transID	ProductID	Desc	Dept	Categ
1	1	40713397700	ΝΟΥΝΟΥ LIGHT 400GR	ΠΑΝΤΟΠΩΛΕΙΟ	ΓΑΛΑ Μ.ΔΙΑΡΚΕΙΑΣ
2	1	40714506000	FITNESS CRUNCHY CARAMEL 6X23.5G	ΠΑΝΤΟΠΩΛΕΙΟ	ΔΗΜΗΤΡΙΑΚΑ-MUSLI
3	1	40722396500	ΜΠΟΥΤΑΚΙΑ ΜΙΜΙΚΟΣ 590ΓΡ+120ΓΡ ΔΩΡΟ	ΚΡΕΟΠΩΛΕΙΟ	ΝΩΠΙΑ ΠΟΥΛΕΡΙΚΑ
4	1	90025680000	ΑΒ ΕΛΛ ΓΗ ΧΥΛΟΠΙΤΑΚΙ ΚΟΡΙΝΘΟΥ 500Γ	ΠΑΝΤΟΠΩΛΕΙΟ	ΖΥΜΑΡΙΚΑ
5	1	90052170000	ΑΒ ΠΕΡΑΣΤΗ ΤΟΜΑΤΑ 500GR	ΠΑΝΤΟΠΩΛΕΙΟ	ΤΟΜΑΤΟΕΙΔΗ
6	1	90117650000	FITNESS ΜΑΥΡΗ ΣΟΚΟΛΑΤΑ 375ΓΡ	ΠΑΝΤΟΠΩΛΕΙΟ	ΔΗΜΗΤΡΙΑΚΑ-MUSLI
7	1	90163610100	LIPTON ΠΥΡΑ.ΑΦΕΨ.ΜΗΛΟ.ΚΑΝ.20ΦΧ2.2Γ	ΠΑΝΤΟΠΩΛΕΙΟ	ΡΟΦΗΜΑΤΑ
33	10	40712946400	ΑΒ ΣΑΛΑΜΙ ΑΕΡΟΣ ΠΙΚΑΝΤΙΚΟ	ΑΛΛΑΝΤΙΚΑ	ΣΑΛΑΜΙΑ ΑΕΡΟΣ
34	10	40714848800	ΣΑΛΑΜΙ ΑΕΡΟΣ ΜΕ ΠΑΡΜΕΖΑΝΑ	ΑΛΛΑΝΤΙΚΑ	ΣΑΛΑΜΙΑ ΑΕΡΟΣ
35	10	40733666600	CAJOLINE ΣΥΜΠ/ΝΟ ELIXIR 64ΜΕΖ(40%)	ΑΠΟΡΡΥΠΑΝΤΙΚΑ - ΚΑΘΑΡΙΣΤΙΚΑ	ΑΠΟΡΡΥΠΑΝΤΙΚΑ ΡΟΥΧΩΝ

Showing 1 to 11 of 477,310 entries

Πίνακας 6 Περιγραφή Δεδομένων

Προτού χρησιμοποιήσουμε τον αλγόριθμο εξόρυξης κανόνων, πρέπει να μετατρέψουμε τα δεδομένα μας, σε τέτοια μορφή, ώστε να μπορεί να τα «διαβάσει» ο αλγόριθμος. Πιο συγκεκριμένα, μετατρέπουμε τα δεδομένα από μορφή πίνακα σε λίστες συναλλαγών έτσι ώστε να έχουμε όλες τις κατηγορίες των προϊόντων που αγοράζονται σε κάθε συναλλαγή σε μία μόνο εγγραφή. Η μορφή που χρειαζόμαστε παρουσιάζεται στην Εικόνα 5 όπου διακρίνονται ενδεικτικά οι 15 από τις 72.543 συναλλαγές.

Categories	
1	ΓΑΛΑ Μ.ΔΙΑΡΚΕΙΑΣ,ΔΗΜΗΤΡΙΑΚΑ-MUSLI,ΝΩΠΑ ΠΟΥΛΕΡΙΚ...
2	ΣΑΛΑΜΙΑ ΑΕΡΟΣ,ΣΑΛΑΜΙΑ ΑΕΡΟΣ,ΑΠΟΡΡΥΠΑΝΤΙΚΑ ΡΟΥΧ...
3	ΑΡΤΟΕΙΔΗ,ΑΠΟΡΡΥΠΑΝΤΙΚΑ ΠΙΑΤΩΝ,ΑΡΤΥΜΑΤΑ,ΑΡΤΥΜΑ...
4	ΑΝΑΨΥΚΤΙΚΑ,ΑΠΟΡΡΥΠΑΝΤΙΚΑ ΠΙΑΤΩΝ,ΑΡΤΟΕΙΔΗ,ΓΙΑΟΥ...
5	ΦΡΕΣΚΑ & ΑΝΑΚΑΤΕΡΓΑΣΜΕΝΑ ΤΥΡΙΑ ΕΙΣΑΓΩΓΗΣ,ΒΟΥΤΥΡ...
6	ΑΝΑΨΥΚΤΙΚΑ,ΦΡΕΣΚΙΑ MOZZARELLA,ΦΡΕΣΚΟΙ ΧΥΜΟΙ,ΑΝΑ...
7	ΤΟΥΡΤΕΣ-ΜΙΓΜΑΤΑ-ΓΛΥΚΑ,ΤΟΜΑΤΟΕΙΔΗ,ΑΡΤΟΕΙΔΗ,ΑΡΤΟΕ...
8	ΣΝΑΚΣ,ΓΑΛΑ Μ.ΔΙΑΡΚΕΙΑΣ,ΓΑΛΑ Μ.ΔΙΑΡΚΕΙΑΣ
9	ΑΠΟΡΡΥΠΑΝΤΙΚΑ ΠΙΑΤΩΝ,ΚΡΕΜΕΣ ΓΑΛ.ΣΑΝΤΥΓΙ,ΖΥΜΑΡΙΚ...
10	SWIFFER,ΣΑΛΑΜΙΑ ΑΕΡΟΣ,ΑΛΛΑΝΤΙΚΑ ΠΟΥΛΕΡΙΚΩΝ,ΝΩΠ...
11	ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΑ,ΦΡΕΣΚΟ & ΥΨΗΛ...
12	ΡΥΖΙΑ,ΝΩΠΑ ΠΟΥΛΕΡΙΚΑ,ΤΟΥΡΤΕΣ-ΜΙΓΜΑΤΑ-ΓΛΥΚΑ,ΜΠΙΣ...
13	ΦΡΟΝΤΙΔΑ ΜΑΛΛΙΩΝ,ΦΡΟΝΤΙΔΑ ΜΑΛΛΙΩΝ,ΑΡΤΟΕΙΔΗ,ΦΡ...
14	ΨΑΡΙΚΑ ΚΟΝΣΕΡΒΑ,ΑΠΟΡΡΥΠΑΝΤΙΚΑ ΡΟΥΧΩΝ,ΦΡΕΣΚΟ & ...
15	ΑΠΟΡΡΥΠΑΝΤΙΚΑ ΡΟΥΧΩΝ,ΣΝΑΚΣ,ΣΝΑΚΣ,ΑΠΟΡΡΥΠΑΝΤΙΚ...

Showing 1 to 15 of 72,543 entries

Εικόνα 5 Υποσύνολο Αρχικού Συνόλου Δεδομένων

Οι κατηγορίες των προϊόντων χωρίζονται με “,” για να γίνονται αντιληπτές ως διαφορετικές κατηγορίες από τον αλγόριθμο και κάθε γραμμή βασίζεται στον μοναδικό αριθμό συναλλαγής. Το μόνο που χρειαζόμαστε είναι οι κατηγορίες των προϊόντων στις συναλλαγές, επομένως δεν μας ενδιαφέρει καμία άλλη στήλη (γνώρισμα) του συνόλου δεδομένων. Αξίζει να σημειωθεί ότι ο αλγόριθμος Apriori, «διαβάζει» την ύπαρξη ή μη ύπαρξη της κατηγορίας σε μια συναλλαγή. Επομένως εάν σε μια συναλλαγή υπάρχουν δύο προϊόντα που ανήκουν στη ίδια κατηγορία, ο αλγόριθμος θα μετρήσει την ύπαρξη της κατηγορίας μία φορά. Τέλος, αντιγράφουμε τις συναλλαγές μας σε ένα αρχείο csv και ελέγχουμε εάν η μετατροπή των δεδομένων έγινε σωστά. Παρουσιάζονται ενδεικτικά οι 21 εγγραφές του αρχείου CSV (Εικόνα 6) και με τη βοήθεια της συνάρτησης «summary», παίρνουμε κάποια βασικά στοιχεία του αρχείου που δημιουργήσαμε (Εικόνα 7)

1	Categories
2	ΓΑΛΛΑ Μ.ΔΙΑΡΚΕΙΑΣ,ΔΗΜΗΤΡΙΑΚΑ-MUSLI,ΝΩΠΑ ΠΟΥΛΕΡΙΚΑ,ΖΥΜΑΡΙΚΑ,ΤΟΜΑΤΟΕΙΔΗ
3	ΣΑΛΑΜΙΑ ΑΕΡΟΣ,ΣΑΛΑΜΙΑ ΑΕΡΟΣ,ΑΠΟΡΡΥΠΑΝΤΙΚΑ ΡΟΥΧΩΝ,ΖΑΧΑΡΗ,ΑΡΤΥΜΑΤΑ,Ζ
4	ΑΡΤΟΕΙΔΗ,ΑΠΟΡΡΥΠΑΝΤΙΚΑ ΠΙΑΤΩΝ,ΑΡΤΥΜΑΤΑ,ΑΡΤΥΜΑΤΑ,ΑΡΤΥΜΑΤΑ,ΤΥΡΙΑ ΕΙΣΑΓ
5	ΑΝΑΨΥΚΤΙΚΑ,ΑΠΟΡΡΥΠΑΝΤΙΚΑ ΠΙΑΤΩΝ,ΑΡΤΟΕΙΔΗ,ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ,ΡΟΦΗΜΑ
6	ΦΡΕΣΚΑ & ΑΝΑΚΑΤΕΡΓΑΣΜΕΝΑ ΤΥΡΙΑ ΕΙΣΑΓΩΓΗΣ,ΒΟΥΤΥΡΑ,ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣ
7	ΑΝΑΨΥΚΤΙΚΑ,ΦΡΕΣΚΙΑ MOZZARELLA,ΦΡΕΣΚΟΙ ΧΥΜΟΙ,ΑΝΑΨΥΚΤΙΚΑ
8	ΤΟΥΡΤΕΣ-ΜΙΓΜΑΤΑ-ΓΛΥΚΑ,ΤΟΜΑΤΟΕΙΔΗ,ΑΡΤΟΕΙΔΗ,ΑΡΤΟΕΙΔΗ,ΠΟΥΡΕΔΕΣ-ΚΥΒΟΙ-ΜΕ
9	ΣΝΑΚΣ,ΓΑΛΛΑ Μ.ΔΙΑΡΚΕΙΑΣ,ΓΑΛΛΑ Μ.ΔΙΑΡΚΕΙΑΣ
10	ΑΠΟΡΡΥΠΑΝΤΙΚΑ ΠΙΑΤΩΝ,ΚΡΕΜΕΣ ΓΑΛ.ΣΑΝΤΥΓΙ,ΖΥΜΑΡΙΚΑ,ΑΠΟΡΡΥΠΑΝΤΙΚΑ ΡΟΥΧ
11	SWIFFER,ΣΑΛΑΜΙΑ ΑΕΡΟΣ,ΑΛΛΑΝΤΙΚΑ ΠΟΥΛΕΡΙΚΩΝ,ΝΩΠΑ ΚΡΕΑΤΑ,ΝΩΠΑ ΠΟΥΛΕΡ
12	ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ,ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ
13	ΡΥΖΙΑ,ΝΩΠΑ ΠΟΥΛΕΡΙΚΑ,ΤΟΥΡΤΕΣ-ΜΙΓΜΑΤΑ-ΓΛΥΚΑ,ΜΠΙΣΚΟΤΑ,ΦΕΤΕΣ,ΚΑΘΑΡΙΣΤΙΚΑ
14	ΦΡΟΝΤΙΔΑ ΜΑΛΛΙΩΝ,ΦΡΟΝΤΙΔΑ ΜΑΛΛΙΩΝ,ΑΡΤΟΕΙΔΗ,ΦΡΟΝΤΙΔΑ ΣΩΜΑΤΟΣ
15	ΨΑΡΙΚΑ ΚΟΝΣΕΡΒΑ,ΑΠΟΡΡΥΠΑΝΤΙΚΑ ΡΟΥΧΩΝ,ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑ
16	ΑΠΟΡΡΥΠΑΝΤΙΚΑ ΡΟΥΧΩΝ,ΣΝΑΚΣ,ΣΝΑΚΣ,ΑΠΟΡΡΥΠΑΝΤΙΚΑ ΡΟΥΧΩΝ,ΑΛΛΑΝΤΙΚΑ ΠΙ
17	ΜΠΙΣΚΟΤΑ,ΚΡΕΜΕΣ ΓΑΛ.ΣΑΝΤΥΓΙ,ΗΜΙΣΚΛΗΡΑ ΕΓΧΩΡΙΑ ΤΥΡΙΑ,ΑΛΛΑΝΤΙΚΑ ΒΡΑΣΤΑ,Ζ
18	ΑΥΓΑ,ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ,ΝΩΠΑ ΚΡΕΑΤΑ,ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ
19	ΜΑΡΓΑΡΙΝΕΣ,ΧΥΜΟΙ,ΠΡΟΣΩΠΙΚΗ ΥΠΕΙΝΗ,ΦΡΟΝΤΙΔΑ ΣΩΜΑΤΟΣ,ΑΠΟΡΡΥΠΑΝΤΙΚΑ Ρ
20	ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ,ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ,ΝΕΙ
21	ΕΝΤΟΜΟΚΤΟΝΑ - ΕΝΤΟΜΟΑΠΩΘΗΤΙΚΑ

Εικόνα 6 Αρχείο CSV

transactions as itemMatrix in sparse format with
72543 rows (elements/itemsets/transactions) and
83 columns (items) and a density of 0.06376941

most frequent items:

ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ	24502	ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ	16922
ΡΟΦΗΜΑΤΑ	16767	ΑΝΑΨΥΚΤΙΚΑ	13577
ΑΡΤΥΜΑΤΑ	13520	(Other)	298672

element (itemset/transaction) length distribution:

sizes	1	2	3	4	5	6	7	8	9	10	11	12
	13	14	15	16	17	18	19	20	21	22	23	24
11177	11303	9985	8174	6509	5361	4087	3268	2517	1950	1558	1173	
	995	799	649	559	447	377	298	272	208	201	170	102
	25	26	27	28	29	30	31	32	33	34	35	36
	37	38	39	43	44	46						
	105	77	57	46	25	17	24	11	17	8	6	3
	1	2	2	1	1	1						
Min.	1.000	2.000	4.000	5.293	7.000	46.000						

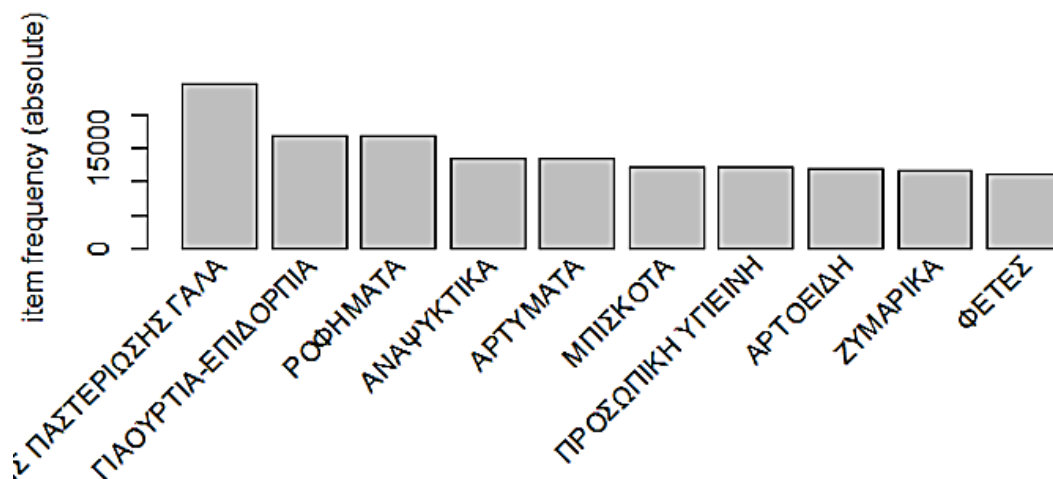
includes extended item information - examples:

	Labels
1	ΕΔΑΜ ΜΠΑΛΑΚΙ ΕΙΣΑΓΩΓΗΣ
2	ΕΜΜΕΝΤΑΛ ΕΙΣΑΓΩΓΗΣ
3	ΕΤΗΝΙΚ FOOD

Εικόνα 7 Σύνοψη Αρχείου CSV

Όπως αναμέναμε, οι συναλλαγές μας είναι 72.543 και ο αριθμός των κατηγοριών είναι 83 στο σύνολο. Η κατηγορία με τη μεγαλύτερη συχνότητα εμφάνισης στις συναλλαγές είναι η «ΦΡΕΣΚΟ ΚΑΙ ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΑ» καθώς υπάρχει σε 24.502 συναλλαγές. Το σύνολο των δεδομένων μας δεν είναι πολύ αραιό, με πυκνότητα μόλις πάνω από 6%. Η πυκνότητα είναι ο λόγος του πλήθους των κατηγοριών που αγοράστηκαν στις συναλλαγές προς το πλήθος των πιθανών κατηγοριών που θα μπορούσαν να είχαν αγοραστεί σε όλες τις συναλλαγές. Κοιτάζοντας το μέγεθος των συναλλαγών, παρατηρούμε ότι 11.177 συναλλαγές περιλάμβαναν μόνο μια κατηγορία, 11.303 συναλλαγές περιλάμβαναν δύο κατηγορίες, 9.985 συναλλαγές περιλάμβαναν τρεις κατηγορίες κοκ.

Τέλος, η κατανομή των δεδομένων είναι λοξή προς τα δεξιά, καθώς ο μέσος όρος είναι μεγαλύτερος της διαμέσου. Μετά την ολοκλήρωση της ανωτέρω διαδικασίας είμαστε σε θέση να δημιουργήσουμε ένα γράφημα, που θα αναπαριστά τις δέκα κατηγορίες προϊόντων με τη μεγαλύτερη συχνότητα εμφάνισης στα καλάθια των καταναλωτών. (Εικόνα 8)



Εικόνα 8 Οι 10 Κατηγορίες με τη Μεγαλύτερη Συχνότητα Εμφάνισης

Έχοντας φέρει τα δεδομένα μας σε τέτοια μορφή ώστε να εισαχθούν στον αλγόριθμο και έχοντας περιγράψει κάποια βασικά χαρακτηριστικά τους, στην επόμενη ενότητα, ορίζουμε ένα κατώτατο όριο υποστήριξης, για να επιστραφούν οι κανόνες συσχέτισης που ικανοποιούν αυτό το όριο.

3.1.3 Εξόρυξη Κανόνων Συσχέτισης Συχνών Στοιχειοσυνόλων σε Επίπεδο Κατηγοριών

Ο αλγόριθμος Apriori παρέχεται μέσω της βιβλιοθήκης «Arules» της R, η οποία παρέχει έτοιμες συναρτήσεις για την εξόρυξη συχνών στοιχειοσυνόλων και κανόνων συσχέτισης. Όπως έχουμε προαναφέρει στο Κεφάλαιο 2, ο αλγόριθμος, χρησιμοποιεί level-wise αναζήτηση για

την εύρεση συχνών στοιχειοσυνόλων και εν συνέχεια χρησιμοποιεί αυτά τα συχνά στοιχειοσύνολα για να εξορύξει ισχυρούς κανόνες συσχέτισης. Ορίζουμε κατώφλι υποστήριξης ίσο με 0.05 και εμπιστοσύνη ίση με 0.05. Οι κανόνες συσχέτισης που θα επιστραφούν ζητάμε να είναι ταξινομημένοι κατά φθίνουσα σειρά με βάση την υποστήριξη.

Ο αλγόριθμος επιστρέφει 106 κανόνες εκ των οποίων μόνο οι 70 είναι αξιολογήσιμοι, γιατί αποτελούνται από 2-στοιχειοσύνολα και από τους 70 οι 35 είναι αντίστροφης κατεύθυνσης.

Με τη βοήθεια της συνάρτησης «summary» της R, λαμβάνουμε μία σύνοψη των κανόνων συσχέτισης και των μετρικών της υποστήριξης, της εμπιστοσύνης και του lift (Εικόνα 9).

Κατόπιν αυτού παρουσιάζονται οι 106 κανόνες συσχέτισης που εξόρυξε ο αλγόριθμος (Εικόνα 10 και Εικόνα 11).

```
rule length distribution (lhs + rhs):sizes
```

```
1 2
36 70
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.00	1.00	2.00	1.66	2.00	2.00

```
summary of quality measures:
```

support		confidence		lift		count	
Min.	:0.05025	Min.	:0.05638	Min.	:1.000	Min.	: 3645
1st Qu.	:0.05601	1st Qu.	:0.15173	1st Qu.	:1.000	1st Qu.	: 4063
Median	:0.06629	Median	:0.23154	Median	:1.270	Median	: 4809
Mean	:0.08315	Mean	:0.25175	Mean	:1.352	Mean	: 6032
3rd Qu.	:0.09149	3rd Qu.	:0.35261	3rd Qu.	:1.461	3rd Qu.	: 6637
Max.	:0.33776	Max.	:0.55185	Max.	:3.344	Max.	:24502

