

ΔΙΕΘΝΕΣ ΠΑΝΕΠΙΣΤΗΜΙΟ ΤΗΣ ΕΛΛΑΔΟΣ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Εξόρυξη Κανόνων Συσχέτισης από Δεδομένα Μεγάλης Κλίμακας:
Πειραματική Σύγκριση Αλγορίθμων και Ανάλυση Καλαθιού Αγορών
κατά την περίοδο COVID-19 και κατά την μεταπανδημική περίοδο**



Της φοιτήτριας
Νίκης Ευαγγελίδου
Αρ. Μητρώου: 03/2023

Επιβλέπων
Ονοματεπώνυμο: Ουγιάρογλου Στέφανος
Βαθμίδα: Επίκουρος καθηγητής

Ημερομηνία: 24 Σεπτεμβρίου 2026

Τίτλος Δ.Ε. “Ανάλυση της καταναλωτικής συμπεριφοράς από την περίοδο πανδημίας COVID-19 μέχρι σήμερα μέσω εξόρυξης κανόνων συσχέτισης”

Κωδικός Δ.Ε. 24292

Ονοματεπώνυμο φοιτήτριας Ευαγγελίδου Νίκη

Ονοματεπώνυμο εισηγητή Ουγιάρογλου Στέφανος

Ημερομηνία ανάληψης Δ.Ε. 29/10/2024

Ημερομηνία περάτωσης Δ.Ε. 24 Σεπτεμβρίου 2026

Βεβαιώνω ότι είμαι η συγγραφέας αυτής της εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, έχω καταγράψει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών, εικόνων και κειμένου, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επιπλέον, βεβαιώνω ότι αυτή η εργασία προετοιμάστηκε από εμένα προσωπικά, ειδικά ως διπλωματική εργασία, στο Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του ΔΙ.ΠΑ.Ε.

Η παρούσα εργασία αποτελεί πνευματική ιδιοκτησία της φοιτήτριας Νίκης Ευαγγελίδου που την εκπόνησε. Στο πλαίσιο της πολιτικής ανοικτής πρόσβασης, η συγγραφέας/δημιουργός εκχωρεί στο Διεθνές Πανεπιστήμιο της Ελλάδος άδεια χρήσης του δικαιώματος αναπαραγωγής, δανεισμού, παρουσίασης στο κοινό και ψηφιακής διάχυσης της εργασίας διεθνώς, σε ηλεκτρονική μορφή και σε οποιοδήποτε μέσο, για διδακτικούς και ερευνητικούς σκοπούς, άνευ ανταλλάγματος. Η ανοικτή πρόσβαση στο πλήρες κείμενο της εργασίας, δεν σημαίνει καθ' οιονδήποτε τρόπο παραχώρηση δικαιωμάτων διανοητικής ιδιοκτησίας της συγγραφέα/δημιουργού, ούτε επιτρέπει την αναπαραγωγή, αναδημοσίευση, αντιγραφή, πώληση, εμπορική χρήση, διανομή, έκδοση, μεταφόρτωση (downloading), ανάρτηση (uploading), μετάφραση, τροποποίηση με οποιονδήποτε τρόπο, τμηματικά ή περιληπτικά της εργασίας, χωρίς τη ρητή προηγούμενη έγγραφη συναίνεση της συγγραφέα/δημιουργού.

Η έγκριση της διπλωματικής εργασίας από το Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του Διεθνούς Πανεπιστημίου της Ελλάδος, δεν υποδηλώνει απαραίτητως και αποδοχή των απόψεων της συγγραφέα, εκ μέρους του Τμήματος.

Στη μικρή μου Ηλέκτρα

Πρόλογος

Η επιλογή της παρούσας διπλωματικής εργασίας, με θέμα «Ανάλυση της καταναλωτικής συμπεριφοράς από την περίοδο πανδημίας COVID-19 μέχρι σήμερα μέσω εξόρυξης κανόνων συσχέτισης» έγινε καθώς με ενδιαφέρει ιδιαίτερα ο λιανεμπορικός κλάδος και ειδικότερα τα σούπερ μάρκετ, αν και στην πράξη εργάζομαι σε έναν λίγο διαφορετικό τομέα δράσης. Η καθημερινή επαφή μου με τον επαγγελματικό χώρο ενός σούπερ μάρκετ αποτέλεσε κίνητρο για να εμβαθύνω στην κατανόηση της αγοραστικής συμπεριφοράς των καταναλωτών αλλά και στο τι άλλαξε στη διάρκεια της πανδημίας και μετά από αυτήν.

Μέσα από την εκπόνηση της συγκεκριμένης εργασίας επιδιώκω να αναπτύξω περαιτέρω τις γνώσεις και τις δεξιότητές μου σε σύγχρονες τεχνικές ανάλυσης δεδομένων, όπως το Market Basket Analysis και η εξόρυξη κανόνων συσχέτισης. Παράλληλα, θεωρώ ότι η διαδικασία αυτή θα συμβάλει τόσο στην ακαδημαϊκή όσο και στην επαγγελματική μου εξέλιξη, προσφέροντάς μου πολύτιμη εμπειρία και νέες προοπτικές.

Περίληψη

Η εργασία συγκρίνει τα αγοραστικά πρότυπα της περιόδου COVID-19 και της μεταπανδημικής περιόδου μέσω εξόρυξης κανόνων συσχέτισης και αξιολογεί την αποδοτικότητα των αλγορίθμων Apriori, FP-Growth και Eclat σε δεδομένα μεγάλης κλίμακας.

Η ανάλυση βασίστηκε σε 9,5 εκατομμύρια πραγματικές συναλλαγές δεκαεπτά καταστημάτων μιας αλυσίδας σούπερ μάρκετ στην Καβάλα. Μετά την προεπεξεργασία, ο Apriori χρησιμοποιήθηκε για τη σύγκριση των περιόδων. Εξετάστηκαν τα supports των 1-itemsets, το μέσο μέγεθος καλαθιού και οι μεταβολές support, confidence και lift στα κοινά στοιχειοσύνολα και στους κοινούς κανόνες. Οι τρεις αλγόριθμοι εφαρμόστηκαν στο ενοποιημένο σύνολο με κοινές παραμέτρους και συγκρίθηκαν ως προς τον χρόνο και τη μέγιστη καταγεγραμμένη χρήση μνήμης.

Στην περίοδο COVID-19 εντοπίστηκαν περισσότερα συχνά στοιχειοσύνολα και κανόνες. Στη μεταπανδημική περίοδο, τα περισσότερα κοινά στοιχειοσύνολα είχαν χαμηλότερο support και το μέσο καλάθι ήταν μικρότερο κατά 9,95%. Το confidence μειώθηκε στους περισσότερους κοινούς κανόνες, ενώ το lift αυξήθηκε. Επομένως, τα περισσότερα πρότυπα δεν συνεπάγονται ισχυρότερες συσχετίσεις και οι διαφοροποιήσεις δεν αποδίδονται αποκλειστικά στην πανδημία.

Οι τρεις αλγόριθμοι παρήγαγαν τον ίδιο αριθμό στοιχειοσυνόλων και κανόνων. Ο Eclat ήταν ταχύτερος και είχε τη χαμηλότερη καταγεγραμμένη χρήση μνήμης. Συνολικά, η εργασία αναδεικνύει τη χρησιμότητα της εξόρυξης κανόνων συσχέτισης για τη μελέτη πραγματικών δεδομένων λιανεμπορίου και τη σημασία επιλογής κατάλληλου αλγορίθμου.

Λέξεις-κλειδιά: Καταναλωτική συμπεριφορά, Market Basket Analysis, Εξόρυξη κανόνων συσχέτισης, Apriori, FP-Growth, Eclat, COVID-19, Λιανεμπόριο, Ανάλυση δεδομένων μεγάλης κλίμακας.

Abstract

Analysis of Consumer Behavior from the COVID-19 Pandemic Period to the Present Through Association Rule Mining

Evangelidou Niki

This study compares purchasing patterns between the COVID-19 and post-pandemic periods using association rule mining and evaluates the computational efficiency of Apriori, FP-Growth, and Eclat on large-scale data.

The analysis used 9.5 million real-world transactions from seventeen supermarket stores in Kavala, Greece. After preprocessing, Apriori was applied to both periods. The comparisons covered 1-itemset support, average basket size, support changes in common itemsets with multiple items, and confidence and lift changes in common rules. All three algorithms were applied to the combined dataset with identical parameters and evaluated by execution time and maximum recorded memory usage.

More frequent itemsets and rules were identified during the COVID-19 period. In the post-pandemic period, most common itemsets had lower support and average basket size was 9.95% smaller. Confidence decreased for most common rules, whereas lift increased. Thus, more patterns do not imply stronger associations, and these differences cannot be attributed solely to the pandemic.

The three algorithms produced the same number of itemsets and rules. Eclat achieved the shortest execution time and the lowest recorded memory usage. Overall, the study highlights the usefulness of association rule mining for examining real-world retail data and the importance of selecting an appropriate algorithm.

Keywords: Consumer Behavior, Market Basket Analysis, Association Rule Mining, Apriori, FP-Growth, Eclat, COVID-19, Retail Analytics, Large-Scale Transactional Data.

Ευχαριστίες

Με την ολοκλήρωση της παρούσας διπλωματικής εργασίας και του μεταπτυχιακού μου κύκλου σπουδών, θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες σε όλους όσους συνέβαλαν, με τον δικό τους τρόπο, στην επίτευξη αυτού του στόχου.

Αρχικά, ευχαριστώ όλους τους καθηγητές του Μεταπτυχιακού Προγράμματος Σπουδών για τις πολύτιμες γνώσεις, την καθοδήγηση και την ακαδημαϊκή τους υποστήριξη καθ' όλη τη διάρκεια των σπουδών μου.

Ιδιαίτερες ευχαριστίες οφείλω στον επιβλέποντα καθηγητή μου, κ. Στέφανο Ουγιάρογλου, για την εμπιστοσύνη που μου έδειξε, την πολύτιμη καθοδήγηση, τις εύστοχες παρατηρήσεις και τη συνεχή υποστήριξή του κατά την εκπόνηση της παρούσας εργασίας. Η συμβολή του υπήρξε καθοριστική για την ολοκλήρωσή της.

Θα ήθελα επίσης να ευχαριστήσω από καρδιάς την οικογένειά μου και τον σύζυγό μου για την αμέριστη αγάπη, την υπομονή και τη στήριξή τους σε κάθε βήμα αυτής της διαδρομής. Η πίστη τους στις δυνατότητές μου αποτέλεσε για μένα πολύτιμο στήριγμα στις απαιτητικές στιγμές των σπουδών μου.

Τέλος, το μεγαλύτερο ευχαριστώ ανήκει στη μικρή μου κόρη, την Ηλέκτρα. Η έλευσή της στη ζωή μου άλλαξε τον τρόπο που βλέπω τον κόσμο, τους στόχους και τα όνειρά μου. Ήταν εκείνη που μου έδωσε το κίνητρο να ολοκληρώσω αυτό το ταξίδι και να διεκδικήσω την ολοκλήρωση των σπουδών μου. Αυτό το πτυχίο είναι αφιερωμένο σε εκείνη, με την ευχή να μεγαλώσει γνωρίζοντας πως υπήρξε η πιο όμορφη αιτία για να μην εγκαταλείψω ποτέ τους στόχους μου και να γίνω η καλύτερη εκδοχή του εαυτού μου.

Σε όλους σας, ένα ειλικρινές και θερμό ευχαριστώ.

Περιεχόμενα

Πρόλογος	3
Περίληψη	4
Abstract	5
Ευχαριστίες	6
Κατάλογος Σχημάτων	10
Κατάλογος Πινάκων	12
Συντομογραφίες	13
1 Εισαγωγή	14
1.1 Market Basket Analysis και Ανάλυση Καταναλωτικής Συμπεριφοράς	14
1.2 Αντικείμενο και Σκοπός της Έρευνας	15
1.3 Κίνητρο της Έρευνας	15
1.4 Συνεισφορά της Εργασίας	16
1.5 Οργάνωση της διπλωματικής εργασίας	16
2 Θεωρητικό Υπόβαθρο	18
2.1 Επίδραση της Πανδημίας COVID-19 στην Καταναλωτική Συμπεριφορά	18
2.1.1 Μεταβολές στις αγοραστικές συνήθειες κατά την πανδημία	18
2.1.2 Μεταβολές μετά την άρση των περιορισμών	20
2.2 Market Basket Analysis	21
2.2.1 Ορισμός και Βασικές Αρχές	21
2.2.2 Εφαρμογές στο λιανεμπόριο	22
2.2.3 Πλεονεκτήματα και περιορισμοί	23
2.3 Εξόρυξη Κανόνων Συσχέτισης (Association Rules)	25
2.3.1 Frequent Itemsets – Συχνά Στοιχειοσύνολα	26
2.3.2 Κανόνες Συσχέτισης (Association Rules)	26
2.4 Μετρικές Αξιολόγησης Κανόνων Συσχέτισης	27
2.4.1 Support	27
2.4.2 Confidence	28
2.4.3 Lift	28

2.4.4	Leverage	29
2.4.5	Conviction	29
2.4.6	Σύγκριση των μετρικών	30
2.5	Αλγόριθμοι Εξόρυξης Συχνών Στοιχειοσυνόλων	31
2.5.1	Αλγόριθμος Apriori	32
2.5.2	Αλγόριθμος FP-Growth	33
2.5.3	Αλγόριθμος Eclat	34
2.5.4	Άλλοι Αλγόριθμοι Εξόρυξης Συχνών Στοιχειοσυνόλων	35
2.5.5	Συγκριτική Παρουσίαση των Αλγορίθμων	36
2.6	Ανασκόπηση Σχετικών Ερευνών	37
2.6.1	Market Basket Analysis στο λιανεμπόριο	38
2.6.2	Association Rules σε Σούπερ Μάρκετ	39
2.6.3	Μελέτες σχετικές με COVID-19	39
2.6.4	Ερευνητικό Κενό	41
3	Πειραματικός Σχεδιασμός και Βελτιστοποίηση Παραμέτρων	42
3.1	Απόκτηση και Περιγραφή Δεδομένων	42
3.2	Προεπεξεργασία και Διαμόρφωση Δεδομένων	43
3.3	Διαχωρισμός των Δεδομένων σε Χρονικές Περιόδους	43
3.4	Διερευνητική Ανάλυση Συχνοτήτων Προϊόντων	44
3.4.1	Σύγκριση των κυριότερων κατηγοριών προϊόντων	45
3.4.2	Σύγκριση των ποσοστών υποστήριξης	46
3.4.3	Μεταβολή του ποσοστού υποστήριξης	46
3.4.4	Συμπληρωματική σύγκριση των κοινών 1-itemsets	47
3.4.5	Σύγκριση του μέσου μεγέθους καλαθιού	49
3.4.6	Σύνδεση με την εξόρυξη κανόνων συσχέτισης	50
3.5	Διερεύνηση της Παραμέτρου Minimum Support	51
3.6	Αρχική Παραμετροποίηση του Αλγορίθμου Apriori	52
3.7	Βελτίωση της Διαδικασίας Φιλτραρίσματος Κανόνων	54
3.8	Διερεύνηση και Αξιολόγηση Παραμέτρων του Αλγορίθμου Apriori	55
3.8.1	Διερεύνηση και Αξιολόγηση Παραμέτρων του Αλγορίθμου Apriori για την Περίοδο COVID-19	55
3.8.2	Διερεύνηση και Αξιολόγηση Παραμέτρων του Αλγορίθμου Apriori για τη Με- ταπανδημική Περίοδο	59
3.8.3	Τελική Επιλογή Μοντέλου Apriori	62
4	Συγκριτική Διερεύνηση της Καταναλωτικής Συμπεριφοράς κατά την Περίοδο COVID- 19 και τη Μεταπανδημική Περίοδο	65
4.1	Αποτελέσματα Apriori για την Περίοδο COVID-19	65
4.1.1	Συχνά Στοιχειοσύνολα (Frequent Itemsets) COVID-19	66
4.1.2	Κανόνες Συσχέτισης (Association Rules)	67
4.1.3	Ερμηνεία Αποτελεσμάτων COVID-19	69
4.2	Αποτελέσματα Apriori για τη Μεταπανδημική Περίοδο	69
4.2.1	Συχνά Στοιχειοσύνολα (Frequent Itemsets) Μεταπανδημικής Περιόδου	70

4.2.2	Κανόνες Συσχέτισης (Association Rules) Μεταπανδημικής Περιόδου	71
4.2.3	Ερμηνεία Αποτελεσμάτων Μεταπανδημικής Περιόδου	72
4.3	Συγκριτική Ανάλυση COVID-19 και Μεταπανδημικής Περιόδου	72
4.3.1	Σύγκριση Συνολικών Αποτελεσμάτων	72
4.3.2	Σύγκριση Frequent Itemsets	73
4.3.3	Σύγκριση Κανόνων Συσχέτισης	76
4.3.4	Συμπέρασμα Συγκριτικής Ανάλυσης Περιόδων με χρήση Apriori	78
4.4	Συγκριτική Αξιολόγηση των Αλγορίθμων Apriori, FP-Growth και Eclat	79
4.4.1	Αποτελέσματα Apriori	80
4.4.2	Αποτελέσματα FP-Growth	82
4.4.3	Αποτελέσματα Eclat	83
4.4.4	Συγκριτική Αξιολόγηση των Αλγορίθμων Apriori, FP-Growth και Eclat	84
5	Συμπεράσματα και Μελλοντική Έρευνα	88
5.1	Συμπεράσματα της Έρευνας	88
5.2	Συνεισφορά της Έρευνας	90
5.3	Περιορισμοί της Μελέτης	91
5.4	Μελλοντική Έρευνα	92
A'	Κώδικες Υλοποίησης	95
A'.1	Μετατροπή του συνόλου δεδομένων σε μορφή καλαθιού	95
A'.2	Ανάλυση ευαισθησίας της παραμέτρου minimum support	96
A'.3	Υλοποίηση του αλγορίθμου Apriori	98
A'.4	Υλοποίηση του αλγορίθμου FP-Growth	101
A'.5	Υλοποίηση του αλγορίθμου Eclat	104
A'.6	Σύγκριση του support των 1-itemsets και του μεγέθους καλαθιού	107
A'.7	Σύγκριση του support των κοινών frequent itemsets	119
A'.8	Σύγκριση του confidence και του lift των κοινών κανόνων	126
B'	Συμπληρωματικοί Πίνακες Κανόνων Συσχέτισης	140
B'.1	Top-100 Κανόνες Συσχέτισης Περιόδου COVID-19	140
B'.2	Top-100 Κανόνες Συσχέτισης Μεταπανδημικής Περιόδου	143
B'.3	Κοινοί κανόνες: αυξήσεις του confidence	147
B'.4	Κοινοί κανόνες: μειώσεις του confidence	148
B'.5	Κοινοί κανόνες: αυξήσεις του lift	149
B'.6	Κοινοί κανόνες: μειώσεις του lift	150

Κατάλογος σχημάτων

3.1	Σύγκριση ποσοστών υποστήριξης (Support %) μεταξύ της περιόδου COVID-19 και της μεταπανδημικής περιόδου	46
3.2	Μεταβολή του support των σημαντικότερων κατηγοριών προϊόντων μεταξύ των δύο περιόδων. Οι τιμές εκφράζονται σε ποσοστιαίες μονάδες και υπολογίζονται ως Support Post-COVID μείον Support COVID-19.	47
3.3	Πλήθος κανόνων ανά πείραμα για την περίοδο COVID-19	57
3.4	Χρόνος εκτέλεσης ανά πείραμα για την περίοδο COVID-19	58
3.5	Κατανάλωση μνήμης ανά πείραμα για την περίοδο COVID-19	58
3.6	Πλήθος κανόνων ανά πείραμα για τη μεταπανδημική περίοδο	60
3.7	Χρόνος εκτέλεσης ανά πείραμα για τη μεταπανδημική περίοδο	61
3.8	Κατανάλωση μνήμης ανά πείραμα για τη μεταπανδημική περίοδο	61
3.9	Σύγκριση αποτελεσμάτων του τελικού μοντέλου P5	63
4.1	Σύγκριση του πλήθους των frequent itemsets και των κανόνων συσχέτισης που εξήχθησαν από τον αλγόριθμο Apriori για την περίοδο COVID-19 και τη μεταπανδημική περίοδο	73
4.2	Σύγκριση του πλήθους των frequent itemsets και των κανόνων συσχέτισης που εξήχθησαν από τους αλγορίθμους Apriori, FP-Growth και Eclat.	85
4.3	Συγκριτική απεικόνιση χρόνου εκτέλεσης των αλγορίθμων Apriori, FP-Growth και Eclat.	86
4.4	Συγκριτική απεικόνιση μέγιστης κατανάλωσης μνήμης των αλγορίθμων Apriori, FP-Growth και Eclat.	87