```
mining info:
```

data	ntransactions	support	confidence
tr2	72543	0.05	0

Εικόνα 9 Σύνοψη Κανόνων

	lhs	rhs	support	confidence	lift	count
[1]	{}	=> {ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ}	0.33775830	0.33775830	1.000000	24502
[2]	{}	=> {ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ}	0.23326854	0.23326854	1.000000	16922
[3]	{}	=> {ΡΟΦΗΜΑΤΑ}	0.23113188	0.23113188	1.000000	16767
[4]	{}	=> {ΑΝΑΨΥΚΤΙΚΑ}	0.18715796	0.18715796	1.000000	13577
[5]	{}	=> {ΑΡΤΥΜΑΤΑ}	0.18637222	0.18637222	1.000000	13520
[6]	{}	=> {ΜΠΙΣΚΟΤΑ}	0.16930648	0.16930648	1.000000	12282
[7]	{}	=> {ΠΡΟΣΩΠΙΚΗ ΥΓΙΕΙΝΗ}	0.16689412	0.16689412	1.000000	12107
[8]	{}	=> {ΑΡΤΟΕΙΔΗ}	0.16416470	0.16416470	1.000000	11909
[9]	{}	=> {ΖΥΜΑΡΙΚΑ}	0.16002922	0.16002922	1.000000	11609
[10]	{}	=> {ΦΕΤΕΣ}	0.15207532	0.15207532	1.000000	11032
[11]	{}	=> {ΑΛΛΑΝΤΙΚΑ ΠΟΥΛΕΡΙΚΩΝ}	0.13768386	0.13768386	1.000000	9988
[12]	{}	=> {ΔΗΜΗΤΡΙΑΚΑ-MUSLI}	0.12946804	0.12946804	1.000000	9392
[13]	{}	=> {ΤΥΡΙΑ ΕΙΣΑΓΩΓΗΣ ΓΙΑ ΤΟΣΤ LOAF & BLOCKS}	0.12555312	0.12555312	1.000000	9108
[14]	{}	=> {ΤΟΜΑΤΟΕΙΔΗ}	0.12446411	0.12446411	1.000000	9029
[15]	{}	=> {ΚΑΘΑΡΙΣΤΙΚΑ ΣΠΙΤΙΟΥ}	0.12275478	0.12275478	1.000000	8905
[16]	{}	=> {ΑΠΟΡΡΥΠΑΝΤΙΚΑ ΡΟΥΧΩΝ}	0.12061812	0.12061812	1.000000	8750
[17]	{}	=> {ΑΛΛΑΝΤΙΚΑ ΒΡΑΣΤΑ}	0.11735109	0.11735109	1.000000	8513
[18]	{ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ}	=> {ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ}	0.11582096	0.49651341	1.470026	8402
[19]	{ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ}	=> {ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ}	0.11582096	0.34291078	1.470026	8402
[20]	{}	=> {ΦΡΟΝΤΙΔΑ ΣΩΜΑΤΟΣ}	0.10987966	0.10987966	1.000000	7971
[21]	{}	=> {ΝΟΠΑ ΚΡΕΑΤΑ}	0.10715024	0.10715024	1.000000	7773
[22]	{}	=> {ΓΑΛΛΑ Μ.ΔΙΑΡΚΕΙΑΣ}	0.10505493	0.10505493	1.000000	7621
[23]	{}	=> {ΣΝΑΚΣ}	0.09952718	0.09952718	1.000000	7220
[24]	{}	=> {ΝΟΠΑ ΠΟΥΛΕΡΙΚΑ}	0.09776271	0.09776271	1.000000	7092
[25]	{}	=> {ΤΟΥΡΤΕΣ-ΜΙΓΜΑΤΑ-ΓΛΥΚΑ}	0.09358587	0.09358587	1.000000	6789
[26]	{}	=> {ΖΑΧΑΡΗ}	0.09291041	0.09291041	1.000000	6740
[27]	{ΡΟΦΗΜΑΤΑ}	=> {ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ}	0.09149056	0.39583706	1.171954	6637
[28]	{ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ}	=> {ΡΟΦΗΜΑΤΑ}	0.09149056	0.27087585	1.171954	6637
[29]	{}	=> {ΦΡΕΣΚΑ & ΑΝΑΚΑΤΕΡΓΑΣΜΕΝΑ ΤΥΡΙΑ ΕΙΣΑΓΩΓΗΣ}	0.08195139	0.08195139	1.000000	5945
[30]	{}	=> {ΝΕΡΑ}	0.07944254	0.07944254	1.000000	5763
[31]	{}	=> {ΜΑΡΓΑΡΙΝΕΣ}	0.07871194	0.07871194	1.000000	5710
[32]	{ΑΡΤΥΜΑΤΑ}	=> {ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ}	0.07718181	0.41412722	1.226105	5599
[33]	{ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ}	=> {ΑΡΤΥΜΑΤΑ}	0.07718181	0.22851196	1.226105	5599
[34]	{}	=> {ΑΥΓΑ}	0.07686476	0.07686476	1.000000	5576
[35]	{ΑΝΑΨΥΚΤΙΚΑ}	=> {ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ}	0.07505893	0.40104589	1.187375	5445
[36]	{ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ}	=> {ΑΝΑΨΥΚΤΙΚΑ}	0.07505893	0.22222676	1.187375	5445
[37]	{}	=> {ΦΡΟΝΤΙΔΑ ΜΑΛΛΙΩΝ}	0.07434211	0.07434211	1.000000	5393
[38]	{ΑΡΤΥΜΑΤΑ}	=> {ΡΟΦΗΜΑΤΑ}	0.07249493	0.38897929	1.682932	5259
[39]	{ΡΟΦΗΜΑΤΑ}	=> {ΑΡΤΥΜΑΤΑ}	0.07249493	0.31365182	1.682932	5259
[40]	{ΜΠΙΣΚΟΤΑ}	=> {ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ}	0.07161270	0.42297671	1.252306	5195
[41]	{ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ}	=> {ΜΠΙΣΚΟΤΑ}	0.07161270	0.21202351	1.252306	5195
[42]	{ΔΗΜΗΤΡΙΑΚΑ-MUSLI}	=> {ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ}	0.07144728	0.55185264	1.633869	5183
[43]	{ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ}	=> {ΔΗΜΗΤΡΙΑΚΑ-MUSLI}	0.07144728	0.21153375	1.633869	5183
[44]	{}	=> {ΑΛΕΥΡΙ - ΣΙΜΙΓΔΑΛΙ}	0.07017907	0.07017907	1.000000	5091
[45]	{ΑΡΤΟΕΙΔΗ}	=> {ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ}	0.06987580	0.42564447	1.260204	5069
[46]	{ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ}	=> {ΑΡΤΟΕΙΔΗ}	0.06987580	0.20688107	1.260204	5069
[47]	{ΖΥΜΑΡΙΚΑ}	=> {ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ}	0.06910384	0.43182014	1.278489	5013
[48]	{ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ}	=> {ΖΥΜΑΡΙΚΑ}	0.06910384	0.20459554	1.278489	5013
[49]	{ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ}	=> {ΡΟΦΗΜΑΤΑ}	0.06845595	0.29346413	1.269683	4966
[50]	{ΡΟΦΗΜΑΤΑ}	=> {ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ}	0.06845595	0.29617701	1.269683	4966
[51]	{ΑΛΛΑΝΤΙΚΑ ΠΟΥΛΕΡΙΚΩΝ}	=> {ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ}	0.06793212	0.49339207	1.460784	4928
[52]	{ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ}	=> {ΑΛΛΑΝΤΙΚΑ ΠΟΥΛΕΡΙΚΩΝ}	0.06793212	0.20112644	1.460784	4928
[53]	{ΦΕΤΕΣ}	=> {ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ}	0.06629172	0.43591371	1.290608	4809
[54]	{ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ}	=> {ΦΕΤΕΣ}	0.06629172	0.19626969	1.290608	4809
[55]	{}	=> {ΡΥΖΙΑ}	0.06607116	0.06607116	1.000000	4793
[56]	{}	=> {ΑΠΟΡΡΥΠΑΝΤΙΚΑ ΠΙΑΤΩΝ}	0.06575410	0.06575410	1.000000	4770
[57]	{ΠΡΟΣΩΠΙΚΗ ΥΓΙΕΙΝΗ}	=> {ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ}	0.06551976	0.39258280	1.162319	4753
[58]	{ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ}	=> {ΠΡΟΣΩΠΙΚΗ ΥΓΙΕΙΝΗ}	0.06551976	0.19398416	1.162319	4753
[59]	{}	=> {ΣΤΟΜΑΤΙΚΗ ΥΓΙΕΙΝΗ}	0.06466510	0.06466510	1.000000	4691
[60]	{}	=> {ΧΥΜΟΙ}	0.06379664	0.06379664	1.000000	4628
[61]	{ΖΥΜΑΡΙΚΑ}	=> {ΡΟΦΗΜΑΤΑ}	0.06221138	0.38875011	1.681941	4513
[62]	{ΡΟΦΗΜΑΤΑ}	=> {ΖΥΜΑΡΙΚΑ}	0.06221138	0.26915966	1.681941	4513
[63]	{ΤΥΡΙΑ ΕΙΣΑΓΩΓΗΣ ΓΙΑ ΤΟΣΤ LOAF & BLOCKS}	=> {ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ}	0.06126022	0.48792271	1.444591	4444
[64]	{ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ}	=> {ΤΥΡΙΑ ΕΙΣΑΓΩΓΗΣ ΓΙΑ ΤΟΣΤ LOAF & BLOCKS}	0.06126022	0.18137295	1.444591	4444
[65]	{ΖΥΜΑΡΙΚΑ}	=> {ΑΡΤΥΜΑΤΑ}	0.06025392	0.37651822	2.020249	4371
[66]	{ΑΡΤΥΜΑΤΑ}	=> {ΖΥΜΑΡΙΚΑ}	0.06025392	0.32329882	2.020249	4371
[67]	{ΑΡΤΥΜΑΤΑ}	=> {ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ}	0.06012985	0.32263314	1.383097	4362
[68]	{ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ}	=> {ΑΡΤΥΜΑΤΑ}	0.06012985	0.25777095	1.383097	4362
[69]	{ΠΡΟΣΩΠΙΚΗ ΥΓΙΕΙΝΗ}	=> {ΡΟΦΗΜΑΤΑ}	0.05956467	0.35690097	1.544144	4321
[70]	{ΡΟΦΗΜΑΤΑ}	=> {ΠΡΟΣΩΠΙΚΗ ΥΓΙΕΙΝΗ}	0.05956467	0.25770859	1.544144	4321

Εικόνα 10 Κανόνες Συσχέτισης

[71]	{ΑΡΤΟΕΙΔΗ}	=>	{ΡΟΦΗΜΑΤΑ}	0.05886164	0.35855236	1.551289	4270
[72]	{ΡΟΦΗΜΑΤΑ}	=>	{ΑΡΤΟΕΙΔΗ}	0.05886164	0.25466691	1.551289	4270
[73]	{}	=>	{ΣΚΛΗΡΑ ΤΥΡΙΑ ΕΙΣΑΓΩΓΗΣ}	0.05833781	0.05833781	1.000000	4232
[74]	{ΤΥΡΙΑ ΕΙΣΑΓΩΓΗΣ ΓΙΑ ΤΟΣΤ LOAF & BLOCKS}	=>	{ΑΛΛΑΝΤΙΚΑ ΠΟΥΛΕΡΙΚΩΝ}	0.05781399	0.46047431	3.344432	4194
[75]	{ΑΛΛΑΝΤΙΚΑ ΠΟΥΛΕΡΙΚΩΝ}	=>	{ΤΥΡΙΑ ΕΙΣΑΓΩΓΗΣ ΓΙΑ ΤΟΣΤ LOAF & BLOCKS}	0.05781399	0.41990388	3.344432	4194
[76]	{ΤΟΜΑΤΟΕΙΔΗ}	=>	{ΖΥΜΑΡΙΚΑ}	0.05694554	0.45752575	2.859014	4131
[77]	{ΖΥΜΑΡΙΚΑ}	=>	{ΤΟΜΑΤΟΕΙΔΗ}	0.05694554	0.35584460	2.859014	4131
[78]	{}	=>	{ΔΙΑΦΟΡΑ ΕΓΧΩΡΙΑ ΣΚΛΗΡΑ ΤΥΡΙΑ}	0.05638035	0.05638035	1.000000	4090
[79]	{ΑΝΑΨΥΚΤΙΚΑ}	=>	{ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ}	0.05600816	0.29925609	1.282882	4063
[80]	{ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ}	=>	{ΑΝΑΨΥΚΤΙΚΑ}	0.05600816	0.24010164	1.282882	4063
[81]	{ΤΟΜΑΤΟΕΙΔΗ}	=>	{ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΑ}	0.05524999	0.44390298	1.314262	4008
[82]	{ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΑ}	=>	{ΤΟΜΑΤΟΕΙΔΗ}	0.05524999	0.16357848	1.314262	4008
[83]	{ΜΠΙΣΚΟΤΑ}	=>	{ΡΟΦΗΜΑΤΑ}	0.05407827	0.31941052	1.381941	3923
[84]	{ΡΟΦΗΜΑΤΑ}	=>	{ΜΠΙΣΚΟΤΑ}	0.05407827	0.23397149	1.381941	3923
[85]	{ΑΝΑΨΥΚΤΙΚΑ}	=>	{ΡΟΦΗΜΑΤΑ}	0.05391285	0.28806069	1.246304	3911
[86]	{ΡΟΦΗΜΑΤΑ}	=>	{ΑΝΑΨΥΚΤΙΚΑ}	0.05391285	0.23325580	1.246304	3911
[87]	{ΜΠΙΣΚΟΤΑ}	=>	{ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ}	0.05378879	0.31770070	1.361953	3902
[88]	{ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ}	=>	{ΜΠΙΣΚΟΤΑ}	0.05378879	0.23058740	1.361953	3902
[89]	{ΤΟΜΑΤΟΕΙΔΗ}	=>	{ΡΟΦΗΜΑΤΑ}	0.05360958	0.43072323	1.863539	3889
[90]	{ΡΟΦΗΜΑΤΑ}	=>	{ΤΟΜΑΤΟΕΙΔΗ}	0.05360958	0.23194370	1.863539	3889
[91]	{ΑΛΛΑΝΤΙΚΑ ΠΟΥΛΕΡΙΚΩΝ}	=>	{ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ}	0.05316847	0.38616340	1.655446	3857
[92]	{ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ}	=>	{ΑΛΛΑΝΤΙΚΑ ΠΟΥΛΕΡΙΚΩΝ}	0.05316847	0.22792814	1.655446	3857
[93]	{ΑΛΛΑΝΤΙΚΑ ΒΡΑΣΤΑ}	=>	{ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΑ}	0.05311333	0.45260190	1.340017	3853
[94]	{ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΑ}	=>	{ΑΛΛΑΝΤΙΚΑ ΒΡΑΣΤΑ}	0.05311333	0.15725247	1.340017	3853
[95]	{ΑΡΤΟΕΙΔΗ}	=>	{ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ}	0.05294791	0.32252918	1.382652	3841
[96]	{ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ}	=>	{ΑΡΤΟΕΙΔΗ}	0.05294791	0.22698263	1.382652	3841
[97]	{ΤΟΜΑΤΟΕΙΔΗ}	=>	{ΑΡΤΥΜΑΤΑ}	0.05243786	0.42130912	2.260579	3804
[98]	{ΑΡΤΥΜΑΤΑ}	=>	{ΤΟΜΑΤΟΕΙΔΗ}	0.05243786	0.28136095	2.260579	3804
[99]	{ΦΕΤΕΣ}	=>	{ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ}	0.05183133	0.34082669	1.461091	3760
[100]	{ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ}	=>	{ΦΕΤΕΣ}	0.05183133	0.22219596	1.461091	3760
[101]	{ΖΥΜΑΡΙΚΑ}	=>	{ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ}	0.05180376	0.32371436	1.387733	3758
[102]	{ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ}	=>	{ΖΥΜΑΡΙΚΑ}	0.05180376	0.22207777	1.387733	3758
[103]	{ΚΑΘΑΡΙΣΤΙΚΑ ΣΠΙΤΙΟΥ}	=>	{ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΑ}	0.05121101	0.41718136	1.235148	3715
[104]	{ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΑ}	=>	{ΚΑΘΑΡΙΣΤΙΚΑ ΣΠΙΤΙΟΥ}	0.05121101	0.15162028	1.235148	3715
[105]	{ΠΡΟΣΩΠΙΚΗ ΥΓΙΕΙΝΗ}	=>	{ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ}	0.05024606	0.30106550	1.290639	3645
[106]	{ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ}	=>	{ΠΡΟΣΩΠΙΚΗ ΥΓΙΕΙΝΗ}	0.05024606	0.21540007	1.290639	3645

Εικόνα 11 Συνέχεια Εικόνας 10

Η υποστήριξη παίρνει τιμές στο διάστημα [0,0502,0,3377], η εμπιστοσύνη παίρνει τιμές στο διάστημα [0,056,0,5518], το μέτρο Lift παίρνει τιμές στο διάστημα [1,00,3,34] και η συχνότητα εμφάνισης των κατηγοριών στις συναλλαγές κυμαίνεται από [3645,24502].

Τη μεγαλύτερη τιμή υποστήριξης την έχει η κατηγορία «ΦΡΕΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΑ», καθώς εμφανίζεται σε 24.502 συναλλαγές με ποσοστό 33.7% των καταναλωτών του καταστήματος που προτιμούν προϊόντα που ανήκουν σε αυτή τη κατηγορία.

Η μεγαλύτερη τιμή της εμπιστοσύνης είναι 55% και παρατηρείται στον κανόνα {ΔΗΜΗΤΡΙΑΚΑ – MUSLI} → {ΦΡΕΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΑ}. Η ερμηνεία έγκειται στο ότι το 55% των φορών που οι καταναλωτές θα αγοράσουν προϊόντα της κατηγορίας «ΔΗΜΗΤΡΙΑΚΑ-MUSLI», θα αγοράσουν και προϊόντα της κατηγορίας «ΦΡΕΚΟ ΚΑΙ ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΑ».

Ο κανόνας που έχει το υψηλότερο lift=3,34 είναι ο {ΤΥΡΙΑ ΕΙΣΑΓΩΓΗΣ ΓΙΑ ΤΟΣΤ LOAF & BLOCKS} → {ΑΛΛΑΝΤΙΚΑ ΠΟΥΛΕΡΙΚΩΝ}. Αυτό σημαίνει ότι ενώ η κατηγορία {ΑΛΛΑΝΤΙΚΑ ΠΟΥΛΕΡΙΚΩΝ} εμφανίζεται σε 9.988 συναλλαγές από τις 72.543, δηλαδή στο 13,76% περίπου των συνολικών συναλλαγών, στον υποπληθυσμό των καταναλωτών που αγοράζουν προϊόντα της κατηγορίας «ΤΥΡΙΑ

ΕΙΣΑΓΩΓΗΣ ΓΙΑ ΤΟΣΤ LOAF & BLOCKS», το ποσοστό των καταναλωτών που αγοράζουν προϊόντα της κατηγορίας «ΑΛΛΑΝΤΙΚΑ ΠΟΥΛΕΡΙΚΩΝ» είναι 3,34 φορές μεγαλύτερο από ότι είναι στον συνολικό πληθυσμό (13,76%).

Τέλος, τη μικρότερη συχνότητα εμφάνισης την έχει ο κανόνας 106, {ΓΙΑΟΥΡΤΙΑ – ΕΠΙΔΟΡΠΙΑ} → {ΠΡΟΣΩΠΙΚΗ ΥΓΙΕΙΝΗ}, καθώς εμφανίζεται σε 3.645 συναλλαγές.

Όπως παρατηρούμε από τα αποτελέσματα του αλγορίθμου, η κατηγορία «ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΑ» συσχετίζεται με τις κατηγορίες: «ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ», «ΑΡΤΥΜΑΤΑ», «ΡΟΦΗΜΑΤΑ», «ΑΝΑΨΥΚΤΙΚΑ», «ΜΠΙΣΚΟΤΑ», «ΔΗΜΗΤΡΙΑΚΑ-MUSLI», «ΑΡΤΟΕΙΔΗ», «ΖΥΜΑΡΙΚΑ», «ΑΛΛΑΝΤΙΚΑ ΠΟΥΛΕΡΙΚΩΝ», «ΦΕΤΕΣ» «ΠΡΟΣΩΠΙΚΗ ΥΓΙΕΙΝΗ», «ΤΥΡΙΑ ΕΙΣΑΓΩΓΗΣ ΓΙΑ ΤΟΣΤ LOAF & BLOCKS», «ΤΟΜΑΤΟΕΙΔΗ», «ΑΛΛΑΝΤΙΚΑ ΒΡΑΣΤΑ», «ΚΑΘΑΡΙΣΤΙΚΑ ΣΠΙΤΙΟΥ».

Βλέπουμε ότι ο κανόνας 18 έχει υποστήριξη ίση με 11% που σημαίνει ότι στο 11% των συναλλαγών τα προϊόντα των κατηγοριών «ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ» και «ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΑ» αγοράζονται μαζί. Το 49% των φορών που οι καταναλωτές αγοράζουν προϊόντα της κατηγορίας «ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ» αγοράζουν και προϊόντα της κατηγορίας ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΑ». Αυτός ο κανόνας έχει αξία γιατί ενώ σε όλα τα καλάθια του καταστήματος τα προϊόντα της κατηγορίας «ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΑ» εμφανίζονται στο 33,7% των συνολικών συναλλαγών, στον υποπληθυσμό των καταναλωτών που αγοράζουν προϊόντα της κατηγορίας «ΓΙΑΟΥΡΤΙΑ - ΕΠΙΔΟΡΠΙΑ», το ποσοστό των καταναλωτών, που αγοράζουν προϊόντα της κατηγορίας «ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΑ» είναι 1,47 φορές μεγαλύτερο από ότι είναι στις συνολικές συναλλαγές.

Ο κανόνας που έχει τη μικρότερη υποστήριξη, περιέχοντας σε κάποιο μέλος του την κατηγορία με την υψηλότερη υποστήριξη, ίση με 5,12%, είναι ο κανόνας 104, {ΚΑΘΑΡΙΣΤΙΚΑ ΣΠΙΤΙΟΥ} → {ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΑ}. Στο 5,12% των συναλλαγών τα προϊόντα των δύο κατηγοριών βρίσκονται μαζί στα καλάθια των καταναλωτών. Σύμφωνα με την εμπιστοσύνη του κανόνα, το 41% των φορών που οι καταναλωτές αγοράζουν προϊόντα της κατηγορίας «ΚΑΘΑΡΙΣΤΙΚΑ ΣΠΙΤΙΟΥ», θα αγοράσουν και προϊόντα της κατηγορίας «ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΑ». Η τιμή του μέτρου lift είναι ίση με 1,23, που σημαίνει ότι η αγορά προϊόντων της κατηγορίας «ΚΑΘΑΡΙΣΤΙΚΑ ΣΠΙΤΙΟΥ», ενισχύει την αγορά προϊόντων της κατηγορίας «ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΑ». Πιο συγκεκριμένα, το ποσοστό των καταναλωτών που αγοράζουν προϊόντα της κατηγορίας «ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΑ» είναι 33.7%. Στον υποπληθυσμό των καταναλωτών που αγοράζουν προϊόντα της κατηγορίας

«ΚΑΘΑΡΙΣΤΙΚΑ ΣΠΙΤΙΟΥ», το ποσοστό των καταναλωτών που αγοράζουν προϊόντα της κατηγορίας «ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ» είναι 1,23 φορές μεγαλύτερο από ότι είναι στον συνολικό πληθυσμό.

Τη μικρότερη τιμή του μέτρου lift ίση με 1, έχουν 34 κατηγορίες, κάποιες από τις οποίες είναι ενδεικτικά οι : «ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ», «ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ», «ΡΟΦΗΜΑΤΑ, ΖΥΜΑΡΙΚΑ», «ΑΠΠΟΥΡΥΠΑΝΤΙΚΑ ΠΙΑΤΩΝ», «ΡΥΖΙΑ, ΑΥΓΑ», «ΔΙΑΦΟΡΑ ΕΓΧΩΡΙΑ ΣΚΛΗΡΑ ΤΥΡΙΑ» κ.α.

Εν συνέχεια θα αναφερθούμε στους κανόνες που προκύπτουν από μία ακόμα κατηγορία με υψηλή συχνότητα εμφάνισης ίση με 23.11%. Αυτή η κατηγορία είναι τα «ΡΟΦΗΜΑΤΑ» η οποία συσχετίζεται με τις κατηγορίες: «ΑΡΤΥΜΑΤΑ», «ΓΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΠΙΑ», «ΖΥΜΑΡΙΚΑ» , «ΠΡΟΣΩΠΙΚΗ ΥΓΙΕΙΝΗ», «ΑΡΤΟΕΙΔΗ», «ΜΠΙΣΚΟΤΑ», «ΤΟΜΑΤΟΕΙΔΗ» και «ΑΝΑΨΥΚΤΙΚΑ».

Ο κανόνας με τη μεγαλύτερη τιμή υποστήριξης είναι ο κανόνας 38, {ΑΡΤΥΜΑΤΑ} → {ΡΟΦΗΜΑΤΑ}, καθώς στο 7,29% των συναλλαγών αυτές οι δύο κατηγορίες προϊόντων υπάρχουν μαζί στις συναλλαγές. Το 38% των φορών που οι καταναλωτές αγοράζουν προϊόντα της κατηγορίας «ΑΡΤΥΜΑΤΑ», θα αγοράσουν και προϊόντα της κατηγορίας «ΡΟΦΗΜΑΤΑ».

Το μέτρο lift είναι ίσο με 1,68. Αυτό σημαίνει ότι παρόλο που η κατηγορία «ΡΟΦΗΜΑΤΑ» εμφανίζεται στο 23,11% των συνολικών συναλλαγών, στον υποπληθυσμό των καταναλωτών που αγοράζουν προϊόντα της κατηγορίας «ΑΡΤΥΜΑΤΑ», το ποσοστό των καταναλωτών που αγοράζουν προϊόντα από την κατηγορία «ΡΟΦΗΜΑΤΑ», είναι 1,68 φορές υψηλότερο από ότι είναι στο συνολικό πληθυσμό (23,11%).

Αντίθετα, ο κανόνας που έχει τη μικρότερη υποστήριξη ίση με 5,36% και περιλαμβάνει στο δεξί του μέλος τη κατηγορία «ΡΟΦΗΜΑΤΑ», είναι ο κανόνας 89, {ΤΟΜΑΤΟΕΙΔΗ} → {ΡΟΦΗΜΑΤΑ}. Το 43% των φορών που οι καταναλωτές προτιμούν τη κατηγορία «ΤΟΜΑΤΟΕΙΔΗ», προτιμούν και τη κατηγορία «ΡΟΦΗΜΑΤΑ», ενώ στον υποπληθυσμό των καταναλωτών που προτιμούν τη κατηγορία «ΤΟΜΑΤΟΕΙΔΗ», το ποσοστό των καταναλωτών που προτιμούν τη κατηγορία «ΡΟΦΗΜΑΤΑ» είναι 1,86 μεγαλύτερο από ότι είναι στον συνολικό πληθυσμό που προτιμά μόνο τη κατηγορία «ΡΟΦΗΜΑΤΑ» και ανέρχεται όπως είπαμε στο 23,11% των συνολικών συναλλαγών.

Ενδεικτικά μπορούμε να σχολιάσουμε ακόμα δύο κανόνες από τους 70, οι οποίοι έχουν υψηλό lift. Αυτοί οι κανόνες είναι ο κανόνας 76, {ΤΟΜΑΤΟΕΙΔΗ} → {ΖΥΜΑΡΙΚΑ} με υποστήριξη ίση με 5,69%, εμπιστοσύνη ίση με 45.75% και lift ίσο με 2,85. Η υποστήριξη της κατηγορίας «ΖΥΜΑΡΙΚΑ» είναι 16% καθώς εμφανίζεται σε 11.609 συναλλαγές από τις 72.543 που υπάρχουν συνολικά. Στον υποπληθυσμό των καταναλωτών που αγοράζουν προϊόντα της

κατηγορίας «TOMATOEIDH», το ποσοστό των καταναλωτών που προτιμούν τη κατηγορία «ZYMARIKA» είναι 2,85 φορές μεγαλύτερο από ότι είναι στον συνολικό πληθυσμό.

Ο τελευταίος κανόνας που θα σχολιάσουμε είναι ο κανόνας 97 {TOMATOEIDH} → {APTYMATATA} ο οποίος έχει lift ίσο με 2,25. Δηλαδή, ενώ μόνο στο 18.6% των συνολικών συναλλαγών εμφανίζεται η κατηγορία «APTYMATATA», στον υποπληθυσμό των καταναλωτών που αγοράζουν προϊόντα της κατηγορίας «TOMATOEIDH», το ποσοστό των καταναλωτών που προτιμούν τη κατηγορία «APTYMATATA» είναι 2.25 φορές υψηλότερο από ότι είναι στο συνολικό πληθυσμό (18,6%).

Στη συνέχεια θα χαμηλώσουμε τη τιμή κατωφλιού υποστήριξης στο 1%, ταξινομώντας τους κανόνες συσχέτισης με το μέτρο lift κατά φθίνουσα σειρά, ώστε να εξετάσουμε εάν παρόλη τη μικρή συχνότητα εμφάνισης των κανόνων, το μέτρο lift έχει υψηλή τιμή. Η περίληψη των κανόνων συσχέτισης με υποστήριξη ίση με 0.01 και εμπιστοσύνη ίση με 0.00 παρουσιάζονται στην Εικόνα 12.

```
setof 5924 rules
```

```
rule length distribution (lhs + rhs):sizes
```

```
  1   2   3   4
66 1372 3570 916
```

```
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
1.000  3.000   3.000   2.901  3.000   4.000
```

```
summary of quality measures:
```

support		confidence		lift		count	
Min.	:0.01001	Min.	:0.01004	Min.	:1.000	Min.	: 726
1st Qu.	:0.01112	1st Qu.	:0.25121	1st Qu.	:1.741	1st Qu.	: 807
Median	:0.01304	Median	:0.34735	Median	:2.130	Median	: 946
Mean	:0.01656	Mean	:0.35209	Mean	:2.233	Mean	: 1202
3rd Qu.	:0.01734	3rd Qu.	:0.46034	3rd Qu.	:2.589	3rd Qu.	: 1258
Max.	:0.33776	Max.	:0.76915	Max.	:6.489	Max.	:24502

```
mining info:
```

```
data ntransactions support confidence
tr2          72543    0.01          0
```

Εικόνα 12 Σύνοψη Κανόνων

Οι κανόνες συσχέτισης που δημιουργήθηκαν είναι 5.924. Εξήντα έξι από αυτούς, είναι μη αξιολογήσιμοι γιατί αποτελούνται από 1-στοιχειοσύνολο, 1.372 κανόνες αποτελούνται από 2-στοιχειοσύνολα κοκ. Η κατανομή των δεδομένων είναι λοξή προς τα αριστερά γιατί η μέση τιμή είναι μικρότερη από τη διάμεσο. Η τιμή της υποστήριξης κυμαίνεται από [0,01, 0,3377], η τιμή της εμπιστοσύνης κυμαίνεται από [0,01004, 0,7691], η συχνότητα εμφάνισης των κανόνων έχει εύρος από [726, 24.502] και τέλος η τιμή του lift κυμαίνεται από [1, 6,489]. Στη

συνέχεια παρουσιάζουμε από τους 5.924 κανόνες που εξόρυξε ο αλγόριθμος τους 10 πρώτους με την υψηλότερη τιμή lift (Εικόνα 13).

	lhs	rhs	support	confidence	lift	count
[1]	{ΖΑΧΑΡΗ, ΤΟΥΡΤΕΣ-ΜΙΓΜΑΤΑ-ΓΛΥΚΑ}	=> {ΑΛΕΥΡΙ - ΣΙΜΙΓΔΑΛΙ}	0.01146906	0.4553914	6.488991	832
[2]	{ΑΡΤΥΜΑΤΑ, ΖΥΜΑΡΙΚΑ, ΤΟΜΑΤΟΕΙΔΗ}	=> {ΡΥΖΙΑ}	0.01050411	0.3690073	5.584998	762
[3]	{ΖΥΜΑΡΙΚΑ, ΤΟΥΡΤΕΣ-ΜΙΓΜΑΤΑ-ΓΛΥΚΑ}	=> {ΑΛΕΥΡΙ - ΣΙΜΙΓΔΑΛΙ}	0.01065575	0.3796660	5.409961	773
[4]	{ΖΑΧΑΡΗ, ΑΛΕΥΡΙ - ΣΙΜΙΓΔΑΛΙ}	=> {ΤΟΥΡΤΕΣ-ΜΙΓΜΑΤΑ-ΓΛΥΚΑ}	0.01146906	0.4896998	5.232625	832
[5]	{ΑΡΤΥΜΑΤΑ, ΤΟΥΡΤΕΣ-ΜΙΓΜΑΤΑ-ΓΛΥΚΑ}	=> {ΑΛΕΥΡΙ - ΣΙΜΙΓΔΑΛΙ}	0.01316461	0.3652008	5.203842	955
[6]	{ΑΡΤΥΜΑΤΑ, ΖΥΜΑΡΙΚΑ, ΡΟΦΗΜΑΤΑ}	=> {ΡΥΖΙΑ}	0.01029734	0.3423465	5.181481	747
[7]	{ΖΑΧΑΡΗ, ΖΥΜΑΡΙΚΑ}	=> {ΡΥΖΙΑ}	0.01013192	0.3365385	5.093576	735
[8]	{ΑΡΤΥΜΑΤΑ, ΖΥΜΑΡΙΚΑ, ΡΥΖΙΑ}	=> {ΤΟΜΑΤΟΕΙΔΗ}	0.01050411	0.6195122	4.977436	762
[9]	{ΖΑΧΑΡΗ, ΖΥΜΑΡΙΚΑ}	=> {ΑΛΕΥΡΙ - ΣΙΜΙΓΔΑΛΙ}	0.01029734	0.3420330	4.873718	747
[10]	{ΡΟΦΗΜΑΤΑ, ΤΟΥΡΤΕΣ-ΜΙΓΜΑΤΑ-ΓΛΥΚΑ}	=> {ΑΛΕΥΡΙ - ΣΙΜΙΓΔΑΛΙ}	0.01272349	0.3418519	4.871137	923

Εικόνα 13 Εμφάνιση 10 Κανόνων

Πράγματι, ο κανόνας 1, $\{ΖΑΧΑΡΗ, ΤΟΥΡΤΕΣ - ΜΙΓΜΑΤΑ - ΓΛΥΚΑ\} \rightarrow \{ΑΛΕΥΡΙ - ΣΙΜΙΓΔΑΛΙ\}$, έχει υποστήριξη ίση με 1,11% περίπου, που σημαίνει ότι οι κατηγορίες «ΖΑΧΑΡΗ», «ΤΟΥΡΤΕΣ-ΜΙΓΜΑΤΑ-ΓΛΥΚΑ» και «ΑΛΕΥΡΙ-ΣΙΜΙΓΔΑΛΙ» υπάρχουν μαζί στο 1,11% των καλαθιών, ενώ το 45,53% των φορών που οι καταναλωτές αγοράζουν προϊόντα της κατηγορίας «ΖΑΧΑΡΗ» και «ΤΟΥΡΤΕΣ-ΜΙΓΜΑΤΑ-ΓΛΥΚΑ», αγοράζουν και προϊόντα της κατηγορίας «ΑΛΕΥΡΙ-ΣΙΜΙΓΔΑΛΙ». Παρόλο που αυτές οι κατηγορίες δεν εμφανίζονται μαζί συχνά στα καλάθια, και η κατηγορία «ΑΛΕΥΡΙ-ΣΙΜΙΓΔΑΛΙ» υπάρχει μόλις σε 5.091 καλάθια από τα 72.543 που υπάρχουν συνολικά, δηλαδή σε ποσοστό 7%, στον υποπληθυσμό των καταναλωτών που αγοράζουν προϊόντα των κατηγοριών «ΖΑΧΑΡΗ» και «ΤΟΥΡΤΕΣ-ΜΙΓΜΑΤΑ-ΓΛΥΚΑ», το ποσοστό των καταναλωτών που αγοράζουν προϊόντα της κατηγορίας «ΑΛΕΥΡΙ-ΣΙΜΙΓΔΑΛΙ» είναι 6,48 φορές υψηλότερο από ότι είναι στον συνολικό πληθυσμό.

Επίσης ο κανόνας 2, δείχνει ότι η κατηγορία «ΡΥΖΙΑ», όταν αγοράζεται μαζί με τις κατηγορίες «ΑΡΤΥΜΑΤΑ», «ΖΥΜΑΡΙΚΑ» και «ΤΟΜΑΤΟΕΙΔΗ», το lift είναι ίσο με 5,58. Αυτό σημαίνει ότι παρόλο που η κατηγορία «ΡΥΖΙΑ» εμφανίζεται στο 6% των συνολικών συναλλαγών, στον υποπληθυσμό των καταναλωτών που αγοράζουν προϊόντα των κατηγοριών «ΑΡΤΥΜΑΤΑ», «ΖΥΜΑΡΙΚΑ» και «ΤΟΜΑΤΟΕΙΔΗ», το ποσοστό των καταναλωτών που προτιμούν τα προϊόντα της κατηγορίας «ΡΥΖΙΑ» είναι 5,58 φορές μεγαλύτερο από ότι είναι στο συνολικό πληθυσμό. Συνοψίζοντας, και οι 10 κανόνες έχουν πολύ μικρή συχνότητα εμφάνισης στις συναλλαγές μόλις 1% κατά μέσο όρο, αλλά η ύπαρξη των κατηγοριών που βρίσκονται στο αριστερό μέλος των κανόνων, ενισχύει τη πιθανότητα προτίμησης της κατηγορίας στο δεξί μέλος των κανόνων.

3.2 Δεύτερο Μέρος Ανάλυσης

3.2.1 Περιγραφή των Δεδομένων της 1^{ης} Εβδομάδας

Το δεύτερο μέρος της ανάλυσης μας, αποτελείται από δύο αρχεία προς επεξεργασία, που το κάθε ένα αφορά δύο διαφορετικές εβδομάδες όπως αναφέραμε στην αρχή του κεφαλαίου. Πρέπει να σημειωθεί, ότι τα δύο αρχεία αναφέρονται στο χρονικό διάστημα 30/10/2017-4/11/2017 και 13/11/2017 – 18/11/2017 αντίστοιχα, σε ένα υποσύνολο της περιόδου του αρχικού αρχείου (Πίνακας 5). Η μόνη διαφορά είναι ότι περιέχονται οι συναλλαγές των καταναλωτών που είναι κάτοχοι καρτών του AB Βασιλόπουλου και έκαναν αγορές τη συγκεκριμένη περίοδο, έχοντας αυτή τη φορά και τη πληροφορία ότι κάποια προϊόντα είναι σε προσφορά.

Πιο συγκεκριμένα, όσον αφορά στην πρώτη εβδομάδα, το αρχείο που επεξεργαστήκαμε, περιέχει κάθε καταναλωτή (κάτοχο κάρτας μέλους) που επισκέφτηκε το κατάστημα (CustomerID), τον αριθμό που χαρακτηρίζει τις διαφορετικές συναλλαγές που πραγματοποίησε (transaction ID), τον κωδικό του κάθε διαφορετικού προϊόντος (Master Code), την περιγραφή του κάθε προϊόντος (Description), το «κανάλι» προωθητικής ενέργειας (TV or LEAFLET), το τμήμα στο οποίο ανήκει το κάθε προϊόν, τη κατηγορία στην οποία ανήκει (Category), τη χρηματική αξία του κάθε προϊόντος (Value) και τη ποσότητα του κάθε προϊόντος σε κάθε συναλλαγή (Quantity). Κάθε προϊόν που είναι σε προσφορά, διαφημίζεται με έναν από τους δύο τρόπους που προαναφέραμε. Παρακάτω παρουσιάζεται η δομή του αρχείου της πρώτης εβδομάδας (Πίνακας 7).

Columns	Labels	Count
1 st Column	CustomerID	6.278
2 nd Column	TransID	9.939
3 rd Column	MasterCode	9.811
4 th Column	Desc	9.801
5 th Column	PromoSKU	
6 th Column	Dept	33
7 th Column	Categ	187
8 th Column	Value	
9 th Column	Quantity	
Rows of the Dataset	100.505	

Πίνακας 7 Περιγραφή Δεδομένων 1^{ης} Εβδομάδας

Όπως παρατηρούμε, οι διαφορετικοί καταναλωτές, κάτοχοι καρτών μέλους που επισκέφτηκαν το κατάστημα είναι 6.278 και πραγματοποίησαν 9.939 συναλλαγές. Αυτό υποδηλώνει, ότι κάποιοι καταναλωτές σε μια εβδομάδα επισκέφτηκαν παραπάνω από μια φορές το κατάστημα. Ενώ τα διαφορετικά προϊόντα ως προς τη περιγραφή τους που βρέθηκαν στα καλάθια των

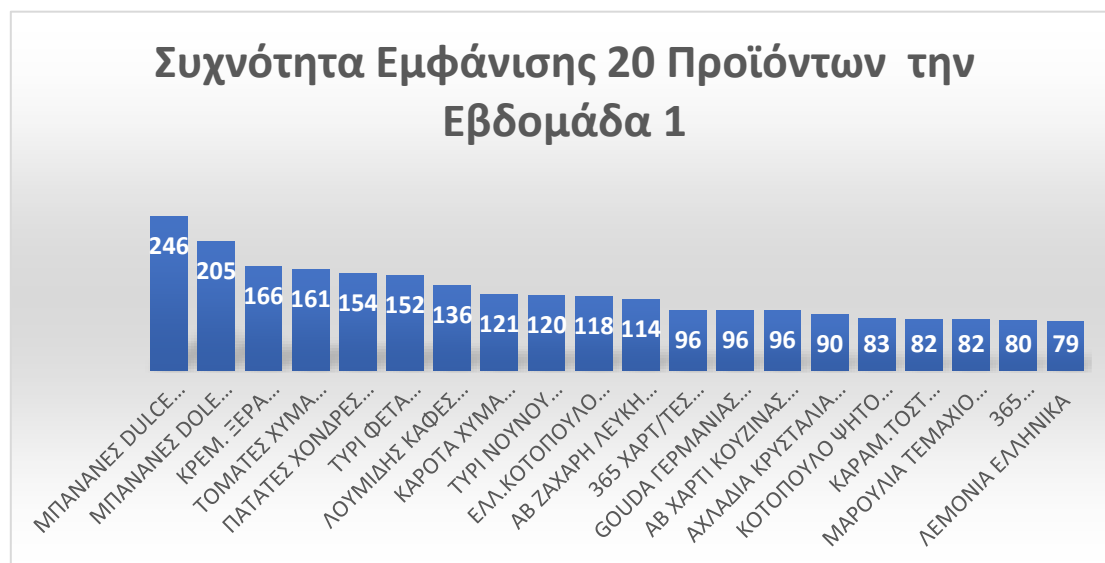
καταναλωτών είναι 9.801 προϊόντα, οι κωδικοί που τα χαρακτηρίζουν είναι περισσότεροι. Αυτό οφείλεται στο γεγονός ότι η περιγραφή ενός προϊόντος αντιστοιχεί σε παραπάνω από έναν κωδικούς. Συνολικά, ο διαφορετικός αριθμός των τμημάτων και των κατηγοριών που βρέθηκαν στα καλάθια είναι 33 τμήματα προϊόντων και 187 κατηγορίες. Όσον αφορά στη προωθητική ενέργεια, τα προϊόντα που δεν είναι σε προσφορά την 1^η εβδομάδα έχουν κενή τη τιμή του γνωρίσματος «PromoSKU», ενώ τα προϊόντα που είναι χαρακτηρισμένα με «PromoSKU="TV" ή PromoSKU="Leaflet"» στη πρώτη εβδομάδα, βρίσκονται σε προσφορά την 2^η εβδομάδα. Να σημειώσουμε, ότι η αξία κάθε προϊόντος είναι υπολογισμένη με βάση τα τεμάχια σε κάθε συναλλαγή. Στην Εικόνα 14, παρουσιάζονται οι 15 πρώτες εγγραφές από τις 100.505 του αρχείου της 1^{ης} εβδομάδας.

CustomerID	transID	MasterCode	Desc	PromoSKU	Dept	Categ	Value	Quantity
1	131157123610	92470960400	KRISPIES ΜΕ ΣΟΥΣΑΜΙ 200GR	NA	ΠΑΝΤΟΠΩΛΕΙΟ	ΑΡΤΟΕΙΔΗ	1.28	1
2	104208133410	40722839300	SOUPLINE ULTRA ΜΑΝΟΛΙΑ 60M(1+1)	NA	ΑΠΟΡΡΥΠΑΝΤΙΚΑ - ΚΑΘΑΡΙΣΤΙΚΑ	ΑΠΟΡΡΥΠΑΝΤΙΚΑ ΡΟΥΧΩΝ	5.82	1
3	104208133410	40724504300	ΦΕΤΑ ΠΟΠ ΑΧΑΪΑΣ ΔΟΧΕΙΟ ΤΥΓ.	NA	ΤΥΡΙΑ	ΦΕΤΕΣ	4.90	1
4	104208133410	40733594400	ΠΑΤΑΤΕΣ ΧΟΝΔΡΕΣ ΕΛΛΗΝΙΚΕΣ	NA	ΟΠΩΡΟΠΩΛΕΙΟ	ΠΑΤΑΤΕΣ	2.03	1
5	104208133410	90441831600	ΒΟΕΙΑ ΣΤΗΘΟΠΛΕΥΡΑ Μ/Ο ΕΛΛΗΝ.ΕΚΤΡ ΑΝ	NA	ΚΡΕΟΠΩΛΕΙΟ	ΝΩΠΑ ΚΡΕΑΤΑ	15.86	1
6	104208133410	90600671300	ΕΠΙΛΕΓΜΕΝΟ Φ.ΓΑΛΛΑ ΟΛΥΜΠΟΣ LIGHT 1.	NA	ΓΑΛΑΚΤΟΚΟΜΙΚΑ-ΑΥΓΑ-ΦΡΕΣΚΟΙ ΧΥΜΟΙ	ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ	1.90	1
7	104208133410	92460012800	ΑΡΤΟΠΙ/ΣΜΑ ΣΙΚΑΛΗΣ 350ΓΡ ΚΑΣΙΜΗ	NA	FRESH BAKERY	ΦΕΤΕΣ ΨΩΜΙ - ΔΙΑΦΟΡΑ	1.00	1
8	1001100131239114	40714262000	PANTENE ΜΑΣΚΑ ΧΡΩΜΑ 2M 300ML Leaflet	Leaflet	ΥΓΙΕΙΝΗ - ΟΜΟΡΦΙΑ	ΦΡΟΝΤΙΔΑ ΜΑΛΛΙΩΝ	8.49	1
9	1001100131239114	40722614400	ZEWA W&W ΧΑΡΤΙ ΚΟΥΖΙΝΑΣ ECONOMY 3P Leaflet	Leaflet	ΧΑΡΤΙΑ TISSUE	ΧΑΡΤΙΑ ΚΟΥΖΙΝΑΣ	4.00	1
10	1001100131239114	40729990300	LENOR GOLD ORCHID 26ΜΕΖ Leaflet	Leaflet	ΑΠΟΡΡΥΠΑΝΤΙΚΑ - ΚΑΘΑΡΙΣΤΙΚΑ	ΑΠΟΡΡΥΠΑΝΤΙΚΑ ΡΟΥΧΩΝ	6.16	2
11	1001100131239114	40730968200	BABYLINO SENS MEGAPACK N4 78T	NA	ΥΓΙΕΙΝΗ - ΟΜΟΡΦΙΑ	ΠΡΟΣΩΠΙΚΗ ΥΓΙΕΙΝΗ	24.80	1
12	1001100131239114	90760260000	BITAM SOFT 250ΓΡ ΑΝΑΛΑΤΟ	NA	ΓΑΛΑΚΤΟΚΟΜΙΚΑ-ΑΥΓΑ-ΦΡΕΣΚΟΙ ΧΥΜΟΙ	ΜΑΡΓΑΡΙΝΕΣ	1.30	1
13	10021002424115061	40729676500	ΚΡΙΣ ΚΡΙΣ ΨΩΜΑΚΙΑ ΣΑΝΤΟΥΤΙΣ 480ΓΡ	NA	ΑΡΤΟΖΑΧΑΡΟΠΛΑΣΤΕΙΟ	ΨΩΜΑΚΙΑ ROLLS	1.88	1
14	10021002424115061	40730129900	ΓΟΥΔΑ ΣΕ ΦΕΤΕΣ 400ΓΡ	NA	ΤΥΡΙΑ	ΦΡΕΣΚΑ & ΑΝΑΚΑΤΕΡΓΑΣΜΕΝΑ ΤΥΡΙΑ ΕΙΣΑΓΩΓΗΣ	3.00	1
15	10021002424115061	40730547100	ΚΑΤΣΕΛΗΣ ΤΟ ΤΟΣΤ ΣΤΑΡ 0.30 720ΓΡ	NA	ΑΡΤΟΖΑΧΑΡΟΠΛΑΣΤΕΙΟ	ΤΟΣΤ	1.82	1

Showing 1 to 16 of 100,505 entries

Εικόνα 14 Σύνολο Δεδομένων 1ης Εβδομάδας

Στην συνέχεια παρουσιάζουμε στο Σχήμα 15, τη συχνότητα των 20 από τα 9.801 προϊόντα, και στην επόμενη ενότητα θα μιλήσουμε για τα δεδομένα της δεύτερης εβδομάδας.



Σχήμα 15 Η Μεγαλύτερη Συχνότητα Εμφάνισης 20 Προϊόντων στις Συναλλαγές την Εβδομάδα 1

3.2.2 Περιγραφή των Δεδομένων της 2^{ης} Εβδομάδας

Όσον αφορά στην δεύτερη εβδομάδα, το αρχείο που επεξεργαστήκαμε, περιέχει τα ίδια γνωρίσματα με το αρχείο της πρώτης εβδομάδας αλλά το περιεχόμενο του αρχείου είναι διαφορετικό. Την δεύτερη εβδομάδα οι καταναλωτές που επισκέφθηκαν το κατάστημα είναι 5.487 και εντοπίστηκαν 8.407 διαφορετικά καλάθια αγορών. Τα διαφορετικά προϊόντα που βρέθηκαν στα καλάθια των καταναλωτών είναι 9.323, ενώ οι κωδικοί που τα χαρακτηρίζουν είναι 9.334. Τα διαφορετικά τμήματα προϊόντων που βρέθηκαν στα καλάθια των καταναλωτών είναι 34 και οι διαφορετικές κατηγορίες των προϊόντων 192. Τέλος οι συνολικές εγγραφές του αρχείου είναι 82.134 (Πίνακας 8).

Columns	Labels	Count
1 st Column	CustomerID	5.487
2 nd Column	TransID	8.407
3 rd Column	MasterCode	9.334
4 th Column	Desc	9.323
5 th Column	PromoSKU	
6 th Column	Dept	34
7 th Column	Categ	192
8 th Column	Value	
9 th Column	Quantity	
Rows of the Dataset	82.134	

Πίνακας 8 Περιγραφή Δεδομένων 2^{ης} Εβδομάδας

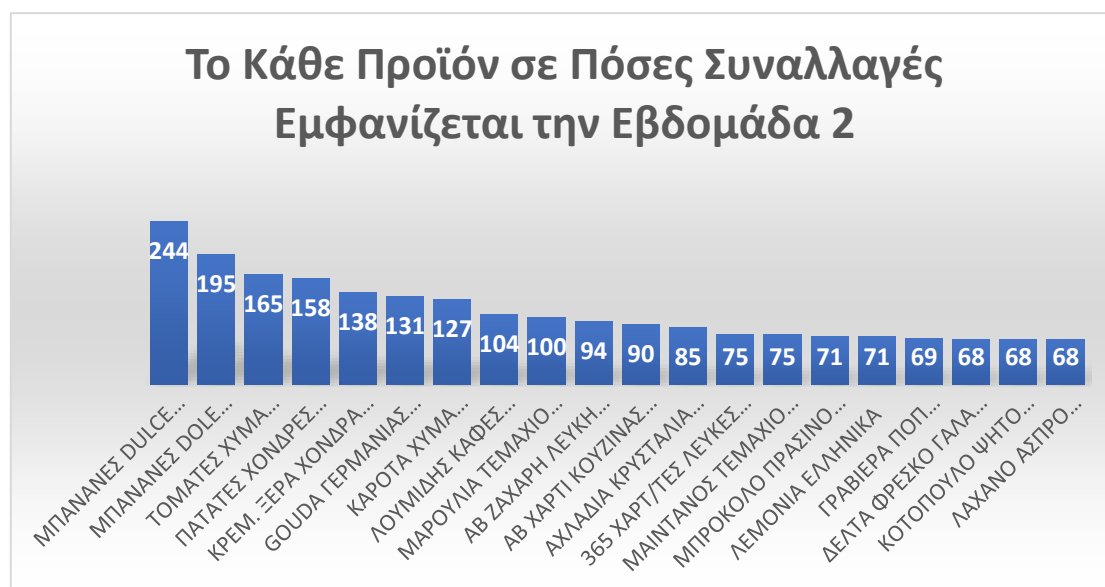
Στην Εικόνα 15, παρουσιάζεται τι περιέχει ενδεικτικά το αρχείο της 2^{ης} εβδομάδας .

▲	CustomerID	transID	MasterCode	Desc	PromoSKU	Dept	Categ	Value	Quantity
1	100	100187711214	40730550200	COLGATE ΟΔ/ΜΑ TR. ACTION 75ML(0.50	NA	ΥΓΙΕΙΝΗ - ΟΜΟΡΦΙΑ	ΣΤΟΜΑΤΙΚΗ ΥΓΙΕΙΝΗ	1.78	1
2	100	100187711214	90033760000	365 ΑΛΑΤΙ ΜΑΓΕΙΡΙΚΟ ΨΙΛΟ 1000ΓΡ.	NA	ΠΑΝΤΟΠΩΛΕΙΟ	ΑΡΤΥΜΑΤΑ	0.37	1
3	100	100187711214	90050400100	ΜΑΙΝΤΑΝΟΣ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΟΣ 80ΓΡ	NA	ΟΠΩΡΟΠΩΛΕΙΟ	ΑΡΩΜΑΤΙΚΑ ΟΠΩΡΟΠΩΛΕΙΟΥ	0.39	1
4	100	100187711214	90050400400	ΑΝΗΘΟΣ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΟΣ 80ΓΡ	NA	ΟΠΩΡΟΠΩΛΕΙΟ	ΑΡΩΜΑΤΙΚΑ ΟΠΩΡΟΠΩΛΕΙΟΥ	0.47	1
5	100	100187711214	90196640000	365 ΤΟΜΑ/ΛΤΟΣ ΔΙΠ. ΣΥΜ/ΚΝΩΣΗΣ 410G	NA	ΠΑΝΤΟΠΩΛΕΙΟ	ΤΟΜΑΤΟΕΙΔΗ	0.90	1
6	100	100187711214	90650440103	ΜΜΜΙΛΚ ΟΙΚΟΓΕΝΕΙΑΚΟ ΠΛΗΡΗΣ 1.5LT 0	NA	ΓΑΛΑΚΤΟΚΟΜΙΚΑ-ΑΥΓΑ-ΦΡΕΣΚΟΙ ΧΥΜΟΙ	ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΙΑΣΤΕΡΙΩΣΗΣ ΓΑΛΛΑ	1.85	1
7	100	100187711214	90680020000	ΑΒ ΚΡΕΜΑ ΓΑΛΑΚΤΟΣ LIGHT 200ML	NA	ΓΑΛΑΚΤΟΚΟΜΙΚΑ-ΑΥΓΑ-ΦΡΕΣΚΟΙ ΧΥΜΟΙ	ΚΡΕΜΕΣ ΓΑΛ.ΣΑΝΤΥΓΙ	1.03	1
8	100	100187711214	90703880000	ΚΡΙ ΚΡΙ ΣΤΡ/ΣΤΟ ΠΙΑΟΥΡΤΙ 10% 1KG	NA	ΓΑΛΑΚΤΟΚΟΜΙΚΑ-ΑΥΓΑ-ΦΡΕΣΚΟΙ ΧΥΜΟΙ	ΠΙΑΟΥΡΤΙΑ-ΕΠΙΔΟΡΤΙΑ	3.38	1
9	1000	1000136911543	40724540600	ΦΕΤΑ ΠΙΝΔΟΥ ΔΟΧΕΙΟ Π.Ο.Π ΤΥΠ	NA	ΤΥΡΙΑ	ΦΕΤΕΣ	11.33	3
10	1000	1000136911543	40733594400	ΠΑΤΑΤΕΣ ΧΟΝΔΡΕΣ ΕΛΛΗΝΙΚΕΣ	NA	ΟΠΩΡΟΠΩΛΕΙΟ	ΠΑΤΑΤΕΣ	0.79	1
11	1000	1000136911543	90040400100	ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ	NA	ΟΠΩΡΟΠΩΛΕΙΟ	ΒΟΛΒΟΕΙΔΗ	0.51	1
12	1000	1000136911543	90050150200	ΚΡΕΜ. ΞΕΡΑ ΧΟΝΔΡΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ	NA	ΟΠΩΡΟΠΩΛΕΙΟ	ΚΡΕΜΜΥΔΙΑ - ΣΚΟΡΔΑ	0.24	1
13	1000	1000136911543	90179760200	ΑΒ ΜΠΕΙΚΟΝ ΚΑΠΝΙΣΤΟ 130ΓΡ	NA	ΑΛΛΑΝΤΙΚΑ	ΑΛΛΑΝΤΙΚΑ ΒΡΑΣΤΑ	1.55	1
14	1000	1000136911543	90630570100	ΓΟΥΔΑ ΓΕΡΜΑΝΙΑΣ ΦΡΑΤΖΟΛΑ	NA	ΤΥΡΙΑ	ΤΥΡΙΑ ΕΙΣΑΓΩΓΗΣ ΓΙΑ ΤΟΣΤ LOAF & BLOCKS	7.05	2
15	1002	10021829216441	40713059600	ΑΒ ΞΙΔΙ 400ML	NA	ΠΑΝΤΟΠΩΛΕΙΟ	ΑΡΤΥΜΑΤΑ	0.49	1

Showing 1 to 16 of 82,134 entries

Εικόνα 15 Σύνολο Δεδομένων 2ης Εβδομάδας

Το Σχήμα 16 παρουσιάζει 20 από τα 9.323 προϊόντα με τη μεγαλύτερη συχνότητα εμφάνισης, που αγοράστηκαν την 2^η εβδομάδα από τους καταναλωτές.



Σχήμα 16 Η Μεγαλύτερη Συχνότητα Εμφάνισης 20 Προϊόντων στις Συναλλαγές την Εβδομάδα 2

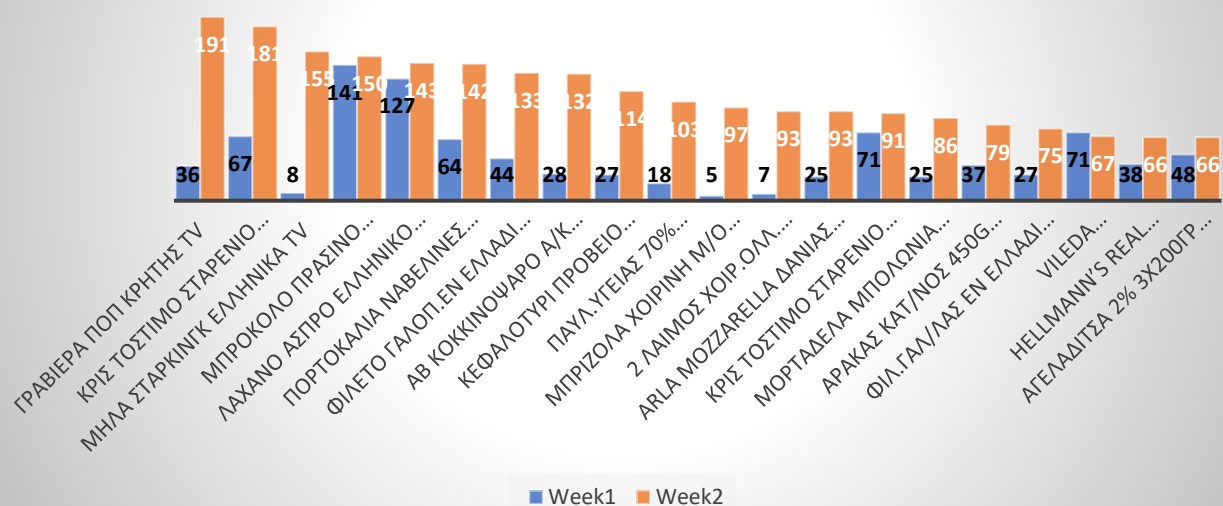
Παρατηρούμε ότι και την δεύτερη εβδομάδα τα προϊόντα που έχουν ως πρώτη επιλογή οι καταναλωτές, είναι οι μπανάνες.

Σημειώνουμε ότι στο τέλος κάθε περιγραφής προϊόντος και για τις δύο εβδομάδες, συνενώσαμε με τη βοήθεια της συνάρτησης «CONCANTENATE» του υπολογιστικού φύλλου Excel, τη σηματοδότηση της στήλης «PromoSKU».

Για να διαπιστώσουμε πόσα προϊόντα είναι σε προσφορά, ελέγχουμε το αρχείο της πρώτης εβδομάδας. Αυτό το κάνουμε, γιατί τη δεύτερη εβδομάδα που τα ίδια προϊόντα είναι πραγματικά σε προσφορά, δεν σημαίνει απαραίτητα ότι οι καταναλωτές θα τα προτιμήσουν. Τα προϊόντα που είναι σε προσφορά είναι 579 στο σύνολο. Παρόλο αυτά την δεύτερη εβδομάδα βρέθηκαν στα καλάθια των καταναλωτών μόνο 526 από τα 579 προϊόντα προσφοράς. Στην συνέχεια εξετάσαμε τις πωλήσεις των προϊόντων που είναι σε προσφορά συγκριτικά και για τις δύο εβδομάδες, ώστε να ελέγξουμε αν υπάρχει κάποια μεταβολή. Αυτό αναπαρίσταται μέσω του Σχήματος 17 για 20 από τα 579 προϊόντα προσφοράς.

Όπως διαπιστώνεται, τα παρακάτω προϊόντα παρουσιάζουν αυξητικές τάσεις την εβδομάδα που είναι σε προσφορά. Πιο συγκεκριμένα, η «ΓΡΑΒΙΕΡΑ ΠΟΠ ΚΡΗΤΗΣ» τη πρώτη εβδομάδα εντοπίστηκε σε 36 καλάθια καταναλωτών, ενώ τη δεύτερη που ήταν σε προσφορά εντοπίστηκε σε 191 καλάθια. Αντίστοιχα το προϊόν «ΚΡΙΣ ΤΟΣΤΙΜΟ ΣΤΑΡΕΝΙΟ», την πρώτη εβδομάδα εντοπίστηκε σε 67 καλάθια, ενώ τη δεύτερη σε 181 καλάθια κοκ.

Πωλήσεις Promo Προϊόντων για την Εβδομάδα 1 και 2



Σχήμα 17 Πωλήσεις Promo Προϊόντων την Εβδομάδα 1 και 2

Εν συνεχεία διαλέγουμε μόνο τους καταναλωτές που επισκέφτηκαν το κατάστημα και τις δύο εβδομάδες, και συγκρίνουμε τα καλάθια τους ως προς το πλήθος των προϊόντων που περιέχουν.

Με απλά λόγια θα παρατηρήσουμε πόσα προϊόντα είχε το καλάθι του ίδιου καταναλωτή την πρώτη και τη δεύτερη εβδομάδα (Σχήμα 18). Κατόπιν αυτού, θα εξετάσουμε αν άλλαξε η χρηματική αξία καλαθιού του ίδιου χρήστη σε σχέση με τη πρώτη και τη δεύτερη εβδομάδα.

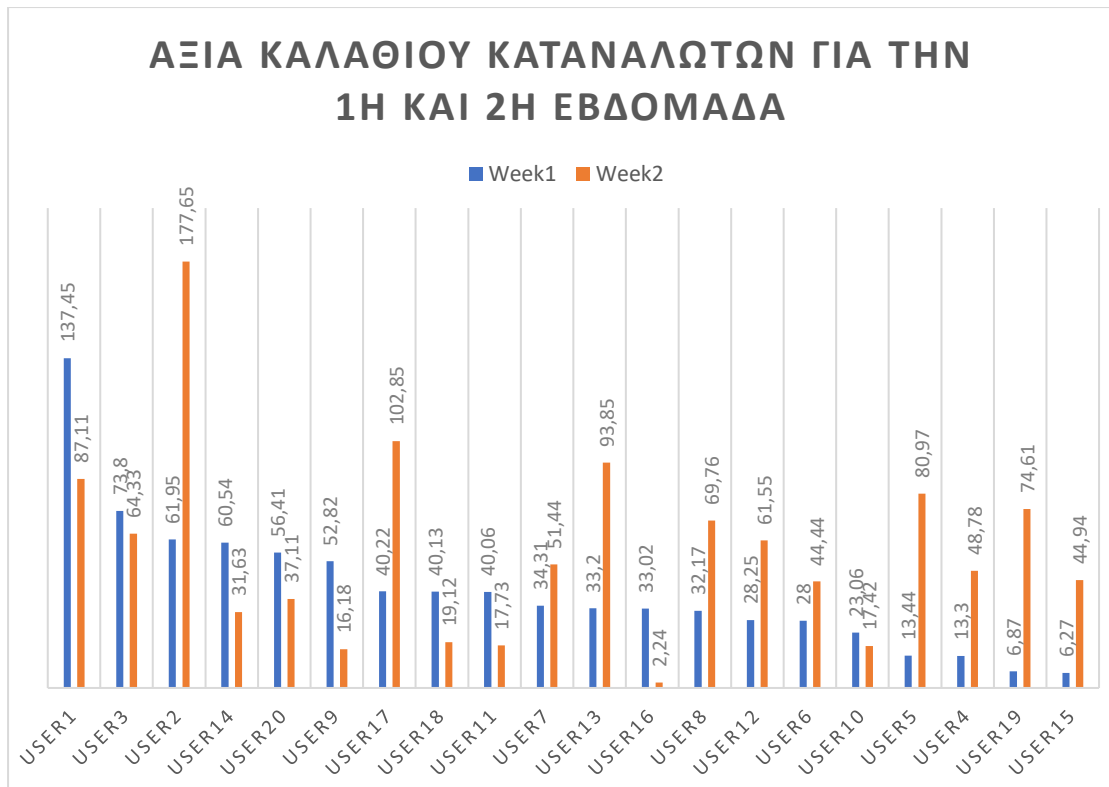
Οι κοινοί καταναλωτές που επισκέφθηκαν το κατάστημα και τις δύο εβδομάδες είναι 1.144. Μέσω του Σχήματος 18 που παρουσιάζει ενδεικτικά τους 20 από τους 1.144 καταναλωτές, φαίνεται η κατανομή των προϊόντων ανά καλάθι, πριν και μετά τις προωθητικές ενέργειες των εκάστοτε προϊόντων.



Σχήμα 18 Πλήθος Προϊόντων ανά Καλάθι

Είναι φανερό ότι οι περισσότεροι καταναλωτές, αγόρασαν περισσότερα προϊόντα την εβδομάδα με προσφορές. Όπως φαίνεται ο Χρήστης 1, την πρώτη εβδομάδα προτίμησε 33 προϊόντα, ενώ τη δεύτερη 48. Αντίστοιχα ο Χρήστης 3, ενώ τη πρώτη εβδομάδα αγόρασε 19 προϊόντα, τη δεύτερη αγόρασε 24 προϊόντα. Προφανώς, δεν είχαν όλοι οι καταναλωτές την ίδια συμπεριφορά. Παραδείγματος χάριν, ο Χρήστης 19, την πρώτη εβδομάδα αγόρασε 41 προϊόντα και τη δεύτερη μόνο 11. Ομοίως και ο Χρήστης 2 του οποίου το καλάθι περιέχε 52 προϊόντα τη πρώτη εβδομάδα και τη δεύτερη εβδομάδα 22.

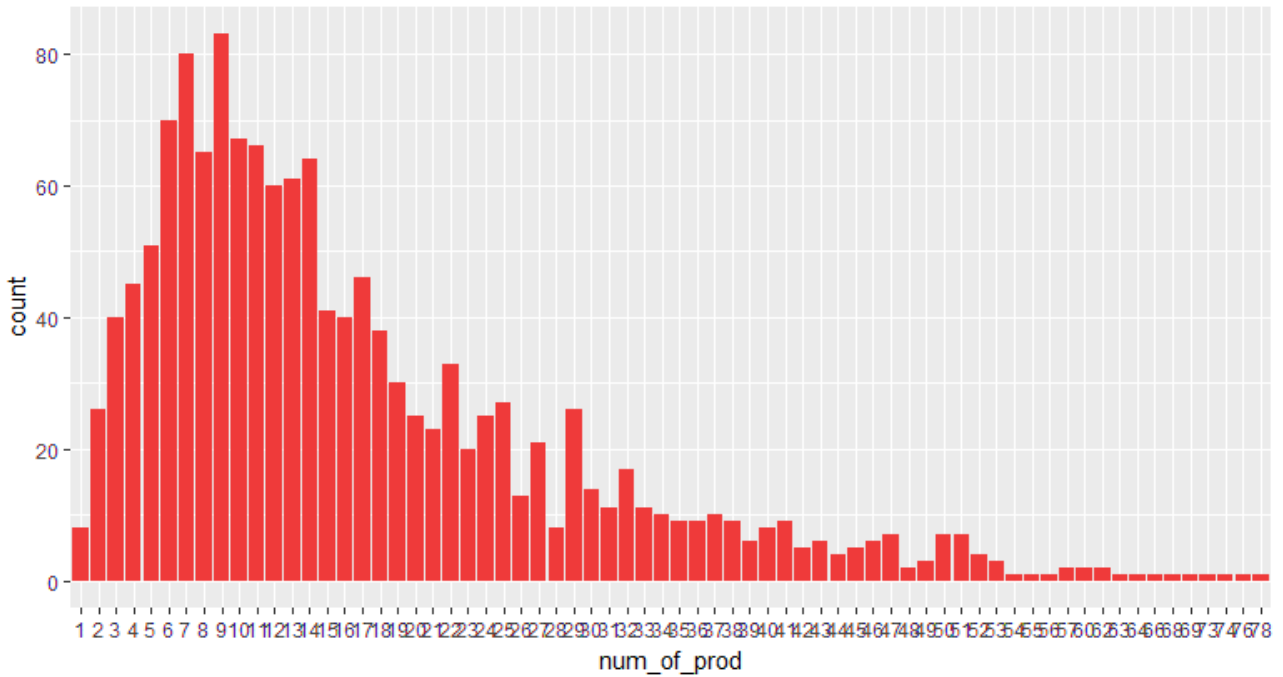
Όσον αφορά τη μεταβολή στη χρηματική αξία του καλαθιού συγκριτικά με τις δύο εβδομάδες που εξετάζουμε, η συνολική αξία των καλαθιών των κοινών καταναλωτών και στις δύο εβδομάδες ανέρχεται στα 66.978,37 ευρώ την πρώτη εβδομάδα και 69.138,34 ευρώ τη δεύτερη εβδομάδα. Πρόκειται για αύξηση αξίας καλαθιού λόγω προωθητικών ενεργειών σε κάποια προϊόντα, της τάξεως του 3.2248%. Στο Σχήμα 19, παρουσιάζονται οι χρηματικές αξίες 20 καλαθιών για 20 καταναλωτές, που επισκέφθηκαν το κατάστημα και τις 2 εβδομάδες.



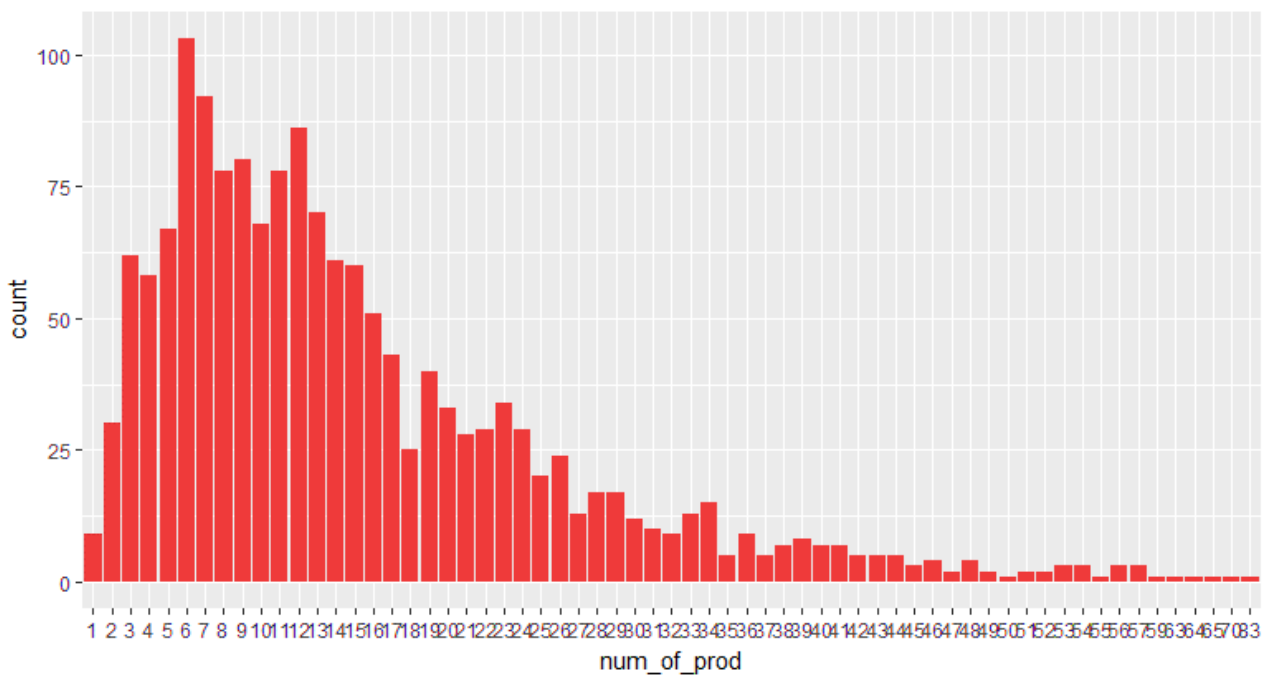
Σχήμα 19 Αξία Καλαθιού Καταναλωτών για την Εβδομάδα 1 και 2

Παρόλο που η χρηματική αξία των αγορών του Χρήστη 1 την πρώτη εβδομάδα ανέρχεται στα 137,45 ευρώ, την δεύτερη εβδομάδα ξόδεψε 87,11 ευρώ για το καλάθι των αγορών του. Ο Χρήστης 2 ξόδεψε 177,66 ευρώ την εβδομάδα που κάποια προϊόντα ήταν σε προσφορά, ενώ την εβδομάδα που δεν ήταν σε προσφορά, ξόδεψε 61,95 ευρώ. Παρατηρούμε λοιπόν ότι δεν έχουν όλοι οι καταναλωτές την ίδια συμπεριφορά.

Τέλος, θα παρουσιάσουμε την κατανομή του πλήθους προϊόντων που πραγματοποιήθηκαν την πρώτη (Σχήμα 20) και τη δεύτερη εβδομάδα (Σχήμα 21) ανά συναλλαγή.



Σχήμα 20 Κατανομή Συναλλαγών ανά Πλήθος Προϊόντων για την Εβδομάδα 1



Σχήμα 21 Κατανομή Συναλλαγών ανά Πλήθος Προϊόντων για την Εβδομάδα 2

Από τα Σχήματα 20 και 21, παρατηρούμε ότι οι 2 κατανομές έχουν κάποιες σημαντικές διαφορές. Η πρώτη και σημαντικότερη διαφορά έγκειται στους άξονες x και y.

Την πρώτη εβδομάδα, 6 προϊόντα εμφανίζονται σε περίπου 70 συναλλαγές, ενώ την εβδομάδα που κάποια προϊόντα ήταν σε προσφορά, 6 προϊόντα εμφανίζονται σε πάνω από 100 συναλλαγές. Ακόμα τη πρώτη εβδομάδα, 78 διαφορετικά προϊόντα βρέθηκαν σε μία μόνο

συναλλαγή, ενώ τη δεύτερη εβδομάδα βρέθηκαν 83 προϊόντα σε μια συναλλαγή. Είναι εμφανές ότι ένα μεγάλο ποσοστό των καταναλωτών την εβδομάδα που κάποια προϊόντα είναι σε προσφορά, κάνει περισσότερες αγορές και καταχωρούνται στη βάση δεδομένων του καταστήματος περισσότερες συναλλαγές.

Αφού παρουσιάσαμε κάποια σημαντικά χαρακτηριστικά των δεδομένων των δύο αρχείων μας, είμαστε έτοιμοι να εισάγουμε στον αλγόριθμο Apriori τα δεδομένα που επιθυμούμε, ώστε ο αλγόριθμος να εξορύξει κανόνες συσχέτισης.

3.2.3 Προετοιμασία Δεδομένων 1^{ης} και 2^{ης} Εβδομάδας

Οι συσχετίσεις που επιθυμούμε να εξάγουμε δεν αφορούν όλο το σύνολο των δεδομένων των δύο εβδομάδων. Αντιθέτως, μας ενδιαφέρει να παρατηρήσουμε την συμπεριφορά των καταναλωτών που επισκέφθηκαν και τις δύο εβδομάδες το κατάστημα και έχουν τουλάχιστον ένα προϊόν στο καλάθι τους το οποίο είναι σε προσφορά. Με άλλα λόγια θα πραγματοποιήσουμε δύο αναλύσεις και στο τέλος θα συγκρίνουμε τα αποτελέσματά τους.

Αρχικά, θα εισάγουμε τον αλγόριθμο Apriori τις συναλλαγές που πραγματοποιήθηκαν τη πρώτη εβδομάδα και περιείχαν τουλάχιστον ένα προϊόν προς προσφορά για την δεύτερη εβδομάδα, ώστε να διαπιστώσουμε με ποια άλλα προϊόντα αγοράζονται μαζί από τους καταναλωτές.

Κατόπιν, θα εισάγουμε τα καλάθια των ίδιων καταναλωτών της δεύτερης εβδομάδας και θα παρατηρήσουμε ξανά τα προϊόντα που είναι σε προσφορά με ποια άλλα προϊόντα αγοράζονται μαζί. Η Εικόνα 16 και η Εικόνα 17, παρουσιάζουν ενδεικτικά τα δύο υποσύνολα των δεδομένων της πρώτης και δεύτερης εβδομάδας αντίστοιχα, που θα χρησιμοποιήσουμε στην ανάλυσή μας. Τα υποσύνολα της 1^{ης} και της 2^{ης} εβδομάδας, αποτελούνται μόνο από τους καταναλωτές που έχουν επισκεφτεί το κατάστημα και τις δύο εβδομάδες και έχουν τουλάχιστον 1 προϊόν προσφοράς στο καλάθι τους. Οι κοινοί καταναλωτές ανέρχονται στους 1.144 και οι συναλλαγές που πραγματοποίησαν την πρώτη εβδομάδα είναι 1.402 και 1.566 την δεύτερη, ενώ τα διαφορετικά προϊόντα που βρέθηκαν στα καλάθια τους την πρώτη εβδομάδα είναι 6.016 και 6.045 τη δεύτερη.

CustomerID	transID	MasterCode	Desc	PromoSKU	Dept	Categ	Value	Quantity	Visits	
1	1002	1002424215071	40712676200	FAIRY ΥΓΡΟ ΠΙΑΤΩΝ ΛΕΜΟΝΙ 900ML	NA	ΑΠΟΡΡΥΠΑΝΤΙΚΑ - ΚΑΘΑΡΙΣΤΙΚΑ	ΑΠΟΡΡΥΠΑΝΤΙΚΑ ΠΙΑΤΩΝ	2.88	1	1
2	1002	1002424215071	40712801600	SANI PANTS MEDIUM NO2 14T	NA	ΥΓΙΕΙΝΗ - ΟΜΟΡΦΙΑ	ΠΡΟΣΩΠΙΚΗ ΥΓΙΕΙΝΗ	25.64	2	1
3	1002	1002424215071	40712822800	SENSODYNE ΟΔΟΝ. COMPLETE PROTECTIO	NA	ΥΓΙΕΙΝΗ - ΟΜΟΡΦΙΑ	ΣΤΟΜΑΤΙΚΗ ΥΓΙΕΙΝΗ	4.21	1	1
4	1002	1002424215071	40714461500	ΑΒ ΚΕΦΑΛΟΥΡΙ ΤΡΙΜΜΕΝΟ 200ΓΡ	NA	ΤΥΡΙΑ	ΔΙΑΦΟΡΑ ΕΓΧΩΡΙΑ ΣΚΛΗΡΑ ΤΥΡΙΑ	2.45	1	1
5	1002	1002424215071	40723234400	ΑΒ ΓΑΛΑ ΥΨ.ΘΕΡΜΙΚΗΣ ΕΠΕΞ. ΠΛΗΡΕΣ 1	NA	ΓΑΛΑΚΤΟΚΟΜΙΚΑ-ΑΥΓΑ-ΦΡΕΣΚΟΙ ΧΥΜΟΙ	ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΑ	3.57	3	1
6	1002	1002424215071	40728268400	ΑΒ ΣΑΛΑΤΑ ΠΡΙΜΑ 125 ΓΡ.	NA	ΟΠΩΡΟΠΩΛΕΙΟ	ΕΤΟΙΜΕΣ ΣΑΛΑΤΕΣ	1.39	1	1
7	1002	1002424215071	40730129900	ΓΟΥΔΑ ΣΕ ΦΕΤΕΣ 400ΓΡ	NA	ΤΥΡΙΑ	ΦΡΕΣΚΑ & ΑΝΑΚΑΤΕΡΓΑΣΜΕΝΑ ΤΥΡΙΑ ΕΙΣΑΓΩΓΗΣ	3.00	1	1
8	1002	1002424215071	40730511800	ΑΒ ΚΡΟΥΤΟΝ ΜΕ ΑΡΩΜΑ ΤΥΡΙΟΥ 100GR	NA	ΠΑΝΤΟΠΩΛΕΙΟ	ΣΟΥΠΕΣ-CROUTONS	1.08	1	1
9	1002	1002424215071	40731011300	LAYS ΤΣΙΠΣ ΑΛΑΤΙ 150GR	NA	ΠΑΝΤΟΠΩΛΕΙΟ	ΣΝΑΚΣ	1.45	1	1