Κατάλογος πινάκων

2.1	Σύγκριση των βασικών μετρικών αξιολόγησης κανόνων συσχέτισης	31
2.2	Συγκριτική παρουσίαση των αλγορίθμων Apriori, FP-Growth και Eclat	37
3.1	Βασικά χαρακτηριστικά των συνόλων δεδομένων	44
3.2	Σύγκριση ποσοστού υποστήριξης των σημαντικότερων κατηγοριών προϊόντων μεταξύ της περιόδου COVID-19 και της μεταπανδημικής περιόδου	45
3.3	Κατηγορίες με τις δέκα μεγαλύτερες αυξήσεις του support μεταξύ των δύο περιόδων	48
3.4	Κατηγορίες με τις δέκα μεγαλύτερες μειώσεις του support μεταξύ των δύο περιόδων	48
3.5	Σύγκριση του μεγέθους καλαθιού μεταξύ των δύο περιόδων	49
3.6	Επίδραση της παραμέτρου minimum support στον αριθμό frequent itemsets και στον χρόνο εκτέλεσης του FP-Growth	51
3.7	Αντιπροσωπευτικοί κανόνες συσχέτισης με υψηλές τιμές lift κατά τη διερεύνηση των παραμέτρων του αλγορίθμου Apriori	52
3.8	Επίδραση της παραμέτρου confidence στον αριθμό των παραγόμενων κανόνων . . .	53
3.9	Συνδυασμοί παραμέτρων που χρησιμοποιήθηκαν στα πειράματα Apriori για την περίοδο COVID-19	55
3.10	Συγκεντρωτικά αποτελέσματα των πειραμάτων Apriori για την περίοδο COVID-19 .	56
3.11	Συνδυασμοί παραμέτρων που χρησιμοποιήθηκαν στα πειράματα Apriori για τη μεταπανδημική περίοδο	59
3.12	Συγκεντρωτικά αποτελέσματα των πειραμάτων Apriori για τη μεταπανδημική περίοδο	59
3.13	Σύγκριση του τελικού μοντέλου P5 μεταξύ της περιόδου COVID-19 και της μεταπανδημικής περιόδου	63
3.14	Τελικές παράμετροι του αλγορίθμου Apriori	64
4.1	Συνοπτικά αποτελέσματα του επιλεγμένου πειραματικού σεναρίου P5 για την περίοδο COVID-19	65
4.2	Στατιστικά χαρακτηριστικά των παραγόμενων κανόνων συσχέτισης στο επιλεγμένο πειραματικό σενάριο P5 για την περίοδο COVID-19	66
4.3	Δεκαπέντε συχνότερα πολυ-στοιχειακά frequent itemsets κατά την περίοδο COVID-19	67
4.4	Κορυφαίοι κανόνες συσχέτισης που εξήχθησαν από το επιλεγμένο πειραματικό σενάριο P5 για την περίοδο COVID-19	68
4.5	Συνοπτικά αποτελέσματα του επιλεγμένου πειραματικού σεναρίου P5 για τη μεταπανδημική περίοδο	69
4.6	Στατιστικά χαρακτηριστικά των παραγόμενων κανόνων συσχέτισης στο επιλεγμένο πειραματικό σενάριο P5 για τη μεταπανδημική περίοδο	69

4.7	Δεκαπέντε συχνότερα πολυ-στοιχειακά frequent itemsets κατά τη μεταπανδημική περίοδο	70
4.8	Κορυφαίοι κανόνες συσχέτισης που εξήχθησαν από το επιλεγμένο πειραματικό σε- νάριο P5 για τη μεταπανδημική περίοδο	71
4.9	Συγκριτικά αποτελέσματα εφαρμογής του αλγορίθμου Apriori στις δύο εξεταζόμενες χρονικές περιόδους	72
4.10	Κοινά frequent itemsets με τις δέκα μεγαλύτερες αυξήσεις του support	74
4.11	Κοινά frequent itemsets με τις δέκα μεγαλύτερες μειώσεις του support	75
4.12	Ενδεικτικοί ισχυροί κανόνες που εμφανίζονται και στις δύο περιόδους	76
4.13	Συνολική σύγκριση confidence και lift στους 124 κοινούς κανόνες	77
4.14	Αποτελέσματα εκτέλεσης Apriori	81
4.15	Αποτελέσματα εκτέλεσης FP-Growth	82
4.16	Αποτελέσματα εκτέλεσης Eclat	83
4.17	Συγκριτική Αξιολόγηση των Αλγορίθμων Apriori, FP-Growth και Eclat	84
B'.1	Top-100 κανόνες συσχέτισης περιόδου COVID-19	140
B'.2	Top-100 κανόνες συσχέτισης μεταπανδημικής περιόδου	143
B'.3	Κοινοί κανόνες με τις δέκα μεγαλύτερες αυξήσεις του confidence	147
B'.4	Κοινοί κανόνες με τις δέκα μεγαλύτερες μειώσεις του confidence	148
B'.5	Κοινοί κανόνες με τις δέκα μεγαλύτερες αυξήσεις του lift	149
B'.6	Κοινοί κανόνες με τις δέκα μεγαλύτερες μειώσεις του lift	150

Συντομογραφίες

Δ.Ε.	Διπλωματική Εργασία
ΔΙΠΑΕ	Διεθνές Πανεπιστήμιο της Ελλάδος
MBA	Market Basket Analysis
ARL	Association Rule Learning
COVID-19	Coronavirus Disease 2019

Κεφάλαιο 1

Εισαγωγή

1.1 Market Basket Analysis και Ανάλυση Καταναλωτικής Συμπεριφοράς

Οι επιχειρήσεις λιανικού εμπορίου επιδιώκουν όλο και περισσότερο να κατανοήσουν τον τρόπο με τον οποίο οι καταναλωτές διαμορφώνουν τις αγοραστικές τους επιλογές. Η συνεχής καταγραφή μεγάλου όγκου δεδομένων των συναλλαγών η οποία πραγματοποιείται μέσω των πληροφοριακών συστημάτων των επιχειρήσεων βοηθά τις επιχειρήσεις να εξάγουν χρήσιμες πληροφορίες σχετικά με τις αγοραστικές συνήθειες των πελατών τους. Για αυτό τον λόγο, το Market Basket Analysis έχει αναδειχθεί ως μία από τις πιο διαδεδομένες τεχνικές εξόρυξης δεδομένων για τη μελέτη των προτύπων αγορών αλλά και για την ανάλυση της σχέσης μεταξύ των προϊόντων που επιλέγουν οι καταναλωτές [1], [2].

Ουσιαστικά, η βασική ιδέα πίσω από το Market Basket Analysis είναι ότι τα προϊόντα που περιλαμβάνονται σε μία συναλλαγή δεν επιλέγονται ανεξάρτητα το ένα από το άλλο, αντιθέτως, πολλές φορές η αγορά ενός προϊόντος σχετίζεται άμεσα με την αγορά κάποιου άλλου, δημιουργώντας συγκεκριμένα πρότυπα κατανάλωσης. Για τον σκοπό αυτό η ανάλυση μεγάλου όγκου συναλλαγών δύναται να εντοπίσει τέτοιες σχέσεις και να αναδείξει συνδυασμούς προϊόντων που εμφανίζονται συχνά μαζί στα καλάθια αγορών των πελατών [2], [3].

Προκειμένου να μπορέσει να πραγματοποιηθεί αυτή η ανάλυση των σχέσεων μεταξύ των προϊόντων χρησιμοποιούνται συνήθως τεχνικές εξόρυξης κανόνων συσχέτισης (Association Rule Mining), οι οποίες επιτρέπουν την ανακάλυψη προτύπων που δεν είναι άμεσα εμφανή στα δεδομένα των συναλλαγών. Οι τεχνικές αυτές αποτελούν βασικό εργαλείο του Market Basket Analysis και αναλύονται εκτενέστερα στο θεωρητικό υπόβαθρο της εργασίας [2].

Συγκεκριμένα για τον κλάδο του λιανεμπορίου η εφαρμογή του Market Basket Analysis είναι σχεδόν επιτακτική. Βοηθά τις επιχειρήσεις να καταλάβουν καλύτερα τόσο τις ανάγκες όσο και τις προτιμήσεις των καταναλωτών προκειμένου να αξιοποιήσουν αυτές τις πληροφορίες, για τον σχεδιασμό προωθητικών ενεργειών, τη βελτίωση της διάταξης των προϊόντων στα καταστήματα, τη δημιουργία συνδυαστικών προσφορών και την αποτελεσματικότερη διαχείριση των αποθεμάτων [1], [4] ενώ ταυτόχρονα συμβάλλει στη λήψη αποφάσεων που βασίζονται σε πραγματικά δεδομένα και όχι αποκλειστικά στην εμπειρική γνώση των στελεχών μιας επιχείρησης.

Πέρα όμως από την επιχειρησιακή του αξία, η χρησιμότητα του Market Basket Analysis έγκειται και στη μελέτη της καταναλωτικής συμπεριφοράς. Οι επιλογές των καταναλωτών επηρεάζονται

από οικονομικούς, κοινωνικούς και περιβαλλοντικούς παράγοντες, οι οποίοι μεταβάλλονται με την πάροδο του χρόνου με αποτέλεσμα οι σχέσεις μεταξύ των προϊόντων και τα πρότυπα αγορών να μην παραμένουν σταθερά αλλά να εξελίσσονται ανάλογα με τις συνθήκες που επικρατούν σε κάθε χρονική περίοδο [1].

Η συνεισφορά της συγκεκριμένης προσέγγισης έγινε ακόμη πιο εμφανής κατά την περίοδο της πανδημίας, διότι τόσο οι συνθήκες της αγοράς όσο και η καθημερινότητα των καταναλωτών μεταβλήθηκαν σε μεγάλο βαθμό. Οι αλλαγές αυτές επηρέασαν τη συχνότητα αγοράς ορισμένων προϊόντων και σε κάποιες περιπτώσεις τους συνδυασμούς προϊόντων που εμφανίζονταν στα καλάθια αγορών. Ως αποτέλεσμα, η διερεύνηση των προτύπων αγορών μέσω τεχνικών εξόρυξης κανόνων συσχέτισης μπορεί να συμβάλει ουσιαστικά στην κατανόηση του τρόπου με τον οποίο εξελίχθηκε η καταναλωτική συμπεριφορά κατά την περίοδο της πανδημίας και μετά το τέλος αυτής.

Στην παρούσα εργασία το MBA χρησιμοποιείται για τη διερεύνηση των μεταβολών στην καταναλωτική συμπεριφορά στον κλάδο των σούπερ μάρκετ. Μέσα από την ανάλυση πραγματικών δεδομένων συναλλαγών εξετάζονται οι σχέσεις μεταξύ κατηγοριών προϊόντων κατά την περίοδο της πανδημίας COVID-19 και μετά από αυτήν, με στόχο την ανάδειξη πιθανών διαφοροποιήσεων στα πρότυπα αγορών και την καλύτερη κατανόηση των αλλαγών που σημειώθηκαν στις καταναλωτικές συνήθειες.

1.2 Αντικείμενο και Σκοπός της Έρευνας

Αντικείμενο της εργασίας είναι η μελέτη της καταναλωτικής συμπεριφοράς στον κλάδο του οργανωμένου λιανεμπορίου τροφίμων, με εφαρμογή τεχνικών Market Basket Analysis και εξόρυξης κανόνων συσχέτισης σε πραγματικά δεδομένα συναλλαγών για την περίοδο της πανδημίας COVID-19 και τη μεταπανδημική περίοδο. Παράλληλα, αντικείμενο της εργασίας αποτελεί η συγκριτική αξιολόγηση της υπολογιστικής αποδοτικότητας των αλγορίθμων Apriori, FP-Growth και Eclat κατά την επεξεργασία δεδομένων μεγάλης κλίμακας.

Σκοπός της εργασίας είναι η διερεύνηση των διαφοροποιήσεων στα πρότυπα αγορών των καταναλωτών μεταξύ των δύο χρονικών περιόδων. Για τον σκοπό αυτό, εκτός από την ανάλυση των συχνών στοιχειοσυνόλων και των κανόνων συσχέτισης, συγκρίνονται τα ποσοστά υποστήριξης των κοινών κατηγοριών προϊόντων και το μέσο μέγεθος του καλάθιού. Επιπλέον, εξετάζονται οι μεταβολές του support στα κοινά συχνά στοιχειοσύνολα, καθώς και οι μεταβολές του confidence και του lift στους κοινούς κανόνες συσχέτισης. Τέλος, οι αλγόριθμοι Apriori, FP-Growth και Eclat συγκρίνονται ως προς τον χρόνο εκτέλεσης και τη μέγιστη καταγεγραμμένη χρήση μνήμης.

1.3 Κίνητρο της Έρευνας

Η πανδημία COVID-19 προκάλεσε σοβαρές κοινωνικές αλλά και οικονομικές επιπτώσεις. Επηρέασε μεταξύ άλλων και τις αγοραστικές συνήθειες των καταναλωτών. Οι περιορισμοί στις μετακινήσεις, η αβεβαιότητα που επικράτησε κατά τη διάρκεια της υγειονομικής κρίσης και οι αλλαγές στην καθημερινότητα οδήγησαν σε διαφοροποιήσεις τόσο στη σύνθεση των καλάθιων αγορών όσο και στη ζήτηση συγκεκριμένων κατηγοριών προϊόντων.

Τα τελευταία χρόνια έχουν δημοσιευθεί αρκετές μελέτες που εξετάζουν την καταναλωτική συμπεριφορά κατά την περίοδο της πανδημίας. Ωστόσο, λιγότερες μελέτες εστιάζουν στο κατά πόσο

οι μεταβολές αυτές διατηρήθηκαν και μετά την επιστροφή στην κανονικότητα. Για αυτό τον λόγο, βασικό κίνητρο της παρούσας εργασίας αποτελεί η συγκριτική διερεύνηση των αγοραστικών προτύπων μεταξύ της περιόδου COVID-19 και της μεταπανδημικής περιόδου. Οι διαφοροποιήσεις που εντοπίζονται εξετάζονται ως μεταβολές μεταξύ δύο διακριτών χρονικών περιόδων, χωρίς να αποδίδονται αποκλειστικά στην πανδημία, καθώς ενδέχεται να συνδέονται και με άλλες κοινωνικές και οικονομικές συνθήκες.

Ένα ακόμη σημαντικό κίνητρο ήταν η δυνατότητα αξιοποίησης ενός ιδιαίτερα μεγάλου συνόλου πραγματικών δεδομένων, το οποίο περιλαμβάνει περισσότερες από 9,5 εκατομμύρια συναλλαγές λιανικής. Τα δεδομένα αυτά αποτυπώνουν πραγματικές αγοραστικές επιλογές και δεν βασίζονται σε ερωτηματολόγια ή περιορισμένα δείγματα, επιτρέποντας μια πληρέστερη διερεύνηση της καταναλωτικής συμπεριφοράς.

Παράλληλα, η επεξεργασία ενός τόσο μεγάλου συνόλου συναλλαγών δημιουργεί ιδιαίτερες υπολογιστικές απαιτήσεις. Για τον λόγο αυτό, πρόσθετο κίνητρο της εργασίας αποτελεί η διερεύνηση της υπολογιστικής αποδοτικότητας διαφορετικών αλγορίθμων εξόρυξης συχνών στοιχειοσυνόλων. Ειδικότερα, οι αλγόριθμοι Apriori, FP-Growth και Eclat αξιολογούνται συγκριτικά ως προς τον χρόνο εκτέλεσης και τη μέγιστη κατανάλωση μνήμης, προκειμένου να εξεταστεί η καταλληλότητά τους για την επεξεργασία πραγματικών δεδομένων λιανεμπορίου μεγάλης κλίμακας.

1.4 Συνεισφορά της Εργασίας

Η συνεισφορά της παρούσας εργασίας έγκειται τόσο στον μεγάλο όγκο των δεδομένων, ο οποίος υπερβαίνει τις 9,5 εκατομμύρια συναλλαγές, όσο και στην πραγματική προέλευσή τους από τον κλάδο του οργανωμένου λιανεμπορίου τροφίμων. Το γεγονός αυτό επιτρέπει τη μελέτη της καταναλωτικής συμπεριφοράς σε πραγματικές συνθήκες αγοράς.

Σε δεύτερο επίπεδο, η εργασία εξετάζει συγκριτικά δύο διακριτές χρονικές περιόδους, την περίοδο COVID-19 και τη μεταπανδημική περίοδο, προκειμένου να αναδειχθούν πιθανές διαφοροποιήσεις στα πρότυπα αγορών των καταναλωτών. Η σύγκριση εμπλουτίζεται με την εξέταση των ποσοστών υποστήριξης των κοινών κατηγοριών προϊόντων, του μέσου μεγέθους του καλαθιού και των μεταβολών του support στα κοινά συχνά στοιχειοσύνολα, καθώς και του confidence και του lift στους κοινούς κανόνες συσχέτισης. Ακόμη, πραγματοποιείται συγκριτική αξιολόγηση των αλγορίθμων Apriori, FP-Growth και Eclat ως προς τον χρόνο εκτέλεσης και τη μέγιστη καταγεγραμμένη χρήση μνήμης κατά την εφαρμογή τους σε δεδομένα μεγάλης κλίμακας.

Τέλος, η έρευνα παρέχει ένα ολοκληρωμένο πλαίσιο εφαρμογής τεχνικών Market Basket Analysis σε πραγματικά δεδομένα λιανεμπορίου, συνδυάζοντας την ανάλυση της καταναλωτικής συμπεριφοράς με την αξιολόγηση διαφορετικών προσεγγίσεων εξόρυξης συχνών στοιχειοσυνόλων και κανόνων συσχέτισης.

1.5 Οργάνωση της διπλωματικής εργασίας

Η παρούσα διπλωματική εργασία αποτελείται από πέντε κεφάλαια.

Στο πρώτο κεφάλαιο παρουσιάζεται το αντικείμενο της εργασίας, ο σκοπός και το κίνητρο της έρευνας, καθώς και η συνεισφορά της μελέτης.

Το δεύτερο κεφάλαιο περιλαμβάνει το θεωρητικό υπόβαθρο της εργασίας. Αρχικά παρουσιάζεται η επίδραση της πανδημίας COVID-19 στην καταναλωτική συμπεριφορά και στη συνέχεια αναλύονται οι βασικές έννοιες του Market Basket Analysis, των κανόνων συσχέτισης και των σημαντικότερων μετρικών αξιολόγησης. Επιπροσθέτως, παρουσιάζονται οι αλγόριθμοι εξόρυξης συχνών στοιχειοσυνόλων που χρησιμοποιούνται στην εργασία, καθώς και σχετικές ερευνητικές μελέτες από τη διεθνή βιβλιογραφία.

Στο τρίτο κεφάλαιο περιγράφεται η πειραματική διαδικασία που ακολουθήθηκε. Παρουσιάζονται τα δεδομένα της έρευνας, τα στάδια προεπεξεργασίας, ο διαχωρισμός των συναλλαγών στις δύο εξεταζόμενες χρονικές περιόδους και η διερευνητική ανάλυση των δεδομένων. Η τελευταία περιλαμβάνει, μεταξύ άλλων, τη σύγκριση των ποσοστών υποστήριξης των κοινών κατηγοριών προϊόντων και του μέσου μεγέθους του καλαθιού μεταξύ των δύο περιόδων. Ταυτόχρονα, αναλύεται η διαδικασία επιλογής και βελτιστοποίησης των παραμέτρων του αλγορίθμου Apriori.

Το τέταρτο κεφάλαιο περιλαμβάνει τα αποτελέσματα της έρευνας. Παρουσιάζονται τα συχνά στοιχειοσύνολα και οι κανόνες συσχέτισης που προέκυψαν για κάθε χρονική περίοδο, ενώ ακολουθεί συγκριτική ανάλυση της καταναλωτικής συμπεριφοράς μεταξύ της περιόδου COVID-19 και της μεταπανδημικής περιόδου. Στο πλαίσιο της σύγκρισης εξετάζονται οι μεταβολές του support στα κοινά συχνά στοιχειοσύνολα, καθώς και οι μεταβολές του confidence και του lift στους κοινούς κανόνες συσχέτισης. Τέλος, πραγματοποιείται συγκριτική αξιολόγηση των αλγορίθμων Apriori, FP-Growth και Eclat στο ενοποιημένο σύνολο δεδομένων, με κοινές παραμέτρους, ως προς τον χρόνο εκτέλεσης και τη μέγιστη καταγεγραμμένη χρήση μνήμης.

Στο πέμπτο κεφάλαιο συνοψίζονται τα βασικά συμπεράσματα της εργασίας, παρουσιάζεται η συνεισφορά της έρευνας, αναφέρονται οι περιορισμοί της μελέτης και διατυπώνονται προτάσεις για μελλοντική έρευνα.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

2.1 Επίδραση της Πανδημίας COVID-19 στην Καταναλωτική Συμπεριφορά

2.1.1 Μεταβολές στις αγοραστικές συνήθειες κατά την πανδημία

Όπως προαναφέραμε, η πανδημία COVID-19 αποτέλεσε μία από τις πιο σημαντικές υγειονομικές και κοινωνικοοικονομικές κρίσεις των τελευταίων δεκαετιών. Τα περιοριστικά μέτρα, η αβεβαιότητα σχετικά με την εξέλιξη της πανδημίας, οι φόβοι για πιθανές ελλείψεις προϊόντων και οι αλλαγές στην καθημερινή ζωή οδήγησαν σε σημαντικές μεταβολές των αγοραστικών συνηθειών. Ο Sheth [5] υποστηρίζει ότι η πανδημία λειτούργησε ως ένας εξωτερικός παράγοντας που διέκοψε τις καθιερωμένες καταναλωτικές συμπεριφορές και ανάγκασε τους καταναλωτές να προσαρμοστούν σε ένα νέο περιβάλλον λήψης αποφάσεων. Ως αποτέλεσμα, εμφανίστηκαν νέες μορφές κατανάλωσης, ενώ ορισμένες προϋπάρχουσες τάσεις αυξήθηκαν σημαντικά [5].