10	1002	1002424215071	40733590700	ΣΙΔΕΡΟΠΛΑΝΟ ΕΣΟ 52Χ130CM	NA	ΟΙΚΙΑΚΑ	ΕΙΔΗ ΚΑΘΑΡΙΣΜΟΥ	6.55	1	1
11	1002	1002424215071	40733594400	ΠΑΤΑΤΕΣ ΧΟΝΔΡΕΣ ΕΛΛΗΝΙΚΕΣ	NA	ΟΠΩΡΟΠΩΛΕΙΟ	ΠΑΤΑΤΕΣ	0.80	1	1
12	1002	1002424215071	90020010100	ΑΧΛΑΔΙΑ ΚΡΥΣΤΑΛΙΑ ΕΛΛΗΝΙΚΑ	NA	ΟΠΩΡΟΠΩΛΕΙΟ	ΦΡΟΥΤΑ	0.88	1	1
13	1002	1002424215071	90020210100	ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet	Leaflet	ΟΠΩΡΟΠΩΛΕΙΟ	ΦΡΟΥΤΑ	1.53	1	1
14	1002	1002424215071	90030160500	ΜΠΑΝΑΝΕΣ DOLE ΕΙΣΑΓΩΓΗΣ SAVENDISH	NA	ΟΠΩΡΟΠΩΛΕΙΟ	ΤΡΟΠΙΚΑ ΦΡΟΥΤΑ	1.67	1	1
15	1002	1002424215071	90057500400	365 ΠΙΠΕΡΙ ΜΑΥΡΟ ΟΛΟΚΛΗΡΟ 50GR.	NA	ΠΑΝΤΟΠΩΛΕΙΟ	ΑΡΤΥΜΑΤΑ	1.21	1	1
16	1002	1002424215071	90110750200	CRUNCH MILK CHOCOLATE 100GR	NA	ΖΑΧΑΡΩΔΗ	ΣΟΚΟΛΑΤΟΕΙΔΗ	2.32	2	1
17	1002	1002424215071	90178380000	NESCAFE DOLCE GUSTO CHOCOCINO 256G	NA	ΠΑΝΤΟΠΩΛΕΙΟ	ΡΟΦΗΜΑΤΑ	4.98	1	1
18	1002	1002424215071	90181755000	ΑΒ ΧΑΡΤ/ΤΑ ΛΕΥΚΗ 30Χ30 100Φ	NA	ΧΑΡΤΙΑ TISSUE	ΧΑΡΤΟΠΕΤΣΕΤΕΣ	1.14	2	1
19	1002	1002424215071	90315011400	ΑΒ ΣΑΛΑΤΑ ΚΑΠΡΙΤΣΙΟΖΑ 160 ΓΡ	NA	ΟΠΩΡΟΠΩΛΕΙΟ	ΕΤΟΙΜΕΣ ΣΑΛΑΤΕΣ	1.49	1	1
20	1002	1002424215071	91580160500	MELISSA ΜΑΚΑΡΟΝΙΑ Ν5 500GR	NA	ΠΑΝΤΟΠΩΛΕΙΟ	ΖΥΜΑΡΙΚΑ	1.96	2	1

Εικόνα 16 Υποσύνολο Εβδομάδας 1

CustomerID	transID	MasterCode	Desc	PromoSKU	Dept	Categ	Value	Quantity	Visits	
1	1002	10021829216441	40713059600	ΑΒ ΕΙΔΙ 400ML	NA	ΠΑΝΤΟΠΩΛΕΙΟ	ΑΡΤΥΜΑΤΑ	0.49	1	1
2	1002	10021829216441	40713323600	NESC.DOLCE GUSTO ESPR.CARAM 83G	NA	ΠΑΝΤΟΠΩΛΕΙΟ	ΡΟΦΗΜΑΤΑ	5.26	1	1
3	1002	10021829216441	40713348200	ΑΒ ΑΥΓΑ ΕΛ.ΒΟΣΚΗΣ 6ΑΔΑ LARGE 63 73G Leaflet	Leaflet	ΓΑΛΑΚΤΟΚΟΜΙΚΑ-ΑΥΓΑ-ΦΡΕΣΚΟΙ ΧΥΜΟΙ	ΑΥΓΑ	2.46	1	1
4	1002	10021829216441	40713927900	ΧΕΝΙΑ ΠΑΣΤΑ ΕΛΙΑΣ ΚΑΛΑΜ. 185GR 35%	NA	ΠΑΝΤΟΠΩΛΕΙΟ	ΕΛΙΕΣ	2.31	1	1
5	1002	10021829216441	40714000100	JOHNSOF ΚΡΟΥΤΟΝ ΤΟΜΑΤΑ ΠΑΠΡΙΚΑ 160G	NA	ΠΑΝΤΟΠΩΛΕΙΟ	ΣΟΥΠΕΣ-CROUTONS	2.62	1	1
6	1002	10021829216441	40722278000	ΣΕΡΤΟΝΑ Β. ΜΩΡΟΜ.ΧΑΜ.64 Τ 0.50Ε	NA	ΥΓΙΕΙΝΗ - ΟΜΟΡΦΙΑ	ΠΡΟΣΩΠΙΚΗ ΥΓΙΕΙΝΗ	2.28	1	1
7	1002	10021829216441	40722281000	ΑΒ ΠΑΤΑΤΑ ΝΑΞΟΥ ΠΓΕ ΣΥΣΚ.	NA	ΟΠΩΡΟΠΩΛΕΙΟ	ΠΑΤΑΤΕΣ	2.37	1	1
8	1002	10021829216441	40722557000	PRODERM ΣΑΜΠ & ΑΦΡ 0 ΕΩΣ 12 ΜΗΝΩΝ	NA	ΥΓΙΕΙΝΗ - ΟΜΟΡΦΙΑ	ΦΡΟΝΤΙΔΑ ΜΑΛΛΙΩΝ	2.92	1	1
9	1002	10021829216441	40723193900	ΑΡΙΣΤΟΝ ΑΛΑΤΙ ΜΕΣΟΛΟΓ. ΦΙΑΛΗ 400ΓΡ	NA	ΠΑΝΤΟΠΩΛΕΙΟ	ΑΡΤΥΜΑΤΑ	1.70	1	1
10	1002	10021829216441	40723234400	ΑΒ ΓΑΛΑ ΥΨ.ΘΕΡΜΙΚΗΣ ΕΠΕΞ. ΠΛΗΡΕΣ 1	NA	ΓΑΛΑΚΤΟΚΟΜΙΚΑ-ΑΥΓΑ-ΦΡΕΣΚΟΙ ΧΥΜΟΙ	ΦΡΕΣΚΟ & ΥΨΗΛΗΣ ΠΑΣΤΕΡΙΩΣΗΣ ΓΑΛΑ	4.76	4	1
11	1002	10021829216441	40724146900	TENA PANTS MEDIUM 14TEM Leaflet	Leaflet	ΥΓΙΕΙΝΗ - ΟΜΟΡΦΙΑ	ΠΡΟΣΩΠΙΚΗ ΥΓΙΕΙΝΗ	26.80	2	1
12	1002	10021829216441	40724540500	ΦΕΤΑ ΤΡΙΠΟΛΕΩΣ ΤΥΠ.	NA	ΤΥΡΙΑ	ΦΕΤΕΣ	4.57	1	1
13	1002	10021829216441	40725191900	TUBORG SODA ΛΕΜΟΝΙ Κ/Τ 6Χ330ML (5+1	NA	LIQUIDE GROCERY FOOD	ΑΝΑΨΥΚΤΙΚΑ	2.85	1	1
14	1002	10021829216441	40725212700	365 ΑΡΑΒΟΣΙΤΕΛΑΙΟ 1LT	NA	ΠΑΝΤΟΠΩΛΕΙΟ	ΣΠΟΡΕΛΑΙΑ - ΜΑΓ.ΛΙΠΗ	1.79	1	1
15	1002	10021829216441	40728360400	AGRINO BELLA ΡΥΖΙ ΕΛΛΗΝΙΚΟ 500GR	NA	ΠΑΝΤΟΠΩΛΕΙΟ	ΡΥΖΙΑ	1.87	1	1
16	1002	10021829216441	40728364300	AGRINO ΛΑΙΣ ΡΥΖΙ ΕΛΛΗΝΙΚΟ 500GR	NA	ΠΑΝΤΟΠΩΛΕΙΟ	ΡΥΖΙΑ	1.73	1	1
17	1002	10021829216441	40730129900	ΓΟΥΔΑ ΣΕ ΦΕΤΕΣ 400ΓΡ	NA	ΤΥΡΙΑ	ΦΡΕΣΚΑ & ΑΝΑΚΑΤΕΡΓΑΣΜΕΝΑ ΤΥΡΙΑ ΕΙΣΑΓΩΓΗΣ	3.00	1	1
18	1002	10021829216441	40730877000	ΑΒ ΚΑΣΤΑΝΗ ΖΑΧΑΡΗ 500ΓΡ	NA	ΠΑΝΤΟΠΩΛΕΙΟ	ΖΑΧΑΡΗ	1.49	1	1
19	1002	10021829216441	40734668800	ΝΟΥΝΟΥ ΕΒΑΓΓΟΡΕ ΓΕΥΣΕΙΣ ΣΠΙΤΙΚΕΣ 400	NA	ΠΑΝΤΟΠΩΛΕΙΟ	ΓΑΛΑ Μ.ΔΙΑΡΚΕΙΑΣ	0.85	1	1
20	1002	10021829216441	90007780000	MAGGI ΜΕΣΟΓΕΙΑΚΟΙ ΖΩΜΟΣ ΛΑΧ.140GR	NA	ΠΑΝΤΟΠΩΛΕΙΟ	ΠΟΥΡΕΔΕΣ-ΚΥΒΟΙ-ΜΕΙΓΜΑΤΑ	2.37	1	1
21	1002	10021829216441	90020080400	ΜΗΛΑ ΖΑΓΟΡΑΣ ΠΗΛΙΟΥ ΠΟΠ ΣΤΑΡΚΙΝΓΚ	NA	ΟΠΩΡΟΠΩΛΕΙΟ	ΦΡΟΥΤΑ	2.13	1	1
22	1002	10021829216441	90030160500	ΜΠΑΝΑΝΕΣ DOLE ΕΙΣΑΓΩΓΗΣ SAVENDISH	NA	ΟΠΩΡΟΠΩΛΕΙΟ	ΤΡΟΠΙΚΑ ΦΡΟΥΤΑ	1.35	1	1

Εικόνα 17 Υποσύνολο Εβδομάδας 2

Ξεκινώντας τη διαδικασία εισαγωγής των δεδομένων μας στον αλγόριθμο Arriori, πρέπει να τα μετατρέψουμε σε μια συγκεκριμένη μορφή, όπως κάναμε και στην προηγούμενη ενότητα.

Χρειαζόμαστε μόνο τα καλάθια των καταναλωτών που είναι κοινοί και στις δύο εβδομάδες και έχουν αγοράσει τουλάχιστον ένα προϊόν προσφοράς. Επομένως τα μόνα δεδομένα που χρειάζεται ο αλγόριθμος για να εξορύξει κανόνες, είναι το περιεχόμενο του κάθε καλάθιού σε επίπεδο προϊόντων. Μετατρέπουμε τα δεδομένα της πρώτης εβδομάδας από μορφή πίνακα σε λίστες συναλλαγών, έτσι ώστε να έχουμε όλα τα προϊόντα που αγοράζονται μαζί σε κάθε συναλλαγή σε μία εγγραφή, όπως κάναμε και στη Ενότητα 3.1.2. Η μορφή που χρειαζόμαστε είναι μία λίστα που αποτελείται από 1.402 εγγραφές, όσες είναι και οι συναλλαγές που πραγματοποιήθηκαν την πρώτη εβδομάδα.

Η Εικόνα 18 παρουσιάζει τις 10 πρώτες εγγραφές της λίστας της πρώτης εβδομάδας που αφορά τις συναλλαγές κάθε καταναλωτή που έχει αγοράσει τουλάχιστον 1 προϊόν προσφοράς. Κάθε γραμμή αναφέρεται σε κάθε καλάθι αγοράς.

	itemsWeek1
1	FAIRY ΥΓΡΟ ΠΙΑΤΩΝ ΛΕΜΟΝΙ 900ML,SANI PANTS MEDIUM ...
2	ΜΠ.ΡΕΤΙΤ ΒΕΥΡΡΕ ΣΟΚΟΛ 2Χ200ΓΡ 0.50Ε,ΜΙΡΑΝΤΑ ΣΟΚΟΛΑ...
3	ΤΥΡΙ ΝΟΥΝΟΥ ΓΟΥΔΑ ΟΛΛΑΝΔ.ΦΡΑΤΖΟΛΑ,ΟΥΖΟ ΠΛΩΜ...
4	LOTUS ΠΑΙΔΙΚΗ ΚΑΛΤΣΑ 207 ΑΓΟΡΙ 31-3,LOTUS ΠΑΙΔΙΚΗ ΚΑ...
5	FAIRY ΚΑΝΟΝΙΚΟ 900ML,ΣΥΝΕΡΓΑΣΙΑ ΧΡΥΣΗ ΚΡΗΤΗ ΜΕΛΙ4...
6	ΑΜΙΤΑ FUN ΚΟΚΤΕΙΛ ΚΑΡΟΤΟ 250ML,ΗΒΗ Φ.Χ.ΠΟΡΤΟΚΑΛΙ ...
7	7DAYS ΚΡΟΥΑΣΑΝ ΒΕΡΙΚΟΚΟ 70ΓΡ,ΒΛΑΧΑΣ ΓΑΛΑ ΕΒΑΠ.ΠΛ...
8	ΤΥΡΙ ΝΟΥΝΟΥ ΓΟΥΔΑ ΟΛΛΑΝΔ.ΦΡΑΤΖΟΛΑ,ΓΑΛΟΠΟΥΛΑ ...
9	ΜΑΣΤΙQUΑ ΑΝΘΡ/ΧΟ ΝΕΡΟ 4Χ330ΜL(3+1Δ,ΚΡΙΣ ΤΟΣΤΙΜΟ ...
10	ΚΟΤΟΠΟΥΛΟ ΨΗΤΟ ΜΕ ΚΑΡΥΚΕΥΜΑΤΑ ΕΛΛΗΝ,ΛΑCΤΑ ΣΟΚ...

Showing 1 to 10 of 1,402 entries

Εικόνα 18 Καλάθια Καταναλωτών Εβδομάδα 1

Την ίδια διαδικασία ακολουθούμε και για τα δεδομένα της δεύτερης εβδομάδας. Μετατρέπουμε σε μία λίστα τις 1.566 συναλλαγές των 1.144 χρηστών και κάθε γραμμή της λίστας αναφέρεται σε ένα καλάθι.

Στην Εικόνα 19 παρουσιάζονται οι 10 πρώτες εγγραφές τις λίστας.

itemsWeek2	
1	ΑΒ ΞΙΔΙ 400ML,NESC.DOLCE GUSTO ESPR.CARAM 83G,ΑΒ ΑΥ...
2	ΜΠ.PETIT BEURRE ΣΟΚΟΛ 2Χ200ΓΡ 0.50Ε,ΣΟΚΟΛΑΤΑ ΓΑΛΑΚ ...
3	ΟΡ.Β ΟΔ/ΤΣΑ COM.SENS.CLEAN 35S Leaflet,M&MS POUCH ...
4	LACTA ΣΟΚ/ΤΑ OREO ΦΡΑΟΥΛΑ 105Γ Leaflet,CRUNCH MILK...
5	ARLA ΤΥΡΙ REGATO ΤΡΙΜΜΑ 150GR,ΚΑΝΑΚΙ MINI CALZONE..
6	ΑΒ ΜΑΚΑΡΟΝΙΑ Ν2 (ΠΑΣΤΙΤΣΙΟ) 500ΓΡ,ΑΒ ΧΑΡΤΟΜ. FACIAL...
7	ΓΛΥΚΟΕΛΙΕΣ ΚΑΛΑΜΩΝ ΛΙΓΟ ΑΛΑΣ ΕΛΛΑΔ,ΒΙΚΟΣ COLA Φ/...
8	LACTA ΓΚΟΦΡΕΤΑ 28.5ΓΡ,ΚΡΙΣ ΤΟΣΤΙΜΟ ΣΤΑΡΕΝΙΟ 800G 0.4...
9	MASTIQUA ΑΝΘΡ/ΧΟ ΝΕΡΟ 4Χ330ML(3+1Δ,LIPTON Π.ΤΣ.Π...
10	ΟΛΥΜΠΟΣ ΤΑΧΙΝΙ 300GR ΒΑΖΟ,SEPTONA MEDICARE ΥΠΟΣ ...

Showing 1 to 10 of 1,566 entries

Εικόνα 19 Καλάθια Καταναλωτών Εβδομάδα 2

Τέλος, για τη κάθε μία εβδομάδα δημιουργούμε ένα αρχείο csv, όπου τα προϊόντα σε κάθε καλάθι χωρίζονται με «,». Η Εικόνα 20 και η Εικόνα 21 δείχνουν τις 15 πρώτες εγγραφές των αρχείων CSV για την εβδομάδα 1 και 2 αντίστοιχα. Έπειτα παρουσιάζουμε τη σύνοψη του αρχείου CSV της πρώτης (Εικόνα 22) και της δεύτερης εβδομάδας (Εικόνα 23).

itemsWeek1
FAIRY ΥΓΡΟ ΠΙΑΤΩΝ ΛΕΜΟΝΙ 900ML,SANI PANTS MEDIU
ΜΠ.PETIT BEURRE ΣΟΚΟΛ 2Χ200ΓΡ 0.50Ε,ΜΙΡΑΝΤΑ ΣΟΚ
ΤΥΡΙ ΝΟΥΝΟΥ ΓΟΥΔΑ ΟΛΛΑΝΔ.ΦΡΑΤΖΟΛΑ,ΟΥΖΟ ΠΛΩΝ
LOTUS ΠΑΙΔΙΚΗ ΚΑΛΤΣΑ 207 ΑΓΟΡΙ 31-3,LOTUS ΠΑΙΔΙΚΗ Ι
FAIRY ΚΑΝΟΝΙΚΟ 900ML,ΣΥΝΕΡΓΑΣΙΑ ΧΡΥΣΗ ΚΡΗΤΗ ΜΕΛ
ΑΜΙΤΑ FUN ΚΟΚΤΕΙΛ ΚΑΡΟΤΟ 250ML,ΗΒΗ Φ.Χ.ΠΟΡΤΟΚΑ
7DAYS ΚΡΟΥΑΣΑΝ ΒΕΡΙΚΟΚΟ 70ΓΡ,ΒΛΑΧΑΣ ΓΑΛΑ ΕΒΑΠ.Γ
ΤΥΡΙ ΝΟΥΝΟΥ ΓΟΥΔΑ ΟΛΛΑΝΔ.ΦΡΑΤΖΟΛΑ,ΓΑΛΟΠΟΥΛΑ
MASTIQUA ΑΝΘΡ/ΧΟ ΝΕΡΟ 4Χ330ML(3+1Δ,ΚΡΙΣ ΤΟΣΤΙΜ
ΚΟΤΟΠΟΥΛΟ ΨΗΤΟ ΜΕ ΚΑΡΥΚΕΥΜΑΤΑ ΕΛΛΗΝ,ΛΑCTA ΣC
NESCAFE CARPUCCINO 10ΦΑΚ.Χ 14G,ΑJAX ULTRA ΛΕΒΑΙ
ΤΥΡΙ ΝΟΥΝΟΥ ΓΟΥΔΑ ΟΛΛΑΝΔ.ΦΡΑΤΖΟΛΑ,SANI PANTS
ΤΥΡΙ ΝΟΥΝΟΥ ΓΟΥΔΑ ΟΛΛΑΝΔ.ΦΡΑΤΖΟΛΑ,COCA COLA
SANI PANTS LARGE ΝΟ3 14Τ,ΦΙΛΕΤΟ ΣΤΗΘΟΣ ΚΟΤ/ΛΟ Ε/
ΔΕΣΜΕΥΤΗΣ ΧΡΩΜΑΤ.& ΒΡΩΜΙΑΣ 10Φ+5ΦΔΩ Leaflet,Κ2Ι

Εικόνα 20 Αρχείο CSV Εβδομάδα 1

itemsWeek2
ΑΒ ΞΙΔΙ 400ML,NESC.DOLCE GUSTO ESPR.CARAM 83G,Α
ΜΠ.PETIT BEURRE ΣΟΚΟΛ 2Χ200ΓΡ 0.50Ε,ΣΟΚΟΛΑΤΑ Γ.
ΟΡ.Β ΟΔ/ΤΣΑ COM.SENS.CLEAN 35S Leaflet,M&MS ΡΟ
LACTA ΣΟΚ/ΤΑ OREO ΦΡΑΟΥΛΑ 105Γ Leaflet,CRUNCH
ARLA ΤΥΡΙ REGATO ΤΡΙΜΜΑ 150GR,ΚΑΝΑΚΙ MINI CALZ
ΑΒ ΜΑΚΑΡΟΝΙΑ Ν2 (ΠΑΣΤΙΤΣΙΟ) 500ΓΡ,ΑΒ ΧΑΡΤΟΜ. F.
ΓΛΥΚΟΕΛΙΕΣ ΚΑΛΑΜΩΝ ΛΙΓΟ ΑΛΑΣ ΕΛΛΑΔ,ΒΙΚΟΣ COL
LACTA ΓΚΟΦΡΕΤΑ 28.5ΓΡ,ΚΡΙΣ ΤΟΣΤΙΜΟ ΣΤΑΡΕΝΙΟ 80C
MASTIQUA ΑΝΘΡ/ΧΟ ΝΕΡΟ 4Χ330ML(3+1Δ,LIPTON Π.
ΟΛΥΜΠΟΣ ΤΑΧΙΝΙ 300GR ΒΑΖΟ,SEPTONA MEDICARE Υ
MOZZARELLA ΑΥΣΤΡΙΑΣ 40% ΦΡΑΝΤΖΟΛΑ,ΑΒ ΦΡΕΣΚΟ
KLEENEX CARITONNE Χ/ΥΓ.(10+2ΔΩΡΟ) Leaflet,ΙΟΝ ΣC
COLGATE SENPROR REPAIR&PREVENT 75M,ARIEL ΣΚΟΙ
ΔΕΛΤΑ ΜΙΚΡΕΣ ΟΙΚ.ΦΑΡΜΕΣ ΕΛΑΦΡΥ 1LT,ΒΟΕΙΟ ΧΤΕΝ
COCA COLA 1LT,LOREAL REVITALIFT FILLER ΗΜΕΡΑΣ 50

Εικόνα 21 Αρχείο CSV Εβδομάδα 2

transactions as itemMatrix in sparse format with
1402 rows (elements/itemsets/transactions) and
5918 columns (items) and a density of 0.002692528

most frequent items:

ΜΠΑΝΑΝΕΣ DULCE ΒΥΟΝΑ ΕΙΣΑΓΩΓΗΣ	ΜΠΑΝΑΝΕΣ DOLE ΕΙΣΑΓΩΓΗΣ	CAVENDISH	ΤΟΜΑΤΕΣ ΧΥΜΑ ΕΛΛΗΝΙΚΕΣ
157	148		113
ΚΡΕΜ. ΞΕΡΑ ΧΟΝΔΡΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ	ΛΟΥΜΙΔΗΣ ΚΑΦΕΣ	ΠΑΠΑΓ.ΠΑΡΑΔΟΣ.194GR	(Other)
110	105		21707

element (itemset/transaction) length distribution:
sizes

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39		
40	41	42	43	44	45	46																																		
11	31	46	53	54	72	85	66	80	66	68	60	60	65	42	39	44	33	28	24	23	36	20	24	25	15	24	7	21	13	12	15	11	11	10	11	10	7	6		
6	7	5	6	4	3	6																																		
47	48	49	50	51	52	53	54	55	57	60	63	64	68	69	76																									
5	2	2	7	4	2	3	1	1	3	1	1	2	1	1	1																									

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.00	7.00	13.00	15.93	22.00	76.00

includes extended item information - examples:

	Labels
1	1 ΚΟΝΤΡΑ ΜΟΣΧ.ΓΑΛΑΚΤ. Μ/Ο ΟΛΛ.ΒΟΧ
2	1 ΟΣΣΟΜΠΟΥΚΟ ΜΟΣΧ.Μ/Ο ΓΑΛΑΚΤ.ΟΛΛ. Leaflet
3	1/2 ΚΟΤΟΠΟΥΛΟ.ΒΙΟΛΟΓΙΚΟ ΝΩΠΟ

Εικόνα 22 Σύνοψη Εβδομάδας 1

transactions as itemMatrix in sparse format with
1566 rows (elements/itemsets/transactions) and
6027 columns (items) and a density of 0.002430209

most frequent items:

ΜΠΑΝΑΝΕΣ DULCE ΒΥΟΝΑ ΕΙΣΑΓΩΓΗΣ	ΜΠΑΝΑΝΕΣ DOLE ΕΙΣΑΓΩΓΗΣ	CAVENDISH	ΤΟΜΑΤΕΣ ΧΥΜΑ ΕΛΛΗΝΙΚΕΣ
185	155		128
ΠΑΤΑΤΕΣ ΧΟΝΔΡΕΣ ΕΛΛΗΝΙΚΕΣ	ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ		(Other)
113	105		22251

element (itemset/transaction) length distribution:
sizes

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	
15	34	67	60	69	104	94	81	80	72	78	85	65	57	58	55	45	27	36	31	26	31	31	27	20	20	14	17	
29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	51	52	53	54	55	56	57	
16	12	9	10	13	12	5	8	6	7	9	7	7	4	5	6	3	3	3	3	3	2	2	2	3	1	1	1	2
59	63	64	65	82																								
1	1	1	1	1																								

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.00	7.00	12.00	14.65	19.00	82.00

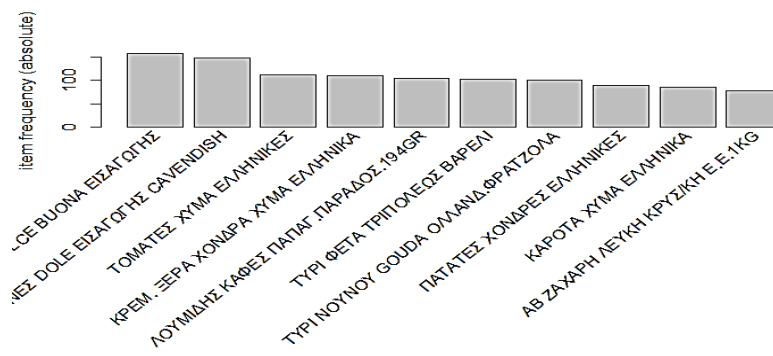
includes extended item information - examples:

	Labels
1	1 ΕΛΙΑ ΝΕΑΡΟΥ ΜΟΣΧ.Α/Ο 12/ΜΗΝ ΟΛΛ.
2	1 ΚΟΝΤΡΑ ΜΟΣΧ.ΓΑΛΑΚΤ. Μ/Ο ΟΛΛ.ΒΟΧ
3	1 ΣΠΑΛΑ ΝΕΑΡΟΥ ΜΟΣΧ Α/Ο 12/ΜΗΝ ΟΛΛ

Εικόνα 23 Σύνοψη Εβδομάδα 2

Παρατηρούμε από τις συνόψεις και των δύο εβδομάδων κάποια βασικά στατιστικά στοιχεία. Πιο συγκεκριμένα, όπως αναμέναμε στην πρώτη εβδομάδα, οι συναλλαγές που πραγματοποίησαν οι 1.144 καταναλωτές είναι 1.402 και στα καλάθια τους υπήρχαν συνολικά 5.918 διαφορετικά προϊόντα. Από την άλλη μεριά την δεύτερη εβδομάδα, οι ίδιοι καταναλωτές πραγματοποίησαν 1.566 συναλλαγές και βρέθηκαν στα καλάθια τους 6.027 διαφορετικά προϊόντα. Το σύνολο των δεδομένων και στις δύο εβδομάδες δεν είναι πολύ αραιό, με πυκνότητα μόλις πάνω από 2%. Η κατανομή των δεδομένων και στις δύο εβδομάδες είναι λοξή προς τα δεξιά γιατί η μέση τιμή είναι μεγαλύτερη από τη διάμεσο.

Παρατηρώντας τη κατανομή των συναλλαγών ανά πλήθος προϊόντων, 15 συναλλαγές περιείχαν μόνο 1 προϊόν, 34 συναλλαγές περιείχαν 2 προϊόντα κοκ. Επιπρόσθετα παρατηρούμε ότι και στις 2 εβδομάδες τα προϊόντα «ΜΠΑΝΑΝΕΣ DULCA BUONA ΕΙΣΑΓΩΓΗΣ», «ΜΠΑΝΑΝΕΣ DOLE ΕΙΣΑΓΩΓΗΣ CAVEDISH» και οι «ΤΟΜΑΤΕΣ ΧΥΜΑ ΕΛΛΗΝΙΚΕΣ» είναι τα 3 προϊόντα με τη μεγαλύτερη συχνότητα εμφάνισης στις συναλλαγές. Παρακάτω, αναπαριστούμε γραφικά τα 10 προϊόντα με τη μεγαλύτερη συχνότητα εμφάνισης την πρώτη εβδομάδα (Εικόνα 24) και τη δεύτερη εβδομάδα(Εικόνα 25).



Εικόνα 24 Προϊόντα Μεγαλύτερης Συχνότητας Εμφάνισης Εβδομάδα 1



Εικόνα 25 Προϊόντα Μεγαλύτερης Συχνότητας Εμφάνισης Εβδομάδα 2

3.2.4 Εξόρυξη Κανόνων Συσχέτισης για τα Δεδομένα της 1^{ης} Εβδομάδας

Έχοντας μετατρέψει τα δεδομένα μας, (Εικόνα 20) και (Εικόνα 21), ο αλγόριθμος θα μας επιστρέψει τους κανόνες συσχέτισης που ικανοποιούν το όριο της υποστήριξης που θέτουμε.

Στη συγκεκριμένη περίπτωση, θέλουμε να μελετήσουμε τα προϊόντα που είναι σε προσφορά με ποια άλλα προϊόντα αγοράζονται την πρώτη και την δεύτερη εβδομάδα που μελετούμε. Για αυτό τον λόγο, θέλουμε το ένα εκ των δύο μελών των κανόνων να περιέχει μόνο προϊόντα προσφοράς.

Ορίζουμε λοιπόν μέσω της συνάρτησης του αλγορίθμου, στο δεξί μέλος των κανόνων, να υπάρχουν μόνο προϊόντα προσφοράς. Ξεκινάμε με την ανάλυση των καλαθιών των καταναλωτών της πρώτης εβδομάδας. Θέτουμε κατώφλι υποστήριξης ίσο με 0,004 και εμπιστοσύνη ίση με 0,00 ταξινομώντας τους κανόνες κατά φθίνουσα σειρά με βάση το μέτρο lift.

Δημιουργούνται 161 κανόνες συσχέτισης από τους οποίους μόνο οι 59 είναι αξιολογήσιμοι γιατί έχουν lift>1. Πιο συγκριμένα οι 55 από τους 59 είναι μήκους 2-στοιχειοσυνόλων και οι 4 είναι μήκους 3-στοιχειοσυνόλων. Η Εικόνα 26 παρουσιάζει κάποια βασικά στοιχεία των κανόνων που εξόρυξε ο αλγόριθμος και οι Εικόνες 27 και 28, παρουσιάζουν τους 161 κανόνες.

set of 161 rules

rule length distribution (lhs + rhs):sizes

1	2	3
102	55	4

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.000	1.000	1.000	1.391	2.000	3.000

summary of quality measures:

support	confidence	lift	count
Min. :0.004280	Min. :0.004280	Min. : 1.000	Min. : 6.00
1st Qu.:0.004993	1st Qu.:0.005706	1st Qu.: 1.000	1st Qu.: 7.00
Median :0.005706	Median :0.011412	Median : 1.000	Median : 8.00
Mean :0.008023	Mean :0.060177	Mean : 2.065	Mean :11.25
3rd Qu.:0.009272	3rd Qu.:0.081395	3rd Qu.: 2.291	3rd Qu.:13.00
Max. :0.047076	Max. :0.818182	Max. :17.380	Max. :66.00

mining info:

data ntransactions	support	confidence
tr1	1402	0.004
		0

Εικόνα 26 Σύνοψη Κανόνων Εβδομάδα 1

	lhs	rhs	support	confidence	lift	count
[1]	{ΛΑΧΑΝΟ ΚΟΚΚΙΝΟ ΕΛΛΗΝΙΚΟ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.006419401	0.818181818	17.380165	9
[2]	{ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ, ΚΡΕΜ. ΞΕΡΑ ΧΟΝΔΡΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ}	=> {ΚΟΥΝΟΥΠΙΔΙ ΕΛΛΗΝΙΚΟ Leaflet}	0.004279601	0.333333333	15.075269	6
[3]	{ΣΠΑΝΑΚΙ ΕΛΛΗΝΙΚΟ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004279601	0.600000000	12.745455	6
[4]	{ΑΝΗΘΟΣ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΟΣ 80ΓΡ, ΜΑΙΝΤΑΝΟΣ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΟΣ 80ΓΡ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004279601	0.375000000	7.965909	6
[5]	{ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ, ΜΑΡΟΥΛΙΑ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΑ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004279601	0.352941176	7.497326	6
[6]	{ΚΡΕΜ. ΞΕΡΑ ΧΟΝΔΡΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ, ΜΠΑΝΑΝΕΣ DOLE ΕΙΣΑΓΩΓΗΣ SAUVENDISH}	=> {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004279601	0.285714286	6.460829	6
[7]	{ΠΡΑΣΣΑ ΕΛΛΗΝΙΚΑ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004279601	0.285714286	6.069264	6
[8]	{ΛΕΜΟΝΙΑ ΕΛΛΗΝΙΚΑ}	=> {ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.004992867	0.142857143	5.563492	7
[9]	{ΑΒ ΧΑΡΤ/ΤΑ ΛΕΥΚΗ 30Χ30 100φ}	=> {ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.004279601	0.136363636	5.310606	6
[10]	{ΑΓΓΟΥΡΙΑ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΑ}	=> {ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.004992867	0.134615385	5.242521	7
[11]	{365 ΧΑΡΤ/ΤΕΣ ΛΕΥΚΕΣ 28Χ30 80φ}	=> {ΑΒ ΦΑΚΕΣ ΥΪΛΕΣ 500G ΕΙΣΑΓΩΓΗΣ Leaflet}	0.004279601	0.077922078	5.202226	6
[12]	{ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.014978602	0.244186047	5.187104	21
[13]	{ΜΑΙΝΤΑΝΟΣ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΟΣ 80ΓΡ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.006419401	0.230769231	4.902098	9
[14]	{ΠΑΤΑΤΕΣ ΧΟΝΔΡΕΣ ΚΥΠΡΟΥ ΣΥΣ/ΝΕΣ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.006419401	0.230769231	4.902098	9
[15]	{ΚΡΕΜ. ΞΕΡΑ ΧΟΝΔΡΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ}	=> {ΚΟΥΝΟΥΠΙΔΙ ΕΛΛΗΝΙΚΟ Leaflet}	0.007845934	0.100000000	4.522581	11
[16]	{ΚΟΥΝΟΥΠΙΔΙ ΕΛΛΗΝΙΚΟ Leaflet}	=> {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004279601	0.193548387	4.376691	6
[17]	{ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	=> {ΚΟΥΝΟΥΠΙΔΙ ΕΛΛΗΝΙΚΟ Leaflet}	0.004279601	0.096774194	4.376691	6
[18]	{ΜΑΡΟΥΛΙΑ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΑ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.008559201	0.200000000	4.248485	12
[19]	{ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ}	=> {ΚΟΥΝΟΥΠΙΔΙ ΕΛΛΗΝΙΚΟ Leaflet}	0.005706134	0.093023256	4.207052	8
[20]	{ΜΗΛΑ ΣΤΑΡΚΙΝΓΚ ΕΛΛΗΝΙΚΑ Ν.ΕΣΟΔΕΙΑΣ}	=> {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004992867	0.175000000	3.957258	7
[21]	{ΤΟΜΑΤΕΣ ΧΥΜΑ ΕΛΛΗΝΙΚΕΣ}	=> {ΛΑΚΤΑ ΣΟΚ. TRIPLE ΚΑΡΑΜΕΛΑ 90G Leaflet}	0.004279601	0.053097345	3.722124	6
[22]	{ΚΡΕΜΜΥΔΙΑ ΧΛΩΡΑ ΕΛΛΗΝΙΚΑ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004992867	0.175000000	3.717424	7
[23]	{ΤΟΜΑΤΕΣ ΧΥΜΑ ΕΛΛΗΝΙΚΕΣ}	=> {ΚΟΥΝΟΥΠΙΔΙ ΕΛΛΗΝΙΚΟ Leaflet}	0.006419401	0.079646018	3.620055	9
[24]	{ΠΑΤΑΤΕΣ ΧΟΝΔΡΕΣ ΕΛΛΗΝΙΚΕΣ}	=> {ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.005706134	0.088888889	3.461728	8
[25]	{ΑΓΓΟΥΡΑΚΙΑ Τ. ΚΝΩΣΟΥ ΧΥΜΑ ΕΛΛΗΝΙΚΑ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004279601	0.162162162	3.444717	6
[26]	{GOUDA ΓΕΡΜΑΝΙΑΣ ΦΡΑΤΖΟΛΑ}	=> {ΚΡΙΣ ΤΟΣΤΙΜΟ ΣΤΑΡΕΝΙΟ 800G 0.45E Leaflet}	0.004279601	0.085714286	3.433469	6
[27]	{ΑΝΗΘΟΣ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΟΣ 80ΓΡ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.005706134	0.160000000	3.398788	8
[28]	{ΚΟΛΟΚΥΘΙΑ ΕΛΛΗΝΙΚΑ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.006419401	0.157894737	3.354067	9
[29]	{ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ}	=> {ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.004992867	0.081395349	3.169897	7
[30]	{ΑΧΛΑΔΙΑ ΚΡΥΣΤΑΛΙΑ ΕΛΛΗΝΙΚΑ}	=> {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.005706134	0.133333333	3.015054	8
[31]	{ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ}	=> {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.007845934	0.127906977	2.892348	11
[32]	{ΑΓΓΟΥΡΙΑ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΑ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004992867	0.134615385	2.859557	7
[33]	{ΚΡΕΜ. ΞΕΡΑ ΧΟΝΔΡΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ}	=> {ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.005706134	0.072727273	2.832323	8
[34]	{ΤΟΜΑΤΕΣ ΧΥΜΑ ΕΛΛΗΝΙΚΕΣ}	=> {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.009272468	0.115044248	2.601484	13
[35]	{ΛΟΥΜΙΔΗΣ ΚΑΦΕΣ ΠΑΠΑΓ.ΠΑΡΑΔΟΣ.194GR}	=> {ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.004992867	0.066666667	2.596296	7
[36]	{ΑΧΛΑΔΙΑ ΚΡΥΣΤΑΛΙΑ ΕΛΛΗΝΙΚΑ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004992867	0.116666667	2.478283	7
[37]	{ΚΡΕΜ. ΞΕΡΑ ΧΟΝΔΡΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ}	=> {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.008559201	0.109090909	2.466862	12
[38]	{ΜΠΑΝΑΝΕΣ DOLE ΕΙΣΑΓΩΓΗΣ SAUVENDISH}	=> {ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.006419401	0.060810811	2.368243	9
[39]	{ΕΛΛ.ΚΟΤΟΠΟΥΛΟ ΝΩΠΟ ΧΥΜΑ PR}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.005706134	0.111111111	2.360269	8
[40]	{ΜΠΑΝΑΝΕΣ DULCE BUONA ΕΙΣΑΓΩΓΗΣ}	=> {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.011412268	0.101910828	2.304500	16
[41]	{ΤΥΡΙ ΦΕΤΑ ΤΡΙΠΟΛΕΩΣ ΒΑΡΕΛΙ}	=> {ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.004279601	0.058823529	2.290850	6
[42]	{ΜΑΡΟΥΛΙΑ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΑ}	=> {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004279601	0.100000000	2.261290	6