Μία από τις πιο χαρακτηριστικές εκδηλώσεις της καταναλωτικής συμπεριφοράς κατά τα πρώτα στάδια της πανδημίας ήταν το φαινόμενο του *panic buying*. Η συμπεριφορά αυτή αναφέρεται στην υπερβολική αγορά προϊόντων εξαιτίας φόβου ή αβεβαιότητας σχετικά με τη μελλοντική διαθεσιμότητά τους. Η συστηματική βιβλιογραφική ανασκόπηση των Billore και Anisimova [6] κατέδειξε ότι το *panic buying* αποτέλεσε παγκόσμιο φαινόμενο, το οποίο παρατηρήθηκε σε διαφορετικές χώρες και πολιτισμικά περιβάλλοντα. Βάσει των θέσεων των συγγραφέων, η συγκεκριμένη συμπεριφορά δεν αποτελεί απλώς μία παρορμητική αντίδραση των καταναλωτών. Αποτελεί ένα σύνθετο φαινόμενο που επηρεάζεται από ψυχολογικούς, κοινωνικούς και περιβαλλοντικούς παράγοντες.

Παρόμοια συμπεράσματα προκύπτουν και από τη μελέτη των Jazemi et al. [7], οι οποίοι επισημαίνουν ότι το *panic buying* αποτελεί επαναλαμβανόμενο φαινόμενο σε περιόδους κρίσεων και έκτακτων καταστάσεων. Οι συγγραφείς αναφέρουν ότι ο φόβος, η αβεβαιότητα και η ανάγκη διατήρησης ελέγχου οδηγούν πολλούς καταναλωτές στο να αγοράζουν μεγαλύτερες ποσότητες προϊόντων από ό,τι χρειάζονται. Η συμπεριφορά αυτή ενισχύεται συχνά από τις πληροφορίες που διακινούνται μέσω των μέσων ενημέρωσης και των μέσων κοινωνικής δικτύωσης, καθώς και από την παρατήρηση αντίστοιχων ενεργειών από άλλους καταναλωτές.

Η βιβλιογραφία συνδέει το *panic buying* κυρίως με ψυχολογικούς και κοινωνικούς παράγοντες. Οι Chua et al. [8] διαπίστωσαν ότι η αντίληψη κινδύνου, ο φόβος ελλείψεων αλλά και η κοινωνική επιρροή αποτελούν τους κύριους παράγοντες που είναι υπεύθυνοι για την εκδήλωση της συγκεκρι-

μένης συμπεριφοράς. Όταν ένας καταναλωτής αντιληφθεί ότι ένα προϊόν ενδέχεται να μην είναι διαθέσιμο στο μέλλον, αυξάνεται η επιθυμία του να πραγματοποιήσει μεγαλύτερες αγορές προκειμένου να μπορέσουν να περιορίσουν την αβεβαιότητα. Αντίστοιχα, οι Omar et al. [9] κατέληξαν στο συμπέρασμα ότι η αβεβαιότητα, η αντιλαμβανόμενη σοβαρότητα της πανδημίας, η αντίληψη έλλειψης αγαθών όπως και το άγχος αποτελούν καθοριστικούς παράγοντες που ενισχύουν την τάση για υπερβολικές αγορές. Τα αποτελέσματα της μελέτης τους φανερώνουν ότι οι καταναλωτές συχνά δημιουργούν αποθέματα προκειμένου να μπορέσουν οι ίδιοι να διαχειριστούν το άγχος τους και να αντιμετωπίσουν τις συνθήκες της αβεβαιότητας.

Στενά συνδεδεμένο με το panic buying ήταν και το φαινόμενο της δημιουργίας αποθεμάτων (stockpiling). Αν και οι δύο έννοιες παρουσιάζουν ομοιότητες, η βιβλιογραφία τις διαχωρίζει εννοιολογικά. Ο Fernando et al. [10] επισημαίνουν ότι η αποθεματοποίηση μπορεί να θεωρηθεί ως μία μορφή ασυνήθιστης συσσώρευσης αγαθών με σκοπό τη μείωση του φόβου απώλειας ή της πιθανότητας μελλοντικών ελλείψεων. Οι συγγραφείς υποστηρίζουν ότι αυτή η συμπεριφορά παρουσιάζεται εξαιτίας της ανάγκης των καταναλωτών για ανάκτηση ενός αισθήματος ελέγχου και ασφάλειας στις περιόδους έντονης αβεβαιότητας. Σε δεύτερο επίπεδο, επισημαίνουν ότι το stockpiling διαφοροποιείται από το panic buying, καθώς συχνά αποτελεί περισσότερο μία προληπτική και οργανωμένη στρατηγική διαχείρισης κινδύνου, ενώ το panic buying συνδέεται με πιο άμεσες και παρορμητικές αντιδράσεις στον φόβο που προκαλεί μια κρίση [10].

Η δημιουργία αποθεμάτων συνδέθηκε άμεσα με την αντίληψη πιθανών ελλείψεων στην αγορά. Οι Schumacher και Micheli [11] υποστηρίζουν ότι οι καταναλωτές δεν αντιδρούσαν μόνο σε πραγματικές ελλείψεις προϊόντων αλλά και στην προσδοκία ότι τέτοιες ελλείψεις ενδέχεται να εμφανιστούν στο μέλλον. Η έννοια της «αναμενόμενης σπανιότητας» (anticipated scarcity) φαίνεται να έπαιξε σημαντικό ρόλο στη διαμόρφωση των αγοραστικών αποφάσεων κατά τη διάρκεια της πανδημίας. Ακόμη και όταν τα προϊόντα είχαν αρκετό απόθεμα, η αντίληψη ότι η πρόσβαση σε αυτά θα μπορούσε να περιοριστεί στο μέλλον ήταν αρκετή ώστε να οδηγήσει σε αυξημένες αγορές και δημιουργία αποθεμάτων.

Οι μεταβολές αυτές αποτυπώθηκαν και στα πρότυπα αγορών των καταναλωτών. Η μελέτη των Żurek και Rudy [12] κατέγραψε σημαντικές αλλαγές στις αγορές τροφίμων κατά την περίοδο της πανδημίας. Οι καταναλωτές περιορίσαν τη συχνότητα των επισκέψεών τους στα καταστήματα προκειμένου να μειώσουν τον κίνδυνο έκθεσης στον ιό. Αυτό συντέλεσε στην αύξηση των ποσοτήτων αγορών ανά επίσκεψη με αποτέλεσμα να παρατηρηθεί σημαντική αύξηση του μεγέθους του καλαθιού αγορών, καθώς και της συνολικής αξίας των συναλλαγών ενώ την ίδια στιγμή καταγράφηκε αυξημένη ζήτηση για βασικά αγαθά και προϊόντα που θεωρούνταν απαραίτητα για την κάλυψη καθημερινών αναγκών.

Από την βιβλιογραφική ανασκόπηση προκύπτει ότι η πανδημία COVID-19 επηρέασε ουσιαστικά την καταναλωτική συμπεριφορά και τις αγοραστικές συνήθειες των καταναλωτών. Η εμφάνιση του panic buying, η εκτεταμένη αποθεματοποίηση προϊόντων, η αυξημένη ζήτηση βασικών αγαθών, η μείωση της συχνότητας επισκέψεων στα καταστήματα και η αύξηση του μεγέθους των καλαθιών αγορών αποτελούν ορισμένες από τις σημαντικότερες μεταβολές που καταγράφηκαν διεθνώς [5],[6],[7],[8],[9],[10],[11],[12]. Παρ'όλα αυτά, παραμένει ανοιχτό το ερώτημα κατά πόσο οι αλλαγές αυτές διατηρήθηκαν και μετά την άρση των περιοριστικών μέτρων, ζήτημα που εξετάζεται στην επόμενη ενότητα.

2.1.2 Μεταβολές μετά την άρση των περιορισμών

Παρότι οι επιπτώσεις της πανδημίας COVID-19 στην καταναλωτική συμπεριφορά έχουν μελετηθεί εκτενώς, η διερεύνηση της διάρκειας αυτών των μεταβολών αποτελεί ένα σχετικά νεότερο πεδίο έρευνας. Μετά την άρση των περιοριστικών μέτρων και τη σταδιακή επαναφορά στην κανονικότητα προέκυψε το ερώτημα κατά πόσο οι αλλαγές που παρατηρήθηκαν κατά τη διάρκεια της υγειονομικής κρίσης ήταν προσωρινές ή αν ορισμένες από αυτές διατηρήθηκαν και στη μεταπανδημική περίοδο. Οι σχετικές μελέτες παρουσιάζουν ότι ενώ αρκετές συμπεριφορές υποχώρησαν με την επιστροφή στην κανονικότητα, ορισμένες άλλες εδραιώθηκαν στις καθημερινές πρακτικές των καταναλωτών.

Ένα από τα πιο χαρακτηριστικά παραδείγματα αποτελεί η χρήση των ηλεκτρονικών αγορών (online shopping). Η πανδημία συντέλεσε σε μεγάλο βαθμό στη στροφή σημαντικού αριθμού καταναλωτών να στραφούν στις ηλεκτρονικές αγορές κυρίως για την αγορά τροφίμων και ειδών καθημερινής χρήσης. Όπως υποστηρίζουν οι Tyrväinen και Karjalainen [13], η εμπειρία χρήσης των ηλεκτρονικών αγορών κατά τη διάρκεια της πανδημίας συνέβαλε στη διαμόρφωση νέων καταναλωτικών συνηθειών, οι οποίες δεν εγκαταλείφθηκαν μετά την άρση των περιορισμών διότι η ευκολία πραγματοποίησης αγορών, η εξοικονόμηση χρόνου και η δυνατότητα παράδοσης κατ' οίκον αναγνωρίστηκαν ως σημαντικά πλεονεκτήματα, με αποτέλεσμα πολλοί καταναλωτές να συνεχίσουν να χρησιμοποιούν ψηφιακά κανάλια ακόμη και όταν είχαν πλέον τη δυνατότητα να επιστρέψουν στις παραδοσιακές μορφές αγορών.

Παρόμοια ευρήματα παρουσιάζουν και οι Galushko και Riabchyk [14], οι οποίοι υποστηρίζουν ότι η πανδημία επιτάχυνε τη διαδικασία ψηφιακού μετασχηματισμού του λιανεμπορίου και βοήθησε σημαντικά στην εξοικείωση των καταναλωτών με τις ηλεκτρονικές υπηρεσίες. Όπως υποστηρίζουν οι συγγραφείς, πολλοί καταναλωτές διατήρησαν τη χρήση των ηλεκτρονικών αγορών και μετά το τέλος της υγειονομικής κρίσης, γεγονός που υποδηλώνει ότι η εμπειρία της πανδημίας λειτούργησε ως καταλύτης για τη διαμόρφωση πιο μόνιμων καταναλωτικών συνηθειών.

Επιπρόσθετα, η μεταπανδημική περίοδος δεν χαρακτηρίζεται από πλήρη αντικατάσταση των φυσικών καταστημάτων από τα ηλεκτρονικά κανάλια. Αντίθετα, η βιβλιογραφία καταγράφει την ανάπτυξη ενός υβριδικού μοντέλου αγορών, στο οποίο οι καταναλωτές αξιοποιούν συνδυαστικά τόσο τα φυσικά όσο και τα ηλεκτρονικά κανάλια. Οι Diaz-Gutierrez et al. [15] διαπίστωσαν ότι οι καταναλωτές επέστρεψαν στα φυσικά καταστήματα μετά την άρση των περιορισμών, χωρίς όμως να εγκαταλείψουν τις ψηφιακές υπηρεσίες που είχαν υιοθετήσει κατά τη διάρκεια της πανδημίας. Η συμπεριφορά αυτή υποδηλώνει ότι τα δύο κανάλια δεν αντιμετωπίζονται πλέον ως εναλλακτικές επιλογές αλλά ως συμπληρωματικά μέσα που εξυπηρετούν διαφορετικές ανάγκες και περιστάσεις.

Εκτός από τις αλλαγές στον τρόπο πραγματοποίησης των αγορών, αρκετές μελέτες αναφέρουν ότι υπήρξαν μεταβολές και στις καταναλωτικές προτεραιότητες. Η πανδημία φαίνεται να ενίσχυσε το ενδιαφέρον των καταναλωτών για ζητήματα που σχετίζονται με την υγεία, την ασφάλεια των προϊόντων, την αξιοπιστία των επιχειρήσεων και τη διασφάλιση της διαθεσιμότητας βασικών αγαθών ενώ η αξία για τα χρήματα και η δυνατότητα σύγκρισης εναλλακτικών επιλογών αναδείχθηκαν σε σημαντικά κριτήρια λήψης αγοραστικών αποφάσεων, ιδιαίτερα σε ένα περιβάλλον που επηρεάστηκε από πιέσεις εξαιτίας του πληθωρισμού και της οικονομικής αβεβαιότητας [13],[14],[15].

Αντίθετα, οι πιο ακραίες συμπεριφορές που παρατηρήθηκαν κατά τα πρώτα στάδια της πανδημίας, όπως το panic buying και η εκτεταμένη αποθεματοποίηση προϊόντων, φαίνεται να υποχώρησαν

σημαντικά μετά την ομαλοποίηση της αγοράς. Η σταδιακή αποκατάσταση της εμπιστοσύνης στις αλυσίδες εφοδιασμού και η μείωση της αβεβαιότητας σχετικά με την επάρκεια των προϊόντων περιόρισαν σημαντικά την ανάγκη δημιουργίας μεγάλων αποθεμάτων, οδηγώντας σε πιο ορθολογικά πρότυπα κατανάλωσης.

Με βάση τις παραπάνω μελέτες υποστηρίζεται ότι η μεταπανδημική περίοδος δεν σηματοδότησε μια πλήρη επιστροφή στις προπανδημικές καταναλωτικές συνήθειες. Αν και ορισμένες συμπεριφορές που συνδέονταν άμεσα με τις συνθήκες της κρίσης υποχώρησαν, άλλες διατηρήθηκαν και ενσωματώθηκαν στην καθημερινότητα των καταναλωτών. Ιδιαίτερα η αυξημένη χρήση των ψηφιακών καναλιών, η ανάπτυξη υβριδικών μορφών αγορών και η μεταβολή ορισμένων καταναλωτικών προτεραιοτήτων αποτελούν ενδείξεις ότι η πανδημία άφησε ένα πιο μακροχρόνιο αποτύπωμα στην καταναλωτική συμπεριφορά.

Από την άλλη πλευρά είναι σημαντικό να αναφερθεί ότι παρά τη σημαντική συμβολή των παραπάνω ερευνών, οι περισσότερες βασίζονται σε ερωτηματολόγια και στις δηλωμένες αντιλήψεις των καταναλωτών. Ως εκ τούτου, παραμένει ανοιχτό το ερώτημα κατά πόσο οι μεταβολές αυτές αποτυπώνονται και στα πραγματικά πρότυπα αγορών που προκύπτουν από δεδομένα συναλλαγών. Η παρούσα εργασία επιχειρεί να συμβάλει στη διερεύνηση αυτού του ζητήματος μέσω της ανάλυσης πραγματικών δεδομένων από τον κλάδο των σούπερ μάρκετ, συγκρίνοντας τα πρότυπα αγορών της περιόδου COVID-19 με εκείνα της μεταπανδημικής περιόδου.

2.2 Market Basket Analysis

2.2.1 Ορισμός και Βασικές Αρχές

Το Market Basket Analysis (MBA) αποτελεί μία από τις πιο διαδεδομένες εφαρμογές που χρησιμοποιούνται για την εξόρυξη δεδομένων στον χώρο του λιανεμπορίου. Στηρίζεται στην ανακάλυψη σχέσεων μεταξύ προϊόντων που εμφανίζονται μαζί στις συναλλαγές των καταναλωτών. Η θεωρητική του βάση είναι άμεσα συνδεδεμένη με την εξόρυξη κανόνων συσχέτισης (Association Rule Mining), η οποία παρουσιάστηκε από τους Agrawal, Imieliński και Swami [16] προκειμένου να μπορέσουν να εντοπιστούν συσχετίσεις μεταξύ αντικειμένων μέσα σε μεγάλες βάσεις δεδομένων συναλλαγών. Η βασική ιδέα της προσέγγισης είναι ότι η συχνή συνύπαρξη συγκεκριμένων προϊόντων σε διαφορετικές αγορές μπορεί να αποκαλύπτει υποκείμενα πρότυπα καταναλωτικής συμπεριφοράς.

Η λειτουργία του Market Basket Analysis βασίζεται στην ανάλυση συναλλαγών (transactions). Κάθε συναλλαγή αντιπροσωπεύει το σύνολο των προϊόντων που αγοράστηκαν από έναν καταναλωτή σε μία συγκεκριμένη αγορά και μπορεί να θεωρηθεί ως ένα «καλάθι αγορών». Το σύνολο των συναλλαγών σχηματίζει τη βάση δεδομένων πάνω στην οποία εφαρμόζονται οι τεχνικές εξόρυξης δεδομένων με στόχο την ανακάλυψη των προτύπων που επαναλαμβάνονται. Σύμφωνα με τους Han, Kamber και Pei [17], η ανάλυση μεγάλων συνόλων συναλλαγών επιτρέπει την εξαγωγή χρήσιμης γνώσης σχετικά με τις σχέσεις που προκύπτουν μεταξύ αντικειμένων, οι οποίες δεν είναι εύκολα αντιληπτές μέσω των συμβατικών μεθόδων ανάλυσης.

Κεντρική έννοια του Market Basket Analysis αποτελούν τα συχνά στοιχειοσύνολα (frequent itemsets). Ως συχνό στοιχειοσύνολο ορίζεται ένα σύνολο προϊόντων που εμφανίζεται ταυτόχρονα σε αριθμό συναλλαγών μεγαλύτερο από ένα προκαθορισμένο όριο συχνότητας. Ο εντοπισμός αυτών

των συνόλων αποτελεί το πρώτο στάδιο της διαδικασίας εξόρυξης γνώσης, καθώς πάνω σε αυτά βασίζεται η δημιουργία των κανόνων συσχέτισης που περιγράφουν τις σχέσεις μεταξύ προϊόντων.

Η παρουσίαση του αλγορίθμου Apriori από τους Agrawal και Srikant [18] έπαιξε σημαντικό ρόλο στην ανάπτυξη του πεδίου. Ο αλγόριθμος αυτός επέτρεψε την αποδοτικότερη αναζήτηση συχνών στοιχειοσυνόλων σε μεγάλες βάσεις δεδομένων και αποτέλεσε τη βάση για την ανάπτυξη μεταγενέστερων αλγορίθμων εξόρυξης κανόνων συσχέτισης. Η συμβολή του Apriori θεωρείται καθοριστική για την ευρεία διάδοση του Market Basket Analysis τόσο στην ακαδημαϊκή έρευνα όσο και στις επιχειρησιακές εφαρμογές.

Σήμερα, η προσέγγιση αυτή αναγνωρίζεται ως μία από τις σημαντικότερες τεχνικές ανάλυσης δεδομένων στον χώρο του λιανεμπορίου. Όπως αναφέρουν οι Alawadh και Barnawi [2], η μέθοδος έχει εξελιχθεί σημαντικά τα τελευταία χρόνια και χρησιμοποιείται σε πληθώρα εφαρμογών που σχετίζονται με την κατανόηση της αγοραστικής συνήθειας αλλά και στη λήψη επιχειρηματικών αποφάσεων. Η δυνατότητα εντοπισμού προτύπων συνύπαρξης προϊόντων καθιστά το MBA ένα ιδιαίτερα χρήσιμο εργαλείο για την εξαγωγή γνώσης από μεγάλα σύνολα δεδομένων συναλλαγών.

Συνεπώς, η τεχνική αυτή δεν περιορίζεται στην απλή καταγραφή των αγορών των καταναλωτών. Αποτελεί μία συστηματική μεθοδολογία ανακάλυψης σχέσεων και προτύπων μέσα από τα δεδομένα συναλλαγών. Η μεθοδολογία αυτή αποτελεί τη βάση για την εξόρυξη κανόνων συσχέτισης, οι οποίοι παρουσιάζονται αναλυτικά στην επόμενη ενότητα.

2.2.2 Εφαρμογές στο λιανεμπόριο

Το Market Basket Analysis αποτελεί μία από τις πιο διαδεδομένες τεχνικές ανάλυσης δεδομένων στον χώρο του λιανεμπορίου, καθώς επιτρέπει την αξιοποίηση των δεδομένων συναλλαγών για την εξαγωγή χρήσιμης γνώσης σχετικά με την καταναλωτική συμπεριφορά. Η συνεχής αύξηση του όγκου των δεδομένων που παράγονται από τις καθημερινές πωλήσεις έχει δημιουργήσει την ανάγκη ανάπτυξης μεθόδων που μπορούν να εντοπίσουν πρότυπα και σχέσεις οι οποίες δεν είναι εύκολα ορατές μέσω της συμβατικής ανάλυσης. Για αυτό τον λόγο, χρησιμοποιείται για την αναγνώριση προϊόντων που αγοράζονται συχνά μαζί, συμβάλλοντας στην καλύτερη κατανόηση των αγοραστικών συνηθειών των πελατών [2].

Από την οπτική των Martinez et al. [4], η ανάλυση των δεδομένων συναλλαγών επιτρέπει στις επιχειρήσεις να εντοπίζουν μοτίβα συσχέτισης μεταξύ προϊόντων αλλά και να κατανοούν καλύτερα τη συμπεριφορά των πελατών τους. Οι πληροφορίες αυτές μπορούν να αξιοποιηθούν κατά τη διαδικασία λήψης επιχειρηματικών αποφάσεων, διότι αποκαλύπτουν σχέσεις που δεν είναι άμεσα εμφανείς μέσα από τους παραδοσιακούς δείκτες πωλήσεων. Η δυνατότητα πρόβλεψης προϊόντων που τείνουν να εμφανίζονται μαζί στις ίδιες συναλλαγές προσφέρει σημαντικό πλεονέκτημα στις επιχειρήσεις λιανικής, ιδιαίτερα σε περιβάλλοντα όπου διαχειρίζονται μεγάλο αριθμό προϊόντων και καθημερινών συναλλαγών.

Μία από τις σημαντικότερες εφαρμογές του Market Basket Analysis αφορά τον σχεδιασμό στρατηγικών προώθησης και διασταυρομένων πωλήσεων (cross-selling). Μέσα από την αναγνώριση προϊόντων που παρουσιάζουν ισχυρές συσχετίσεις, οι επιχειρήσεις μπορούν να αναπτύξουν στοχευμένες προωθητικές ενέργειες αλλά και συνδυαστικές προσφορές. Όπως αναφέρουν οι Alawadh και Barnawi [2], η εξόρυξη κανόνων συσχέτισης έχει χρησιμοποιηθεί εκτενώς για την υποστήριξη ενεργειών marketing, καθώς συμβάλλει στην καλύτερη κατανόηση των αναγκών των πελατών ενώ συμ-

βάλλει και στη βελτίωση της αποτελεσματικότητας των εμπορικών στρατηγικών.

Τα αποτελέσματα του Market Basket Analysis μπορούν να αξιοποιηθούν στη διαχείριση αποθεμάτων. Η γνώση των προϊόντων που αγοράζονται συχνά από κοινού επιτρέπει την αποτελεσματικότερη παρακολούθηση της ζήτησης και τη βελτίωση του προγραμματισμού προμηθειών. Οι Martinez et al. [4] επισημαίνουν ότι η αναγνώριση προϊόντων που εμφανίζονται συστηματικά στις ίδιες συναλλαγές μπορεί να υποστηρίξει σημαντικά τη διαδικασία διαχείρισης αποθεμάτων, καθώς παρέχει πληροφορίες σχετικά με τη συμπεριφορά της αγοράς καθώς και τις μεταβολές στη ζήτηση των προϊόντων.

Αξίζει επίσης να σημειωθεί πως το MBA χρησιμοποιείται ευρέως ως εργαλείο μελέτης της καταναλωτικής συμπεριφοράς, διότι σε αντίθεση με μεθόδους που βασίζονται σε ερωτηματολόγια ή δηλωμένες προθέσεις των καταναλωτών, το MBA αξιοποιεί πραγματικά δεδομένα συναλλαγών και επομένως αποτυπώνει την πραγματική αγοραστική δραστηριότητα των πελατών. Μέσα από την ανάλυση των συσχετίσεων μεταξύ προϊόντων είναι δυνατό να αναγνωριστούν καταναλωτικές συνήθειες, προτιμήσεις και επαναλαμβανόμενα μοτίβα αγορών, τα οποία αντανακλούν τις πραγματικές επιλογές των καταναλωτών [2].

Η δυνατότητα αυτή καθιστά το MBA ένα εργαλείο ιδιαίτερα χρήσιμο σε ερευνητικές μελέτες που επιδιώκουν να διερευνήσουν μεταβολές στην καταναλωτική συμπεριφορά. Μέσω της ανάλυσης μεγάλων όγκων δεδομένων συναλλαγών μπορούν να εντοπιστούν αλλαγές στις αγοραστικές συνήθειες διαφορετικών χρονικών περιόδων, καθώς και διαφοροποιήσεις στις σχέσεις μεταξύ προϊόντων που σχετίζονται με εξωτερικούς παράγοντες, όπως οικονομικές κρίσεις, κοινωνικές μεταβολές ή έκτακτα γεγονότα. Εξαιτίας όσων προαναφέρθηκαν, η μέθοδος έχει χρησιμοποιηθεί εκτενώς τόσο από επιχειρήσεις όσο και από ερευνητές που επιδιώκουν να κατανοήσουν εν τω βάθει τη συμπεριφορά των καταναλωτών.

Στο πλαίσιο της παρούσας εργασίας, το Market Basket Analysis χρησιμοποιείται προκειμένου να διερευνηθούν τα αγοραστικά πρότυπα που εμφανίζονται στις συναλλαγές των καταναλωτών κατά την περίοδο της πανδημίας COVID-19 και μετά από αυτήν. Μέσω της ανάλυσης πραγματικών δεδομένων συναλλαγών επιδιώκεται η αναγνώριση πιθανών μεταβολών στις σχέσεις μεταξύ κατηγοριών προϊόντων και η αξιολόγηση του κατά πόσο οι αλλαγές που καταγράφονται στη διεθνή βιβλιογραφία αποτυπώνονται και στην πραγματική αγοραστική συμπεριφορά των καταναλωτών.

2.2.3 Πλεονεκτήματα και περιορισμοί

Το Market Basket Analysis έχει καθιερωθεί ως μία από τις πιο σημαντικές τεχνικές ανάλυσης δεδομένων στο λιανεμπόριο. Αυτό συμβαίνει επειδή παρέχει τη δυνατότητα για εξαγωγή χρήσιμης γνώσης από μεγάλα σύνολα δεδομένων συναλλαγών. Η ευρεία εφαρμογή του οφείλεται κυρίως στην ικανότητά του να αποκαλύπτει πρότυπα αγορών και σχέσεις μεταξύ προϊόντων που δεν είναι άμεσα εμφανείς μέσω της συμβατικής ανάλυσης πωλήσεων. Όπως κάθε αναλυτική μέθοδος, έτσι και αυτή, παρουσιάζει σημαντικά πλεονεκτήματα αλλά και ορισμένους περιορισμούς που πρέπει να λαμβάνονται υπόψη κατά την εφαρμογή και την ερμηνεία των αποτελεσμάτων του.

Αρχικά, ένα από τα σημαντικότερα πλεονεκτήματά του είναι ότι βασίζεται σε πραγματικά δεδομένα συναλλαγών και όχι σε δηλωμένες προτιμήσεις ή απόψεις των καταναλωτών. Αυτό έχει ως αποτέλεσμα τα ευρήματα που προκύπτουν να αντανακλούν την πραγματική αγοραστική συμπεριφορά και όχι τις αντιλήψεις ή τις προθέσεις των καταναλωτών. Όπως υποστηρίζουν οι Alawadh και

Barnawi [2], η αξιοποίηση πραγματικών δεδομένων επιτρέπει την ανακάλυψη προτύπων και σχέσεων οι οποίες δύσκολα μπορούν να εντοπιστούν μέσω παραδοσιακών ερευνητικών μεθόδων, όπως τα ερωτηματολόγια ή οι συνεντεύξεις.

Στα πλεονεκτήματα συγκαταλέγεται και η δυνατότητα διαχείρισης και ανάλυσης πολύ μεγάλων βάσεων δεδομένων. Οι σύγχρονες επιχειρήσεις λιανικής συλλέγουν καθημερινά τεράστιο όγκο δεδομένων πωλήσεων, γεγονός που καθιστά απαραίτητη τη χρήση αυτοματοποιημένων τεχνικών εξόρυξης γνώσης. Όπως αναφέρουν οι Han, Kamber και Pei [17], οι τεχνικές εξόρυξης δεδομένων επιτρέπουν την αναγνώριση επαναλαμβανόμενων προτύπων σε σύνολα δεδομένων τα οποία θα ήταν αδύνατο να αναλυθούν αποτελεσματικά με συμβατικές στατιστικές προσεγγίσεις.

Επίσης, το Market Basket Analysis παρέχει τη δυνατότητα εντοπισμού μη προφανών σχέσεων μεταξύ προϊόντων. Μέσω της εξόρυξης κανόνων συσχέτισης οι επιχειρήσεις έχουν την δυνατότητα να ανακαλύψουν συνδυασμούς προϊόντων που εμφανίζονται συστηματικά στις ίδιες συναλλαγές, ακόμη και όταν οι σχέσεις αυτές δεν είναι εμφανείς στους αναλυτές ή στα στελέχη των επιχειρήσεων. Η δυνατότητα αυτή καθιστά τη μέθοδο ιδιαίτερα χρήσιμη για την κατανόηση των καταναλωτικών προτύπων και την υποστήριξη επιχειρηματικών αποφάσεων [4].

Παρά τα σημαντικά πλεονεκτήματά του, το Market Basket Analysis παρουσιάζει και ορισμένους περιορισμούς. Ο σημαντικότερος ίσως περιορισμός είναι ότι οι κανόνες που παράγονται περιγράφουν σχέσεις συσχέτισης και όχι σχέσεις αιτιότητας. Το γεγονός ότι δύο προϊόντα αγοράζονται συχνά μαζί δεν σημαίνει απαραίτητα ότι το ένα προκαλεί την αγορά του άλλου. Οι κανόνες συσχέτισης αποκαλύπτουν την ύπαρξη στατιστικών σχέσεων μεταξύ προϊόντων, χωρίς όμως να μπορούν να ερμηνεύσουν πλήρως τους λόγους που οδηγούν τους καταναλωτές στις συγκεκριμένες αγοραστικές επιλογές [17].

Ένας ακόμη περιορισμός σχετίζεται με τον μεγάλο αριθμό κανόνων που μπορεί να παραχθεί κατά τη διαδικασία ανάλυσης. Σε μεγάλες βάσεις δεδομένων, ιδιαίτερα όταν περιλαμβάνουν εκατοντάδες ή χιλιάδες διαφορετικά προϊόντα, η εφαρμογή των αλγορίθμων εξόρυξης μπορεί να οδηγήσει στη δημιουργία πολύ μεγάλου πλήθους κανόνων συσχέτισης. Η ύπαρξη μεγάλου αριθμού αποτελεσμάτων δυσκολεύει τη διαδικασία ερμηνείας και καθιστά αναγκαία τη χρήση κατάλληλων μέτρων αξιολόγησης για τον εντοπισμό των σημαντικότερων κανόνων [18].

Δεν θα πρέπει να παραβλεφθεί ότι τα αποτελέσματα του Market Basket Analysis επηρεάζονται σημαντικά από την επιλογή των παραμέτρων που χρησιμοποιούνται κατά τη διαδικασία εξόρυξης. Η μεταβολή των ορίων συχνότητας εμφάνισης ή άλλων κριτηρίων αξιολόγησης μπορεί να οδηγήσει σε σημαντικά διαφορετικά αποτελέσματα. Για αυτό, η επιλογή κατάλληλων παραμέτρων αποτελεί κρίσιμο στάδιο για την αναλυτική διαδικασία και επηρεάζει άμεσα την ποιότητα αλλά και τη χρησιμότητα των κανόνων που προκύπτουν [2].

Τέλος, η αποτελεσματικότητα της μεθόδου εξαρτάται σε μεγάλο βαθμό από την ποιότητα των διαθέσιμων δεδομένων. Δεδομένα τα οποία είναι ελλιπή, λανθασμένα ή μη αντιπροσωπευτικά μπορούν να οδηγήσουν σε παραπλανητικά συμπεράσματα και να μειώσουν την αξιοπιστία των αποτελεσμάτων. Για αυτό τον λόγο, η προεπεξεργασία και η διασφάλιση της ποιότητας των δεδομένων αποτελούν απαραίτητες προϋποθέσεις για την επιτυχή εφαρμογή του [17].

Συνολικά, το Market Basket Analysis προσφέρει σημαντικές δυνατότητες κατανόησης της αγοραστικής συμπεριφοράς μέσω πραγματικών δεδομένων συναλλαγών. Οι περιορισμοί του καθιστούν αναγκαία την προσεκτική επιλογή των παραμέτρων και τη συνδυαστική αξιολόγηση των παραγόμενων κανόνων. Κατά συνέπεια, η επόμενη ενότητα παρουσιάζει αναλυτικά την έννοια των κανόνων

συσχέτισης και τις βασικές μετρικές που χρησιμοποιούνται για την αξιολόγηση της ισχύος και της σημασίας τους.

2.3 Εξόρυξη Κανόνων Συσχέτισης (Association Rules)

Η εξόρυξη κανόνων συσχέτισης (Association Rule Mining – ARM) αποτελεί μία από τις σημαντικότερες τεχνικές που χρησιμοποιούνται για την εξόρυξη δεδομένων και συνιστά τον βασικό μηχανισμό λειτουργίας του Market Basket Analysis. Σκοπός της είναι η ανακάλυψη σχέσεων μεταξύ αντικειμένων που εμφανίζονται συχνά μαζί σε ένα σύνολο συναλλαγών. Αυτό επιτρέπει την εξαγωγή χρήσιμης γνώσης μέσα από μεγάλους όγκους δεδομένων. Οι κανόνες συσχέτισης περιγράφουν σχέσεις μεταξύ συνόλων αντικειμένων (itemsets) που συνυπάρχουν συχνά στις συναλλαγές και αποκαλύπτουν πρότυπα τα οποία δεν είναι άμεσα εμφανή μέσω της συμβατικής ανάλυσης δεδομένων [19].

Η έννοια της εξόρυξης κανόνων συσχέτισης παρουσιάστηκε αρχικά από τους Agrawal, Imieliński και Swami [16], οι οποίοι διερεύνησαν το πρόβλημα της ανακάλυψης σχέσεων μεταξύ προϊόντων σε μεγάλες βάσεις δεδομένων συναλλαγών. Η εργασία τους αποτέλεσε το θεμέλιο για την ανάπτυξη του Market Basket Analysis αλλά και για τη δημιουργία αλγορίθμων που επιτρέπουν την αποδοτική αναζήτηση συσχετίσεων σε πολύ μεγάλα σύνολα δεδομένων.

Στο πλαίσιο του λιανεμπορίου, κάθε συναλλαγή αντιστοιχεί σε ένα σύνολο προϊόντων που αγοράστηκαν ταυτόχρονα από έναν καταναλωτή. Η ανάλυση μεγάλου αριθμού τέτοιων συναλλαγών επιτρέπει τον εντοπισμό επαναλαμβανόμενων προτύπων αγορών και τη διερεύνηση των σχέσεων μεταξύ διαφορετικών προϊόντων ή κατηγοριών προϊόντων. Σύμφωνα με τους Malliaridis, Ougiaroglou και Dervos [19], η εξόρυξη κανόνων συσχέτισης χρησιμοποιείται για την ανακάλυψη γνώσης που σχετίζεται με τη συχνή συνύπαρξη προϊόντων και μπορεί να εφαρμοστεί σε πλήθος επιχειρηματικών και ερευνητικών πεδίων.

Η διαδικασία εξόρυξης κανόνων συσχέτισης πραγματοποιείται σε δύο διαδοχικά στάδια. Στο πρώτο στάδιο εντοπίζονται τα συχνά σύνολα αντικειμένων (frequent itemsets), δηλαδή οι συνδυασμοί προϊόντων που εμφανίζονται με υψηλή συχνότητα στις συναλλαγές ενώ στο δεύτερο στάδιο δημιουργούνται κανόνες συσχέτισης από τα frequent itemsets που έχουν ήδη εντοπιστεί. Η προσέγγιση αυτή αποτελεί τη βασική διαδικασία που ακολουθείται από τους περισσότερους αλγορίθμους εξόρυξης κανόνων συσχέτισης [19].

Η αποτελεσματική εξόρυξη κανόνων συσχέτισης είναι ιδιαίτερα σημαντική σε μεγάλα σύνολα δεδομένων, όπου ο αριθμός των πιθανών συνδυασμών προϊόντων αυξάνεται εκθετικά. Αυτό εξηγεί τον λόγο για τον οποίο έχουν αναπτυχθεί εξειδικευμένοι αλγόριθμοι, όπως ο Apriori, ο FP-Growth και ο Eclat, οι οποίοι επιτρέπουν τον αποδοτικό εντοπισμό συχνών συνόλων αντικειμένων και την παραγωγή αξιόπιστων κανόνων συσχέτισης ακόμη και σε βάσεις δεδομένων που περιλαμβάνουν εκατομμύρια συναλλαγές [20].

Στη συγκεκριμένη εργασία η εξόρυξη κανόνων συσχέτισης αξιοποιείται για τη μελέτη των αγοραστικών προτύπων που εμφανίζονται σε πραγματικά δεδομένα συναλλαγών από τον κλάδο των σούπερ μάρκετ. Στόχος είναι η αναγνώριση σχέσεων μεταξύ κατηγοριών προϊόντων και η διερεύνηση πιθανών διαφοροποιήσεων στην καταναλωτική συμπεριφορά μεταξύ της περιόδου της πανδημίας COVID-19 και της μεταπανδημικής περιόδου.

2.3.1 Frequent Itemsets – Συχνά Στοιχειοσύνολα

Τα συχνά στοιχειοσύνολα (frequent itemsets) αποτελούν τη βάση της εξόρυξης κανόνων συσχέτισης. Ως itemset ορίζεται οποιοδήποτε σύνολο αντικειμένων ή προϊόντων που εμφανίζονται μέσα σε μία συναλλαγή. Ένα itemset χαρακτηρίζεται ως συχνό όταν η συχνότητα εμφάνισής του στο σύνολο των συναλλαγών υπερβαίνει ένα προκαθορισμένο όριο υποστήριξης (minimum support threshold) [20].

Η αναζήτηση frequent itemsets αποτελεί το πρώτο στάδιο της διαδικασίας εξόρυξης γνώσης. Βάσει των Malliaridis και Ougiaroglou [20], το Frequent Itemset Mining αποτελεί μία από τις πιο σημαντικές εργασίες της εξόρυξης δεδομένων, καθώς από τα αποτελέσματά του εξαρτάται η ποιότητα των κανόνων συσχέτισης που θα παραχθούν στη συνέχεια. Η σημασία του έγκειται στο γεγονός ότι επιτρέπει τον εντοπισμό προϊόντων που εμφανίζονται συστηματικά μαζί στις συναλλαγές των καταναλωτών.

Για παράδειγμα, εάν οι κατηγορίες προϊόντων «Τυριά» και «Αλλαντικά» εμφανίζονται ταυτόχρονα σε μεγάλο αριθμό συναλλαγών, τότε το σύνολο Τυριά, Αλλαντικά μπορεί να χαρακτηριστεί ως frequent itemset. Αντίστοιχα, μεγαλύτερα σύνολα προϊόντων μπορούν επίσης να θεωρηθούν συχνά, εφόσον ικανοποιούν το ελάχιστο όριο συχνότητας που έχει οριστεί από τον αναλυτή.

Η διαδικασία αναζήτησης frequent itemsets είναι υπολογιστικά απαιτητική. Αυτό συμβαίνει διότι ο αριθμός των πιθανών συνδυασμών αυξάνεται σημαντικά όσο αυξάνεται ο αριθμός των διαθέσιμων προϊόντων. Για τον λόγο αυτό έχουν αναπτυχθεί ειδικοί αλγόριθμοι και τεχνικές περικοπής (pruning techniques), οι οποίοι μειώνουν τον χώρο αναζήτησης και καθιστούν εφικτή την ανάλυση μεγάλων συνόλων δεδομένων [20].

2.3.2 Κανόνες Συσχέτισης (Association Rules)

Μετά τον εντοπισμό των συχνών συνόλων αντικειμένων (frequent itemsets), σειρά έχει η παραγωγή κανόνων συσχέτισης (association rules). Οι κανόνες συσχέτισης αποτελούν το τελικό αποτέλεσμα της διαδικασίας εξόρυξης γνώσης και χρησιμοποιούνται προκειμένου να περιγράψουν τις σχέσεις μεταξύ αντικειμένων που εμφανίζονται συχνά μαζί στις συναλλαγές. Σύμφωνα με τους Malliaridis, Ougiaroglou και Dervos [19], οι κανόνες συσχέτισης αποκαλύπτουν σχέσεις μεταξύ συνόλων αντικειμένων που συνυπάρχουν συχνά στα δεδομένα και επιτρέπουν την εξαγωγή χρήσιμων συμπερασμάτων σχετικά με τα πρότυπα που εμφανίζονται στις συναλλαγές.

Ένας κανόνας συσχέτισης εκφράζει μία σχέση μεταξύ δύο συνόλων αντικειμένων και έχει γενικά τη μορφή:

$$X \rightarrow Y$$

όπου το σύνολο X ονομάζεται πρόδρομο σύνολο (*antecedent*) και το σύνολο Y ονομάζεται επακόλουθο σύνολο (*consequent*). Το πρόδρομο σύνολο αντιπροσωπεύει το σύνολο των αντικειμένων που παρατηρούνται σε μία συναλλαγή, ενώ το επακόλουθο σύνολο αντιπροσωπεύει τα αντικείμενα που εμφανίζονται με αυξημένη πιθανότητα όταν εμφανίζεται το πρόδρομο σύνολο [19].