[43]	{ΤΟΜΑΤΕΣ ΧΥΜΑ ΕΛΛΗΝΙΚΕΣ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.008559201	0.106194690	2.255833	12
[44]	{ΜΠΑΝΑΝΕΣ DOLE ΕΙΣΑΓΩΓΗΣ SAUVENDISH}	=> {ΚΟΥΝΟΥΠΙΔΙ ΕΛΛΗΝΙΚΟ Leaflet}	0.004992867	0.047297297	2.139058	7
[45]	{ΚΑΡΑΜ.ΤΟΣΤ ΣΤΑΡΕΝΙΟ 680G 0.60E ΦΘΗΝ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004279601	0.100000000	2.124242	6
[46]	{ΤΟΜΑΤΕΣ ΧΥΜΑ ΕΛΛΗΝΙΚΕΣ}	=> {ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.004279601	0.053097345	2.067847	6
[47]	{ΜΠΑΝΑΝΕΣ DULCE BUONA ΕΙΣΑΓΩΓΗΣ}	=> {ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.005706134	0.050955414	1.984430	8
[48]	{ΚΡΕΜ. ΞΕΡΑ ΧΟΝΔΡΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.007132668	0.090909091	1.931129	10
[49]	{ΠΑΤΑΤΕΣ ΧΟΝΔΡΕΣ ΕΛΛΗΝΙΚΕΣ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.005706134	0.088888889	1.888215	8
[50]	{ΤΥΡΙ ΦΕΤΑ ΤΡΙΠΟΛΕΩΣ ΒΑΡΕΛΙ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.006419401	0.088235294	1.874332	9
[51]	{GOUDA ΓΕΡΜΑΝΙΑΣ ΦΡΑΤΖΟΛΑ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004279601	0.085714286	1.820779	6
[52]	{ΠΑΤΑΤΕΣ ΧΟΝΔΡΕΣ ΕΛΛΗΝΙΚΕΣ}	=> {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004992867	0.077777778	1.758781	7
[53]	{ΑΒ ΖΑΧΑΡΗ ΛΕΥΚΗ ΚΡΥΣ/ΚΗ Ε.Ε.ΙΚG}	=> {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004279601	0.075949367	1.717436	6
[54]	{ΜΠΑΝΑΝΕΣ DULCE BUONA ΕΙΣΑΓΩΓΗΣ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.008559201	0.076433121	1.623625	12
[55]	{ΛΟΥΜΙΔΗΣ ΚΑΦΕΣ ΠΑΠΑΓ.ΠΑΡΑΔΟΣ.194GR}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.005706134	0.076190476	1.618470	8
[56]	{ΤΥΡΙ ΦΕΤΑ ΤΡΙΠΟΛΕΩΣ ΒΑΡΕΛΙ}	=> {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004992867	0.068627451	1.551866	7
[57]	{ΛΟΥΜΙΔΗΣ ΚΑΦΕΣ ΠΑΠΑΓ.ΠΑΡΑΔΟΣ.194GR}	=> {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004992867	0.066666667	1.507527	7
[58]	{ΜΠΑΝΑΝΕΣ DOLE ΕΙΣΑΓΩΓΗΣ SAUVENDISH}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.007132668	0.067567568	1.435299	10
[59]	{ΜΠΑΝΑΝΕΣ DOLE ΕΙΣΑΓΩΓΗΣ SAUVENDISH}	=> {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.006419401	0.060810811	1.375109	9
[60]	{}	=> {GILLET ΕΥΡ.ΜΗΧ.ΒLUE II SLALOM ST. TV}	0.004279601	0.004279601	1.000000	6
[61]	{}	=> {FRULITE 1LT ΦΡΟΥΤ/ΤΟ ΜΑΝΤΑΡ/ΣΑΓΚΝΙ Leaflet}	0.004279601	0.004279601	1.000000	6
[62]	{}	=> {ΑΓΕΛΑΔΙΤΣΑ ΦΑΓΕ 3Χ200ΓΡ Leaflet}	0.004992867	0.004992867	1.000000	7
[63]	{}	=> {LENOR MOONLIGHT HARMONY 26ΜΕΖ Leaflet}	0.004279601	0.004279601	1.000000	6
[64]	{}	=> {ΤΥΡΙ MILNER ΣΕ ΦΕΤΕΣ 175ΓΡ Leaflet}	0.004992867	0.004992867	1.000000	7
[65]	{}	=> {ΠΑΥΛΙΔΗΣ ΥΓΕΙΑΣ ΓΕΜ. ΚΑΚΑΟ 100GR Leaflet}	0.007845934	0.007845934	1.000000	11

Εικόνα 27 Κανόνες Συσχέτισης Εβδομάδα 1

[66]	{}	=> {ΚΡΙΣ ΚΡΙΣ ΤΟΣΤΙΜΟ ΜΠΡΙΟΣ 400G Leaflet}	0.004279601	0.004279601	1.000000	6
[67]	{}	=> {ΠΑΥΛΙΔΗΣ ΣΟΚ/ΤΑ ΥΓΕΙΑΣ ΛΕΜΟΝΙ 100G Leaflet}	0.004992867	0.004992867	1.000000	7
[68]	{}	=> {NATURA ΚΡ/ΝΟ ΥΓ.ΠΡΟΣ.750ML(0.55E) Leaflet}	0.005706134	0.005706134	1.000000	8
[69]	{}	=> {ΣΠΑΛΑ Μ/Ο ΧΟΙΡΙΝΗ ΕΛΛΗΝΙΚΗ Leaflet}	0.004279601	0.004279601	1.000000	6
[70]	{}	=> {ΙΟΝ BREAK ΣΟΚΟΛΑΤΑ ΦΟΥΝΤΟΥΚΙ 85GR Leaflet}	0.004992867	0.004992867	1.000000	7
[71]	{}	=> {LACTA ΣΟΚ. ΦΡΑΟΥΛΑ ΓΙΑΟΥΡΤΙ 100GR Leaflet}	0.005706134	0.005706134	1.000000	8
[72]	{}	=> {WETTEX ΣΦΟΥΓΓΑΡΙΣΤΡΑ Leaflet}	0.004279601	0.004279601	1.000000	6
[73]	{}	=> {LURPAK LIGHTER SOFT ME ΕΛΑΙΟΛΑ 225g Leaflet}	0.004279601	0.004279601	1.000000	6
[74]	{}	=> {NEPO ΚΟΡΠΗ 500ML 6PACK Leaflet}	0.004992867	0.004992867	1.000000	7
[75]	{}	=> {OB 30W LUXE GLAM.SHINE 75ML(1+1) Leaflet}	0.004992867	0.004992867	1.000000	7
[76]	{}	=> {ΠΑΥΛΙΔΗΣ ΣΟΚ.ΥΓΕΙΑΣ ΚΑΠΟΥΤΣΙΝΟ 100G Leaflet}	0.004279601	0.004279601	1.000000	6
[77]	{}	=> {ZEWA W&W EXTRA LANG X/K 2PLY 2ΡΟΛ Leaflet}	0.004279601	0.004279601	1.000000	6
[78]	{}	=> {LURPAK LIGHTER SOFT ΑΝΑΛΑΤΟ 225GR Leaflet}	0.006419401	0.006419401	1.000000	9
[79]	{}	=> {ΤΣΙΚΛΑ ORBIT ΔΥΟΣΜΟΣ ΜΠΟΥΚΑΛΙ 64GR Leaflet}	0.004279601	0.004279601	1.000000	6
[80]	{}	=> {AB ΤΖΑΤΖΙΚΙ 250GR Leaflet}	0.005706134	0.005706134	1.000000	8
[81]	{}	=> {AJAX ULTRA ΛΕΜΟΝΙ 1LT(+CRYSTAL) Leaflet}	0.004992867	0.004992867	1.000000	7
[82]	{}	=> {JOHNNIE WALKER 70CL RED LABEL SECUR Leaflet}	0.005706134	0.005706134	1.000000	8
[83]	{}	=> {ΛΑΤΙΣ SOFT 250GR Leaflet}	0.007845934	0.007845934	1.000000	11
[84]	{}	=> {ΑΦΕΛΑΔΙΤΣΑ 0% 3Χ200GR Leaflet}	0.005706134	0.005706134	1.000000	8
[85]	{}	=> {GILLETTE ΞΥΡ.Μ.ΒLUE II FIXED 5TM TV}	0.006419401	0.006419401	1.000000	9
[86]	{}	=> {ΣΟΛΟΜΟΣ ΦΕΤΕΣ ΥΔΑΤ.ΝΟΡΒΗΓΙΑΣ ΝΩΠΟΣ Leaflet}	0.009985735	0.009985735	1.000000	14
[87]	{}	=> {TOTAL 500GR Leaflet}	0.005706134	0.005706134	1.000000	8
[88]	{}	=> {LACTA TRIPLE ΣΟΚΟΛΑΤΑ 90G Leaflet}	0.006419401	0.006419401	1.000000	9
[89]	{}	=> {AB ΑΥΓΑ ΕΛ.ΒΟΣΚΗΣ 6ΑΔΑ LARGE 63 73G Leaflet}	0.009272468	0.009272468	1.000000	13
[90]	{}	=> {CREAM CRACKERS ΣΙΤΟΥ Ν018 140GR Leaflet}	0.004992867	0.004992867	1.000000	7
[91]	{}	=> {LACTA ΣΟΚΟΛΑΤΑ ΚΡΑΚΕΡ 87GR Leaflet}	0.006419401	0.006419401	1.000000	9
[92]	{}	=> {BARILLA ΜΑΚΕΡΟΤΣΙΝΙ Ν10 500GR Leaflet}	0.004279601	0.004279601	1.000000	6
[93]	{}	=> {ΠΑΥΛ.ΥΓΕΙΑΣ 70% ΑΛΑΤ.ΑΜΥΓΔ. 100G Leaflet}	0.004992867	0.004992867	1.000000	7
[94]	{}	=> {ΙΟΝ BREAK ΣΟΚΟΛΑΤΑ ΛΕΥΚΗ 85GR Leaflet}	0.005706134	0.005706134	1.000000	8
[95]	{}	=> {AB ΒΙΟ ΚΑΡΟΤΑ ΕΛΛΗΝΙΚΑ Leaflet}	0.007845934	0.007845934	1.000000	11
[96]	{}	=> {ΦΙΛ.ΓΑΛ/ΛΑΣ ΕΝ ΕΛΛΑΔΙ ΒΡΑΣΤΟ ΤΟΣΤ Leaflet}	0.007132668	0.007132668	1.000000	10
[97]	{}	=> {ΑΜΙΣΟ ΠΑΝΤΕΛΟΝΙ ΦΟΡΜΑΣ UNISEX TV}	0.009272468	0.009272468	1.000000	13
[98]	{}	=> {LENOR EMERALD&IVORY FLOWER 26ΜΕΖ Leaflet}	0.004992867	0.004992867	1.000000	7
[99]	{}	=> {AJAX ULTRA ΛΕΒΑΝΤΑ 1LT(+CRYSTAL) Leaflet}	0.006419401	0.006419401	1.000000	9
[100]	{}	=> {SWIFFER ΑΝΤΑΛΛΑΚΤΙΚΑ ΠΑΝΑΚΙΑ 16Τ Leaflet}	0.004992867	0.004992867	1.000000	7
[101]	{}	=> {ΟΛΥΜΠΟΣ ΦΥΣΙΚΟΣ ΧΥΜΟΣ ΠΟΡΤΟΚ. 1.5L Leaflet}	0.007845934	0.007845934	1.000000	11
[102]	{}	=> {BARILLA ΣΠΑΓΓΕΤΟΝΙ Ν7 500GR Leaflet}	0.005706134	0.005706134	1.000000	8
[103]	{}	=> {ΚΑΡΜΟΛΟΓΚΟΣ ΑΡΓΟΥΜΕΝΟ ΧΩΡΙΑΤΙΚΟ 7 Leaflet}	0.007132668	0.007132668	1.000000	10
[104]	{}	=> {ΚΥΚΝΟΣ ΤΟΜΑΤΕΣ ΤΡΙΜΜΕΝΕΣ 680GR Leaflet}	0.004992867	0.004992867	1.000000	7
[105]	{}	=> {LENOR OCEAN ESCAPE 26ΜΕΖ Leaflet}	0.007845934	0.007845934	1.000000	11
[106]	{}	=> {ΟΛΥΜΠΟΣ ΑΜΥΓΔΑΛΟ ΧΩΡΙΣ ΖΑΧΑΡΗ 1LT Leaflet}	0.005706134	0.005706134	1.000000	8
[107]	{}	=> {BARILLA ΣΑΛΤΣΑ PESTO GENOVEZE 190G Leaflet}	0.005706134	0.005706134	1.000000	8
[108]	{}	=> {NATURA ΚΡΕΜ/ΝΟ ΑΜΥΓΔΑΛΟ 750ML (0.5 Leaflet)}	0.004279601	0.004279601	1.000000	6
[109]	{}	=> {SWIFFER DUSTER(ΛΑΒΗ+5 REF.ΠΑΝΑΚΙΑ) Leaflet}	0.004992867	0.004992867	1.000000	7
[110]	{}	=> {MOLTO ΚΡΟΥΑΣΑΝ ΠΡΑΙΝΑ 5Χ40GR Leaflet}	0.004279601	0.004279601	1.000000	6
[111]	{}	=> {ΦΕΤΑ ΗΠΕΙΡΟΣ ΛΙΓΟΤΕΡΟ ΑΛΑΤΙ ΣΕ ΑΜΜΗ Leaflet}	0.004992867	0.004992867	1.000000	7
[112]	{}	=> {FLORA SOFT 225GR ΚΑΙ ΔΩΡ 45GR Leaflet}	0.004992867	0.004992867	1.000000	7
[113]	{}	=> {FLORA ΠΑΚΕΤΟ 250GR Leaflet}	0.005706134	0.005706134	1.000000	8
[114]	{}	=> {AB ΕΛΑΙΟΛΑΔΟ 1LT Leaflet}	0.008559201	0.008559201	1.000000	12
[115]	{}	=> {KINDER ΕΚΠΛΗΞΗ T3 Leaflet}	0.007132668	0.007132668	1.000000	10
[116]	{}	=> {ΚΕΦΑΛΟΥΤΥΡΙ ΠΡΟΒΕΙΟ ΕΛΛΑΣΣΟΝΑΣ ΤΥΠ. Leaflet}	0.005706134	0.005706134	1.000000	8
[117]	{}	=> {ΠΑΥΛΙΔΗΣ ΣΟΚ.ΥΓΕΙΑΣ ΑΜΥΓΔΑΛΟΥ 100G Leaflet}	0.007845934	0.007845934	1.000000	11
[118]	{}	=> {ΑΦΕΛΑΔΙΤΣΑ 2% 3Χ200GR Leaflet}	0.012125535	0.012125535	1.000000	17
[119]	{}	=> {AB ΚΟΚΚΙΝΟΥΑΡΟ Α/Κ ΔΙΧΤΑΚΙ Leaflet}	0.008559201	0.008559201	1.000000	12
[120]	{}	=> {ΚΕΦΑΛΟΥΤΥΡΙ ΠΡΟΒΕΙΟ ΕΛΛΑΣΣΟΝΑΣ. Leaflet}	0.007845934	0.007845934	1.000000	11
[121]	{}	=> {BARILLA GRANCEREALE ΜΠΙΣΚ.ΣΟΚ.230G Leaflet}	0.007132668	0.007132668	1.000000	10
[122]	{}	=> {LACTA ΣΟΚ/ΤΑ ΟΡΕΟ ΦΡΑΟΥΛΑ 105G Leaflet}	0.011412268	0.011412268	1.000000	16
[123]	{}	=> {ΜΟΡΤΑΔΕΛΛΑ ΜΠΟΛΩΝΙΑ ΥΦΑΝΤΗΣ Leaflet}	0.008559201	0.008559201	1.000000	12
[124]	{}	=> {ΝΟΥΝΟΥ ΚΙΔ 6Χ400GR Leaflet}	0.009272468	0.009272468	1.000000	13
[125]	{}	=> {ΑΛΛΑΤΙΝΗ ΠΤΙ ΜΠΕΡ ΚΛΑΣ. 225GR (0.1 Leaflet)}	0.010699001	0.010699001	1.000000	15
[126]	{}	=> {BARILLA ΜΑΚΑΡΟΝΙΑ Ν3 500GR Leaflet}	0.005706134	0.005706134	1.000000	8
[127]	{}	=> {LENOR DIAMOND&LOTUS FL.26ΜΕΖ Leaflet}	0.005706134	0.005706134	1.000000	8
[128]	{}	=> {ΦΡΥΓ.ELITE ΣΙΚΑΛΗΣ 90GR 3+1ΔΩΡΟ Leaflet}	0.010699001	0.010699001	1.000000	15
[129]	{}	=> {ARLA MOZZARELLA ΔΑΝΙΑΣ 40% ΦΡΑΤΖΟΛΑ Leaflet}	0.009985735	0.009985735	1.000000	14
[130]	{}	=> {ΑΜΙΤΑ ΦΡΟΥΤΟΠΟΤΟ ΠΜΒ 1.5LT Leaflet}	0.007132668	0.007132668	1.000000	10
[131]	{}	=> {ΚΡΙΣ ΚΡΙΣ ΤΟΣΤΙΜΟ ΟΛΙΚΗΣ ΑΛ. 400G Leaflet}	0.011412268	0.011412268	1.000000	16
[132]	{}	=> {ΔΕΣΜΕΥΤΗ ΧΡΩΜΑΤ.& ΒΡΩΜΙΑΣ 10φ+5φΔΩ Leaflet}	0.009272468	0.009272468	1.000000	13
[133]	{}	=> {LIPTON ΧΑΜΟΜΗΛΙ 10φΑΚ.ΧΙGR Leaflet}	0.005706134	0.005706134	1.000000	8
[134]	{}	=> {ΤΥΡΙ REGATO ΤΡΙΜΜΕΝΟ 400GR Leaflet}	0.009985735	0.009985735	1.000000	14
[135]	{}	=> {LENOR AMETHYST&FLORAL B.26ΜΕΖ Leaflet}	0.012125535	0.012125535	1.000000	17
[136]	{}	=> {BARILLA LINGUINE LA COLLEZIONE500G Leaflet}	0.007132668	0.007132668	1.000000	10
[137]	{}	=> {ΜΙΡΑΝΤΑ ΜΠΙΣΚΟΤΑ 2Χ250GR Leaflet}	0.009272468	0.009272468	1.000000	13
[138]	{}	=> {TOTAL 2% 500GR Leaflet}	0.012125535	0.012125535	1.000000	17
[139]	{}	=> {ΓΡΑΒΙΕΡΑ ΠΟΠ ΚΡΗΤΗΣ TV}	0.016405136	0.016405136	1.000000	23
[140]	{}	=> {ΑΡΑΚΑΣ ΚΑΤ/ΝΟΣ 450G ΜΠ.ΣΤΑΘΗ Leaflet}	0.012125535	0.012125535	1.000000	17
[141]	{}	=> {LACTA ΣΟΚ/ΤΑ LOV IT ΟΡΕΟ 105GR Leaflet}	0.009985735	0.009985735	1.000000	14
[142]	{}	=> {ΓΡΑΒΙΕΡΑ ΚΡΗΤΗΣ ΤΥΠ/ΝΗ TV}	0.011412268	0.011412268	1.000000	16
[143]	{}	=> {ΠΑΥΛΙΔΗΣ ΣΟΚ. ΥΓΕΙΑΣ ΠΟΡΤΟΚΑΛΙ 100 Leaflet}	0.012838802	0.012838802	1.000000	18
[144]	{}	=> {LACTA ΣΟΚΟΛ.ΓΑΛΑΚΤΟΣ 85G+30% ΔΩΡΕΑ Leaflet}	0.014978602	0.014978602	1.000000	21
[145]	{}	=> {HELLMANN'S REAL ΜΑΓΙΟΝΕΖΑ 450ML Leaflet}	0.010699001	0.010699001	1.000000	15
[146]	{}	=> {ΣΟΚ. ΠΑΥΛΙΔΗΣ ΥΓΕΙΑΣ 70% 100G Leaflet}	0.017118402	0.017118402	1.000000	24
[147]	{}	=> {WETTEX ΣΠΟΓΓΟΠΕΤΣΕΤΑ Ν01(2+1) Leaflet}	0.009985735	0.009985735	1.000000	14
[148]	{}	=> {CREAM CRACKERS ΣΙΚΑΛΗΣ 175GR Leaflet}	0.015691869	0.015691869	1.000000	22
[149]	{}	=> {LACTA ΣΟΚ..TRIPLE ΚΑΡΑΜΕΛΑ 90G Leaflet}	0.014265335	0.014265335	1.000000	20
[150]	{}	=> {LURPAK SOFT ΑΝΑΛΑΤΟ 225GR Leaflet}	0.010699001	0.010699001	1.000000	15
[151]	{}	=> {MILNER ΦΕΤΕΣ 300GR Leaflet}	0.012838802	0.012838802	1.000000	18
[152]	{}	=> {ΦΙΛΕΤΟ ΓΑΛΟΠ.ΕΝ ΕΛΛΑΔΙ ΚΑΠΝ ΤΟΣΤ Leaflet}	0.013552068	0.013552068	1.000000	19
[153]	{}	=> {VILEDA CLAS.ΣΦΟΥΓΓ.ΚΟΥΖΙΝΑΣ 2+1Τ.Δ TV}	0.019258203	0.019258203	1.000000	27
[154]	{}	=> {AB ΦΑΚΕΣ ΨΙΑΕΣ 500G ΕΙΣΑΓΩΓΗΣ Leaflet}	0.014978602	0.014978602	1.000000	21
[155]	{}	=> {LENOR GOLD ORCHID 26ΜΕΖ Leaflet}	0.013552068	0.013552068	1.000000	19
[156]	{}	=> {ΚΡΙΣ ΤΟΣΤΙΜΟ ΣΤΑΡΕΝΙΟ 400G 0.30E Leaflet}	0.024964337	0.024964337	1.000000	35
[157]	{}	=> {ΚΟΥΝΟΥΠΙΔΙ ΕΛΛΗΝΙΚΟ Leaflet}	0.022111270	0.022111270	1.000000	31
[158]	{}	=> {ΚΡΙΣ ΤΟΣΤΙΜΟ ΣΤΑΡΕΝΙΟ 800G 0.45E Leaflet}	0.024964337	0.024964337	1.000000	35
[159]	{}	=> {ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.025677603	0.025677603	1.000000	36
[160]	{}	=> {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.044222539	0.044222539	1.000000	62
[161]	{}	=> {ΑΛΧΑΛΟ ΑΞΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.047075606	0.047075606	1.000000	66

Σύμφωνα με την Εικόνα 26, η τιμή της υποστήριξης κυμαίνεται από [0,004279 – 0,047076] μονάδες, η τιμή της εμπιστοσύνης από [0,004280 – 0,047076] και η τιμή του μέτρου lift από [1,00 – 17,38] μονάδες. Όσον αφορά τη συχνότητα εμφάνισης των προϊόντων, η μικρότερη είναι 6 και η μεγαλύτερη είναι 66.

Τη μεγαλύτερη υποστήριξη ίση με 0.047 την έχει το προϊόν «ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ», ενώ τη μικρότερη τιμή ίση με 0.004279 την έχουν παραπάνω από ένα προϊόντα και παραπάνω από έναν κανόνες. Κάποια από τα προϊόντα είναι τα : «MOLTO ΚΡΟΥΑΣΑΝ ΠΡΑΛΙΝΑ 5×40 GR leaflet», «GILLET ΞΥΡ.ΜΗΧ.ΒΛΕΠΕ II SLALOM 5T TV.», «LURPAK LIGHTER SOFT ΜΕ ΕΛΑΙΟΛ 225g leaflet», «LENOR MOONLIGHT HARMONY 26MEZ leaflet» και κάποιοι από τους κανόνες είναι οι : {ΣΠΑΝΑΚΙ ΕΛΛΗΝΙΚΟ} → [ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ leaflet], {TOMATEΣ ΧΥΜΑ ΕΛΛΗΝΙΚΕΣ} → {LACTA ΣΟΚ..TRIPLE ΚΑΡΑΜΕΛΑ 90Γ leaflet}, {365 ΧΑΡΤ/ΤΕΣ ΛΕΥΚΕΣ 28X30 80Φ} → {ΑΒ ΦΑΚΕΣ ΨΙΛΕΣ 500G ΕΙΣΑΓΩΓΗΣ leaflet}, {ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ, ΚΡΕΜ. ΞΕΡΑ ΧΟΝΔΡΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ} → {ΚΟΥΝΟΥΠΙΔΙ ΕΛΛΗΝΙΚΟ leaflet} κα.

Τη μεγαλύτερη τιμή εμπιστοσύνης την έχει ο κανόνας {ΛΑΧΑΝΟ ΚΟΚΚΙΝΟ ΕΛΛΗΝΙΚΟ} → {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ} ίση με 0.8181, που σημαίνει ότι το 81% των φορών που οι καταναλωτές θα αγοράσουν «ΛΑΧΑΝΟ ΚΟΚΚΙΝΟ ΕΛΛΗΝΙΚΟ», θα αγοράσουν και «ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ».

Τη μικρότερη τιμή εμπιστοσύνης ίση με 0.0042 μονάδες έχουν παραπάνω από ένα προϊόντα τα οποία ενδεικτικά είναι τα: «MOLTO ΚΡΟΥΑΣΑΝ ΠΡΑΛΙΝΑ 5×40 GR leaflet», «NATURA ΚΡΕΜ/ΝΟ ΑΜΥΓΔΑΛΟ 750ML leaflet», «ΤΣΙΚΛΑ ORBIT ΔΥΟΣΜΟΣ ΜΠΟΥΚΑΛΙ 64ΓΡ leaflet», «ΚΡΙΣ ΚΡΙΣ ΤΟΣΤΙΜΟ ΜΠΡΙΟΣ 400G leaflet», «FRULITE 1LT ΦΡΟΥΤ/ΤΟ ΜΑΝΤΑΡ/ΣΑΓΚΝΙ leaflet».

Την υψηλότερη τιμή lift ίση με 17,38 την έχει ο κανόνας {ΛΑΧΑΝΟ ΚΟΚΚΙΝΟ ΕΛΛΗΝΙΚΟ} → {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ}. Παρόλο που το προϊόν «ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ» εμφανίζεται στο 4.7% των συνολικών συναλλαγών, στον υποπληθυσμό των καταναλωτών που αγοράζουν «ΛΑΧΑΝΟ ΚΟΚΚΙΝΟ ΕΛΛΗΝΙΚΟ», το ποσοστό των καταναλωτών που θα το αγοράσουν «ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ» είναι 17.38 φορές υψηλότερο από ότι είναι στον συνολικό πληθυσμό (4.7%).

Ακόμη, 102 προϊόντα έχουν τη μικρότερη τιμή του μέτρου lift ίση με 1, και κάποια από αυτά είναι τα : «ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet», «ΠΟΡΤΟΚΑΛΙΑ NABELINEΣ ΕΛΛΗΝΙΚΑ Leaflet», «ΚΟΥΝΟΥΠΙΔΙ ΕΛΛΗΝΙΚΟ Leaflet», «ΓΡΑΒΙΕΡΑ ΠΟΠ ΚΡΗΤΗΣ TV», «CREAM CRACKERS ΣΙΚΑΛΗΣ 175ΓΡ Leaflet», «TOTAL 2% 500ΓΡ Leaflet» κα.

Όπως παρατηρούμε από την Εικόνα 27 και την Εικόνα 28, οι κανόνες που δημιουργήθηκαν την πρώτη εβδομάδα περιέχουν μόνο χαρακτηρισμένα προϊόντα στο δεξί μέλος τους και πιο

συγκεκριμένα 26 κανόνες περιέχουν το προϊόν «ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ», 7 κανόνες περιέχουν το «ΚΟΥΝΟΥΠΙΔΙ ΕΛΛΗΝΙΚΟ», 15 κανόνες περιέχουν το «ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ», 12 κανόνες περιέχουν τα «ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ», και από 2 κανόνες έχουν τα προϊόντα «ΑΒ ΦΑΚΕΣ ΨΙΛΕΣ 500G ΕΙΣΑΓΩΓΗΣ», «LACTA ΣΟΚ. . TRIPLE ΚΑΡΑΜΕΛΑ 90Γ» και «ΚΡΙΣ ΤΟΣΤΙΜΟ ΣΤΑΡΕΝΙΟ 800G 0.45Ε».

Καθώς δεν μπορούμε να σχολιάσουμε κάθε έναν κανόνα ξεχωριστά, θα σχολιάσουμε 10 από τους 59 κανόνες που είναι αξιολογήσιμοι και έχουν το υψηλότερο lift (Εικόνα 29)

	lhs	rhs	support	confidence	lift	count
[1]	{ΛΑΧΑΝΟ ΚΟΚΚΙΝΟ ΕΛΛΗΝΙΚΟ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.006419401	0.818181818	17.380165	9
[2]	{ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ, ΚΡΕΜ. ΞΕΡΑ ΧΟΝΔΡΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ}	=> {ΚΟΥΝΟΥΠΙΔΙ ΕΛΛΗΝΙΚΟ Leaflet}	0.004279601	0.333333333	15.075269	6
[3]	{ΣΠΑΝΑΚΙ ΕΛΛΗΝΙΚΟ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004279601	0.600000000	12.745455	6
[4]	{ΑΝΗΘΟΣ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΟΣ 80ΓΡ, ΜΑΙΝΤΑΝΟΣ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΟΣ 80ΓΡ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004279601	0.375000000	7.965909	6
[5]	{ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ, ΜΑΡΟΥΛΙΑ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΑ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004279601	0.352941176	7.497326	6
[6]	{ΚΡΕΜ. ΞΕΡΑ ΧΟΝΔΡΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ, ΜΠΑΝΑΝΕΣ DOLE ΕΙΣΑΓΩΓΗΣ CAVENDISH}	=> {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004279601	0.285714286	6.460829	6
[7]	{ΠΡΑΣΣΑ ΕΛΛΗΝΙΚΑ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004279601	0.285714286	6.069264	6
[8]	{ΛΕΜΟΝΙΑ ΕΛΛΗΝΙΚΑ}	=> {ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.004992867	0.142857143	5.563492	7
[9]	{ΑΒ ΧΑΡΤ/ΤΑ ΛΕΥΚΗ 30Χ30 100Φ}	=> {ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.004279601	0.136363636	5.310606	6
[10]	{ΑΓΓΟΥΡΙΑ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΑ}	=> {ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.004992867	0.134615385	5.242521	7

Εικόνα 29 Κανόνες με το Υψηλότερο Lift

Ο πρώτος κανόνας που παρουσιάζεται στην Εικόνα 29, έχει lift=17.380165 και είναι ο {ΛΑΧΑΝΟ ΚΟΚΚΙΝΟ ΕΛΛΗΝΙΚΟ} → {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ} η ερμηνεία του οποίου δόθηκε παραπάνω. Όπως παρατηρούμε το προϊόν «ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ» αγοράζεται μαζί με προϊόντα «ΣΠΑΝΑΚΙ ΕΛΛΗΝΙΚΟ», «ΑΝΗΘΟΣ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΟΣ 80ΓΡ, ΜΑΙΝΤΑΝΟΣ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΟΣ 80ΓΡ», «ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ, ΜΑΡΟΥΛΙΑ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΑ» και «ΠΡΑΣΣΑ ΕΛΛΗΝΙΚΑ».

Ο δεύτερος κανόνας με υψηλή τιμή lift ίση με 15,075269, είναι ο κανόνας {ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ, ΚΡΕΜ. ΞΕΡΑ ΧΟΝΔΡΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ} → {ΚΟΥΝΟΥΠΙΔΙ ΕΛΛΗΝΙΚΟ}. Η υποστήριξη του κανόνα είναι πολύ μικρή της τάξεως του 0,4%, και οι καταναλωτές που αγοράζουν «ΚΟΥΝΟΥΠΙΔΙ ΕΛΛΗΝΙΚΟ» αποτελούν το 2,2% των καταναλωτών. Στον υποπληθυσμό των καταναλωτών που αγοράζουν «ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ, ΚΡΕΜ. ΞΕΡΑ ΧΟΝΔΡΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ», το ποσοστό των καταναλωτών που αγοράζουν «ΚΟΥΝΟΥΠΙΔΙ ΕΛΛΗΝΙΚΟ» είναι περίπου 15 φορές μεγαλύτερο από ότι είναι στο συνολικό πληθυσμό (2,2%).

Το προϊόν «ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ» συσχετίζεται με τα «ΚΡΕΜ. ΞΕΡΑ ΧΟΝΔΡΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ, ΜΠΑΝΑΝΕΣ DOLE ΕΙΣΑΓΩΓΗΣ CADEVISH». Η τιμή του μέτρου lift είναι ίση με 6,46, που σημαίνει ότι η αγορά των προϊόντων που βρίσκονται στο

αριστερό μέλος, ενθαρρύνει την αγορά του προϊόντος προφοράς στο δεξί μέλος του κανόνα. Πιο συγκεκριμένα ενώ το 4,4% των καταναλωτών αγοράζουν «ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ», στον υποπληθυσμό των καταναλωτών που αγοράζουν «ΚΡΕΜ. ΞΕΡΑ ΧΟΝΔΡΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ, ΜΠΑΝΑΝΕΣ DOLE ΕΙΣΑΓΩΓΗΣ CADEVISH», το ποσοστό των καταναλωτών που αγοράζουν «ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ», είναι περίπου 6,46 φορές μεγαλύτερο από ότι είναι στον συνολικό πληθυσμό (4,4%).

Τέλος το προϊόν «ΠΟΡΤΟΚΑΛΙΑ NABELINEΣ ΕΛΛΗΝΙΚΑ» αγοράζεται μαζί με τα προϊόντα «ΛΕΜΟΝΙΑ ΕΛΛΗΝΙΚΑ», «ΑΒ ΧΑΡΤ/ΤΑ ΛΕΥΚΗ 30x30 100Φ» και «ΑΓΓΟΥΡΙΑ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΑ».

Από τους κανόνες 8, 9 και 10, ο κανόνας 8 έχει τη μεγαλύτερη τιμή lift ίση με 5,56 που σημαίνει ότι ενώ το ποσοστό των καταναλωτών που αγοράζουν «ΠΟΡΤΟΚΑΛΙΑ NABELINEΣ ΕΛΛΗΝΙΚΑ» είναι 2,5%, στον υποπληθυσμό των καταναλωτών που αγοράζουν «ΛΕΜΟΝΙΑ ΕΛΛΗΝΙΚΑ», το ποσοστό των καταναλωτών που αγοράζουν «ΠΟΡΤΟΚΑΛΙΑ NABELINEΣ ΕΛΛΗΝΙΚΑ» είναι περίπου 5,5 φορές υψηλότερο από στο συνολικό πληθυσμό (2,5%).

3.2.5 Εξόρυξη Κανόνων Συσχέτισης για τα Δεδομένα της 2^{ης} Εβδομάδας

Στη συνέχεια θέτουμε κατώτατο όριο υποστήριξης ίσο με 0,004 και εμπιστοσύνη ίση με 0,00 για το αρχείο της δεύτερης εβδομάδας ταξινομώντας τους κανόνες με φθίνουσα σειρά με βάση το μέτρο lift. Επαναλαμβάνουμε την ίδια διαδικασία με πριν, δηλαδή ζητάμε στο δεξί μέλος των κανόνων να υπάρχουν μόνο προϊόντα που είναι σε προσφορά.

Η Εικόνα 30 δείχνει τα στατιστικά των κανόνων συσχέτισης που προκύπτουν για τα προϊόντα της δεύτερης εβδομάδας και οι Εικόνες 31, 32 και 33 δείχνουν τους 216 κανόνες που εξόρυξε ο αλγόριθμος.