Λαμβάνοντας υπόψη τη βιβλιογραφία, κάθε κανόνας συσχέτισης αποτελείται από δύο μέρη: το αριστερό τμήμα (*Left-Hand Side – LHS*) και το δεξί τμήμα (*Right-Hand Side – RHS*). Το αριστερό τμήμα αντιστοιχεί στο antecedent, ενώ το δεξί τμήμα αντιστοιχεί στο consequent. Η σχέση μεταξύ των

δύο τμημάτων περιγράφει μία στατιστική συσχέτιση που έχει εντοπιστεί μέσα από την ανάλυση των συναλλαγών και όχι μία σχέση αιτίου-αποτελέσματος [16].

Για παράδειγμα, ο κανόνας:

$$\{\text{Τυριά}\} \rightarrow \{\text{Αλλαντικά}\}$$

υποδηλώνει ότι οι καταναλωτές που αγοράζουν προϊόντα της κατηγορίας «Τυριά» παρουσιάζουν αυξημένη πιθανότητα να αγοράσουν και προϊόντα της κατηγορίας «Αλλαντικά». Η πληροφορία αυτή προκύπτει από την ανάλυση των πραγματικών συναλλαγών και αποτυπώνει μία επαναλαμβανόμενη καταναλωτική συμπεριφορά που παρατηρείται στο εξεταζόμενο σύνολο δεδομένων.

Είναι σημαντικό να τονιστεί ότι ένας κανόνας συσχέτισης δεν υποδηλώνει σχέση αιτιότητας μεταξύ των αντικειμένων που συμμετέχουν σε αυτόν. Το γεγονός ότι δύο κατηγορίες προϊόντων εμφανίζονται συχνά μαζί δεν σημαίνει ότι η αγορά της μίας προκαλεί την αγορά της άλλης. Ο κανόνας εκφράζει μια στατιστική συσχέτιση και όχι σχέση αιτιότητας. Η απόκλιση της σχέσης από τη στατιστική ανεξαρτησία αξιολογείται, μεταξύ άλλων, μέσω της μετρικής *lift*, η οποία παρουσιάζεται στην επόμενη ενότητα [16].

Η αξία των κανόνων συσχέτισης έγκειται στην ικανότητά τους να μετατρέπουν μεγάλους όγκους δεδομένων συναλλαγών σε αξιοποιήσιμη γνώση σχετικά με τις αγοραστικές συνήθειες των καταναλωτών. Μέσω της ανάλυσης των κανόνων αυτών είναι δυνατό να εντοπιστούν πρότυπα κατανάλωσης, σχέσεις μεταξύ προϊόντων αλλά και μεταβολές στην καταναλωτική συμπεριφορά, στοιχεία που μπορούν να αξιοποιηθούν τόσο σε επιχειρηματικές όσο και σε ερευνητικές εφαρμογές.

Ωστόσο, δεν έχουν όλοι οι κανόνες την ίδια σημασία ή αξιοπιστία. Κατά συνέπεια, χρησιμοποιούνται συγκεκριμένες μετρικές αξιολόγησης, οι οποίες επιτρέπουν την ποσοτική εκτίμηση της ισχύος και της χρησιμότητας κάθε κανόνα. Οι σημαντικότερες από αυτές είναι το *support*, το *confidence*, το *lift*, το *leverage* και το *conviction*, τα οποία παρουσιάζονται αναλυτικά στην επόμενη ενότητα.

2.4 Μετρικές Αξιολόγησης Κανόνων Συσχέτισης

Η εξαγωγή των κανόνων συσχέτισης από ένα σύνολο δεδομένων μπορεί να οδηγήσει στη δημιουργία μεγάλου αριθμού πιθανών κανόνων. Παρά το γεγονός αυτό, δεν παρουσιάζουν όλοι οι κανόνες την ίδια σημασία ή χρησιμότητα. Ως αποτέλεσμα χρησιμοποιούνται συγκεκριμένες μετρικές αξιολόγησης, οι οποίες επιτρέπουν την ποσοτική εκτίμηση της ισχύος, της αξιοπιστίας και της σημαντικότητας των κανόνων που παράγονται. Κατά τους συγγραφείς Malliaridis, Ougiaroglou και Dervos [19], οι σημαντικότερες μετρικές που χρησιμοποιούνται στην εξόρυξη κανόνων συσχέτισης είναι το *support*, το *confidence*, το *lift*, το *leverage* και το *conviction*.

2.4.1 Support

Η υποστήριξη (*support*) εκφράζει τη συχνότητα εμφάνισης ενός συνόλου αντικειμένων ή ενός κανόνα στο σύνολο των συναλλαγών. Πρόκειται για μία από τις σημαντικότερες μετρικές της εξόρυξης κανόνων συσχέτισης, καθώς χρησιμοποιείται για τον εντοπισμό των συχνών συνόλων αντικειμένων (*frequent itemsets*) και για την απομάκρυνση σπάνιων συνδυασμών προϊόντων που ενδέχεται να μην έχουν πρακτική σημασία [19, 20].

Για έναν κανόνα της μορφής:

$$X \rightarrow Y$$

το support υπολογίζεται ως:

$$Support(X \rightarrow Y) = P(X \cup Y)$$

όπου $P(X \cup Y)$ είναι η πιθανότητα να εμφανίζονται ταυτόχρονα τα σύνολα X και Y σε μία συναλλαγή.

Υψηλές τιμές support υποδηλώνουν ότι ο κανόνας εμφανίζεται συχνά στο σύνολο των δεδομένων και επομένως αφορά σημαντικό τμήμα των συναλλαγών. Αντίθετα, χαμηλές τιμές support υποδηλώνουν ότι ο κανόνας εμφανίζεται σε περιορισμένο αριθμό συναλλαγών και ενδέχεται να μην είναι αντιπροσωπευτικός του συνολικού πληθυσμού των δεδομένων [20].

2.4.2 Confidence

Η εμπιστοσύνη (*confidence*) εκφράζει την πιθανότητα εμφάνισης του επακόλουθου συνόλου Y όταν έχει ήδη εμφανιστεί το πρόδρομο σύνολο X . Η μετρική αυτή χρησιμοποιείται για την αξιολόγηση της αξιοπιστίας ενός κανόνα συσχέτισης και δείχνει πόσο συχνά επαληθεύεται ο κανόνας στα δεδομένα [19].

Για έναν κανόνα της μορφής:

$$X \rightarrow Y$$

το confidence υπολογίζεται ως:

$$Confidence(X \rightarrow Y) = \frac{Support(X \cup Y)}{Support(X)}$$

Η μετρική αυτή εκφράζει το ποσοστό των συναλλαγών που περιέχουν το σύνολο X και περιέχουν ταυτόχρονα και το σύνολο Y . Υψηλές τιμές confidence υποδηλώνουν ότι όταν εμφανίζεται το antecedent, το consequent εμφανίζεται επίσης με μεγάλη συχνότητα. Ωστόσο, υψηλό confidence δεν συνεπάγεται απαραίτητα ισχυρή συσχέτιση μεταξύ των δύο συνόλων αντικειμένων, καθώς η τιμή του μπορεί να επηρεάζεται από τη συνολική συχνότητα εμφάνισης του consequent [19].

2.4.3 Lift

Το lift αποτελεί μία από τις σημαντικότερες μετρικές αξιολόγησης κανόνων συσχέτισης, καθώς εξετάζει κατά πόσο η ταυτόχρονη εμφάνιση δύο συνόλων αντικειμένων είναι μεγαλύτερη από αυτή που θα αναμενόταν εάν τα δύο σύνολα ήταν στατιστικά ανεξάρτητα [19].

Ο υπολογισμός του πραγματοποιείται ως εξής:

$$Lift(X \rightarrow Y) = \frac{Confidence(X \rightarrow Y)}{Support(Y)}$$

Η τιμή του lift επιτρέπει την αξιολόγηση της πραγματικής ισχύος της σχέσης μεταξύ των δύο συνόλων αντικειμένων. Η ερμηνεία της μετρικής έχει ως εξής:

- $Lift > 1$: θετική συσχέτιση μεταξύ των συνόλων αντικειμένων.

- Lift = 1: στατιστική ανεξαρτησία μεταξύ των συνόλων.
- Lift < 1: αρνητική συσχέτιση μεταξύ των συνόλων.

Έτσι το lift χρησιμοποιείται συχνά ως βασικό κριτήριο επιλογής ισχυρών κανόνων συσχέτισης, καθώς επιτρέπει τη διάκριση μεταξύ πραγματικών σχέσεων και τυχαίων συνυπάρξεων προϊόντων [19].

2.4.4 Leverage

Το leverage μετρά τη διαφορά μεταξύ της πραγματικής συχνότητας συνύπαρξης δύο συνόλων αντικειμένων και της συχνότητας που θα αναμενόταν εάν τα δύο σύνολα ήταν στατιστικά ανεξάρτητα [19].

Ορίζεται ως:

$$Leverage(X \rightarrow Y) = Support(X \cup Y) - Support(X) \times Support(Y)$$

Η συγκεκριμένη μετρική επιτρέπει την αξιολόγηση του πραγματικού βαθμού συσχέτισης μεταξύ δύο συνόλων αντικειμένων. Θετικές τιμές leverage υποδηλώνουν ότι τα προϊόντα εμφανίζονται μαζί συχνότερα από το αναμενόμενο, ενώ τιμές κοντά στο μηδέν υποδηλώνουν ασθενή ή μηδενική συσχέτιση.

Σε αντίθεση με το lift, το leverage αποτυπώνει την πραγματική διαφορά μεταξύ της παρατηρούμενης και της αναμενόμενης συχνότητας εμφάνισης και συνεπώς παρέχει μία πιο άμεση εκτίμηση της πρακτικής σημασίας ενός κανόνα. Στην έρευνα αυτή το leverage χρησιμοποιήθηκε ως ένα από τα βασικά κριτήρια φιλτραρίσματος των κανόνων συσχέτισης, συμβάλλοντας στον περιορισμό κανόνων που εμφανίζουν υψηλό confidence αλλά περιορισμένη πραγματική συσχέτιση [19].

2.4.5 Conviction

Το conviction αποτελεί μία επιπλέον μετρική αξιολόγησης η οποία εξετάζει τον βαθμό εξάρτησης μεταξύ antecedent και consequent. Σε αντίθεση με το confidence, το conviction λαμβάνει υπόψη και τις περιπτώσεις στις οποίες ο κανόνας δεν επαληθεύεται, προσφέροντας μία διαφορετική οπτική στην αξιολόγηση της αξιοπιστίας ενός κανόνα [19].

Υπολογίζεται ως:

$$Conviction(X \rightarrow Y) = \frac{1 - Support(Y)}{1 - Confidence(X \rightarrow Y)}$$

Όσο μεγαλύτερη είναι η τιμή του conviction, τόσο ισχυρότερη θεωρείται η σχέση μεταξύ των δύο συνόλων αντικειμένων. Τιμή conviction ίση με 1 υποδηλώνει στατιστική ανεξαρτησία μεταξύ antecedent και consequent, ενώ μεγαλύτερες τιμές υποδηλώνουν αυξημένο βαθμό εξάρτησης.

Η μετρική conviction χρησιμοποιείται συχνά συμπληρωματικά προς το confidence και το lift, καθώς λαμβάνει υπόψη όχι μόνο τις επιτυχημένες αλλά και τις αποτυχημένες εμφανίσεις ενός κανόνα. Με τον τρόπο αυτό συμβάλλει σε μία πιο ολοκληρωμένη αξιολόγηση της αξιοπιστίας των κανόνων συσχέτισης [19].

2.4.6 Σύγκριση των μετρικών

Οι μετρικές αξιολόγησης των κανόνων συσχέτισης εξετάζουν διαφορετικές πτυχές της σχέσης μεταξύ του antecedent και του consequent με αποτέλεσμα η καθεμία να παρέχει διαφορετική πληροφορία σχετικά με την ποιότητα ενός κανόνα. Για τον λόγο αυτό, η αξιολόγηση των κανόνων δεν βασίζεται συνήθως σε μία μόνο μετρική αλλά σε συνδυασμό πολλών, ώστε να εξασφαλίζεται η επιλογή κανόνων που είναι ταυτόχρονα συχνοί, αξιόπιστοι και ουσιαστικοί [19].

Το support αποτελεί τη βασικότερη μετρική συχνότητας, καθώς αποτυπώνει το ποσοστό των συναλλαγών στις οποίες εμφανίζεται ένας κανόνας. Παρότι είναι απαραίτητο για τον εντοπισμό των frequent itemsets, δεν παρέχει πληροφορίες σχετικά με την ισχύ της σχέσης μεταξύ των αντικειμένων. Ένας κανόνας μπορεί να παρουσιάζει υψηλό support απλώς επειδή τα προϊόντα που συμμετέχουν είναι ιδιαίτερα δημοφιλή στο σύνολο των συναλλαγών.

Το confidence αξιολογεί την αξιοπιστία ενός κανόνα εκφράζοντας την πιθανότητα εμφάνισης του consequent όταν έχει ήδη εμφανιστεί το antecedent. Η συγκεκριμένη μετρική όμως επηρεάζεται σημαντικά από τη συνολική συχνότητα εμφάνισης του consequent. Αυτό μπορεί να οδηγήσει σε υψηλές τιμές ακόμη και όταν η πραγματική συσχέτιση μεταξύ των αντικειμένων είναι περιορισμένη.

Για την αντιμετώπιση αυτού του περιορισμού χρησιμοποιείται το lift, το οποίο συγκρίνει την πραγματική συνύπαρξη δύο συνόλων αντικειμένων με την αναμενόμενη συνύπαρξή τους υπό συνθήκες στατιστικής ανεξαρτησίας. Το lift θεωρείται ιδιαίτερα χρήσιμο για τον εντοπισμό ουσιαστικών συσχετίσεων, καθώς επιτρέπει τη διάκριση μεταξύ πραγματικών σχέσεων και τυχαίων συνυπάξεων προϊόντων.

Το leverage παρέχει μία διαφορετική προσέγγιση αξιολόγησης, καθώς μετρά την πραγματική διαφορά μεταξύ της παρατηρούμενης και της αναμενόμενης συχνότητας συνύπαρξης δύο συνόλων αντικειμένων. Η μετρική αυτή είναι ιδιαίτερα χρήσιμη σε εφαρμογές λιανεμπορίου, καθώς αποτυπώνει με πιο άμεσο τρόπο τη συνεισφορά ενός κανόνα και βοηθά στον εντοπισμό σχέσεων με μεγαλύτερη πρακτική σημασία.

Τέλος, το conviction εστιάζει στις περιπτώσεις όπου ένας κανόνας δεν επαληθεύεται και αξιολογεί τον βαθμό εξάρτησης μεταξύ antecedent και consequent. Η συγκεκριμένη μετρική παρέχει συμπληρωματική πληροφορία σε σχέση με το confidence και το lift, επιτρέποντας μία πιο ολοκληρωμένη εκτίμηση της αξιοπιστίας ενός κανόνα [19].

Συνεπώς, καμία μετρική δεν μπορεί να θεωρηθεί επαρκής από μόνη της για την πλήρη αξιολόγηση των κανόνων συσχέτισης. Η συνδυαστική χρήση των support, confidence, lift, leverage και conviction επιτρέπει την εξαγωγή κανόνων που παρουσιάζουν όχι μόνο υψηλή συχνότητα εμφάνισης αλλά και ουσιαστική στατιστική και πρακτική σημασία. Με βάση όλα τα παραπάνω, στη σύγχρονη βιβλιογραφία αλλά και στην παρούσα μελέτη, η επιλογή των κανόνων βασίζεται στον συνδυασμό πολλαπλών μετρικών και όχι μόνο σε ένα κριτήριο αξιολόγησης.

Πίνακας 2.1: Σύγκριση των βασικών μετρικών αξιολόγησης κανόνων συσχέτισης

Μετρική	Πλεονέκτημα	Περιορισμός
Support	Μετρά τη συχνότητα εμφάνισης	Δεν αξιολογεί τη δύναμη της σχέσης
Confidence	Μετρά την αξιοπιστία του κανόνα	Επηρεάζεται από δημοφιλή προϊόντα
Lift	Εντοπίζει πραγματικές συσχετίσεις	Μπορεί να ευνοεί σπάνιους κανόνες
Leverage	Αποτυπώνει την πραγματική συνεισφορά της σχέσης	Δυσκολότερη ερμηνεία
Conviction	Αξιολογεί τον βαθμό εξάρτησης	Λιγότερο διαισθητική ερμηνεία

2.5 Αλγόριθμοι Εξόρυξης Συχνών Στοιχειοσυνόλων

Η εξόρυξη συχνών στοιχειοσυνόλων (Frequent Itemset Mining – FIM) αποτελεί μία από τις βασικότερες διαδικασίες εξόρυξης δεδομένων και το πρώτο στάδιο για την παραγωγή κανόνων συσχέτισης. Στόχος της είναι ο εντοπισμός ομάδων αντικειμένων που εμφανίζονται συχνά μαζί σε ένα σύνολο συναλλαγών. Τα συχνά αυτά στοιχειοσύνολα αποτελούν τη βάση για την ανακάλυψη σχέσεων μεταξύ προϊόντων και χρησιμοποιούνται ευρέως σε εφαρμογές όπως η ανάλυση του καλαθιού αγορών (Market Basket Analysis), τα συστήματα συστάσεων και η ανάλυση καταναλωτικής συμπεριφοράς [20].

Καθώς ο όγκος των διαθέσιμων δεδομένων αυξάνεται συνεχώς, η αποτελεσματική εξόρυξη συχνών στοιχειοσυνόλων γίνεται ολοένα και πιο απαιτητική. Για αυτό έχουν αναπτυχθεί διάφοροι αλγόριθμοι οι οποίοι έχουν ως στόχο τη μείωση του υπολογιστικού κόστους και τη βελτίωση της αποδοτικότητας της διαδικασίας. Οι αλγόριθμοι αυτοί μπορούν να ταξινομηθούν σε δύο βασικές κατηγορίες: τους αλγορίθμους που βασίζονται σε ένα προκαθορισμένο ελάχιστο όριο υποστήριξης (minimum support) και τους αλγορίθμους Top-K, οι οποίοι εξάγουν έναν προκαθορισμένο αριθμό συχνών στοιχειοσυνόλων χωρίς να απαιτείται ο καθορισμός συγκεκριμένης τιμής support [20].

Αλγόριθμοι Βασισμένοι στο Ελάχιστο Όριο Υποστήριξης (Min Support Based)

Οι αλγόριθμοι που βασίζονται στο ελάχιστο όριο υποστήριξης αποτελούν την πλέον διαδεδομένη κατηγορία τεχνικών εξόρυξης συχνών στοιχειοσυνόλων. Στην προσέγγιση αυτή, ο αναλυτής ορίζει εκ των προτέρων μία ελάχιστη τιμή υποστήριξης (minimum support threshold), η οποία χρησιμοποιείται ως κριτήριο για τον διαχωρισμό των συχνών από τα μη συχνά στοιχειοσύνολα. Όσα στοιχειοσύνολα εμφανίζονται με συχνότητα μεγαλύτερη ή ίση από το συγκεκριμένο όριο θεωρούνται σημαντικά και διατηρούνται για περαιτέρω επεξεργασία, ενώ τα υπόλοιπα απορρίπτονται [20].

Η συγκεκριμένη προσέγγιση έχει κυριαρχήσει στη βιβλιογραφία για περισσότερες από τρεις δεκαετίες και αποτελεί τη βάση λειτουργίας πολλών γνωστών αλγορίθμων εξόρυξης δεδομένων. Χαρακτηριστικά παραδείγματα αποτελούν οι αλγόριθμοι Apriori, FP-Growth και Eclat, οι οποίοι παρουσιάζονται αναλυτικά στις επόμενες ενότητες. Παρότι οι αλγόριθμοι αυτοί διαφέρουν ως προς τον τρόπο αναζήτησης των συχνών στοιχειοσυνόλων, όλοι βασίζονται στη χρήση ενός προκαθορισμένου

κατωφλίου support για τον περιορισμό του χώρου αναζήτησης και τη βελτίωση της αποδοτικότητας της διαδικασίας [20].

Ένα από τα σημαντικότερα πλεονεκτήματα της συγκεκριμένης κατηγορίας αλγορίθμων είναι η δυνατότητα ελέγχου του πλήθους των παραγόμενων αποτελεσμάτων μέσω της παραμέτρου support. Παρόλα αυτά, η επιλογή της κατάλληλης τιμής δεν είναι πάντοτε εύκολη. Πολύ υψηλές τιμές ενδέχεται να αποκλείσουν ενδιαφέροντα πρότυπα, ενώ πολύ χαμηλές τιμές μπορεί να οδηγήσουν στην παραγωγή εξαιρετικά μεγάλου αριθμού στοιχειοσυνόλων, αυξάνοντας σημαντικά το υπολογιστικό κόστος και τις απαιτήσεις σε μνήμη [20].