```
set of 216 rules
```

```
rule length distribution (lhs + rhs):sizes
```

```
  1  2  3
126 81  9
```

```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000  1.000  1.000  1.458  2.000  3.000
```

```
summary of quality measures:
```

support	confidence	lift	count
Min. :0.004470	Min. :0.004470	Min. : 0.8844	Min. : 7.00
1st Qu.:0.004470	1st Qu.:0.007024	1st Qu.: 1.0000	1st Qu.: 7.00
Median :0.006386	Median :0.017880	Median : 1.0000	Median :10.00
Mean :0.008757	Mean :0.063414	Mean : 2.0569	Mean :13.71
3rd Qu.:0.008940	3rd Qu.:0.104549	3rd Qu.: 2.7044	3rd Qu.:14.00
Max. :0.043423	Max. :0.500000	Max. :11.6866	Max. :68.00

```
mining info:
```

```
data ntransactions support confidence
tr2          1566  0.004          0
```

Εικόνα 30 Σύνοψη Κανόνων Εβδομάδα 2

Δημιουργούνται 216 κανόνες και μόνο οι 90 είναι αξιολογήσιμοι, καθώς οι 121 είναι μήκους 1-στοιχειοσύνολο, οι 81 είναι μήκους 2-στοιχειοσυνόλων και 9 κανόνες είναι μήκους 3-στοιχειοσυνόλων. Πράγματι οι κανόνες που δημιουργούνται την δεύτερη εβδομάδα είναι περισσότεροι από την πρώτη, καθώς τη δεύτερη εβδομάδα κάποια προϊόντα είναι σε προσφορά και υπάρχουν περισσότερα διαφορετικά προϊόντα στα καλάθια των καταναλωτών.

	lhs	rhs	support	confidence	lift	count
[1]	{ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet, ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004469987	0.500000000	11.6865672	7
[2]	{ΑΒ ΧΑΡΤΙ ΥΓΕΙΑΣ 10ΑΡΙ 4ΡΛΥ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.005108557	0.421052632	9.8413197	8
[3]	{ΓΟΥΔΑ ΣΕ ΦΕΤΕΣ 400ΓΡ}	=> {ΚΡΙΣ ΤΟΣΤΙΜΟ ΣΤΑΡΕΝΙΟ 800G 0.45E Leaflet}	0.004469987	0.350000000	8.1805970	7
[4]	{ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ, ΤΟΜΑΤΕΣ ΧΥΜΑ ΕΛΛΗΝΙΚΕΣ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004469987	0.333333333	7.7910448	7
[5]	{ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ, ΜΠΑΝΑΝΕΣ DULCE BUONA ΕΙΣΑΓΩΓΗΣ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004469987	0.333333333	7.7910448	7
[6]	{ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ, ΚΡΕΜ. ΞΕΡΑ ΧΟΝΔΡΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004469987	0.304347826	7.1135626	7
[7]	{ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ, ΠΑΤΑΤΕΣ ΧΟΝΔΡΕΣ ΕΛΛΗΝΙΚΕΣ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.005108557	0.275862069	6.4477612	8
[8]	{ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.017879949	0.266666667	6.2328358	28
[9]	{ΑΒ ΠΑΤΑΤΑ ΝΑΞΟΥ ΠΓΕ ΣΥΣΚ.}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004469987	0.259259259	6.0597015	7
[10]	{ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ, ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	=> {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004469987	0.250000000	5.7573529	7
[11]	{ΜΠΑΝΑΝΕΣ DULCE BUONA ΕΙΣΑΓΩΓΗΣ, ΤΟΜΑΤΕΣ ΧΥΜΑ ΕΛΛΗΝΙΚΕΣ}	=> {ΜΗΛΑ ΣΤΑΡΚΙΝΓΚ ΕΛΛΗΝΙΚΑ TV}	0.004469987	0.205882353	5.6563467	7
[12]	{ΜΑΡΟΥΛΙΑ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΑ}	=> {ΚΟΥΝΟΥΠΙΔΙ ΕΛΛΗΝΙΚΟ Leaflet}	0.004469987	0.093333333	5.4133333	7
[13]	{ΓΟΥΔΑ ΓΕΡΜΑΝΙΑΣ ΦΡΑΤΖΟΛΑ}	=> {2 ΛΑΙΜΟΣ ΧΟΙΡ.ΟΛΛ. ΚΙΒ.ΧΥΜΑ TV}	0.005108557	0.076923077	5.2374582	8
[14]	{365 ΒΑΝΙΛΙΝΗ ΣΤΕΜ}	=> {ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.004469987	0.212121212	5.1104895	7
[15]	{ΚΟΛΟΚΥΘΙΑ ΕΛΛΗΝΙΚΑ}	=> {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.007024266	0.211538462	4.8716063	11
[16]	{ΤΥΡΙ ΦΕΤΑ ΤΡΙΠΟΛΕΩΣ ΒΑΡΕΛΙ}	=> {ΦΙΛΕΤΟ ΓΑΛΟΠ.ΕΝ ΕΛΛΑΔΙ ΚΑΠΝ ΤΟΣΤ Leaflet}	0.004469987	0.175000000	4.8078947	7
[17]	{ΜΠΑΝΑΝΕΣ DULCE BUONA ΕΙΣΑΓΩΓΗΣ, ΤΟΜΑΤΕΣ ΧΥΜΑ ΕΛΛΗΝΙΚΕΣ}	=> {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004469987	0.205882353	4.7413495	7
[18]	{ΤΥΡΙ ΝΟΥΝΟΥ ΓΟΥΔΑ ΟΛΛΑΝΔ.ΦΡΑΤΖΟΛΑ}	=> {ΦΙΛΕΤΟ ΓΑΛΟΠ.ΕΝ ΕΛΛΑΔΙ ΚΑΠΝ ΤΟΣΤ Leaflet}	0.006385696	0.172413793	4.7368421	10
[19]	{ΤΥΡΙ MILNER ΟΛΛΑΝΔΙΑΣ ΦΡΑΤΖΟΛΑ 17%}	=> {ΦΙΛΕΤΟ ΓΑΛΟΠ.ΕΝ ΕΛΛΑΔΙ ΚΑΠΝ ΤΟΣΤ Leaflet}	0.005747126	0.163636364	4.4956938	9
[20]	{ΚΡΕΜ. ΞΕΡΑ ΧΟΝΔΡΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ, ΠΑΤΑΤΕΣ ΧΟΝΔΡΕΣ ΕΛΛΗΝΙΚΕΣ}	=> {ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.004469987	0.184210526	4.4380567	7
[21]	{ΑΓΓΟΥΡΙΑ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΑ}	=> {ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.007024266	0.183333333	4.4169231	11
[22]	{ΜΑΝΤΑΡΙΝΙΑ ΚΑΗΜΕΝΤΙΝΕΣ ΕΛΛΗΝΙΚΑ}	=> {ΜΗΛΑ ΣΤΑΡΚΙΝΓΚ ΕΛΛΗΝΙΚΑ TV}	0.005108557	0.156862745	4.3095975	8
[23]	{ΠΑΤΑΤΕΣ ΧΟΝΔΡΕΣ ΚΥΠΡΟΥ ΣΥΣ/ΝΕΣ}	=> {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004469987	0.179487179	4.1334842	7
[24]	{ΜΑΝΤΑΡΙΝΙΑ ΚΑΗΜΕΝΤΙΝΕΣ ΕΛΛΗΝΙΚΑ}	=> {ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.005108557	0.156862745	3.7791855	8
[25]	{ΓΟΥΔΑ ΓΕΡΜΑΝΙΑΣ ΦΡΑΤΖΟΛΑ}	=> {ΜΟΡΤΑΔΕΛΑ ΜΠΟΛΩΝΙΑ ΥΦΑΝΤΗΣ Leaflet}	0.005108557	0.076923077	3.7644231	8
[26]	{ΓΟΥΔΑ ΓΕΡΜΑΝΙΑΣ ΦΡΑΤΖΟΛΑ}	=> {ΦΙΛ.ΓΑΛ/ΛΑΣ ΕΝ ΕΛΛΑΔΙ ΒΡΑΣΤΟ ΤΟΣΤ Leaflet}	0.005108557	0.076923077	3.7644231	8
[27]	{ΜΑΙΝΤΑΝΟΣ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΟΣ 80ΓΡ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.005747126	0.160714286	3.7563966	9
[28]	{ΣΣΛΙΝΟ ΕΛΛΗΝΙΚΟ}	=> {ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.004469987	0.152173913	3.6662207	7
[29]	{ΜΑΝΤΑΡΙΝΙΑ ΚΑΗΜΕΝΤΙΝΕΣ ΕΛΛΗΝΙΚΑ}	=> {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.005108557	0.156862745	3.6124567	8
[30]	{ΜΠΑΝΑΝΕΣ DOLE ΕΙΣΑΓΩΓΗΣ SAVENDISH}	=> {ΑΒ ΒΙΟ ΚΑΡΟΤΑ ΕΛΛΗΝΙΚΑ Leaflet}	0.006385696	0.064516129	3.6082949	10
[31]	{ΑΒ ΣΑΛΑΤΑ ΚΑΠΡΙΤΣΙΟΖΑ 160 ΓΡ}	=> {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004469987	0.152173913	3.5044757	7
[32]	{ΜΠΑΝΑΝΕΣ DULCE BUONA ΕΙΣΑΓΩΓΗΣ}	=> {ΠΑΥΛΙΔΗΣ ΥΓΕΙΑΣ ΓΕΜ. ΚΑΚΑΟ 100GR Leaflet}	0.004469987	0.037837838	3.4855326	7
[33]	{ΓΑΛΟΠΟΥΛΑ ΚΑΠΝΙΣΤΗ ΤΟΣΤΑΚΙ}	=> {ΚΡΙΣ ΤΟΣΤΙΜΟ ΣΤΑΡΕΝΙΟ 800G 0.45E Leaflet}	0.004469987	0.148936170	3.4811051	7
[34]	{ΜΑΙΝΤΑΝΟΣ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΟΣ 80ΓΡ}	=> {ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.005108557	0.142857143	3.4417582	8
[35]	{ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	=> {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.006385696	0.149253731	3.4372256	10
[36]	{ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.006385696	0.147058824	3.4372256	10
[37]	{ΛΕΜΟΝΙΑ ΕΛΛΗΝΙΚΑ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.005108557	0.145454545	3.3997286	8
[38]	{ΑΧΛΑΔΙΑ ΚΡΥΣΤΑΛΙΑ ΕΛΛΗΝΙΚΑ}	=> {ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.006385696	0.138888889	3.3461538	10
[39]	{ΤΥΡΙ ΝΟΥΝΟΥ ΓΟΥΔΑ ΟΛΛΑΝΔ.ΦΡΑΤΖΟΛΑ}	=> {ΓΡΑΒΙΕΡΑ ΠΟΠ ΚΡΗΤΗΣ TV}	0.005108557	0.137931034	3.2238806	8
[40]	{ΤΥΡΙ ΝΟΥΝΟΥ ΓΟΥΔΑ ΟΛΛΑΝΔ.ΦΡΑΤΖΟΛΑ}	=> {ΚΡΙΣ ΤΟΣΤΙΜΟ ΣΤΑΡΕΝΙΟ 800G 0.45E Leaflet}	0.005108557	0.137931034	3.2238806	8
[41]	{ΜΑΝΤΑΡΙΝΙΑ ΚΑΗΜΕΝΤΙΝΕΣ ΕΛΛΗΝΙΚΑ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004469987	0.137254902	3.2080773	7
[42]	{ΜΠΑΝΑΝΕΣ DOLE ΕΙΣΑΓΩΓΗΣ SAVENDISH}	=> {ΦΙΛ.ΓΑΛ/ΛΑΣ ΕΝ ΕΛΛΑΔΙ ΒΡΑΣΤΟ ΤΟΣΤ Leaflet}	0.006385696	0.064516129	3.1572581	10
[43]	{ΚΟΛΟΚΥΘΙΑ ΕΛΛΗΝΙΚΑ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004469987	0.134615385	3.1463835	7
[44]	{ΤΟΜΑΤΕΣ ΧΥΜΑ ΕΛΛΗΝΙΚΕΣ}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.010855683	0.132812500	3.1042444	17
[45]	{ΑΓΓΟΥΡΙΑ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΑ}	=> {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.005108557	0.133333333	3.0705882	8
[46]	{ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ}	=> {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.008939974	0.133333333	3.0705882	14
[47]	{ΜΑΡΟΥΛΙΑ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΑ}	=> {ΜΗΛΑ ΣΤΑΡΚΙΝΓΚ ΕΛΛΗΝΙΚΑ TV}	0.005108557	0.106666667	2.9305263	8
[48]	{ΜΗΛΑ ΣΤΑΡΚΙΝΓΚ ΕΛΛΗΝΙΚΑ TV}	=> {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004469987	0.122807018	2.8703849	7
[49]	{ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	=> {ΜΗΛΑ ΣΤΑΡΚΙΝΓΚ ΕΛΛΗΝΙΚΑ TV}	0.004469987	0.104477612	2.8703849	7
[50]	{ΜΠΑΝΑΝΕΣ DOLE ΕΙΣΑΓΩΓΗΣ SAVENDISH}	=> {ΑΓΕΛΑΔΙΤΣΑ 2% 3Χ200ΓΡ Leaflet}	0.004469987	0.045161290	2.8289032	7

Εικόνα 31 Κανόνες Συσχέτισης Εβδομάδα 2

[51]	{ΜΗΛΑ ΣΤΑΡΚΙΝΓΚ ΕΛΛΗΝΙΚΑ TV}	=>	{ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004469987	0.122807018	2.8281734	7
[52]	{ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	=>	{ΜΗΛΑ ΣΤΑΡΚΙΝΓΚ ΕΛΛΗΝΙΚΑ TV}	0.004469987	0.102941176	2.8281734	7
[53]	{ΠΑΤΑΤΕΣ ΧΟΝΔΡΕΣ ΕΛΛΗΝΙΚΕΣ}	=>	{ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.008301405	0.115044248	2.7716814	13
[54]	{ΑΓΓΟΥΡΙΑ ΤΕΜΑΧΙΟ ΕΛΛΗΝΙΚΑ}	=>	{ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.004469987	0.116666667	2.7268657	7
[55]	{ΓΟΥΔΑ ΓΕΡΜΑΝΙΑΣ ΦΡΑΤΖΟΛΑ}	=>	{ΚΡΙΣ ΤΟΣΤΙΜΟ ΣΤΑΡΕΝΙΟ 800G 0.45E Leaflet}	0.007662835	0.115384615	2.6969001	12
[56]	{ΤΟΜΑΤΕΣ ΧΥΜΑ ΕΛΛΗΝΙΚΕΣ}	=>	{ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.008939974	0.109375000	2.6350962	14
[57]	{ΑΧΛΑΔΙΑ ΚΡΥΣΤΑΛΙΑ ΕΛΛΗΝΙΚΑ}	=>	{ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.005108557	0.111111111	2.5970149	8
[58]	{ΛΟΥΜΙΔΗΣ ΚΑΦΕΣ ΠΑΠΑΓ.ΠΑΡΑΔΟΣ.194GR}	=>	{ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.005747126	0.108433735	2.5344363	9
[59]	{ΜΠΑΝΑΝΕΣ DOLE ΕΙΣΑΓΩΓΗΣ SAVENDISH}	=>	{ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.010855683	0.109677419	2.5258065	17
[60]	{ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ}	=>	{ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.007024266	0.104761905	2.5239560	11
[61]	{ΤΟΜΑΤΕΣ ΧΥΜΑ ΕΛΛΗΝΙΚΕΣ}	=>	{ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.008939974	0.109375000	2.5188419	14
[62]	{ΜΠΑΝΑΝΕΣ DULCE ΒΥΟΝΑ ΕΙΣΑΓΩΓΗΣ}	=>	{ΚΟΥΝΟΥΠΙΔΙ ΕΛΛΗΝΙΚΟ Leaflet}	0.005108557	0.043243243	2.5081081	8
[63]	{ΚΡΕΜ. ΞΕΡΑ ΧΟΝΔΡΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ}	=>	{ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.007024266	0.106796117	2.4961600	11
[64]	{ΜΠΑΝΑΝΕΣ DULCE ΒΥΟΝΑ ΕΙΣΑΓΩΓΗΣ}	=>	{ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.012771392	0.108108108	2.4896661	20
[65]	{ΜΠΑΝΑΝΕΣ DOLE ΕΙΣΑΓΩΓΗΣ SAVENDISH}	=>	{ΜΗΛΑ ΣΤΑΡΚΙΝΓΚ ΕΛΛΗΝΙΚΑ TV}	0.008939974	0.090322581	2.4814941	14
[66]	{ΠΑΤΑΤΕΣ ΧΟΝΔΡΕΣ ΕΛΛΗΝΙΚΕΣ}	=>	{ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.007662835	0.106194690	2.4456012	12
[67]	{ΓΡΑΒΙΕΡΑ ΠΟΠ ΚΡΗΤΗΣ TV}	=>	{ΚΡΙΣ ΤΟΣΤΙΜΟ ΣΤΑΡΕΝΙΟ 800G 0.45E Leaflet}	0.004469987	0.104477612	2.4419693	7
[68]	{ΚΡΙΣ ΤΟΣΤΙΜΟ ΣΤΑΡΕΝΙΟ 800G 0.45E Leaflet}	=>	{ΓΡΑΒΙΕΡΑ ΠΟΠ ΚΡΗΤΗΣ TV}	0.004469987	0.104477612	2.4419693	7
[69]	{ΜΠΑΝΑΝΕΣ DOLE ΕΙΣΑΓΩΓΗΣ SAVENDISH}	=>	{ΚΡΙΣ ΤΟΣΤΙΜΟ ΣΤΑΡΕΝΙΟ 400G 0.30E Leaflet}	0.005747126	0.058064516	2.3928693	9
[70]	{ΤΟΜΑΤΕΣ ΧΥΜΑ ΕΛΛΗΝΙΚΕΣ}	=>	{ΜΗΛΑ ΣΤΑΡΚΙΝΓΚ ΕΛΛΗΝΙΚΑ TV}	0.007024266	0.085937500	2.3610197	11
[71]	{ΜΠΑΝΑΝΕΣ DOLE ΕΙΣΑΓΩΓΗΣ SAVENDISH}	=>	{ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.009578544	0.096774194	2.3315136	15
[72]	{ΜΠΑΝΑΝΕΣ DULCE ΒΥΟΝΑ ΕΙΣΑΓΩΓΗΣ}	=>	{ΜΗΛΑ ΣΤΑΡΚΙΝΓΚ ΕΛΛΗΝΙΚΑ TV}	0.009578544	0.081081081	2.2275960	15
[73]	{ΛΟΥΜΙΔΗΣ ΚΑΦΕΣ ΠΑΠΑΓ.ΠΑΡΑΔΟΣ.194GR}	=>	{ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.005108557	0.096385542	2.2197023	8
[74]	{ΜΠΑΝΑΝΕΣ DULCE ΒΥΟΝΑ ΕΙΣΑΓΩΓΗΣ}	=>	{ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.010855683	0.091891892	2.2138877	17
[75]	{ΑΒ ΖΑΧΑΡΗ ΛΕΥΚΗ ΚΡΥΣ/ΚΗ Ε.Ε.1ΚG}	=>	{ΚΡΙΣ ΤΟΣΤΙΜΟ ΣΤΑΡΕΝΙΟ 800G 0.45E Leaflet}	0.004469987	0.094594595	2.2109722	7
[76]	{ΓΟΥΔΑ ΓΕΡΜΑΝΙΑΣ ΦΡΑΤΖΟΛΑ}	=>	{ΦΙΑΛΕΤΟ ΓΑΛΟΠ.ΕΝ ΕΛΛΑΔΙ ΚΑΠΝ ΤΟΣΤ Leaflet}	0.005108557	0.076923077	2.1133603	8
[77]	{ΚΡΕΜ. ΞΕΡΑ ΧΟΝΔΡΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ}	=>	{ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.005747126	0.087378641	2.1051531	9
[78]	{ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ}	=>	{ΜΗΛΑ ΣΤΑΡΚΙΝΓΚ ΕΛΛΗΝΙΚΑ TV}	0.005108557	0.076190476	2.0932331	8
[79]	{ΜΠΑΝΑΝΕΣ DOLE ΕΙΣΑΓΩΓΗΣ SAVENDISH}	=>	{ΑΡΑΧΙΣ ΚΑΤ/ΝΟΣ 450G ΜΠ.ΣΤΑΘΗ Leaflet}	0.004469987	0.045161290	2.0800759	7
[80]	{ΠΑΤΑΤΕΣ ΧΟΝΔΡΕΣ ΕΛΛΗΝΙΚΕΣ}	=>	{ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.006385696	0.088495575	2.0684190	10
[81]	{ΚΡΕΜ. ΞΕΡΑ ΧΟΝΔΡΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ}	=>	{ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.005747126	0.087378641	2.0122787	9
[82]	{ΜΠΑΝΑΝΕΣ DULCE ΒΥΟΝΑ ΕΙΣΑΓΩΓΗΣ}	=>	{ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.009578544	0.081081081	1.8951190	15
[83]	{ΚΡΕΜ. ΞΕΡΑ ΧΟΝΔΡΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ}	=>	{ΦΙΑΛΕΤΟ ΓΑΛΟΠ.ΕΝ ΕΛΛΑΔΙ ΚΑΠΝ ΤΟΣΤ Leaflet}	0.004469987	0.067961165	1.8671436	/
[84]	{ΤΟΜΑΤΕΣ ΧΥΜΑ ΕΛΛΗΝΙΚΕΣ}	=>	{ΚΡΙΣ ΤΟΣΤΙΜΟ ΣΤΑΡΕΝΙΟ 800G 0.45E Leaflet}	0.006385696	0.078125000	1.8260261	10
[85]	{ΓΟΥΔΑ ΓΕΡΜΑΝΙΑΣ ΦΡΑΤΖΟΛΑ}	=>	{ΓΡΑΒΙΕΡΑ ΠΟΠ ΚΡΗΤΗΣ TV}	0.005108557	0.076923077	1.7979334	8
[86]	{ΜΠΑΝΑΝΕΣ DOLE ΕΙΣΑΓΩΓΗΣ SAVENDISH}	=>	{ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.007024266	0.070967742	1.6587386	11
[87]	{ΚΡΕΜ. ΞΕΡΑ ΧΟΝΔΡΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ}	=>	{ΚΡΙΣ ΤΟΣΤΙΜΟ ΣΤΑΡΕΝΙΟ 800G 0.45E Leaflet}	0.004469987	0.067961165	1.5884654	7
[88]	{ΜΠΑΝΑΝΕΣ DOLE ΕΙΣΑΓΩΓΗΣ SAVENDISH}	=>	{ΓΡΑΒΙΕΡΑ ΠΟΠ ΚΡΗΤΗΣ TV}	0.005747126	0.058064516	1.3571497	9
[89]	{ΜΠΑΝΑΝΕΣ DULCE ΒΥΟΝΑ ΕΙΣΑΓΩΓΗΣ}	=>	{ΦΙΑΛΕΤΟ ΓΑΛΟΠ.ΕΝ ΕΛΛΑΔΙ ΚΑΠΝ ΤΟΣΤ Leaflet}	0.004469987	0.037837838	1.0395448	7
[90]	{}	=>	{ΟΛΥΜΠΟΣ ΦΥΣΙΚΟΣ ΧΥΜΟΣ ΠΟΡΤΟΚ. 1.5L Leaflet}	0.005108557	0.005108557	1.0000000	8
[91]	{}	=>	{ΣΟΛΟΜΟΣ ΦΕΤΕΣ ΥΔΑΤ.ΝΟΡΒΗΓΙΑΣ ΝΩΠΟΣ Leaflet}	0.004469987	0.004469987	1.0000000	7
[92]	{}	=>	{LENOR OCEAN ESCAPE 26MEZ Leaflet}	0.004469987	0.004469987	1.0000000	7
[93]	{}	=>	{ΑΒ ΑΥΓΑ ΕΛΕΥΘ. ΒΟΣΚΗΣ 6ΑΔΑ MEDIUM Leaflet}	0.004469987	0.004469987	1.0000000	7
[94]	{}	=>	{ΑΒ ΦΕΤΑ ΠΟΠ ΧΕΛΜΟΥ ΣΕ ΑΛΜΗ 400 ΓΡ. Leaflet}	0.004469987	0.004469987	1.0000000	7
[95]	{}	=>	{FRUITTE 1LT ΦΡΟΥΤ/ΤΟ ΜΑΝΤΑΡ/ΣΑΓΚΝΙ Leaflet}	0.004469987	0.004469987	1.0000000	7
[96]	{}	=>	{ΠΑΥΛΙΔΗΣ ΣΟΚ.ΥΓΕΙΑΣ ΚΑΠΟΥΤΣΙΝΟ 100Γ Leaflet}	0.004469987	0.004469987	1.0000000	7
[97]	{}	=>	{LAYS ΤΣΙΠΕΣ ΑΛΑΤΙ 150GR (2+1 ΔΩΡΟ) Leaflet}	0.005108557	0.005108557	1.0000000	8
[98]	{}	=>	{ΦΥΛΛΟ ΚΡΟΥΣΤΑΣ ΠΟΛΙΤΙΚΟ 450G ALFA Leaflet}	0.004469987	0.004469987	1.0000000	7
[99]	{}	=>	{ΤΣΙΚΛΑ ΟΡΒΙΤ WHITE ΜΠΟΥΚ. ΔΥΟΣΜΟΣ Leaflet}	0.005747126	0.005747126	1.0000000	9
[100]	{}	=>	{PRINGLES HOT PAPRIKA 165GR Leaflet}	0.004469987	0.004469987	1.0000000	7
[101]	{}	=>	{ION BREAK ΣΟΚΟΛΑΤΑ ΑΜΥΓΔΑΛΟΥ 85GR Leaflet}	0.005108557	0.005108557	1.0000000	8
[102]	{}	=>	{LEPTON Μ.ΤΣΑΙ ΒΑΝ/ΚΑΡΑΜ.ΠΥΡ.20Χ1.7 Leaflet}	0.004469987	0.004469987	1.0000000	7
[103]	{}	=>	{MINI PICK CRACKERS BARBECUE 70GR Leaflet}	0.005747126	0.005747126	1.0000000	9
[104]	{}	=>	{ΠΑΡΑΔΟΣΙΑΚΗ ΠΙΤΑ ΤΥΡΙ ΜΕΤΣΟΒΟΥ 850 Leaflet}	0.004469987	0.004469987	1.0000000	7
[105]	{}	=>	{ION BREAK ΜΥΡΤΙΛΑ ΑΜΥΓΔΑΛΑ 85Γ Leaflet}	0.004469987	0.004469987	1.0000000	7
[106]	{}	=>	{BARILLA ΜΑΚΕΡΟΤΣΙΝΙ Ν10 500GR Leaflet}	0.004469987	0.004469987	1.0000000	7
[107]	{}	=>	{ZEWA W&W EXTRA LANG Χ/Κ 2PLY 2ΡΟΛ Leaflet}	0.005108557	0.005108557	1.0000000	8
[108]	{}	=>	{ΩΜΟΠΑΛΗ ΧΟΙΡΙΝΗ ΒΡΑΣΤΗ ΥΘΑΝΤΗΣ Leaflet}	0.004469987	0.004469987	1.0000000	7
[109]	{}	=>	{JOHNNIE WALKER 70CL RED LABEL SECUR Leaflet}	0.006385696	0.006385696	1.0000000	10
[110]	{}	=>	{ΠΑΠΑΔ/ΛΟΥ COOKIES ΚΑΚΑΟΣΟΚ.180ΓΡ Leaflet}	0.005108557	0.005108557	1.0000000	8
[111]	{}	=>	{ΟΛΥΜΠΟΣ ΑΜΥΓΔΑΛΟ 1LT Leaflet}	0.004469987	0.004469987	1.0000000	7
[112]	{}	=>	{BARILLA GRANCEREALE ΜΠΙΣΚ.ΣΟΚ.230G Leaflet}	0.005747126	0.005747126	1.0000000	9
[113]	{}	=>	{ΑΓΕΛΑΔΙΤΣΑ ΦΑΓΕ 3Χ200ΓΡ Leaflet}	0.005108557	0.005108557	1.0000000	8
[114]	{}	=>	{VILEDA ΣΦΟΥΓΓΑΡΙΣΤΡΑ 3ΑΚΤΙΟΝ TV}	0.005108557	0.005108557	1.0000000	8
[115]	{}	=>	{BARILLA ΣΠΑΓΓΕΤΙ Ν5 3Χ500GR 0.65E Leaflet}	0.004469987	0.004469987	1.0000000	7
[116]	{}	=>	{SWIFFER ΑΝΤΑΛΛΑΚΤΙΚΑ ΠΑΝΑΚΙΑ 16Τ Leaflet}	0.004469987	0.004469987	1.0000000	7
[117]	{}	=>	{ΕΨΑ ΠΟΡΤΟΚΑΛΑΔΑ LIGHT 1.5LT ΕΚΠ.0 Leaflet}	0.004469987	0.004469987	1.0000000	7
[118]	{}	=>	{LIPTON ΧΑΜΟΜΗΛΙ 10ΦΑΚ.Χ1GR Leaflet}	0.005108557	0.005108557	1.0000000	8
[119]	{}	=>	{BARILLA ΣΠΑΓΓΕΤΙ Ν5 500GR Leaflet}	0.004469987	0.004469987	1.0000000	7
[120]	{}	=>	{OREO ΜΠΙΣΚΟΤΑ CHEESECAKE154Γ Leaflet}	0.005108557	0.005108557	1.0000000	8
[121]	{}	=>	{ΟΛΥΜΠΟΣ ΑΜΥΓΔΑΛΟ ΧΩΡΙΣ ΖΑΧΑΡΗ 1LT Leaflet}	0.005108557	0.005108557	1.0000000	8
[122]	{}	=>	{BARILLA GRANCEREALE ΜΠΙΣΚ.ΚΑΛΣ.250G Leaflet}	0.005747126	0.005747126	1.0000000	9
[123]	{}	=>	{ΜΙΜΙΚΟΣ ΦΙΑΛΕΤΟ ΑΠΟ ΣΤΗΘΟΣ 650ΓΡ Leaflet}	0.004469987	0.004469987	1.0000000	7
[124]	{}	=>	{HELLMANN.S ΚΕΤΣΑΠ ΦΙΑΛΗ 340ML Leaflet}	0.004469987	0.004469987	1.0000000	7
[125]	{}	=>	{CREAM CRACKERS ΣΙΚΑΛΗΣ 175ΓΡ Leaflet}	0.009578544	0.009578544	1.0000000	15
[126]	{}	=>	{ΚΑΡΑΜΟΛΕΓΚΟΣ ΑΡΓΟΨΗΜΕΝΟ ΧΩΡΙΑΤΙΚΟ 7 Leaflet}	0.007662835	0.007662835	1.0000000	12
[127]	{}	=>	{ARLA SAMSOE ΗΜΙΣΚ. ΤΥΡΙ ΦΡΑΤΖΟΛΑ Leaflet}	0.007024266	0.007024266	1.0000000	11
[128]	{}	=>	{ΑΒ ΖΑΜΠΟΝ ΒΡΑΣΤΟ ΦΕΤΕΣ 160G Leaflet}	0.008301405	0.008301405	1.0000000	13
[129]	{}	=>	{RIO MARE ΤΟΝΟΣ ΣΕ ΑΛΑΔΙ 2Χ104GR Leaflet}	0.006385696	0.006385696	1.0000000	10

[130]	{}	=>	{FRULITE ΚΑΡΟΤ/ΠΟΡΤ/ΜΑΝΓΚΟ 1LT Leaflet}	0.007662835	0.007662835	1.0000000	12
[131]	{}	=>	{ΜΠΙΣΚΟΤΑ ΟΡΕΟ ΓΕΜΙΣΤΑ ΒΑΝΙΛΙΑ 154Γ Leaflet}	0.004469987	0.004469987	1.0000000	7
[132]	{}	=>	{ION BREAK ΣΟΚΟΛΑΤΑ ΛΕΥΚΗ 85GR Leaflet}	0.005747126	0.005747126	1.0000000	9
[133]	{}	=>	{LACTA ΣΟΚΟΛΑΤΑ ΚΡΑΚΕΡ 87Γ Leaflet}	0.008301405	0.008301405	1.0000000	13
[134]	{}	=>	{ΣΠΙΤΙΚΗ ΚΑΣΕΡΟΠΙΤΑ 850G Χ.ΖΥΜΗ Leaflet}	0.007662835	0.007662835	1.0000000	12
[135]	{}	=>	{ΑΒ ΠΙΠΕΡΙ ΜΑΥΡΟ ΤΡΙΜΜΕΝΟ 40Γ. Leaflet}	0.005108557	0.005108557	1.0000000	8
[136]	{}	=>	{LACTA ΣΟΚ/ΤΑ ΟΡΕΟ ΦΡΑΟΥΛΑ 105Γ Leaflet}	0.009578544	0.009578544	1.0000000	15
[137]	{}	=>	{KLEENEX CAPITONNE Χ/ΥΓ.(10+2ΔΩΡΟ) Leaflet}	0.005747126	0.005747126	1.0000000	9
[138]	{}	=>	{ΝΙΚΑΣ ΝΙΤΣΙΑΚΟΣ ΚΟΤΟΠΟΥΛΟ ΠΑΡΙΖΑ Τ Leaflet}	0.005747126	0.005747126	1.0000000	9
[139]	{}	=>	{ION BREAK ΣΟΚΟΛΑΤΑ ΦΟΥΝΤΟΥΚΙ 85GR Leaflet}	0.008301405	0.008301405	1.0000000	13
[140]	{}	=>	{WETTEX ΣΦΟΥΓΓΑΡΙΣΤΡΑ Leaflet}	0.004469987	0.004469987	1.0000000	7
[141]	{}	=>	{LIPTON ΠΥΡΑ.ΑΦΕΥ.ΜΗΛΟ.ΚΑΝ.20x2.2Γ Leaflet}	0.005747126	0.005747126	1.0000000	9
[142]	{}	=>	{ΤΣΙΚΛΑ ΟΡΙΒΤ ΔΥΟΣΜΟΣ ΜΠΟΥΚΑΛΙ 64Γ Leaflet}	0.004469987	0.004469987	1.0000000	7
[143]	{}	=>	{WETTEX ΑΠΟΡΟΦ/ΚΗ ΠΕΤΣΕΤΑ ΡΟΛΟ Leaflet}	0.007024266	0.007024266	1.0000000	11
[144]	{}	=>	{ΠΑΥΛΙΔΗΣ ΥΓΕΙΑΣ ΜΕ ΦΟΥΝΤΟΥΚΙΑ 100Γ Leaflet}	0.009578544	0.009578544	1.0000000	15
[145]	{}	=>	{ΦΡΥΓ.ΕΛΙΤΕ ΣΙΚΑΛΗΣ 90GR 3+1ΔΩΡΟ Leaflet}	0.008301405	0.008301405	1.0000000	13
[146]	{}	=>	{LACTA ΣΟΚΟΛΑΤΑ ΜΕΡΕΝΤΑ 90Γ Leaflet}	0.005747126	0.005747126	1.0000000	9
[147]	{}	=>	{FLORA ΠΑΚΕΤΟ 250GR Leaflet}	0.008301405	0.008301405	1.0000000	13
[148]	{}	=>	{LACTA ΣΟΚ..TRIPLE ΚΑΡΑΜΕΛΑ 90Γ Leaflet}	0.008301405	0.008301405	1.0000000	13