Top-K

Οι αλγόριθμοι Top-K αναπτύχθηκαν με στόχο την αντιμετώπιση των δυσκολιών που σχετίζονται με την επιλογή κατάλληλης τιμής support. Στην περίπτωση αυτή, ο χρήστης δεν καθορίζει ένα ελάχιστο όριο υποστήριξης αλλά τον αριθμό των επιθυμητών αποτελεσμάτων. Ο αλγόριθμος επιστρέφει τα K σημαντικότερα ή συχνότερα στοιχειοσύνολα, προσαρμόζοντας δυναμικά τα εσωτερικά κριτήρια αναζήτησης [20].

Η συγκεκριμένη προσέγγιση θεωρείται ιδιαίτερα χρήσιμη σε μεγάλα σύνολα δεδομένων, όπου η επιλογή κατάλληλης τιμής support απαιτεί συνήθως πολλαπλές δοκιμές και πειραματισμούς. Τα τελευταία χρόνια έχουν προταθεί διάφορες τεχνικές Top-K, όπως οι TKFIM, BOMO και TFP, οι οποίες στοχεύουν στη βελτίωση της αποδοτικότητας της διαδικασίας και στη μείωση του αριθμού των παραγόμενων αποτελεσμάτων [20].

Σύμφωνα με τους Malliaridis και Ougiaroglou [20], οι Top-K προσεγγίσεις έχουν προσελκύσει αυξανόμενο ερευνητικό ενδιαφέρον, καθώς προσφέρουν μεγαλύτερη ευελιξία και επιτρέπουν την εξαγωγή χρήσιμης γνώσης χωρίς να απαιτείται εκ των προτέρων γνώση της κατάλληλης τιμής support. Παρά τα πλεονεκτήματά τους, οι περισσότεροι αλγόριθμοι που χρησιμοποιούνται στην πράξη εξακολουθούν να βασίζονται στην κλασική προσέγγιση του minimum support, γεγονός που εξηγεί τη μεγάλη διάδοση αλγορίθμων όπως ο Apriori, ο FP-Growth και ο Eclat.

2.5.1 Αλγόριθμος Apriori

Ο αλγόριθμος Apriori αποτελεί έναν από τους πιο γνωστούς και ευρέως χρησιμοποιούμενους αλγορίθμους εξόρυξης συχνών στοιχειοσυνόλων και κανόνων συσχέτισης. Παρουσιάστηκε από τους Agrawal και Srikant το 1994 και αποτέλεσε μία από τις πρώτες αποδοτικές προσεγγίσεις για την ανακάλυψη των συχνών στοιχειοσυνόλων σε μεγάλες βάσεις δεδομένων συναλλαγών [18]. Η δημοσίευσή του θεωρείται ορόσημο στον χώρο της εξόρυξης δεδομένων, καθώς έθεσε τις βάσεις για την ανάπτυξη μεταγενέστερων αλγορίθμων εξόρυξης συχνών στοιχειοσυνόλων.

Η βασική ιδέα του Apriori στηρίζεται στην παρατήρηση ότι εάν ένα σύνολο αντικειμένων είναι συχνό, τότε όλα τα υποσύνολά του θα πρέπει επίσης να είναι συχνά. Η ιδιότητα αυτή είναι γνωστή ως Apriori Principle ή Apriori Property και χρησιμοποιείται για τη σημαντική μείωση του χώρου αναζήτησης κατά τη διάρκεια της εξόρυξης [18].

Πιο συγκεκριμένα, εάν ένα στοιχειοσύνολο δεν ικανοποιεί το ελάχιστο όριο υποστήριξης, τότε κανένα υπερσύνολό του δεν μπορεί να θεωρηθεί συχνό. Η ιδιότητα αυτή επιτρέπει στον αλγόριθμο να απορρίπτει μεγάλο αριθμό υποψήφιων συνδυασμών χωρίς να απαιτείται η αξιολόγησή τους, περιορίζοντας έτσι σημαντικά το υπολογιστικό κόστος της διαδικασίας.

Η λειτουργία του Apriori πραγματοποιείται επαναληπτικά και βασίζεται σε δύο κύρια στάδια: τη δημιουργία υποψήφιων στοιχειοσυνόλων (candidate generation) και τη διαδικασία περικοπής (pruning). Αρχικά εντοπίζονται όλα τα συχνά στοιχειοσύνολα μεγέθους ενός αντικειμένου (1-itemsets). Στη συνέχεια, συνδυάζονται μεταξύ τους για τη δημιουργία υποψήφιων στοιχειοσυνόλων μεγαλύτερου μεγέθους. Μετά από κάθε επανάληψη, υπολογίζεται το support των νέων υποψηφίων συνόλων και διατηρούνται μόνο όσα ικανοποιούν το προκαθορισμένο κατώφλι υποστήριξης [18].

Η διαδικασία συνεχίζεται μέχρι να μην είναι δυνατή η παραγωγή νέων συχνών στοιχειοσυνόλων. Με τον τρόπο αυτό ο αλγόριθμος εξερευνά σταδιακά τον χώρο αναζήτησης ξεκινώντας από τα απλούστερα σύνολα και προχωρώντας σε ολοένα μεγαλύτερους συνδυασμούς αντικειμένων.

Για την καλύτερη κατανόηση της λειτουργίας του αλγορίθμου, ας θεωρήσουμε ότι τα προϊόντα A, B και Γ εμφανίζονται συχνά στις συναλλαγές. Εάν το στοιχειοσύνολο A, B δεν θεωρείται συχνό, τότε σύμφωνα με την Apriori Property κανένα στοιχειοσύνολο που περιέχει ταυτόχρονα τα A και B, όπως τα A, B, Γ ή A, B, Δ, δεν χρειάζεται να εξεταστεί. Η δυνατότητα αυτή συμβάλλει σημαντικά στη μείωση του αριθμού των συνδυασμών που πρέπει να αξιολογηθούν.

Παρά τη μεγάλη διάδοσή του, ο Apriori παρουσιάζει ορισμένους περιορισμούς. Ο σημαντικότερος σχετίζεται με τον μεγάλο αριθμό υποψήφιων στοιχειοσυνόλων που δημιουργούνται κατά τη διαδικασία αναζήτησης. Καθώς αυξάνεται το πλήθος των αντικειμένων ή μειώνεται η τιμή του support, ο αριθμός των υποψήφιων συνδυασμών μπορεί να αυξηθεί εκθετικά, οδηγώντας σε σημαντική επιβάρυνση τόσο του χρόνου εκτέλεσης όσο και της απαιτούμενης μνήμης [20].

Επίσης, ο αλγόριθμος απαιτεί πολλαπλές διελεύσεις (database scans) πάνω στο σύνολο των δεδομένων προκειμένου να υπολογίσει το support των υποψήφιων στοιχειοσυνόλων. Το χαρακτηριστικό αυτό καθιστά τον Apriori λιγότερο αποδοτικό σε ιδιαίτερα μεγάλα σύνολα δεδομένων συγκριτικά με νεότερες προσεγγίσεις, όπως ο FP-Growth και ο Eclat, οι οποίοι αναπτύχθηκαν ακριβώς για την αντιμετώπιση αυτών των περιορισμών [20].

Παρά τα μειονεκτήματά του, ο Apriori εξακολουθεί να χρησιμοποιείται ευρέως τόσο στην έρευνα όσο και στη βιομηχανία, κυρίως λόγω της απλότητας, της ευκολίας κατανόησης και της υψηλής ερμηνευσιμότητας των αποτελεσμάτων του. Επιπλέον, αποτελεί σημείο αναφοράς για τη σύγκριση νέων αλγορίθμων εξόρυξης συχνών στοιχειοσυνόλων και συχνά χρησιμοποιείται ως αλγόριθμος αναφοράς (baseline algorithm) σε πειραματικές μελέτες [20].

Στην ανάλυση που ακολουθεί ο Apriori χρησιμοποιείται ως ο βασικός αλγόριθμος εξόρυξης κανόνων συσχέτισης για την ανάλυση των δεδομένων των δύο εξεταζόμενων περιόδων. Η επιλογή του πραγματοποιήθηκε λόγω της ευρείας αποδοχής του στη βιβλιογραφία, της δυνατότητας εύκολης ερμηνείας των αποτελεσμάτων και της καταλληλότητάς του για τη διερεύνηση μεταβολών στην καταναλωτική συμπεριφορά μέσω κανόνων συσχέτισης.

2.5.2 Αλγόριθμος FP-Growth

Ο αλγόριθμος FP-Growth (Frequent Pattern Growth) προτάθηκε από τους Han, Pei και Yin το 2000 ως μία πιο αποδοτική εναλλακτική προσέγγιση στην εξόρυξη συχνών στοιχειοσυνόλων σε σχέση με τον Apriori [21]. Ο βασικός στόχος της ανάπτυξής του ήταν η αντιμετώπιση των σημαντικότερων περιορισμών του Apriori, οι οποίοι σχετίζονται με τη δημιουργία μεγάλου αριθμού υποψήφιων στοιχειοσυνόλων και τις επαναλαμβανόμενες διελεύσεις πάνω στα δεδομένα.

Σε αντίθεση με τον Apriori, ο FP-Growth δεν βασίζεται στη διαδικασία παραγωγής υποψήφιων

συνδυασμών (candidate generation). Αντίθετα, χρησιμοποιεί μία ειδική δομή δεδομένων γνωστή ως FP-Tree (Frequent Pattern Tree), η οποία επιτρέπει τη συμπίεσμένη αναπαράσταση των συναλλαγών και τη σημαντική μείωση του χώρου αναζήτησης [21].

Η λειτουργία του αλγορίθμου πραγματοποιείται σε δύο βασικά στάδια. Αρχικά πραγματοποιείται μία σάρωση του συνόλου δεδομένων προκειμένου να υπολογιστεί η συχνότητα εμφάνισης κάθε αντικειμένου και να απορριφθούν όσα δεν ικανοποιούν το ελάχιστο όριο υποστήριξης. Στη συνέχεια κατασκευάζεται το FP-Tree, στο οποίο αποθηκεύονται οι συναλλαγές με τρόπο που επιτρέπει την κοινή αναπαράσταση επαναλαμβανόμενων προτύπων. Με τον τρόπο αυτό μειώνεται σημαντικά ο όγκος των δεδομένων που πρέπει να επεξεργαστεί ο αλγόριθμος [21].

Αφού ολοκληρωθεί η κατασκευή του FP-Tree, ο αλγόριθμος εφαρμόζει μία διαδικασία αναδρομικής εξόρυξης γνωστή ως pattern growth. Αντί να δημιουργεί υποψήφια στοιχειοσύνολα και να υπολογίζει επαναληπτικά το support τους, ο FP-Growth αναζητά απευθείας συχνά πρότυπα μέσα στο δέντρο. Η προσέγγιση αυτή επιτρέπει την εξαγωγή συχνών στοιχειοσυνόλων με μικρότερο υπολογιστικό κόστος σε σχέση με τον Apriori [21].

Ένα από τα σημαντικότερα πλεονεκτήματα του FP-Growth είναι η υψηλή αποδοτικότητά του σε μεγάλα σύνολα δεδομένων. Επειδή αποφεύγει τη δημιουργία μεγάλου αριθμού υποψήφιων στοιχειοσυνόλων, απαιτεί συνήθως λιγότερο χρόνο εκτέλεσης και λιγότερες προσπελάσεις της βάσης δεδομένων. Ως αποτέλεσμα θεωρείται ένας από τους πλέον αποδοτικούς αλγορίθμους εξόρυξης συχνών στοιχειοσυνόλων και χρησιμοποιείται ευρέως σε εφαρμογές λιανεμπορίου και ανάλυσης καταναλωτικής συμπεριφοράς [20].

Παρά τα πλεονεκτήματά του, ο FP-Growth παρουσιάζει και ορισμένες προκλήσεις. Η κατασκευή και η διαχείριση του FP-Tree μπορεί να καταστεί απαιτητική σε περιπτώσεις όπου τα δεδομένα είναι ιδιαίτερα αραιά ή περιέχουν μεγάλο αριθμό διαφορετικών αντικειμένων. Επίσης, η δομή του αλγορίθμου είναι πιο σύνθετη συγκριτικά με τον Apriori καθιστώντας δυσκολότερη την υλοποίηση αλλά και την κατανόησή του [20].

Εν κατακλείδι, ο FP-Growth θεωρείται μία από τις σημαντικότερες εξελίξεις στον χώρο της εξόρυξης συχνών στοιχειοσυνόλων, καθώς επιτυγχάνει σημαντική βελτίωση της αποδοτικότητας χωρίς να θυσιάζει την ποιότητα των αποτελεσμάτων. Για αυτόν τον λόγο χρησιμοποιείται συχνά ως σημείο αναφοράς σε συγκριτικές μελέτες αλγορίθμων εξόρυξης δεδομένων και συγκαταλέγεται μεταξύ των πλέον δημοφιλών τεχνικών Frequent Itemset Mining [20].

2.5.3 Αλγόριθμος Eclat

Ο αλγόριθμος Eclat (Equivalence Class Clustering and Bottom-Up Lattice Traversal) προτάθηκε από τον Zaki το 2000 και αποτελεί μία από τις σημαντικότερες εναλλακτικές προσεγγίσεις για την εξόρυξη συχνών στοιχειοσυνόλων [22]. Σε αντίθεση με τους περισσότερους κλασικούς αλγορίθμους που χρησιμοποιούν οριζόντια αναπαράσταση των δεδομένων, ο Eclat βασίζεται σε μία κάθετη αναπαράσταση (vertical data format), η οποία επιτρέπει την αποδοτικότερη αναζήτηση συχνών στοιχειοσυνόλων σε μεγάλες βάσεις δεδομένων.

Η βασική ιδέα του αλγορίθμου είναι ότι κάθε αντικείμενο συνδέεται με μία λίστα αναγνωριστικών συναλλαγών (Transaction ID List – TID List), η οποία περιλαμβάνει όλες τις συναλλαγές στις οποίες εμφανίζεται το συγκεκριμένο αντικείμενο. Αντί να πραγματοποιεί επαναλαμβανόμενες σαρώσεις της βάσης δεδομένων για τον υπολογισμό του support, ο Eclat υπολογίζει τη συχνότητα εμ-

φάνισης των στοιχειοσυνόλων μέσω πράξεων τομής (set intersections) μεταξύ των αντίστοιχων TID Lists [22].

Για παράδειγμα, εάν το αντικείμενο A εμφανίζεται στις συναλλαγές 1, 2, 5, 8 και το αντικείμενο B στις συναλλαγές 2, 5, 7, 8, τότε το στοιχειοσύνολο A, B εμφανίζεται στις συναλλαγές 2, 5, 8. Η συχνότητα εμφάνισής του μπορεί να υπολογιστεί άμεσα από το πλήθος των κοινών στοιχείων των δύο λιστών, χωρίς να απαιτείται νέα σάρωση της βάσης δεδομένων.

Η διαδικασία εξόρυξης πραγματοποιείται αναδρομικά και ακολουθεί μία στρατηγική αναζήτησης βάθους (depth-first search). Ξεκινώντας από τα συχνά στοιχειοσύνολα ενός αντικειμένου, ο αλγόριθμος επεκτείνει σταδιακά τα σύνολα μέσω τομών των αντίστοιχων λιστών συναλλαγών, δημιουργώντας μεγαλύτερα συχνά στοιχειοσύνολα όταν αυτά ικανοποιούν το προκαθορισμένο όριο υποστήριξης [22].

Ένα από τα σημαντικότερα πλεονεκτήματα του Eclat είναι η υψηλή αποδοτικότητά του σε περιπτώσεις όπου οι TID Lists μπορούν να αποθηκευτούν αποτελεσματικά στη μνήμη. Επειδή ο υπολογισμός του support πραγματοποιείται μέσω πράξεων τομής και όχι μέσω επαναλαμβανόμενων σαρώσεων της βάσης δεδομένων, ο αλγόριθμος μπορεί να επιτύχει ιδιαίτερα γρήγορους χρόνους εκτέλεσης. Για αυτό θεωρείται μία από τις πλέον αποδοτικές τεχνικές εξόρυξης συχνών στοιχειοσυνόλων και χρησιμοποιείται συχνά σε συγκριτικές μελέτες μαζί με τους Apriori και FP-Growth [20].

Η απόδοση του Eclat εξαρτάται σημαντικά από το μέγεθος των λιστών συναλλαγών. Σε περιπτώσεις όπου τα δεδομένα είναι ιδιαίτερα πυκνά ή ο αριθμός των συχνών στοιχειοσυνόλων είναι πολύ μεγάλος, οι TID Lists μπορεί να καταλαμβάνουν σημαντικό χώρο μνήμης, επηρεάζοντας αρνητικά την αποδοτικότητα του αλγορίθμου. Αξίζει να αναφερθεί ότι η κάθετη αναπαράσταση των δεδομένων είναι λιγότερο διαισθητική σε σχέση με τις κλασικές οριζόντιες προσεγγίσεις αυξάνοντας τη σχετική πολυπλοκότητα υλοποίησης [20].

Παρά τους περιορισμούς αυτούς, ο Eclat αποτελεί μία ιδιαίτερα αποτελεσματική τεχνική εξόρυξης συχνών στοιχειοσυνόλων και συγκαταλέγεται μεταξύ των πλέον διαδεδομένων αλγορίθμων του πεδίου. Η χρήση κάθετης αναπαράστασης και πράξεων τομής του επιτρέπει να επιτυγχάνει υψηλές επιδόσεις σε πολλές κατηγορίες προβλημάτων, καθιστώντας ιδανικό για συγκριτική αξιολόγηση με άλλους κλασικούς αλγορίθμους εξόρυξης δεδομένων [20].

Στην παρούσα έρευνα ο Eclat χρησιμοποιείται στο πλαίσιο της συγκριτικής αξιολόγησης αλγορίθμων εξόρυξης συχνών στοιχειοσυνόλων. Η απόδοσή του συγκρίνεται με τους Apriori και FP-Growth ως προς τον χρόνο εκτέλεσης, την κατανάλωση μνήμης και την ικανότητα διαχείρισης μεγάλων συνόλων δεδομένων λιανεμπορίου.

2.5.4 Άλλοι Αλγόριθμοι Εξόρυξης Συχνών Στοιχειοσυνόλων

Η συνεχής εξέλιξη του πεδίου της εξόρυξης συχνών στοιχειοσυνόλων έχει οδηγήσει στην ανάπτυξη μεγάλου αριθμού αλγορίθμων, οι οποίοι έχουν ως στόχο τη βελτίωση της αποδοτικότητας της διαδικασίας εξόρυξης ή την αντιμετώπιση συγκεκριμένων περιορισμών των κλασικών προσεγγίσεων. Παρότι οι Apriori, FP-Growth και Eclat αποτελούν τους πλέον διαδεδομένους αλγορίθμους της βιβλιογραφίας, έχουν προταθεί και πολλές άλλες τεχνικές που αξιοποιούνται σε εξειδικευμένες εφαρμογές [20].

Μία χαρακτηριστική κατηγορία αποτελείται από αλγορίθμους που επεκτείνουν τη λογική του FP-Growth. Ενδεικτικό παράδειγμα αποτελεί ο FP-Max, ο οποίος επικεντρώνεται στην εξαγωγή

μόνο των μέγιστων συχνών στοιχειοσυνόλων (maximal frequent itemsets), μειώνοντας σημαντικά τον αριθμό των αποτελεσμάτων που παράγονται. Με τον τρόπο αυτό επιτυγχάνεται ταχύτερη επεξεργασία και απλούστερη ανάλυση των προτύπων που εντοπίζονται στα δεδομένα [20].

Μία ακόμη γνωστή προσέγγιση είναι ο H-Mine, ο οποίος σχεδιάστηκε για την αποδοτική διαχείριση μεγάλων και αραιών συνόλων δεδομένων. Ο αλγόριθμος χρησιμοποιεί ειδικές δομές αποθήκευσης που επιτρέπουν τη μείωση των απαιτήσεων σε μνήμη, διατηρώντας ταυτόχρονα υψηλή αποδοτικότητα κατά τη διαδικασία εξόρυξης [20].

Ακόμη, έχουν αναπτυχθεί αλγόριθμοι που βασίζονται στη λογική των Top-K προσεγγίσεων, όπως οι TKFIM, TFP και BOMO. Οι τεχνικές αυτές στοχεύουν στην εξαγωγή ενός προκαθορισμένου αριθμού σημαντικών στοιχειοσυνόλων χωρίς να απαιτείται ο καθορισμός ελάχιστου ορίου υποστήριξης. Η προσέγγιση αυτή θεωρείται ιδιαίτερα χρήσιμη σε εφαρμογές όπου δεν είναι γνωστή εκ των προτέρων η κατάλληλη τιμή support ή όταν ο αναλυτής επιθυμεί να περιορίσει τον αριθμό των αποτελεσμάτων σε ένα διαχειρίσιμο επίπεδο [20].

Επιπρόσθετα, στη βιβλιογραφία συναντώνται και άλλες τεχνικές όπως οι Relim, LCM (Linear time Closed itemset Miner) και Charm, οι οποίες επικεντρώνονται είτε στην ταχύτερη εξόρυξη κλειστών συχνών στοιχειοσυνόλων (closed frequent itemsets) είτε στη μείωση του πλήθους των παραγόμενων αποτελεσμάτων μέσω ειδικών στρατηγικών αναζήτησης και περικοπής [20].