[149]	{}	=>	{CREAM CRACKERS ΣΙΤΟΥ Ν018 140Γ Leaflet}	0.010217114	0.010217114	1.0000000	16
[150]	{}	=>	{LENOR ΑΜΕΤΗΥΣΤ&FLORAL Β.26ΜΕΖ Leaflet}	0.005747126	0.005747126	1.0000000	9
[151]	{}	=>	{ΑΛΛΑΤΙΝΗ ΠΤΙ ΜΠΕΡ ΚΛΑΣ. 225ΓΡ (0.1 Leaflet}	0.007662835	0.007662835	1.0000000	12
[152]	{}	=>	{SWIFFER DUSTER(ΑΑΒΗ+5 REF.ΠΑΝΑΚΙΑ) Leaflet}	0.005108557	0.005108557	1.0000000	8
[153]	{}	=>	{VILEDA ΣΦΟΥΓΓΑΡΑΚΙ COMBO 3ΤΕΜ TV}	0.007024266	0.007024266	1.0000000	11
[154]	{}	=>	{MOLTO ΚΡΟΥΑΣΑΝ ΠΡΑΛΙΝΑ 5Χ40GR Leaflet}	0.006385696	0.006385696	1.0000000	10
[155]	{}	=>	{LURPAK LIGHTER SOFT ΑΝΑΛΑΤΟ 225ΓΡ Leaflet}	0.008939974	0.008939974	1.0000000	14
[156]	{}	=>	{ΑΜΙΤΑ ΦΡΟΥΤΟΠΟΤΟ ΠΜΒ 1.5LT Leaflet}	0.008301405	0.008301405	1.0000000	13
[157]	{}	=>	{WETTEX ΣΠΟΓΓΟΠΕΤΣΕΤΑ Ν01(2+1) Leaflet}	0.007024266	0.007024266	1.0000000	11
[158]	{}	=>	{ΚΙΧΙ ΤΑΥΙ ΠΙΤΑ ΤΥΡΙ 800Γ Κ/Ψ Leaflet}	0.007024266	0.007024266	1.0000000	11
[159]	{}	=>	{KINDER ΕΚΠΑΗΗ Τ3 Leaflet}	0.008301405	0.008301405	1.0000000	13
[160]	{}	=>	{ΤΥΡΙ REGATO ΤΡΙΜΜΕΝΟ 400Γ Leaflet}	0.009578544	0.009578544	1.0000000	15
[161]	{}	=>	{ΑΒ ΑΥΓΑ ΕΛ.ΒΟΣΚΗΣ 6ΑΔΑ LARGE 63 73G Leaflet}	0.008939974	0.008939974	1.0000000	14
[162]	{}	=>	{TOTAL 500ΓΡ Leaflet}	0.008301405	0.008301405	1.0000000	13
[163]	{}	=>	{SANITAS ULTRA ΜΕΓΑΛΕΣ 10Τ 1+1 Leaflet}	0.005747126	0.005747126	1.0000000	9
[164]	{}	=>	{ΚΕΦΑΛΟΤΥΡΙ ΠΡΟΒΕΙΟ ΕΛΑΣΣΟΝΑΣ ΤΥΠ. Leaflet}	0.007662835	0.007662835	1.0000000	12
[165]	{}	=>	{BARILLA GRANCERALE ΜΠΙΣΚ.ΤΡΑΓ.230 Leaflet}	0.007662835	0.007662835	1.0000000	12
[166]	{}	=>	{LENOR GOLD ORCHID 26ΜΕΖ Leaflet}	0.008939974	0.008939974	1.0000000	14
[167]	{}	=>	{ΣΟΥΒΛΑΚΙ ΧΟΙΡΙΝΟ ΟΛΛΑΝΔΙΑΣ TV}	0.009578544	0.009578544	1.0000000	15
[168]	{}	=>	{ΠΑΥΛΙΔΗΣ ΥΓΕΙΑΣ ΓΕΜ. ΚΑΚΑΟ 100GR Leaflet}	0.010855683	0.010855683	1.0000000	17
[169]	{}	=>	{ΕΛΙΑ ΝΕΑΡΟΥ ΜΟΣΧΟΥ Μ/Ο 12ΜΗΝΟΝ ΟΛΗ Leaflet}	0.010217114	0.010217114	1.0000000	16
[170]	{}	=>	{ΓΡΑΒΙΕΡΑ ΚΡΗΤΗΣ ΤΥΠ/ΝΗ TV}	0.010855683	0.010855683	1.0000000	17
[171]	{}	=>	{ΚΥΚΝΟΣ ΤΟΜΑΤΕΣ ΤΡΙΜΜΕΝΕΣ 680GR Leaflet}	0.008301405	0.008301405	1.0000000	13
[172]	{}	=>	{NOYNOY KID 6Χ400GR Leaflet}	0.007662835	0.007662835	1.0000000	12
[173]	{}	=>	{OREO ΜΠΙΣΚΟΤΑ BROWNIE 154Γ Leaflet}	0.008939974	0.008939974	1.0000000	14
[174]	{}	=>	{BARILLA ΣΠΑΓΓΕΤΙΝΙ Ν7 3Χ500G 0.65E Leaflet}	0.007662835	0.007662835	1.0000000	12
[175]	{}	=>	{VILEDA CLAS.ΣΦΟΥΓΓ.ΚΟΥΖΙΝΑΣ 1Τ TV}	0.011494253	0.011494253	1.0000000	18
[176]	{}	=>	{ΚΡΙΣ ΚΡΙΣ ΤΟΣΤΙΜΟ ΟΛΙΚΗΣ ΑΛ. 400G Leaflet}	0.010217114	0.010217114	1.0000000	16
[177]	{}	=>	{SANITAS ΣΑΚ.ΑΠΟΜ.ΑΡΟΜ.ΜΙΝΙ30Τ 1+1 Leaflet}	0.006385696	0.006385696	1.0000000	10
[178]	{}	=>	{MIPANTA ΜΠΙΣΚΟΤΑ 2Χ250ΓΡ Leaflet}	0.008939974	0.008939974	1.0000000	14
[179]	{}	=>	{BARILLA ΜΑΚΑΡΟΝΙΑ Ν3 3Χ500GR 0.65E Leaflet}	0.008301405	0.008301405	1.0000000	13
[180]	{}	=>	{EVA ΠΟΡΤΟΚΑΛΑΔΑ ΚΟΚΚ.1.5LT(0.30ΕΚΠ Leaflet}	0.008301405	0.008301405	1.0000000	13
[181]	{}	=>	{BARILLA ΠΕΝΕΣ ΡΙΓΚΑΤ.2Χ500GR(0.40E Leaflet}	0.007662835	0.007662835	1.0000000	12
[182]	{}	=>	{LACTA TRIPLE ΣΟΚΟΛΑΤΑ 90Γ Leaflet}	0.010855683	0.010855683	1.0000000	17
[183]	{}	=>	{ΠΑΥΛΙΔΗΣ ΣΟΚ.ΥΓΕΙΑΣ ΑΜΥΓΔΑΛΟΥ 100G Leaflet}	0.009578544	0.009578544	1.0000000	15
[184]	{}	=>	{LURPAK SOFT ΑΝΑΛΑΤΟ 225ΓΡ Leaflet}	0.010217114	0.010217114	1.0000000	16
[185]	{}	=>	{2 ΛΑΙΜΟΣ ΧΟΙΡ.ΟΛΛ. ΚΙΒ.ΧΥΜΑ TV}	0.014687101	0.014687101	1.0000000	23
[186]	{}	=>	{RIO MARE ΤΟΝΟΣ ΣΕ ΝΕΡΟ 2Χ112GR Leaflet}	0.008301405	0.008301405	1.0000000	13
[187]	{}	=>	{TOTAL 2% 500ΓΡ Leaflet}	0.014687101	0.014687101	1.0000000	23
[188]	{}	=>	{ΑΒ ΦΑΚΕΣ ΨΙΛΕΣ 500G ΕΙΣΑΓΩΓΗΣ Leaflet}	0.010217114	0.010217114	1.0000000	16
[189]	{}	=>	{LACTA ΣΟΚ/ΤΑ LOV IT OREO 105GR Leaflet}	0.014048531	0.014048531	1.0000000	22
[190]	{}	=>	{HELLMANN'S REAL ΜΑΓΓΙΟΝΕΖΑ 450ML Leaflet}	0.015964240	0.015964240	1.0000000	25
[191]	{}	=>	{ΣΟΚ. ΠΑΥΛΙΔΗΣ ΥΓΕΙΑΣ 70% 100G Leaflet}	0.017879949	0.017879949	1.0000000	28
[192]	{}	=>	{VILEDA CLAS.ΣΦΟΥΓΓ.ΚΟΥΖΙΝΑΣ 2+1Τ.Δ TV}	0.016602810	0.016602810	1.0000000	26
[193]	{}	=>	{ΑΓΓΕΛΑΔΙΤΣΑ 2% 3Χ200ΓΡ Leaflet}	0.015964240	0.015964240	1.0000000	25
[194]	{}	=>	{ΔΕΣΜΕΥΤΗΣ ΧΡΩΜΑΤ.& ΒΡΩΜΙΑΣ 10φ+5φΔΩ Leaflet}	0.011494253	0.011494253	1.0000000	18
[195]	{}	=>	{ΠΑΥΛΙΔΗΣ ΣΟΚ. ΥΓΕΙΑΣ ΠΟΡΤΟΚΑΛΙ 100 Leaflet}	0.015325670	0.015325670	1.0000000	24
[196]	{}	=>	{ΚΟΥΝΟΥΠΙΔΙ ΕΛΛΗΝΙΚΟ Leaflet}	0.017241379	0.017241379	1.0000000	27
[197]	{}	=>	{ΜΠΡΙΣΟΛΑ ΧΟΙΡΙΝΗ Μ/Ο ΟΛΛΑΝΔΙΑΣ TV}	0.019795658	0.019795658	1.0000000	31
[198]	{}	=>	{MILNER ΦΕΤΕΣ 300GR Leaflet}	0.015964240	0.015964240	1.0000000	25
[199]	{}	=>	{ΑΒ ΚΟΚΚΙΝΟΥΦΑΡΟ Α/Κ ΔΙΧΤΑΚΙ Leaflet}	0.021711367	0.021711367	1.0000000	34
[200]	{}	=>	{ΜΟΡΤΑΔΕΛΑ ΜΠΟΛΟΝΙΑ ΥΦΑΝΤΗΣ Leaflet}	0.020434227	0.020434227	1.0000000	32
[201]	{}	=>	{ΠΑΥΛ.ΥΓΕΙΑΣ 70% ΑΛΑΤ.ΑΜΥΓΔ. 100Γ Leaflet}	0.022349936	0.022349936	1.0000000	35
[202]	{}	=>	{ARLA MOZZARELLA ΔΑΝΙΑΣ 40% ΦΡΑΤΖΟΛΑ Leaflet}	0.019795658	0.019795658	1.0000000	31
[203]	{}	=>	{ΑΡΑΚΑΣ ΚΑΤ/ΝΟΣ 450G ΜΠ.ΣΤΑΘΗ Leaflet}	0.021711367	0.021711367	1.0000000	34
[204]	{}	=>	{ΚΕΦΑΛΟΤΥΡΙ ΠΡΟΒΕΙΟ ΕΛΑΣΣΟΝΑΣ. Leaflet}	0.022988506	0.022988506	1.0000000	36
[205]	{}	=>	{ΦΙΛ.ΓΑΛ/ΛΑΣ ΕΝ ΕΛΛΑΔΙ ΒΡΑΣΤΟ ΤΟΣΤ Leaflet}	0.020434227	0.020434227	1.0000000	32
[206]	{}	=>	{ΑΒ ΒΙΟ ΚΑΡΟΤΑ ΕΛΛΗΝΙΚΑ Leaflet}	0.017879949	0.017879949	1.0000000	28
[207]	{}	=>	{NUTELLA Β READY ΜΠΙΣΚΟΤΑ Τ6 132Γ Leaflet}	0.022988506	0.022988506	1.0000000	36
[208]	{}	=>	{ΚΡΙΣ ΤΟΣΤΙΜΟ ΣΤΑΡΕΝΙΟ 400G 0.30E Leaflet}	0.024265645	0.024265645	1.0000000	38
[209]	{}	=>	{ΜΗΛΑ ΣΤΑΡΚΙΝΓΚ ΕΛΛΗΝΙΚΑ TV}	0.036398467	0.036398467	1.0000000	57
[210]	{}	=>	{ΦΙΛΕΤΟ ΓΑΛΟΠ.ΕΝ ΕΛΛΑΔΙ ΚΑΠΝ ΤΟΣΤ Leaflet}	0.036398467	0.036398467	1.0000000	57
[211]	{}	=>	{ΓΡΑΒΙΕΡΑ ΠΟΠ ΚΡΗΤΗΣ TV}	0.042784163	0.042784163	1.0000000	67
[212]	{}	=>	{ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}	0.041507024	0.041507024	1.0000000	65
[213]	{}	=>	{ΚΡΙΣ ΤΟΣΤΙΜΟ ΣΤΑΡΕΝΙΟ 800G 0.45E Leaflet}	0.042784163	0.042784163	1.0000000	67
[214]	{}	=>	{ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.042784163	0.042784163	1.0000000	67
[215]	{}	=>	{ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}	0.043422733	0.043422733	1.0000000	68
[216]	{ΜΠΑΝΑΝΕΣ DULCE BUONA ΕΙΣΑΓΩΓΗΣ}	=>	{ΚΡΙΣ ΤΟΣΤΙΜΟ ΣΤΑΡΕΝΙΟ 800G 0.45E Leaflet}	0.004469987	0.037837838	0.8843889	7

Η τιμή της υποστήριξης κυμαίνεται από [0,004,0,043] μονάδες. Την υψηλότερη τιμή υποστήριξης ίση με 4,34%, την έχει ο κανόνας {ΛΑΧΑΝΟ ΚΟΚΚΙΝΟ ΕΛΛΗΝΙΚΟ} → {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ}, ενώ τη χαμηλότερη τιμή υποστήριξης ίση με 0,4699 την έχουν 22 προϊόντα και 33 κανόνες. Κάποια από αυτά τα προϊόντα είναι τα : «ΤΣΙΚΛΑ ORBIT ΔΥΟΣΜΟΣ ΜΠΟΥΚΑΛΙ 64ΓΡ Leaflet», «ΜΠΙΣΚΟΤΑ OREO ΓΕΜΙΣΤΑ ΒΑΝΙΛΙΑ 154Γ Leaflet», «HELLMANN.S ΚΕΤΣΑΠ ΦΙΑΛΗ 340ML Leaflet», «ΜΙΜΙΚΟΣ ΦΙΛΕΤΟ ΑΠΟ ΣΤΗΘΟΣ 650ΓΡ Leaflet», «BARILLA ΣΠΑΓΓΕΤΙ Ν5 500GR Leaflet», «WETTEX ΣΦΟΥΓΓΑΡΙΣΤΡΑ Leaflet», «ΕΨΑ ΠΟΡΤΟΚΑΛΑΔΑ LIGHT 1.5LT ΕΚΠΤ.0 Leaflet» και κάποιοι κανόνες είναι οι : {GOUDA ΣΕ ΦΕΤΕΣ 400ΓΡ} → {ΚΡΙΣ ΤΟΣΤΙΜΟ ΣΤΑΡΕΝΙΟ 800G 0.45E Leaflet}, {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ leaflet, ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ} → [ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ leaflet}, {ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ, ΤΟΜΑΤΕΣ ΧΥΜΑ ΕΛΛΗΝΙΚΕΣ} → {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}, ΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ, ΜΠΑΝΑΝΕΣ DULCE BUONA ΕΙΣΑΓΩΓΗΣ} → {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet}, {ΣΕΛΙΝΟ ΕΛΛΗΝΙΚΟ} → {ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ Leaflet}, {ΜΗΛΑ ΣΤΑΡΚΙΝΓΚ ΕΛΛΗΝΙΚΑ TV} → {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ Leaflet}, {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ Leaflet} → {ΜΗΛΑ ΣΤΑΡΚΙΝΓΚ ΕΛΛΗΝΙΚΑ TV} κα.

Την υψηλότερη τιμή εμπιστοσύνης ίση με 50% την έχει ο κανόνας {ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ, ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ} → {ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ}, που σημαίνει ότι το 50% των φορών που οι καταναλωτές προτιμούν τα προϊόντα «ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ, ΚΑΡΟΤΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ», προτιμούν και το προϊόν προσφοράς «ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ». Επίσης, ο ίδιος κανόνας έχει και την υψηλότερη τιμή lift, ίση με 11,68.

Η χαμηλότερη τιμή εμπιστοσύνης είναι ίση με 0,4469% και την έχουν τα προϊόντα «ΣΟΛΟΜΟΣ ΦΕΤΕΣ ΥΔΑΤ.ΝΟΡΒΗΓΙΑΣ ΝΩΠΙΟΣ Leaflet», «ΑΒ ΑΥΓΑ ΕΛΕΥΘ. ΒΟΣΚΗΣ 6ΑΔΑ MEDIUM Leaflet», «ΑΒ ΦΕΤΑ ΠΟΠ ΧΕΛΜΟΥ ΣΕ ΑΛΜΗ 400 ΓΡ. Leaflet», «ΦΥΛΛΟ ΚΡΟΥΣΤΑΣ ΠΟΛΙΤΙΚΟ 450G ALFA Leaflet», «LIPTON Μ.ΤΣΑΙ ΒΑΝ/ΚΑΡΑΜ.ΠΥΡ.20Χ1.7 Leaflet», «ION BREAK ΜΥΡΤΙΛΑ ΑΜΥΓΔΑΛΑ 85Γ Leaflet» κα.

Τη χαμηλότερη τιμή lift ίση με 0,88, την έχει ο κανόνας 216 {ΜΠΑΝΑΝΕΣ DULCE BUONA ΕΛΛΗΝΙΚΕΣ} → {ΚΡΙΣ ΤΟΣΤΙΜΟ ΣΤΑΡΕΝΙΟ 800G 0.45E}. Αυτό συνεπάγεται ότι η αγορά του προϊόντος που βρίσκεται στο αριστερό μέλος του κανόνα δεν ενθαρρύνει την αγορά του προϊόντος που βρίσκεται στο δεξί μέλος. Είναι φανερό ότι την δεύτερη εβδομάδα δημιουργήθηκαν περισσότεροι κανόνες δύο και τριών στοιχειοσυνόλων, που έχουν το δεξί μέλος τους περισσότερα προϊόντα προσφοράς.

Ο Πίνακας 9, παρουσιάζει στη πρώτη στήλη τα προϊόντα προσφοράς που βρίσκονται στο δεξί μέλος των κανόνων και τις 2 εβδομάδες και στη δεύτερη στήλη ποια είναι η προωθητική τους ενέργεια.

RHS	PromoSKU
ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ	Leaflet
ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ	Leaflet
ΠΟΡΤΟΚΑΛΙΑ ΝΑΒΕΛΙΝΕΣ ΕΛΛΗΝΙΚΑ	Leaflet
ΚΡΙΣ ΤΟΣΤΙΜΟ ΣΤΑΡΕΝΙΟ	Leaflet
ΚΟΥΝΟΥΠΙΔΙ ΕΛΛΗΝΙΚΟ	Leaflet

Πίνακας 9 Προϊόντα που Εμφανίζονται και τις Δύο Εβδομάδες στο Δεξί Μέλος των Κανόνων

Τα προϊόντα που είναι σκιαγραφημένα με γκρι χρώμα, υπάρχουν στο δεξί μέλος των κανόνων που δημιουργήθηκαν και την πρώτη εβδομάδα. Εξετάζοντας τους κανόνες που έχουν στο δεξί μέλος τα σκιαγραφημένα προϊόντα τη πρώτη και τη δεύτερη εβδομάδα, μπορούμε να κάνουμε κάποιες παρατηρήσεις.

Το προϊόν «ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ» αγοράζεται και τις δύο εβδομάδες με το προϊόν «ΛΟΥΜΙΔΗΣ ΚΑΦΕΣ ΠΑΠΑΓ. ΠΑΡΑΔΟΣ.194GR».

Παρόλο που η υποστήριξη του κανόνα { ΛΟΥΜΙΔΗΣ ΚΑΦΕΣ ΠΑΠΑΓ. ΠΑΡΑΔΟΣ.194GR } → { ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ } είναι πολύ μικρής τάξεως 0,5% στην πρώτη και στην δεύτερη εβδομάδα, στη πρώτη εβδομάδα το 7% των φορών που οι καταναλωτές θα αγοράσουν το προϊόν «ΛΟΥΜΙΔΗΣ ΚΑΦΕΣ ΠΑΠΑΓ. ΠΑΡΑΔΟΣ.194GR» θα αγοράσουν και το προϊόν «ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ», ενώ τη δεύτερη εβδομάδα το 10% των φορών που οι καταναλωτές θα αγοράσουν το προϊόν «ΛΟΥΜΙΔΗΣ ΚΑΦΕΣ ΠΑΠΑΓ. ΠΑΡΑΔΟΣ.194GR» θα αγοράσουν και το προϊόν «ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ» που είναι σε προσφορά. Κάτι τέτοιο επαληθεύεται και από τον μέτρο lift.

Στην πρώτη εβδομάδα, το ποσοστό των καταναλωτών που αγοράζουν το προϊόν «ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ», είναι περίπου 4,7%. Στον υποπληθυσμό των καταναλωτών που αγοράζουν το προϊόν «ΛΟΥΜΙΔΗΣ ΚΑΦΕΣ ΠΑΠΑΓ. ΠΑΡΑΔΟΣ.194GR», το ποσοστό των καταναλωτών που αγοράζουν το προϊόν «ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ», είναι 1,61 φορές μεγαλύτερο από ότι είναι στον συνολικό πληθυσμό (4,7%).

Αντίστοιχα στην δεύτερη εβδομάδα, όπου το προϊόν «ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ» είναι σε προσφορά, το ποσοστό των καταναλωτών που αγοράζουν «ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ» είναι 4,2%. Στον υποπληθυσμό των καταναλωτών που αγοράζουν το προϊόν «ΛΟΥΜΙΔΗΣ ΚΑΦΕΣ ΠΑΠΑΓ. ΠΑΡΑΔΟΣ.194GR», το ποσοστό των καταναλωτών που αγοράζουν «ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ» είναι 2.53 φορές υψηλότερο από το συνολικό πληθυσμό (4,2%). Επίσης, το προϊόν «ΛΑΧΑΝΟ ΑΣΠΡΟ ΕΛΛΗΝΙΚΟ» αγοράζεται ξεχωριστά με δύο ακόμη προϊόντα προσφοράς. Αυτά είναι τα προϊόντα «ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ» που είναι σε προσφορά μέσω του φυλλαδίου του καταστήματος και τα «ΜΗΛΑ ΣΤΑΡΚΙΝΓΚ ΕΛΛΗΝΙΚΑ» που διαφημίζονται μέσω της τηλεόρασης.

Όσον αφορά το προϊόν «ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ» που είναι σε προσφορά, αγοράζεται και τις δύο εβδομάδες με το προϊόν «ΠΑΤΑΤΕΣ ΧΟΝΔΡΕΣ ΕΛΛΗΝΙΚΕΣ» αλλά τη δεύτερη εβδομάδα, το τελευταίο προϊόν ενισχύει περισσότερο από τη πρώτη εβδομάδα την αγορά του προϊόντος «ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ». Παρόλο που το «ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ» και τις δύο εβδομάδες βρίσκεται περίπου στο 4,4% των συνολικών συναλλαγών, την πρώτη εβδομάδα στον υποπληθυσμό των καταναλωτών που αγοράζουν το προϊόν «ΠΑΤΑΤΕΣ ΧΟΝΔΡΕΣ ΕΛΛΗΝΙΚΕΣ», το ποσοστό των καταναλωτών που αγοράζουν το προϊόν «ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ» είναι 1,75 φορές μεγαλύτερο από ότι είναι στον συνολικό πληθυσμό (περίπου 4,4%).

Τη δεύτερη εβδομάδα, στον υποπληθυσμό των καταναλωτών που αγοράζουν «ΠΑΤΑΤΕΣ ΧΟΝΔΡΕΣ ΕΛΛΗΝΙΚΕΣ», το ποσοστό των καταναλωτών που αγοράζουν «ΜΠΡΟΚΟΛΟ ΠΡΑΣΙΝΟ ΕΛΛΗΝΙΚΟ», είναι 2,44 φορές μεγαλύτερο από ότι είναι στον συνολικό πληθυσμό (περίπου 4,3%).

Όσον αφορά στο προϊόν «ΚΡΙΣ ΤΟΣΤΙΜΟ ΣΤΑΡΕΝΙΟ 800G 0.45E», διακρίνουμε σημαντικές διαφορές την εβδομάδα που το προϊόν αυτό είναι σε προσφορά.

Την πρώτη εβδομάδα αγοράζεται μόνο με το προϊόν «GOUDA ΓΕΡΜΑΝΙΑΣ ΦΡΑΤΖΟΛΑ» κι εμφανίζονται μαζί στο 0,4% των συναλλαγών.

Τη δεύτερη εβδομάδα, αγοράζεται με τα προϊόντα «GOUDA ΣΕ ΦΕΤΕΣ 400ΓΡ», «ΓΑΛΟΠΟΥΛΑ ΚΑΠΝΙΣΤΗ ΤΟΣΤΑΚΙ», «ΤΥΡΙ ΝΟΥΝΟΥ GOUDA ΟΛΛΑΝΔ.ΦΡΑΤΖΟΛΑ», «GOUDA ΓΕΡΜΑΝΙΑΣ ΦΡΑΤΖΟΛΑ», «ΓΡΑΒΙΕΡΑ ΠΟΠ ΚΡΗΤΗΣ », «ΑΒ ΖΑΧΑΡΗ ΛΕΥΚΗ ΚΡΥΣ/ΚΗ Ε.Ε 1KG», «ΤΟΜΑΤΕΣ ΧΥΜΑ ΕΛΛΗΝΙΚΕΣ», «ΚΡΕΜΜΥΔΙΑ. ΞΕΡΑ ΧΟΝΔΡΑ ΧΥΜΑ ΕΛΛΗΝΙΚΑ» και «ΜΠΙΑΝΑΝΕΣ DULCE BUONA ΕΛΛΗΝΙΚΕΣ».

Ακόμη, το προϊόν «ΚΟΥΝΟΥΠΙΔΙ ΕΛΛΗΝΙΚΟ» εμφανίζεται μαζί με περισσότερα προϊόντα την εβδομάδα που δεν είναι σε προσφορά παρά την εβδομάδα που είναι. Πιο συγκεκριμένα οι

κανόνες που δημιουργούνται τη πρώτη εβδομάδα με το προϊόν «ΚΟΥΝΟΥΠΙΔΙ ΕΛΛΗΝΙΚΟ» στο δεξί μέλος των κανόνων είναι 6 στο σύνολο, ενώ τη δεύτερη εβδομάδα είναι μόνο 2.

Τέλος, το προϊόν «ΠΟΡΤΟΚΑΛΙΑ NABEΛΙΝΕΣ ΕΛΛΗΝΙΚΑ» την πρώτη εβδομάδα εμφανίζεται στο 2,5% των συναλλαγών, ενώ τη δεύτερη εβδομάδα εμφανίζεται στο 4,1% των συναλλαγών. Και τις δύο εβδομάδες, το προϊόν «ΠΟΡΤΟΚΑΛΙΑ NABEΛΙΝΕΣ ΕΛΛΗΝΙΚΑ» βρέθηκε να αγοράζεται με το προϊόν «ΤΟΜΑΤΕΣ ΧΥΜΑ ΕΛΛΗΝΙΚΕΣ» και με το προϊόν «ΜΠΑΝΑΝΕΣ DULCE BUONA ΕΙΣΑΓΩΓΗΣ». Η διαφορά έγκειται στο ότι τη δεύτερη εβδομάδα που τα «ΠΟΡΤΟΚΑΛΙΑ NABEΛΙΝΕΣ ΕΛΛΗΝΙΚΑ» βρίσκονται σε προσφορά, τα μέτρα της υποστήριξης της εμπιστοσύνης και του lift είναι υψηλότερα σε σχέση με τη πρώτη εβδομάδα.

ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην εργασία αυτή μελετήθηκε η εφαρμογή της ανάλυσης συσχέτισης σε δεδομένα συναλλαγών αγοράς η οποία μπορεί να αναγνωρίσει σχέσεις μεταξύ προϊόντων και μεταξύ κατηγοριών. Τα συμπεράσματα μας αφορούν κυρίως γενικές παρατηρήσεις που εξήχθησαν από τον αλγόριθμο Apriori.

Στο πρώτο μέρος της ανάλυσης που πραγματοποιήθηκε, οι κατηγορίες που εμφανίζονταν πολλές φορές στις συναλλαγές μας, ήταν αυτές οι οποίες εμφανίζονταν και στους κανόνες συσχέτισης. Αυτό συμβαίνει καθώς το πλήθος των κατηγοριών είναι μικρό σε σχέση με το πλήθος των συναλλαγών. Σε αντίθεση, το πλήθος των προϊόντων μας είναι πολύ μεγάλο σε σχέση με το πλήθος των συναλλαγών, οπότε δεν θα είχε αξία να προσπαθήσουμε να βρούμε συσχετίσεις μεταξύ προϊόντων καθώς η υποστήριξη τους θα ήταν πάρα πολύ χαμηλή. Ως εκ τούτου, ανάλογα με τα δεδομένα του προβλήματος πρέπει να εκτιμούμε ανά περίπτωση την επιλογή του υποσυνόλου ή του συνόλου δεδομένων που θα επεξεργαστούμε. Επιπλέον, παρατηρήσαμε ότι όταν χαμηλώσαμε τη τιμή του προκαθορισμένου κατωφλιού του μέτρου υποστήριξης, το μέτρο lift αυξήθηκε. Αυτό δείχνει, ότι παρόλο που τα υποσύνολα των κανόνων δεν εμφανίζονται συχνά μαζί στις συναλλαγές, στο υποσύνολο του αριστερού μέλους του κανόνα, η εμφάνιση του δεξιού μέλους του κανόνα είναι έντονη. Τέλος, ένα ακόμα συμπέρασμα που προκύπτει από την πρώτη ανάλυσή μας είναι ότι πρέπει να είμαστε προσεκτικοί ως προς την ερμηνεία του μέτρου εμπιστοσύνης, καθώς ένας κανόνας συσχέτισης με υψηλή εμπιστοσύνη αλλά με χαμηλή υποστήριξη, μπορεί να οφείλεται σε τυχαιότητα.

Στο δεύτερο μέρος της ανάλυσής μας, εξετάσαμε τα καλάθια των καταναλωτών για μια περίοδο όπου κάποια προϊόντα βρίσκονταν σε προσφορά προσπαθώντας να αναγνωρίσουμε συσχετίσεις αναμεσα στα προϊόντα προσφοράς της δεύτερης εβδομάδας σε σχέση με την πρώτη. Αρχικά, από τα περιγραφικά στατιστικά είδαμε ότι η χρηματική αξία των καλαθιών των καταναλωτών που επισκέφτηκαν το κατάστημα και τις δύο εβδομάδες, αυξήθηκε τη δεύτερη εβδομάδα όπου κάποια προϊόντα ήταν σε προσφορά. Ακόμη, αυξήθηκε η κατανάλωση των προϊόντων που ήταν σε προσφορά τη δεύτερη εβδομάδα σε σχέση με τη πρώτη. Όσον αφορά τους κανόνες συσχέτισης της πρώτης εβδομάδας, τους συγκρίναμε με τους κανόνες συσχέτισης της δεύτερης. Σχετικά με τους κανόνες που δημιουργήθηκαν και τις δύο εβδομάδες, εστίασαμε μόνο στους κανόνες που στο δεξί τους μέλος έχουν κοινά προϊόντα προσφοράς. Παρατηρήσαμε ότι κάποιοι καταναλωτές δεν άλλαξαν συμπεριφορά ως προς την επιλογή των προϊόντων που τοποθετούν στο καλάθι τους. Οι περισσότερες όμως συσχετίσεις που αφορούν τον ίδιο κανόνα της πρώτης και δεύτερης εβδομάδας, είχαν υψηλότερο lift τη δεύτερη

εβδομάδα σε σχέση με τη πρώτη, δείχνοντας ότι κάποια προϊόντα ενίσχυσαν τη παρουσία τους στα καλάθια που ούτως ή άλλως εμφανίζονταν. Αντίθετα, υπήρχε ένας κανόνας όπου το μέτρο lift είχε μικρότερη τιμή από 1 που σημαίνει ότι το αριστερό μέλος του κανόνα, είχε αρνητική επίδραση στο δεξί. Ακόμη, οι προωθητικές ενέργειες συνέβαλαν θετικά στις περισσότερες αγορές προϊόντων καθώς κάποια προϊόντα εμφανίστηκαν σε περισσότερα καλάθια τη δεύτερη εβδομάδα. Μια ακόμη σημαντική παρατήρηση είναι ότι η επιρροή προς τη κατανάλωση σε προϊόντα με προσφορά δεν αφορά μόνο το προϊόν που είναι σε προσφορά, αλλά επηρεάζει θετικά και την κατανάλωση άλλων προϊόντων (βλέπε κανόνα 55 - Εικόνα 27 και κανόνα 58 - Εικόνα 32). Επιπλέον, τη δεύτερη εβδομάδα βρέθηκαν πολλοί κανόνες οι οποίοι είχαν και στα δύο μέλη τους προϊόντα προσφοράς, γεγονός που δείχνει μια αμφίδρομη σχέση όσον αφορά στις αγορές των καταναλωτών.

Συνοψίζοντας τα αποτελέσματα της ανάλυσης μας, επαληθεύονται από τον αλγόριθμο Apriori και από τα εργαλεία της ανάλυσης συσχέτισης, καθώς μετά την εφαρμογή τους σε πραγματικά δεδομένα βρέθηκαν προϊόντα και κατηγορίες που αγοράζονται μαζί από τους καταναλωτές ακόμα και αν σε γενικές γραμμές δίνουν ένα αναμενόμενο αποτέλεσμα.

ΚΩΔΙΚΕΣ