Παρά την ύπαρξη μεγάλου αριθμού εναλλακτικών αλγορίθμων, οι Apriori, FP-Growth και Eclat εξακολουθούν να αποτελούν τις πλέον καθιερωμένες και ευρέως χρησιμοποιούμενες προσεγγίσεις τόσο στην έρευνα όσο και στις επιχειρηματικές εφαρμογές. Επομένως, επιλέχθηκαν στο πλαίσιο της μελέτης αυτής ως αντιπροσωπευτικοί αλγόριθμοι εξόρυξης συχνών στοιχειοσυνόλων και χρησιμοποιούνται στη συγκριτική αξιολόγηση που παρουσιάζεται στα επόμενα κεφάλαια.

2.5.5 Συγκριτική Παρουσίαση των Αλγορίθμων

Οι αλγόριθμοι Apriori, FP-Growth και Eclat αποτελούν τρεις από τις πλέον διαδεδομένες τεχνικές εξόρυξης συχνών στοιχειοσυνόλων και χρησιμοποιούνται ευρέως τόσο στην ακαδημαϊκή έρευνα όσο και σε πραγματικές επιχειρηματικές εφαρμογές. Παρότι έχουν κοινό στόχο τον εντοπισμό συχνών στοιχειοσυνόλων και την υποστήριξη της εξόρυξης κανόνων συσχέτισης, διαφοροποιούνται σημαντικά ως προς τη στρατηγική αναζήτησης, τις δομές δεδομένων που χρησιμοποιούν και τις απαιτήσεις τους σε υπολογιστικούς πόρους [20].

Ο Apriori αποτελεί την παλαιότερη και περισσότερο καθιερωμένη προσέγγιση. Η λειτουργία του βασίζεται στη δημιουργία υποψήφιων στοιχειοσυνόλων και στην επαναληπτική αξιολόγησή τους μέσω πολλαπλών σαρώσεων της βάσης δεδομένων. Η προσέγγιση αυτή είναι ιδιαίτερα εύκολη στην κατανόηση και την υλοποίηση, όμως μπορεί να οδηγήσει σε σημαντική αύξηση του χρόνου εκτέλεσης όταν ο αριθμός των υποψήφιων συνδυασμών γίνεται πολύ μεγάλος [18].

Ο FP-Growth αναπτύχθηκε με στόχο την αντιμετώπιση των περιορισμών του Apriori. Μέσω της χρήσης της δομής FP-Tree και της αποφυγής δημιουργίας υποψήφιων στοιχειοσυνόλων, επιτυγχάνει σημαντική μείωση του υπολογιστικού κόστους. Έτσι θεωρείται ιδιαίτερα αποτελεσματικός σε μεγάλα σύνολα δεδομένων και συχνά εμφανίζει καλύτερες επιδόσεις σε σχέση με τον Apriori ως προς τον χρόνο εκτέλεσης [21].

Από την άλλη πλευρά, ο Eclat ακολουθεί μία διαφορετική προσέγγιση, βασιζόμενος στην κάθετη αναπαράσταση των δεδομένων και στη χρήση λιστών συναλλαγών (TID Lists). Ο υπολογισμός

της συχνότητας των στοιχειοσυνόλων πραγματοποιείται μέσω πράξεων τομής μεταξύ των λιστών αυτών. Αυτό μπορεί να οδηγήσει σε ιδιαίτερα υψηλή απόδοση σε συγκεκριμένες κατηγορίες προβλημάτων. Η αποτελεσματικότητά του εξαρτάται σε μεγάλο βαθμό από τα χαρακτηριστικά των δεδομένων και το μέγεθος των λιστών συναλλαγών που πρέπει να διαχειριστεί [22].

Παρά τις διαφορές τους, και οι τρεις αλγόριθμοι ανήκουν στην κατηγορία των μεθόδων που βασίζονται στο ελάχιστο όριο υποστήριξης και έχουν χρησιμοποιηθεί εκτενώς στη βιβλιογραφία για την εξόρυξη συχνών στοιχειοσυνόλων και κανόνων συσχέτισης. Η επιλογή του καταλληλότερου αλγορίθμου εξαρτάται συνήθως από το μέγεθος του συνόλου δεδομένων, τη διαθέσιμη υπολογιστική ισχύ και τους στόχους της ανάλυσης [20].

Ο Πίνακας 2.2 παρουσιάζει συνοπτικά τα βασικά χαρακτηριστικά των αλγορίθμων Apriori, FP-Growth και Eclat.

Πίνακας 2.2: Συγκριτική παρουσίαση των αλγορίθμων Apriori, FP-Growth και Eclat

Χαρακτηριστικό	Apriori	FP-Growth	Eclat
Έτος παρουσίασης	1994	2000	2000
Βασική προσέγγιση	Candidate Generation	Pattern Growth	Vertical Mining
Δομή δεδομένων	Οριζόντια βάση συναλλαγών	FP-Tree	TID Lists
Candidate Generation	Ναι	Όχι	Όχι
Database Scans	Πολλαπλές	Δύο σαρώσεις	Περιορισμένες
Χρήση μνήμης	Υψηλή	Μέτρια	Εξαρτάται από τα TID
Ταχύτητα εκτέλεσης	Μέτρια	Υψηλή	Υψηλή
Ευκολία κατανόησης	Υψηλή	Μέτρια	Μέτρια
Καταλληλότητα για μεγάλα δεδομένα	Περιορισμένη	Πολύ καλή	Πολύ καλή

Παρότι ο Apriori θεωρείται λιγότερο αποδοτικός σε ιδιαίτερα μεγάλα σύνολα δεδομένων συγκριτικά με νεότερες προσεγγίσεις όπως ο FP-Growth και ο Eclat, εξακολουθεί να χρησιμοποιείται ευρέως στη βιβλιογραφία λόγω της απλότητας και της υψηλής ερμηνευσιμότητας των αποτελεσμάτων του. Επιπλέον, τα δεδομένα έχουν μετασχηματιστεί σε επίπεδο κατηγοριών προϊόντων, περιορίζοντας σημαντικά το πλήθος των μοναδικών αντικειμένων που συμμετέχουν στην ανάλυση. Ως εκ τούτου, η εφαρμογή του Apriori κρίνεται εφικτή και κατάλληλη για την εξαγωγή κανόνων συσχέτισης, ενώ η υπολογιστική του απόδοση αξιολογείται συγκριτικά με τους αλγορίθμους FP-Growth και Eclat στο πειραματικό μέρος της εργασίας.

2.6 Ανασκόπηση Σχετικών Ερευνών

Τόσο η εξόρυξη των κανόνων συσχέτισης όσο και το Market Basket Analysis αποτελούν αντικείμενα εκτεταμένης ερευνητικής δραστηριότητας εδώ και αρκετές δεκαετίες έχοντας εφαρμογές σε πλήθος τομέων και ιδιαίτερα στο λιανεμπόριο. Σε δεύτερο επίπεδο, η πανδημία COVID-19 δημιούργησε νέες προκλήσεις και ερευνητικά ερωτήματα σχετικά με τις μεταβολές της καταναλωτικής συμπεριφοράς και των αγοραστικών συνηθειών των καταναλωτών. Σε αυτή την ενότητα παρουσιάζονται αντιπροσωπευτικές μελέτες που αφορούν την εφαρμογή του Market Basket Analysis στο λιαν-

νεμπόριο, τη χρήση κανόνων συσχέτισης σε περιβάλλοντα σούπερ μάρκετ, καθώς και έρευνες που εξετάζουν τις επιπτώσεις της πανδημίας στην καταναλωτική συμπεριφορά. Τέλος, αναδεικνύεται το ερευνητικό κενό που επιχειρεί να καλύψει η παρούσα εργασία.

2.6.1 Market Basket Analysis στο λιανεμπόριο

Το Market Basket Analysis (MBA) αποτελεί μία από τις πιο διαδεδομένες εφαρμογές της εξόρυξης δεδομένων στον κλάδο του λιανεμπορίου, καθώς επιτρέπει την ανακάλυψη προτύπων αγοραστικής συμπεριφοράς μέσω της ανάλυσης των συναλλαγών των καταναλωτών. Η βασική ιδέα της προσέγγισης είναι ο εντοπισμός προϊόντων που εμφανίζονται συχνά μαζί στο ίδιο καλάθι αγορών. Αυτό παρέχει πολύτιμες πληροφορίες για τη λήψη επιχειρηματικών αποφάσεων που αφορούν τη διάταξη των προϊόντων, τις προωθητικές ενέργειες και τη διαχείριση των αποθεμάτων.

Η αξιοποίηση του Market Basket Analysis στο λιανεμπόριο έχει αποτελέσει αντικείμενο εκτεταμένης ερευνητικής δραστηριότητας τα τελευταία χρόνια. Ο Ünvan [3] εφάρμοσε τεχνικές εξόρυξης κανόνων συσχέτισης σε δεδομένα σούπερ μάρκετ με σκοπό την αναγνώριση προτύπων αγορών καθώς και την υποστήριξη των επιχειρηματικών αποφάσεων. Τα αποτελέσματα της μελέτης έδειξαν ότι οι κανόνες συσχέτισης μπορούν να αξιοποιηθούν αποτελεσματικά με σκοπό η διάταξη των προϊόντων στα ράφια να είναι η βέλτιστη αλλά και για την ανάπτυξη στοχευμένων στρατηγικών προώθησης. Τα παραπάνω συμβάλλουν στη βελτίωση της εμπειρίας των καταναλωτών και στην αύξηση των πωλήσεων.

Μία άλλη μελέτη αποτελεί αυτή των Alawadh και Barnawi [2]. Σε αυτή την μελέτη πραγματοποιήσαν εκτενή ανασκόπηση των μεθόδων και εφαρμογών του Market Basket Analysis, εξετάζοντας τόσο τις κλασικές τεχνικές εξόρυξης κανόνων συσχέτισης όσο και νεότερες προσεγγίσεις που αξιοποιούν τεχνικές τεχνητής νοημοσύνης καταλήγοντας στο συμπέρασμα ότι το MBA εξακολουθεί να αποτελεί ένα από τα σημαντικότερα εργαλεία ανάλυσης δεδομένων στο λιανεμπόριο, καθώς παρέχει γνώση που μπορεί να αξιοποιηθεί άμεσα για την υποστήριξη στρατηγικών marketing και διαχείρισης πελατών.

Σημαντική είναι επίσης η συμβολή της μελέτης των Rana και Mondal [1] όπου διερευνήθηκε ο ρόλος της εποχικότητας και της ιεραρχικής οργάνωσης των προϊόντων στην εξαγωγή κανόνων συσχέτισης. Τα αποτελέσματα έδειξαν ότι οι σχέσεις μεταξύ προϊόντων μεταβάλλονται ανάλογα με τη χρονική περίοδο και τα χαρακτηριστικά των προϊόντων, υπογραμμίζοντας τη σημασία της χρονικής διάστασης στην ανάλυση καταναλωτικής συμπεριφοράς.

Εκτός από τις εφαρμογές σε πραγματικά δεδομένα λιανεμπορίου, η βιβλιογραφία έχει εστιάσει και στην ανάπτυξη εργαλείων που διευκολύνουν την εφαρμογή τεχνικών εξόρυξης κανόνων συσχέτισης. Χαρακτηριστικό παράδειγμα αποτελεί το WebApriori των Malliaridis και Ougiaroglou [19], μία διαδικτυακή εφαρμογή που επιτρέπει την εύκολη εκτέλεση αλγορίθμων εξόρυξης κανόνων συσχέτισης και την οπτικοποίηση των αποτελεσμάτων. Η εργασία αυτή αναδεικνύει τη συνεχή εξέλιξη των εργαλείων Market Basket Analysis και τη σημασία τους για την αξιοποίηση των δεδομένων που παράγουν σε επιχειρηματικά περιβάλλοντα.

Από την βιβλιογραφία προκύπτει ότι το MBA αποτελεί ένα ευρέως χρησιμοποιούμενο εργαλείο ανάλυσης στο λιανεμπόριο. Οι περισσότερες μελέτες συγκλίνουν στο γεγονός ότι η αξιοποίηση δεδομένων συναλλαγών μπορεί να προσφέρει ουσιαστική γνώση σχετικά με τις αγοραστικές συνήθειες των καταναλωτών. Επίσης μπορεί να υποστηρίξει αποτελεσματικότερες επιχειρηματικές αποφά-

σεις. Επίσης, αναδεικνύεται η ανάγκη συνεχούς διερεύνησης των μεταβολών της καταναλωτικής συμπεριφοράς σε διαφορετικές χρονικές περιόδους και επιχειρηματικά περιβάλλοντα.

2.6.2 Association Rules σε Σούπερ Μάρκετ

Η εφαρμογή των κανόνων συσχέτισης σε δεδομένα λιανεμπορίου και ιδιαίτερα σε περιβάλλοντα σούπερ μάρκετ αποτελεί μία από τις πλέον διαδεδομένες χρήσεις της εξόρυξης δεδομένων. Όπως προαναφέραμε, οι κανόνες συσχέτισης επιτρέπουν την ανακάλυψη σχέσεων μεταξύ προϊόντων που εμφανίζονται συχνά μαζί στις συναλλαγές των καταναλωτών. Χάρη σε αυτό οι επιχειρήσεις αντλούν χρήσιμες γνώσεις για τη διάταξη των προϊόντων, τις προωθητικές ενέργειες και τη διαχείριση των αποθεμάτων τους.

Η απαρχή της συγκεκριμένης ερευνητικής περιοχής βρίσκεται στην εργασία των Agrawal, Imieliński και Swami [16], οι οποίοι εισήγαγαν για πρώτη φορά το πρόβλημα της εξόρυξης κανόνων συσχέτισης με δεδομένα που αντλούνται από μεγάλες βάσεις δεδομένων συναλλαγών. Οι συγγραφείς χρησιμοποίησαν ως χαρακτηριστικό παράδειγμα δεδομένα πωλήσεων σούπερ μάρκετ και έδειξαν ότι είναι δυνατόν να εντοπιστούν συσχετίσεις μεταξύ προϊόντων που αγοράζονται ταυτόχρονα από τους καταναλωτές. Η εργασία αυτή θεωρείται θεμελιώδης για ολόκληρο το πεδίο του Association Rule Mining [16].

Στη συνέχεια, οι Agrawal και Srikant [18] παρουσίασαν τον αλγόριθμο Apriori, ο οποίος αποτέλεσε την πρώτη ευρέως αποδεκτή και αποδοτική μέθοδο εξόρυξης συχνών στοιχειοσυνόλων και κανόνων συσχέτισης. Η συμβολή της συγκεκριμένης εργασίας υπήρξε καθοριστική, διότι κατέστησε εφικτή την εφαρμογή των κανόνων συσχέτισης σε μεγάλα σύνολα δεδομένων λιανικών πωλήσεων και αποτέλεσε το θεμέλιο για την ανάπτυξη μεταγενέστερων αλγορίθμων εξόρυξης δεδομένων.

Η ερευνητική δραστηριότητα στον χώρο συνεχίζεται μέχρι σήμερα. Οι Kholod et al. [23] αξιοποίησαν σύγχρονα δεδομένα λιανικών συναλλαγών που προέρχονται από συστήματα Point-of-Sale (POS) και τεχνολογίες RFID, εφαρμόζοντας τεχνικές εξόρυξης κανόνων συσχέτισης με στόχο την αναγνώριση προτύπων αγοραστικής συμπεριφοράς. Τα αποτελέσματα αυτής της μελέτης επιβεβαιώνουν ότι οι κανόνες συσχέτισης εξακολουθούν να αποτελούν ένα αποτελεσματικό εργαλείο ανάλυσης δεδομένων λιανεμπορίου, ακόμη και σε σύγχρονα περιβάλλοντα που χαρακτηρίζονται από μεγάλο όγκο και υψηλή πολυπλοκότητα δεδομένων.

Παράλληλα, η ανάπτυξη εργαλείων όπως το WebApriori των Malliaridis και Ougiaroglou [19] διευκολύνει την εφαρμογή τεχνικών Association Rule Mining σε πραγματικά δεδομένα, παρέχοντας ένα εύχρηστο περιβάλλον για την εξαγωγή και ανάλυση κανόνων συσχέτισης. Η ύπαρξη τέτοιων εργαλείων συμβάλλει στη διάδοση των τεχνικών αυτών τόσο στην ερευνητική κοινότητα όσο και στον επιχειρηματικό χώρο.

Η βιβλιογραφία καταδεικνύει ότι οι κανόνες συσχέτισης αποτελούν ένα από τα σημαντικότερα εργαλεία ανάλυσης δεδομένων σούπερ μάρκετ και λιανεμπορίου. Η δυνατότητά τους να αποκαλύπτουν πρότυπα συνδυαστικής αγοράς προϊόντων εξηγεί τη διαχρονική παρουσία τους στη βιβλιογραφία και τη συνεχιζόμενη αξιοποίησή τους σε πραγματικές εφαρμογές.

2.6.3 Μελέτες σχετικές με COVID-19

Η πανδημία COVID-19 αποτέλεσε ένα πρωτοφανές γεγονός που επηρέασε σημαντικά την καταναλωτική συμπεριφορά παγκοσμίως, οδηγώντας σε μεταβολές στις αγοραστικές συνήθειες και

στις προτιμήσεις των καταναλωτών. Ως αποτέλεσμα, η διερεύνηση των επιπτώσεων της πανδημίας στον τομέα του λιανεμπορίου προσέλκυσε έντονο ερευνητικό ενδιαφέρον και οδήγησε τελικά στην παραγωγή σημαντικού αριθμού σχετικών μελετών.

Μία από τις πρώτες και πλέον αναφερόμενες εργασίες είναι αυτή του Sheth [5], ο οποίος διερεύνησε τις επιπτώσεις της πανδημίας στην καταναλωτική συμπεριφορά και υποστήριξε ότι πολλές από τις αλλαγές που παρατηρήθηκαν κατά τη διάρκεια της υγειονομικής κρίσης ενδέχεται να διατηρηθούν και μετά το τέλος της. Ο συγγραφέας επισημαίνει ότι οι περιορισμοί στις μετακινήσεις, η αβεβαιότητα και η αυξημένη χρήση ψηφιακών τεχνολογιών επηρέασαν σημαντικά τον τρόπο με τον οποίο οι καταναλωτές πραγματοποιούν τις αγορές τους.

Στο πλαίσιο της διερεύνησης ειδικότερων φαινομένων, οι Omar et al. [9] μελέτησαν τη συμπεριφορά panic buying κατά την περίοδο της πανδημίας. Τα αποτελέσματα της έρευνας έδειξαν ότι η αντίληψη της έλλειψης των προϊόντων, η αβεβαιότητα αλλά και το άγχος επηρέασαν σημαντικά τις αγοραστικές αποφάσεις των καταναλωτών, οδηγώντας σε αυξημένες αγορές βασικών αγαθών και δημιουργία αποθεμάτων.

Ιδιαίτερο ενδιαφέρον παρουσιάζουν και οι μελέτες που εστιάζουν στις αγορές τροφίμων και ειδών παντοπωλείου (online grocery shopping). Οι Tyrväinen και Karjaluoto [13] εξέτασαν τη συμπεριφορά των καταναλωτών στις ηλεκτρονικές αγορές τροφίμων πριν και κατά τη διάρκεια της πανδημίας μέσα από την ανάλυση της σχετικής βιβλιογραφίας. Τα αποτελέσματά τους έδειξαν ότι η πανδημία έπαιξε σημαντικό ρόλο στην υιοθέτηση των ηλεκτρονικών αγορών για τα είδη παντοπωλείου, καθώς οι καταναλωτές στράφηκαν σε διαδικτυακά κανάλια για την προμήθεια βασικών αγαθών, όπως τρόφιμα και προϊόντα καθημερινής χρήσης. Συγχρόνως, διαπιστώθηκε ότι αυτή η συμπεριφορά δεν αποτέλεσε αποκλειστικά μια προσωρινή αντίδραση στην υγειονομική κρίση, αλλά σε αρκετές περιπτώσεις διατηρήθηκε και μετά τη χαλάρωση των περιοριστικών μέτρων υποδηλώνοντας τη διαμόρφωση νέων καταναλωτικών συνθηκών.

Αντίστοιχα, οι Diaz-Gutierrez et al. [15] διερεύνησαν τις επιπτώσεις της πανδημίας τόσο στις ηλεκτρονικές όσο και στις φυσικές αγορές. Οι συγγραφείς διαπίστωσαν ότι η πανδημία επιτάχυνε τον ψηφιακό μετασχηματισμό του λιανεμπορίου και επηρέασε σημαντικά τον τρόπο με τον οποίο οι καταναλωτές επιλέγουν τα κανάλια αγορών τους. Παράλληλα και οι ίδιοι υποστήριξαν ότι ορισμένες από τις αλλαγές που παρατηρήθηκαν κατά την περίοδο της πανδημίας συνέχισαν να υφίστανται και στη μεταπανδημική περίοδο.