```
#####ΠΡΩΤΟ ΜΕΡΟΣ ΑΝΑΛΥΣΗΣ#####

gc() # clear cache

library(tidyverse)

library(readxl)

library(knitr)

library(ggplot2)

library(lubridate)

library(arules)

library(arulesViz)

library(plyr)

library(readxl)

library(sqldf)

library(ArArRedux)

options(max.print=999999) # max_print

Data <- read_excel("C:/Users/Marianna/Desktop/Shop Egaleo October November
2017_DELETE_KENA_DUPLICATES.xlsx")

Data <- Data[,1:5]

colnames(Data) <- c("transID","ProductID","Desc","Dept","Categ") # Labels of
Columns

length(unique(Data$`transID`)) # Number of Transactions
length(unique(Data$`ProductID`)) # Number of Products Code
length(unique(Data$`Desc`)) # Number of Descriptions
length(unique(Data$`Dept`)) # Number of Departments
```

```

length(unique(Data$`Categ`))      # Number of Categories

Data$transID<- trimws(Data$transID, which = c("both"))  # Trim spaces before and
After TransID

Data$ProductID <- trimws(Data$ProductID, which = c("both")) # Trim spaces before
and After ProductID

Data$Desc <- trimws(Data$Desc, which = c("both"))      # Trim spaces before and After
Description

Data$Dept <- trimws(Data$Dept, which = c("both"))      # Trim spaces before and After
Departments

Data$Categ <- trimws(Data$Categ, which = c("both"))    # Trim spaces before and After
Categories

### check_duplicate_description_for_one_product_code

check_dupl_desc <- sqldf("
    select ProductID,count(distinct Desc) as cnt
    from Data
    group by ProductID
    having cnt>1
    ")

Query<- sqldf("select ProductID, Desc, count(transID) as plthos
    from Data group by Desc, ProductID
    ")

#### check_duplicate_code_for_one_product_description

check_dupl_code <- sqldf("select ProductID, Desc, plthos FROM Query WHERE Desc
IN (select Desc from
    (select ProductID,Desc from Data group by Desc,ProductID)
    group by Desc having count(ProductID)>1) order by Desc,plthos asc")

```



```
##### Descriptive Statistics For Data #####
```

```
str(Data)
```

```
Data$ProductID <- as.factor(Data$ProductID)
```

```
Data$Desc <- as.character(Data$Desc)
```

```
Data$transID <- as.integer(Data$transID)
```

```
#Πλήθος Προϊόντων ανά συναλλαγή
```

```
Plot1 <- sqldf("
```

```
    select transID,count(ProductID) as num_of_prod
```

```
    from Data
```

```
    group by transID
```

```
    ")
```

```
#Take "num_of_prod" from Plot1
```

```
Plotn<- sqldf("
```

```
    select num_of_prod
```

```
    from Plot1
```

```
    ")
```

```
#Create a distribution of number of products per transaction
```

```
Plotn$num_of_prod <- as.factor(Plotn$num_of_prod)
```

```
a %>
```

```
ggplot(Plotn,aes(x=num_of_prod)) +
```

```
geom_histogram(stat="count",fill="indianred")
```

#Πόσα προϊόντα περιέχονται σε πόσες συναλλαγές

```
Plot2 <- sqldf("
    select plithos_prod,count(transID) as count_trans
    from Plot1
    group by plithos_prod
    ")
```

#Το κάθε προϊόν σε πόσες συναλλαγές υπάρχει

```
Plot3 <- sqldf("
    select Desc,count(*) as plithos_trans
    from Data
    group by Desc
    ")
```

#Πόσες κατηγορίες έχει η κάθε συναλλαγή

```
Plot4 <- sqldf("
    select transID,count (distinct(Categ)) as num_of_categ
    from Data
    group by transID
    ")
```

Take the "num_of_categ" from Plot4

```
plot4n <- sqldf("
    select num_of_categ
    from Plot4
    ")
```

#Create a distribution of number of categories per transaction

```
plot4n$num_of_categ <- as.factor(Plotn$num_of_categ)
a %
ggplot(plot4n,aes(x=num_of_categ)) +
  geom_histogram(stat="count",fill="indianred")
```

#Πόσα τμήματα έχει η κάθε συναλλαγή

```
Plot5 <- sqldf("
  select transID,count( distinct (Dept)) as plithos_dept
  from Data
  group by transID order by transID desc
  ")
```

#Πλήθος Προϊόντων ανά Κατηγορία

```
Plot6 <- sqldf("
  select Categ,count (distinct(ProductID)) as plithos_prod
  from Data
  group by Categ
  ")
```

Πλήθος Προϊόντων ανά Τμήμα

```
Plot8 <- sqldf("
  select Dept,count (distinct(ProductID)) as plithos_prod
  from Data
  group by Dept
  ")
```

Πλήθος Κατηγοριών ανά Τμήμα

```
Plot9 <- sqldf("
  select Dept,count (distinct(Categ)) as plithos_Categ
  from Data
  group by Dept
  ")
```

On how many transactions each product appears in average

```
products_in_trans1 <- sqldf ("select Desc, count(Desc) as plithos
  from Data group by Desc order by plithos desc
  ")
```

```
mean(products_in_trans1$plithos)
```

On how many transactions each category appears in average

```
Categories_in_trans1 <- sqldf ("select Categ, count(Categ) as plithos
  from Data group by Categ order by plithos desc")
```

```
mean(Categories_in_trans1$plithos)
```

Select only 9 of 35 Departments

```
Dept <- sqldf("select distinct Dept
  from Data where Dept = 'LIQUIDE GROCERY FOOD' or
  Dept = 'ΓΑΛΑΚΤΟΚΟΜΙΚΑ-ΑΥΓΑ-ΦΡΕΣΚΟΙ ΧΥΜΟΙ'
  or Dept = 'KABA' or Dept = 'ΚΡΕΟΠΩΛΕΙΟ'
  or Dept = 'ΠΑΝΤΟΠΩΛΕΙΟ' or Dept = 'ΥΓΙΕΙΝΗ - ΟΜΟΡΦΙΑ'or Dept =
  'ΑΛΛΑΝΤΙΚΑ'
  or Dept = 'ΤΥΡΙΑ' or Dept = 'ΑΠΟΡΡΥΠΑΝΤΙΚΑ - ΚΑΘΑΡΙΣΤΙΚΑ'
  ")
```

Dataset_FinalData with only 9 specific Departments

```
FinalData <- sqldf("select * from Data where Dept IN (select * from Dept)")
```

```
colnames(FinalData) <- c("transID","ProductID","Desc","Dept","Categ") # Labels of FinalData
```

```
#View(FinalData)
```

```
length(unique(FinalData$`transID`)) # Number of Transactions
```

```
length(unique(FinalData$`ProductID`)) # Number of Products Code
```

```
length(unique(FinalData$`Desc`)) # Number of Descriptions
```

```
length(unique(FinalData$`Dept`)) # Number of Departments
```

```
length(unique(FinalData$`Categ`)) # Number of Categories
```

On how many transactions each product appears in average_FinalData

```
products_in_trans2 <- sqldf ("select Desc, count(Desc) as plithos
                             from Deptfinal group by Desc order by plithos desc
                             ")
```

```
mean(products_in_trans2$plithos)
```

On how many transactions each category appears in average_FinalData

```
Categories_in_trans2 <- sqldf ("select Categ, count(Categ) as plithos
                               from FinalData group by Categ order by plithos desc")
```

```
mean(Categories_in_trans2$plithos)
```

Format_FinalData_to_insert_in_Apriori_Algorithm

```
data_sorted <- FinalData[order(FinalData$transID),]
```

Create a list contains Categories per transaction

```

itemList2 <- ddply(FinalData,c("transID"),
  function(df1)paste(df1$Categ,
    collapse = ","))
View(itemList2)

```

```

itemList2$transID <- NULL
colnames(itemList2) <- c("Categories")

```

Create_CSV_FILE

```

write.csv(itemList2,"C:/Users/Marianna/Desktop/MariannaTest/market_basket2.csv",
  quote = FALSE, row.names = FALSE)

```

Apriori_reads_baskets as to product categories

```

tr2 <-
read.transactions('C:/Users/Marianna/Desktop/MariannaTest/market_basket2.csv',
  format = 'basket', sep=',',skip=1)

```

Summary of CSV

```
summary(tr2)
```

Plot_the_TOP_10_Categories

```
itemFrequencyPlot(tr2, topN=10, type='absolute')
```

Apriori Association Rules_for_Categories_sort_by_support

```

rules_categ <- apriori(tr2, parameter = list(supp=0.05, conf=0.00, target= 'rules'))
rules_categ <- sort(rules_categ, by='support', decreasing = TRUE)
summary(rules_categ)
inspect(rules_categ)

```

```

### Apriori Association Rules__for_Categories_sort_by_lift

rules_categ <- apriori(tr2, parameter = list(supp=0.01, conf=0.00, target= 'rules'),
appearance = list(rhs = 'TOMATOEIΔH'))

rules_categ <- sort(rules_categ, by='lift', decreasing = TRUE)

summary(rules_categ)

inspect(rules_categ)

#####ΔΕΥΤΕΡΟ ΜΕΡΟΣ ΑΝΑΛΥΣΗΣ#####

# LOAD THE SAME LIBRARIES WITH THE Α ANALYSIS PART

# Dataset of Week1

notPromo <- read_excel("C:/Users/Marianna/Desktop/Unique Transaction ID_30 Oct-4
Nov 2017_New.xlsx")

View(notPromo)

# Delete Column "Επισκέψεις" from Week1

notPromo <- notPromo[-grep('Επισκέψεις', colnames(notPromo))]

# Labels of Columns_Week1

colnames(notPromo) <- c("CustomerID",
"transID","MasterCode","Desc","PromoSKU","Dept", "Categ","Value"
,"Quantity"
)

# Dataset of Week2

Promo <- read_excel("C:/Users/Marianna/Desktop/Unique Transaction ID_13-18 Nov
2017_New.xlsx")

View(Promo)

# Delete Column "Επισκέψεις" from Week2

Promo <- Promo[-grep('Επισκέψεις', colnames(Promo))]

```

Labels of Columns_Week2

```
colnames(Promo) <-c("CustomerID",
"transID","MasterCode","Desc","PromoSKU","Dept", "Categ","Value"
,"Quantity"
)
```

```
length(unique(notPromo$`CustomerID`)) # Number of Customers in Week_1
```

```
length(unique(notPromo$`transID`)) # Number of Transactions in Week_1
```

```
length(unique(notPromo$`MasterCode`)) # Number of Products Code in Week_1
```

```
length(unique(notPromo$`Desc`)) # Number of Product Descriptions in Week_1
```

```
length(unique(notPromo$`Dept`)) # Number of Departments in Week_1
```

```
length(unique(notPromo$`Categ`)) # Number of Categories in Week_1
```

TRIM SPACES BEFORE AND AFTER FIELDS OF WEEK_1

```
notPromo$CustomerID<- trimws(notPromo$CustomerID, which = c("both"))
```

```
notPromo$transID <- trimws(notPromo$transID, which = c("both"))
```

```
notPromo$MasterCode <- trimws(notPromo$MasterCode, which = c("both"))
```

```
notPromo$Desc <- trimws(notPromo$Desc, which = c("both"))
```

```
notPromo$PromoSKU <- trimws(notPromo$PromoSKU,which = c("both"))
```

```
notPromo$Dept<- trimws(notPromo$Dept, which = c("both"))
```

```
notPromo$Categ<- trimws(notPromo$Categ, which = c("both"))
```

```
notPromo$Value<- trimws(notPromo$Value, which = c("both"))
```

```
notPromo$Quantity<- trimws(notPromo$Quantity, which = c("both"))
```

```
notPromo$Visits<- trimws(notPromo$Visits, which = c("both"))
```

```
length(unique(Promo$`CustomerID`)) # Number of Customers in Week_2
```

```
length(unique(Promo$`transID`)) # Number of Transactions in Week_2
```

```
length(unique(Promo$`MasterCode`)) # Number of Products Code in Week_2
```

```
length(unique(Promo$`Desc`)) # Number of Product Descriptions in Week_2
```



```

length(unique(Promo$`Dept`))    # Number of Departments in Week_2
length(unique(Promo$`Categ`))  # Number of Categories in Week_2

### TRIM SPACES BEFORE AND AFTER FIELDS OF WEEK_2

Promo$CustomerID<- trimws(Promo$CustomerID, which = c("both"))
Promo$transID <- trimws(Promo$transID, which = c("both"))
Promo$MasterCode <- trimws(Promo$MasterCode, which = c("both"))
Promo$Desc <- trimws(Promo$Desc, which = c("both"))
Promo$PromoSKU <- trimws(Promo$PromoSKU,which = c("both"))
Promo$Dept<- trimws(Promo$Dept, which = c("both"))
Promo$Categ<- trimws(Promo$Categ, which = c("both"))
Promo$Value<- trimws(Promo$Value, which = c("both"))
Promo$Quantity<- trimws(Promo$Quantity, which = c("both"))
Promo$Visits<- trimws(Promo$Visits, which = c("both"))

#### check_duplicate_description_for_one_product_code in WEEK2
check_dupl_desc <- sqldf("
    select MasterCode,count(distinct Desc) as cnt
    from Promo
    group by MasterCode
    having cnt>1
    ")

###check_duplicate_product_code_for_one_description in WEEK2
Q1<- sqldf("select MasterCode, Desc, count(transID) as plthos from Promo group by
Desc, MasterCode")

check_dupl_code <- sqldf("select MasterCode, Desc, plthos FROM Q1 WHERE Desc
IN (select Desc from (select MasterCode,Desc from Promo group by Desc,MasterCode)
group by Desc having count(MasterCode)>1) order by Desc,plthos asc
    ")

```

check_duplicate_description_for_one_product_code in WEEK1

```
check_dupl_desc <- sqldf("
    select MasterCode,count(distinct Desc) as cnt
    from notPromo
    group by MasterCode
    having cnt>1
    ")
```

check_duplicate_product_code_for_one_description in WEEK1

```
Q1<- sqldf("select MasterCode, Desc, count(transID) as plthos from notPromo group
by Desc, MasterCode")
check_dupl_code <- sqldf("select MasterCode, Desc, plthos FROM Q1 WHERE Desc
IN (select Desc from
    (select MasterCode,Desc from notPromo group by Desc,MasterCode)
    group by Desc having count(MasterCode)>1) order by Desc,plthos asc
    ")
```

###Sales of Promo Products in Both Weeks

```
promo_codes <- sqldf("select distinct MasterCode from notPromo where PromoSKU is
not null")
```

```
test1 <- sqldf("select MasterCode, Desc, Count(MasterCode) as Week1 from notPromo
where PromoSKU is not null group by MasterCode, Desc")
```

```
test2 <- sqldf("select MasterCode, Desc, Count(MasterCode) as Week2 from Promo
where MasterCode IN promo_codes group by MasterCode, Desc")
```

How many promo products were sold in both weeks?

```
sales<- sqldf("select p.Desc, n.Week1 ,p.Week2
    From test1 as n INNER JOIN test2 as p ON p.MasterCode = n.MasterCode group
by p.desc ")
```

Get the transactions ID with at least one promo product in both weeks

```
trans_with_promo1 <- sqldf("select distinct transID, CustomerID from notPromo where
PromoSKU is NOT NULL")
```

```
trans_with_promo2 <- sqldf("select distinct transID, CustomerID from Promo where
PromoSKU is NOT NULL")
```

```
promo1_list <- sqldf("select transID from trans_with_promo1")
```

```
promo2_list <- sqldf("select transID from trans_with_promo2")
```

Get customerID with at least one Promo product in both weeks

```
SAME_USERS <- sqldf("select distinct p.CustomerID
                    from trans_with_promo1 as p inner Join trans_with_promo2 as n ON
p.CustomerID= n.CustomerID
                    ")
```

Get the whole baskets with at least one promo product in both weeks

```
promo_trans_week1 <- sqldf("select * from notPromo where (CustomerID IN
SAME_USERS and transID in promo1_list) ")
```

```
promo_trans_week2 <- sqldf("select * from Promo where CustomerID IN
SAME_USERS and transID in promo2_list")
```

Basket Values related to both weeks

```
Basket_USERS1 <- sqldf("select CustomerID, sum(Value) as sum_p
                        from promo_trans_week1
                        group by CustomerID")
```

```
Basket_USERS2 <- sqldf("select CustomerID, sum(Value) as sum_n
                        from promo_trans_week2
                        group by CustomerID")
```

```
Basket_users1_2 <- sqldf("select p.CustomerID, sum_p, sum_n
                        from Basket_USERS1 as p inner join Basket_USERS2 as n where
                        p.CustomerID= n.CustomerID
                        group by p.CustomerID")
```

Summary Customers Basket Value in Week 1

```
sum_not_promo <- sqldf("select sum(sum_n) as nP from Basket_users1_2")
View(sum_not_promo)
```

Summary Customers Basket Value in Week 2

```
sum_promo <- sqldf("select sum(sum_p) as P from Basket_users1_2")
View(sum_promo)
```

Basket change rate from week 1 to week2

```
percentage <- (sum_promo - sum_not_promo)/sum_not_promo*100
View(percentage)
```

Dataset which contains the whole baskets with at least one promo product in week 1

```
colnames(promo_trans_week1) <-c("CustomerID", "transID", "MasterCode", "Desc",
                                "PromoSKU", "Dept", "Categ", "Value",
                                "Quantity", "Visits")
```

Dataset which contains the whole baskets with at least one promo product in week 2

```
colnames(promo_trans_week2) <- c("CustomerID", "transID", "MasterCode", "Desc",
                                "PromoSKU", "Dept", "Categ", "Value",
                                "Quantity", "Visits")
```

length of the Data in promo_trans_week1

```
length(unique(promo_trans_week1$`CustomerID`))
```

```
length(unique(promo_trans_week1$`transID`))
```

```
length(unique(promo_trans_week1$`MasterCode`))
```

```
length(unique(promo_trans_week1$`Desc`))
```

```
length(unique(promo_trans_week1$`Dept`))
```

```
length(unique(promo_trans_week1$`Categ`))
```

length of the Data in promo_trans_week2

```
length(unique(promo_trans_week2$`CustomerID`))
```

```
length(unique(promo_trans_week2$`transID`))
```

```
length(unique(promo_trans_week2$`MasterCode`))
```

```
length(unique(promo_trans_week2$`Desc`))
```

```
length(unique(promo_trans_week2$`Dept`))
```

```
length(unique(promo_trans_week2$`Categ`))
```

FORMAT Promo_trans_week1 to insert in Apriori in Algorithm

```
data_sorted1 <- promo_trans_week1[order(promo_trans_week1$transID),]
```

```
# Create a list with products in Week1
```

```
itemList1 <- ddply(promo_trans_week1,c("transID"),
```

```
  function(df1)paste(df1$Desc,
```

```
    collapse = ",")
```

```
View(itemList1)
```

```
itemList1$transID <- NULL
```

```
colnames(itemList1) <- c("itemsWeek1")
```

```
write.csv(itemList1,"C:/Users/Marianna/Desktop/MariannaTest2/market_basket1.csv",
quote = FALSE, row.names = FALSE)
```

```

tr1<-
read.transactions('C:/Users/Marianna/Desktop/MariannaTest2/market_basket1.csv',
format = 'basket', sep=',',skip=1)

tr1

summary(tr1)

### Plot the TOP 10 Products in Week1

itemFrequencyPlot(tr1, topN=10, type='absolute')

### Apriori Association Rules as to Product Description in Week1

rules_desc1 <- apriori(tr1, parameter = list(supp=0.002, conf=0.00))
#,appearance = list(rhs ='^AXANO ΑΣΠΠΟ ΕΛΛΗΝΙΚΟ Leaflet')
rules_desc1 <- sort(rules_desc1, by='lift' , decreasing = TRUE)

# inspect rules that have only promo products in rhs in Week1
rules_subset <- subset(rules_desc1, (rhs %in% "*(Leaflet|TV)+"))
summary(rules_subset) # Summary of Rules
inspect(rules_subset)

### Format Promo_trans_week2 to insert in Apriori in Algorithm

data_sorted2 <- promo_trans_week2[order(promo_trans_week2$transID),]

# Create a list with products in Week2
itemList2 <- ddply(promo_trans_week2,c("transID"),
  function(df1)paste(df1$Desc,
    collapse = ","))
View(itemList2)
itemList2$transID <- NULL
colnames(itemList2) <- c("itemsWeek2")

```

```

write.csv(itemList2,"C:/Users/Marianna/Desktop/MariannaTest2/market_basket2.csv",
quote = FALSE, row.names = FALSE)

tr2 <- read.transactions('C:/Users/Marianna/Desktop/MariannaTest2/market_basket2.csv',
format = 'basket', sep=',',skip=1)

summary(tr2)

### Plot the TOP 10 Products in Week2
itemFrequencyPlot(tr2, topN=10, type='absolute')

### Apriori Association Rules as to Product Description in Week2

rules_desc2 <- apriori(tr2, parameter = list(supp=0.004, conf=0.0))
#, appearance = list(rhs = 'ΚΟΥΝΟΥΠΙΑΙ ΕΛΛΗΝΙΚΟ Leaflet'))
rules_desc2 <- sort(rules_desc2, by='lift' , decreasing = TRUE)
summary(rules_desc2)
inspect(rules_desc2)

# inspect rules that have only promo products in rhs in Week2
rules_subset2 <- subset(rules_desc2, (rhs %pin% "*(Leaflet|TV)+"))
summary(rules_subset2)
inspect(rules_subset2)

##### Descriptive Statistics #####

# Πλήθος Προϊόντων ανά Συναλλαγή την Εβδομάδα 1 (Για τους κοινούς χρήστες)
Plot1 <- sqldf("
    select transID,count(MasterCode) as num_of_prod
    from promo_trans_week1
    group by transID
    ")

```

Take the “num_of_prod” from Plot1

```
plot1n <- sqldf("select num_of_prod
                from Plot1")
```

Distribution of num_of_prod Per Transaction Week1

```
plot1n$num_of_prod <- as.factor(plot1n$num_of_prod)
a %
ggplot(plot1n,aes(x=num_of_prod)) +
  geom_histogram(stat="count",fill="indianred")
```

Πλήθος Προϊόντων ανά Συναλλαγή την Εβδομάδα 2 (Για τους κοινούς χρήστες) (Για τους κοινούς χρήστες)

```
Plot2 <- sqldf("
                select transID,count(MasterCode) as num_of_prod
                from promo_trans_week2
                group by transID
                ")
```

Take the “num_of_prod” from Plot2

```
plot2n <- sqldf("select num_of_prod
                from Plot2")
```

Distribution of num_of_prod Per Transaction Week2

```
plot2n$num_of_prod <- as.factor(plot2n$num_of_prod)
a %
ggplot(plot2n,aes(x=num_of_prod)) +
  geom_histogram(stat="count",fill="indianred")
```


Το κάθε προϊόν σε πόσες συναλλαγές υπάρχει την Εβδομάδα 1 (Για τους κοινούς χρήστες)

```
Plot3 <- sqldf("
    select Desc,count(*) as num_of_trans
    from promo_trans_week1
    group by Desc
    ")
```

Το κάθε προϊόν σε πόσες συναλλαγές υπάρχει την Εβδομάδα 2 (Για τους κοινούς χρήστες)

```
Plot4 <- sqldf("
    select Desc,count(*) as num_of_trans
    from promo_trans_week2
    group by Desc
    ")
```

ΑΝΑΦΟΡΕΣ

- [1] Divya Bansal and Lekha Bhambhu (2013), “Execution of Apriori Algorithm of Data Mining directed towards tumultuous crimes against women”, International Journal of Advanced Research in Computer Science and Software Engineering 3(9), Sept-2013, pp.54-62.
- [2] Shrawan Ram and Amit Doegar (2015) , “A Comparative Study of Data Mining Techniques for Predicting Disease Using Statlog Heart Disease Database”, International Journal of Advanced Research in Computer Science and Software Engineering , Volume 5, Issue 6, June 2015 5(6), June- 2015, pp. 1202-1210.
- [3] Hong Yu and Xiaolei Huang (2010), “A Comparative Study on Data Mining Algorithms for Individual Credit Risk Evaluation”, IEEE International Conference on Management of e-Commerce and e-Government.
- [4] Piatetsky-Shapiro, Gregory (1991), Discovery, analysis, and presentation of strong rules, in Piatetsky-Shapiro, Gregory; and Frawley, William J.; eds., Knowledge Discovery in Databases, AAAI/MIT Press, Cambridge, MA.
- [5] Agrawal, R.; Imieliński, T.; Swami, A. (1993). "Mining association rules between sets of items in large databases". Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93. p. 207.
- [6] Arun K Pujari “Data Mining Techniques”, Edition 2001, pp 69-109
- [7] Hahsler M, Grun B, Hornik K (2005a). “a rules – A Computational Environment for Mining “ Association Rules and Frequent Item Sets.” Journal of Statistical Software
- [8] Hipp, J.; Güntzer, U.; Nakhaeizadeh, G. (2000). "Algorithms for association rule mining - - a general survey and comparison". *ACM SIGKDD Explorations Newsletter*.
- [9] Webb, Geoffrey (1989). "A Machine Learning Approach to Student Modelling". Proceedings of the Third Australian Joint Conference on Artificial Intelligence (AI 89): 195–205.
- [10] Tan, Pang-Ning; Michael, Steinbach; Kumar, Vipin (2005). "Chapter 6. Association Analysis: Basic Concepts and Algorithms" (PDF). Introduction to Data Mining. Addison-Wesley.

- [11] Thomas, S. and S. Chakravarthy, (1998), "Incremental Mining of Constrained Associations". In the 7th International Conference of High Performance Computing (HiPC).
- [12] Agrawal, R.; Imieliński, T. and Swami, A. (1993), Mining Associations between Sets of Items in Massive Databases. In Proc. of the 1993 ACM-SIGMOD Int'l Conf. on Management of Data, 207-216.
- [13] Zaki MJ (2000). "Scalable Algorithms for Association Mining." IEEE Transactions on Knowledge and Data Engineering, 1993
- [14] Agrawal R, Srikant R (1994). "Fast Algorithms for Mining Association Rules." In JB Bocca, M Jarke, C Zaniolo (eds.), Proc. 20th Int. Conf. Very Large Data Bases, VLDB, pp. 487– 499. Morgan Kaufmann
- [15] Omiecinski, Edward R. (2003); Alternative interest measures for mining associations in databases, IEEE Transactions on Knowledge and Data Engineering, 15(1):57-69.
- [16] Agrawal, R.; Imieliński, T.; Swami, A. (1993). "Mining association rules between sets of items in large databases". Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93. p. 207.
- [17] B. A. Davey, H.A. Priestly (2002), "Introduction to Lattices and order", 2nd ed., Cambridge Uni. Press.
- [18] S.Brin, R.Motwani, J.D. Ullman, and S.Tsur, (1997), " Dynamic itemset counting and implication rules for market basket data", in the proceedings for ACM SIGMOD international conference on Management of Data.
- [19] Bayardo R. J.(1998) Efficiently Mining Long Patterns from Databases." Proceedings of the ACM SIGMOD.
- [20] Pang-Ning; Michael, Steinbach; Kumar, Vipin (2005). "Chapter 6. Association Analysis: Basic Concepts and Algorithms" (PDF). Introduction to Data Mining. Addison-Wesley , pp 331- 351.
- [21] J Pei., J.Han., R.Mao (2000), " CLOSET: An Efficient Algorithm for mining Frequent closed Itemsets", in proceedings of the 2000ACM SIGMOD international conference on Management of Data, Dallas, Texas, USA p 21-30.
- [22] Pasquier N, Bastide Y, Taouil R, Lakhal L (1999). "Discovering Frequent Closed Item sets for Association Rules." In Proceeding of the 7th International Conference on Database Theory, Lecture Notes In Computer Science (LNCS 1540), pp. 398–416.

- [23] Zaki MJ, Parthasarathy S, Ogihara M, Li W (1997b). "New Algorithms for Fast Discovery of Association Rules." Technical Report 651, Computer Science Department, University of Rochester, Rochester, NY 14627.
- [24] Zhao, Q. & Bhowmick, S. S. (2003). Association Rule Mining: A survey (Technical report: No 2003116). Singapore: CAIS, Nanyang Technological University.
- [25] Han (2000). "Mining Frequent Patterns Without Candidate Generation". Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data.
- [26] J. Han, H. Pei, and Y. Yin. Mining Frequent Patterns without Candidate Generation. In: Proc. Conf. on the Management of Data (SIGMOD'00, Dallas, TX). ACM Press, New York, NY, USA 2000.
- [27] Mohammed Javeed Zaki, Srinivasan Parthasarathy, Mitsunori Ogihara, Wei Li (1997): Parallel Algorithms for Discovery of Association Rules. *Data Min. Knowl. Discov.* 1(4): pp 343-373.
- [28] Brin, Sergey; Motwani, Rajeev; Ullman, Jeffrey D.; and Tsur, Shalom (1997), Dynamic itemset counting and implication rules for market basket data, in SIGMOD 1997, Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD 1997), Tucson, Arizona, USA, pp 256-276.
- [29] Sergey Brin, Rajeev Motwani, Jeffrey D. Ullman, and Shalom Turk. Dynamic itemset counting and implication rules for market basket data. In SIGMOD 1997, Proceedings ACM SIGMOD International Conference on Management of Data, pages 255-264, Tucson, Arizona, USA, May 1997.
- [30] Aggarwal, Charu C.; and Yu, Philip S.; A new framework for itemset generation, in PODS 98, Symposium on Principles of Database Systems, Seattle, WA, USA, 1998, pp. 18-24.
- [31] Brin S., Motwani R. and Silverstein C. (1997), "Beyond Market Baskets: Generalizing Association Rules to Correlations." Proceedings of the ACM SIGMOD. (25)
- [32] Aggarwal C. C., and Yu P. S. (1998), "A new framework for itemset generation." IBM Research Report, RC-21064.
- [33] G.D.Ramkumar, Sanjay Ranka, and Shalom Tsur (1998), "Weighted Association Rules: Model and Algorithm" .
- [34] C.H. Cai, A.W.C. Fu, C.H. Cheng, and W.W. Kwong, (1998), "Mining Association Rules with Weighted Items," Proc. IEEE Int'l Database Eng. and Applications Symp. (IDEAS '98), pp. 68-77.

- [35] F. Tao, F. Murtagh, and M. Farid, (2003), "Weighted Association Rule Mining Using Weighted Support and Significance Framework," Proc. ACM SIGKDD '03, pp. 661-666.
- [36] Tan, Pang-Ning; Kumar, Vipin; and Srivastava, Jaideep; (2004), Selecting the right objective measure for association analysis, *Information Systems*, 29(4):293-313.
- [37] Pearson, K. (1895), *Royal Society Proceedings*, 58, 241. (1920), "Notes on the History of Correlation," *Biometrika*.
- [38] P. Tan, V. Kumar, and J. Srivastava (2002.), Selecting the right interestingness measure for association patterns
- [39] Witten, Frank, Hall: *Data mining practical machine learning tools and techniques*, 3rd edition.
- [40] Webb, Geoffrey I. (2007); *Discovering Significant Patterns*, *Machine Learning* 68(1), Netherlands: Springer, pp. 1-33.
- [41] Gionis, Aristides; Mannila, Heikki; Mielikäinen, Taneli; and Tsaparas, Panayiotis (2007), *Assessing Data Mining Results via Swap Randomization*, *ACM Transactions on Knowledge Discovery from Data (TKDD)*, Volume 1, Issue 3 (December 2007), Article No. 14.