Πιο πρόσφατα, οι Żurek και Rudy [12] πραγματοποίησαν συστηματική βιβλιογραφική ανασκόπηση σχετικά με τις επιπτώσεις της πανδημίας στις αγοραστικές συνήθειες και στις διατροφικές επιλογές των καταναλωτών. Οι συγγραφείς κατέληξαν στο συμπέρασμα ότι η πανδημία επηρέασε ουσιαστικά τη δομή της ζήτησης πολλών κατηγοριών προϊόντων, ενώ οδήγησε σε αλλαγές στις καταναλωτικές προτεραιότητες και στις διαδικασίες λήψης αγοραστικών αποφάσεων.

Συμπερασματικά, η υπάρχουσα βιβλιογραφία συμφωνεί ότι η πανδημία COVID-19 προκάλεσε σημαντικές μεταβολές στην καταναλωτική συμπεριφορά και στις αγοραστικές συνήθειες των καταναλωτών. Ωστόσο, οι περισσότερες μελέτες επικεντρώνονται κυρίως στις άμεσες επιπτώσεις της πανδημίας ή στις μεταβολές που παρατηρήθηκαν κατά τη διάρκειά της. Αντίθετα, περιορισμένος αριθμός ερευνών εξετάζει σε βάθος κατά πόσο οι μεταβολές αυτές διατηρήθηκαν και μετά την άρση των περιοριστικών μέτρων, δημιουργώντας ένα ενδιαφέρον πεδίο περαιτέρω διερεύνησης.

2.6.4 Ερευνητικό Κενό

Η βιβλιογραφία που παρουσιάστηκε στις προηγούμενες ενότητες καταδεικνύει ότι το Market Basket Analysis και οι κανόνες συσχέτισης αποτελούν καθιερωμένες τεχνικές που χρησιμοποιούνται για την ανάλυση δεδομένων λιανεμπορίου. Πλήθος μελετών έχουν χρησιμοποιήσει δεδομένα συναλλαγών προκειμένου να αναγνωρίσουν πρότυπα αγοραστικής συμπεριφοράς και την υποστήριξη επιχειρηματικών αποφάσεων, ενώ σημαντικός αριθμός ερευνών έχει εξετάσει τις επιπτώσεις της πανδημίας COVID-19 στις καταναλωτικές συνήθειες [24],[25].

Παρά την εκτενή ερευνητική δραστηριότητα στους παραπάνω τομείς, παρατηρούνται ορισμένα κενά στη σχετική βιβλιογραφία. Αρχικά, οι περισσότερες μελέτες που αφορούν την πανδημία επικεντρώνονται κυρίως στην καταγραφή και ερμηνεία των αλλαγών που σημειώθηκαν κατά τη διάρκειά της, χωρίς να εξετάζουν συστηματικά εάν οι μεταβολές αυτές διατηρήθηκαν και μετά την άρση των περιοριστικών μέτρων. Ως αποτέλεσμα, παραμένει περιορισμένη η γνώση σχετικά με το κατά πόσο οι αλλαγές που παρατηρήθηκαν στην καταναλωτική συμπεριφορά είχαν προσωρινό ή πιο μόνιμο αποτύπωμα στους καταναλωτές.

Ακόμη, μεγάλο μέρος της σχετικής βιβλιογραφίας βασίζεται σε έρευνες ερωτηματολογίων, σε δεδομένα ηλεκτρονικών αγορών ή σε μικρότερης κλίμακας σύνολα δεδομένων. Αντίθετα, είναι σχετικά περιορισμένες οι μελέτες που αξιοποιούν πραγματικά δεδομένα συναλλαγών λιανεμπορίου μεγάλης κλίμακας για τη διερεύνηση των μεταβολών της αγοραστικής συμπεριφοράς μεταξύ διαφορετικών χρονικών περιόδων.

Αντίστοιχα, η βιβλιογραφία που συγκρίνει την απόδοση διαφορετικών αλγορίθμων εξόρυξης συχνών στοιχειοσυνόλων σε πραγματικά δεδομένα λιανεμπορίου παραμένει περιορισμένη, ιδιαίτερα όταν πρόκειται για σύνολα δεδομένων που περιλαμβάνουν εκατομμύρια συναλλαγές. Ως εκ τούτου, εξακολουθεί να υπάρχει ενδιαφέρον για τη διερεύνηση της αποδοτικότητας διαφορετικών αλγορίθμων σε συνθήκες που προσεγγίζουν περισσότερο τα πραγματικά επιχειρηματικά περιβάλλοντα.

Η παρούσα εργασία επιχειρεί να συμβάλει στην κάλυψη των παραπάνω κενών αξιοποιώντας περισσότερες από 9,5 εκατομμύρια πραγματικές συναλλαγές σούπερ μάρκετ από δύο διακριτές χρονικές περιόδους, την περίοδο της πανδημίας COVID-19 και τη μεταπανδημική περίοδο. Μέσω της εφαρμογής τεχνικών εξόρυξης κανόνων συσχέτισης επιδιώκεται η διερεύνηση πιθανών μεταβολών στην καταναλωτική συμπεριφορά, ενώ ταυτόχρονα πραγματοποιείται συγκριτική αξιολόγηση των αλγορίθμων Apriori, FP-Growth και Eclat ως προς την απόδοσή τους σε δεδομένα μεγάλης κλίμακας.

Κεφάλαιο 3

Πειραματικός Σχεδιασμός και Βελτιστοποίηση Παραμέτρων

3.1 Απόκτηση και Περιγραφή Δεδομένων

Τα δεδομένα που χρησιμοποιήθηκαν στην παρούσα διπλωματική εργασία προήλθαν από μία ελληνική αλυσίδα σούπερ μάρκετ. Η πρόσβαση στα δεδομένα πραγματοποιήθηκε κατόπιν υπογραφής δήλωσης εμπιστευτικότητας (Non-Disclosure Agreement – NDA) μεταξύ της εταιρείας, της συγγραφέως της παρούσας εργασίας και του επιβλέποντος καθηγητή. Η συμφωνία προέβλεπε ότι τα δεδομένα θα χρησιμοποιηθούν αποκλειστικά για τους σκοπούς της συγκεκριμένης ακαδημαϊκής έρευνας.

Το σύνολο δεδομένων προέρχεται από δεκαεπτά καταστήματα της εταιρείας που λειτουργούν στον νομό Καβάλας και αποτυπώνει πραγματικές συναλλαγές λιανικής πώλησης. Παρ'όλα αυτά, πριν από την επεξεργασία και ανάλυσή τους εφαρμόστηκαν διαδικασίες ανωνυμοποίησης και γενίκευσης, ώστε να διασφαλιστεί η προστασία τόσο των πελατών όσο και των εμπορικών πληροφοριών της εταιρείας.

Συγκεκριμένα, τα δεδομένα δεν περιλαμβάνουν κανένα προσωπικό στοιχείο πελάτη ούτε οποιαδήποτε πληροφορία που θα μπορούσε να οδηγήσει σε άμεση ή έμμεση ταυτοποίηση φυσικού προσώπου. Επιπλέον, οι αρχικοί κωδικοί συναλλαγών αντικαταστάθηκαν από μοναδικά αναγνωριστικά συναλλαγών (Transaction IDs) που δημιουργήθηκαν αποκλειστικά για τις ανάγκες της ανάλυσης.

Τα προϊόντα δεν αναλύονται σε επίπεδο συγκεκριμένου εμπορικού κωδικού (SKU), αλλά μετατρέπονται στη γενική κατηγορία στην οποία ανήκουν. Για παράδειγμα, ένα προϊόν όπως το «Skip Υγρό Απορρυπαντικό 30 Μεζούρες» καταγράφεται στο σύνολο δεδομένων ως «Απορρυπαντικά». Με τον τρόπο αυτό αποτρέπεται η αποκάλυψη εμπορικά ευαίσθητων πληροφοριών που σχετίζονται με συγκεκριμένα προϊόντα ή προμηθευτές.

Τέλος, από το σύνολο δεδομένων έχουν αφαιρεθεί πληροφορίες που αφορούν την ποσότητα των αγοραζόμενων προϊόντων και τη χρηματική αξία των συναλλαγών. Η ανάλυση βασίζεται αποκλειστικά στην παρουσία ή απουσία μιας κατηγορίας προϊόντων σε κάθε συναλλαγή, επιτρέποντας την εξαγωγή κανόνων συσχέτισης χωρίς να εκτίθενται οικονομικά στοιχεία ή επιχειρηματικά δεδομένα της εταιρείας.

Συνεπώς, το τελικό σύνολο δεδομένων είναι πλήρως ανωνυμοποιημένο και κατάλληλο για ερευνητική χρήση, διασφαλίζοντας τόσο την προστασία προσωπικών δεδομένων όσο και την εμπιστευ-

τικότητα των εμπορικών πληροφοριών της επιχείρησης.

3.2 Προεπεξεργασία και Διαμόρφωση Δεδομένων

Πριν από την εφαρμογή των αλγορίθμων εξόρυξης κανόνων συσχέτισης, πραγματοποιήθηκε κατάλληλη προεπεξεργασία των δεδομένων ώστε να μετατραπούν στη μορφή που απαιτείται για την εκτέλεση τεχνικών Market Basket Analysis. Τα αρχικά σύνολα δεδομένων περιείχαν τις επιμέρους εγγραφές προϊόντων ανά συναλλαγή, όπου κάθε γραμμή αντιστοιχούσε σε ένα προϊόν που αγοράστηκε στο πλαίσιο μιας συγκεκριμένης συναλλαγής. Ωστόσο, οι αλγόριθμοι εξόρυξης κανόνων συσχέτισης απαιτούν κάθε συναλλαγή να αναπαρίσταται σε μία μοναδική γραμμή η οποία να περιλαμβάνει το σύνολο των προϊόντων που αγοράστηκαν ταυτόχρονα.

Για τον σκοπό αυτό αναπτύχθηκε πρόγραμμα σε Python, το οποίο ομαδοποιεί τις εγγραφές βάσει του μοναδικού αναγνωριστικού συναλλαγής (Transaction ID), συγκεντρώνει όλα τα προϊόντα που ανήκουν στην ίδια συναλλαγή και δημιουργεί ένα νέο αρχείο δεδομένων σε μορφή basket, αφαιρεί τυχόν διπλές εμφανίσεις της ίδιας κατηγορίας προϊόντος μέσα στην ίδια συναλλαγή και ταξινομεί τα προϊόντα με ενιαίο τρόπο, ώστε να εξασφαλίζεται ομοιομορφία στα δεδομένα εισόδου. Μετά την ολοκλήρωση της διαδικασίας δημιουργήθηκαν νέα αρχεία δεδομένων σε μορφή basket, στα οποία κάθε γραμμή αντιστοιχεί σε μία πλήρη συναλλαγή και οι κατηγορίες προϊόντων διαχωρίζονται με το σύμβολο «;». Η μορφή αυτή αποτελεί την κατάλληλη είσοδο για την εφαρμογή τεχνικών εξόρυξης συχνών συνόλων αντικειμένων και κανόνων συσχέτισης.

3.3 Διαχωρισμός των Δεδομένων σε Χρονικές Περιόδους

Βασικός στόχος της παρούσας εργασίας αποτελεί η διερεύνηση των πιθανών μεταβολών στην καταναλωτική συμπεριφορά κατά την περίοδο της πανδημίας αλλά και μετά από αυτή. Προκειμένου να γίνει η συγκεκριμένη σύγκριση, το συνολικό σύνολο δεδομένων χωρίστηκε σε δύο ανεξάρτητα υποσύνολα.

Η πρώτη χρονική περίοδος αφορά την περίοδο της πανδημίας και περιλαμβάνει συναλλαγές από τον Μάρτιο του 2020, χρονικό σημείο κατά το οποίο εφαρμόστηκε το πρώτο επίσημο lockdown στην Ελλάδα, έως και την 31η Αυγούστου 2022, όταν είχαν πλέον αρθεί τα τελευταία περιοριστικά μέτρα που σχετίζονταν με την πανδημία. Το αντίστοιχο σύνολο δεδομένων περιλαμβάνει συνολικά 4.408.323 μοναδικές συναλλαγές.

Η δεύτερη χρονική περίοδος αφορά τη μεταπανδημική (Post-COVID) εποχή και περιλαμβάνει συναλλαγές από την 1η Σεπτεμβρίου 2022 έως και την 31η Δεκεμβρίου 2024. Ως χρονικό όριο δεν επιλέχθηκε η τρέχουσα ημερομηνία, αλλά το τέλος του έτους 2024, έτσι ώστε οι δύο εξεταζόμενες περίοδοι να έχουν παρόμοια χρονική διάρκεια προκειμένου η σύγκριση των αποτελεσμάτων να είναι πιο αξιόπιστη. Το αντίστοιχο σύνολο δεδομένων περιλαμβάνει συνολικά 5.101.828 μοναδικές συναλλαγές.

Τα δεδομένα προέρχονται από πραγματικές εμπορικές συναλλαγές που πραγματοποιήθηκαν σε δίκτυο δεκαεπτά (17) καταστημάτων του νομού Καβάλας. Κάθε γραμμή των τελικών αρχείων αντιστοιχεί σε μία μοναδική συναλλαγή και περιλαμβάνει το σύνολο των κατηγοριών προϊόντων που αγοράστηκαν κατά τη συγκεκριμένη επίσκεψη του πελάτη. Ο διαχωρισμός των δεδομένων στις δύο

χρονικές περιόδους επιτρέπει την ανεξάρτητη μελέτη των αγοραστικών προτύπων κάθε περιόδου, καθώς και τη μετέπειτα συγκριτική αξιολόγηση των αποτελεσμάτων.

Πίνακας 3.1: Βασικά χαρακτηριστικά των συνόλων δεδομένων

Περίοδος	Χρονικό Διάστημα	Πλήθος Συναλλαγών
COVID-19	03/2020 – 08/2022	4.408.323
Μετά COVID-19	09/2022 – 12/2024	5.101.828

3.4 Διερευνητική Ανάλυση Συχνοτήτων Προϊόντων

Πριν από την εφαρμογή των τεχνικών εξόρυξης κανόνων συσχέτισης πραγματοποιήθηκε μια αρχική διερευνητική ανάλυση των δεδομένων, με σκοπό την κατανόηση της γενικής δομής των συναλλαγών και την αναγνώριση των κατηγοριών προϊόντων που αγοράζονταν συχνότερα κατά τις δύο εξεταζόμενες χρονικές περιόδους.

Η ανάλυση αυτή παρέχει μια πρώτη εικόνα των προϊόντων που εμφανίζονται συχνότερα στα καλάθια των καταναλωτών και επιτρέπει τον εντοπισμό κατηγοριών των οποίων η συμμετοχή στις συναλλαγές διαφοροποιήθηκε μεταξύ των δύο περιόδων.

Για τον λόγο αυτό, προσδιορίστηκαν οι δεκατρείς (13) κατηγορίες προϊόντων με τη μεγαλύτερη συχνότητα εμφάνισης στις συναλλαγές κάθε περιόδου. Η αξιολόγηση βασίστηκε τόσο στον αριθμό των συναλλαγών στις οποίες εμφανίζεται κάθε κατηγορία όσο και στο αντίστοιχο ποσοστό υποστήριξης (support), το οποίο εκφράζει το ποσοστό των συναλλαγών στις οποίες συμμετέχει η συγκεκριμένη κατηγορία.

Πέρα από την παραπάνω αρχική καταγραφή των συχνότερων κατηγοριών, πραγματοποιήθηκε στη συνέχεια ξεχωριστή συγκριτική ανάλυση των 1-itemsets, με σκοπό να εξεταστούν αναλυτικότερα οι μεταβολές του support μεταξύ των δύο περιόδων. Στη συγκεκριμένη ανάλυση, κάθε κατηγορία προσμετρήθηκε μόνο μία φορά ανά συναλλαγή, ανεξάρτητα από το πλήθος των τεμαχίων που αγοράστηκαν.

Από τη συμπληρωματική αυτή ανάλυση εξαιρέθηκαν οι κατηγορίες «Είδη Συσκευασίας προς Πώληση» και «Απαράδεκτα Εμπορεύματα», καθώς αφορούν κυρίως σακούλες μεταφοράς, παλέτες και κενές φιάλες και δεν αντιπροσωπεύουν ουσιαστικές καταναλωτικές επιλογές. Μετά την αφαίρεσή τους, 42 συναλλαγές της περιόδου COVID-19 και 5 συναλλαγές της μεταπανδημικής περιόδου δεν περιείχαν καμία άλλη κατηγορία προϊόντος και αποκλείστηκαν από τη συγκεκριμένη διαδικασία. Συνεπώς, η συγκριτική ανάλυση βασίστηκε σε 4.408.281 συναλλαγές για την περίοδο COVID-19 και σε 5.101.823 συναλλαγές για τη μεταπανδημική περίοδο.

Στο πλαίσιο της ίδιας συμπληρωματικής διερεύνησης υπολογίστηκε και το μέσο μέγεθος του καλαθιού για κάθε περίοδο. Ως μέγεθος καλαθιού ορίστηκε το πλήθος των διαφορετικών κατηγοριών που περιλαμβάνονται σε μία συναλλαγή και όχι το πλήθος των τεμαχίων. Η σύγκριση αυτή πραγματοποιήθηκε προκειμένου να εξεταστεί κατά πόσο ο μεγαλύτερος αριθμός συχνών στοιχειοσυνόλων κατά την περίοδο COVID-19 μπορεί να σχετίζεται και με μια γενικότερη διαφορά στη σύνθεση ή στο μέγεθος των καλαθιών.

3.4.1 Σύγκριση των κυριότερων κατηγοριών προϊόντων

Ο Πίνακας 3.2 παρουσιάζει την ένωση των δεκατριών συχνότερα εμφανιζόμενων κατηγοριών προϊόντων κάθε χρονικής περιόδου, καθώς και τη μεταβολή του ποσοστού υποστήριξής τους. Επειδή οι δύο κατατάξεις δεν ταυτίζονται πλήρως, από την ένωσή τους προκύπτουν συνολικά δεκατέσσερις διακριτές κατηγορίες. Για λόγους πληρότητας, στον πίνακα διατηρείται επίσης η κατηγορία «Είδη Συσκευασίας προς Πώληση», χωρίς όμως να λαμβάνεται υπόψη στην ουσιαστική ερμηνεία των καταναλωτικών επιλογών.

Πίνακας 3.2: Σύγκριση ποσοστού υποστήριξης των σημαντικότερων κατηγοριών προϊόντων μεταξύ της περιόδου COVID-19 και της μεταπανδημικής περιόδου

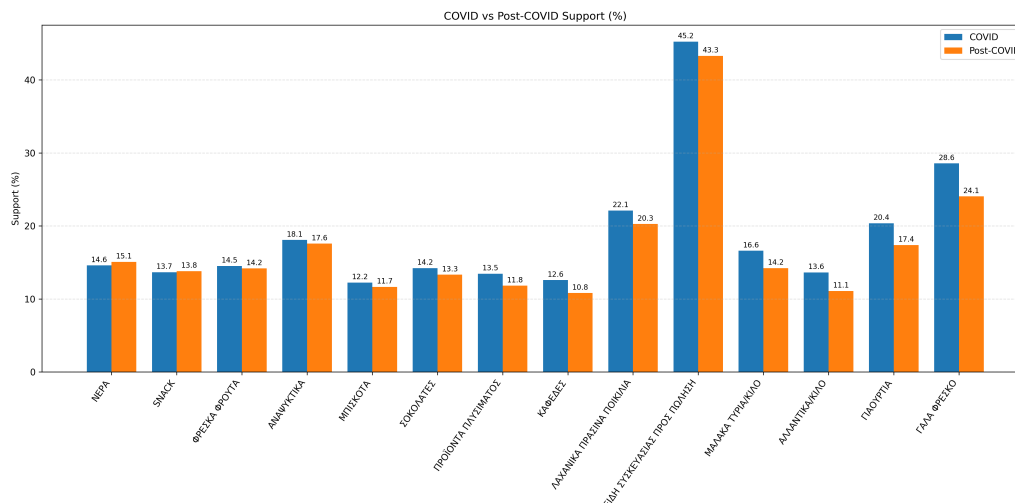
Κατηγορία	Support COVID (%)	Support Post-COVID (%)	Δ Support (π.μ.)
ΝΕΡΑ	14,59	15,08	+0,49
SNACK	13,66	13,81	+0,15
ΦΡΕΣΚΑ ΦΡΟΥΤΑ	14,52	14,18	-0,33
ΑΝΑΨΥΚΤΙΚΑ	18,08	17,58	-0,50
ΜΠΙΣΚΟΤΑ	12,25	11,65	-0,60
ΣΟΚΟΛΑΤΕΣ	14,22	13,32	-0,90
ΠΡΟΪΟΝΤΑ ΠΛΥΣΙΜΑΤΟΣ	13,46	11,82	-1,64
ΚΑΦΕΔΕΣ	12,59	10,82	-1,77
ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	22,11	20,27	-1,84
ΕΙΔΗ ΣΥΣΚΕΤΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	45,25	43,30	-1,95
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	16,60	14,23	-2,37
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	13,63	11,09	-2,54
ΓΙΑΟΥΡΤΙΑ	20,37	17,38	-2,99
ΓΑΛΑ ΦΡΕΣΚΟ	28,57	24,05	-4,52

Από τα αποτελέσματα προκύπτει ότι αρκετές κατηγορίες προϊόντων διατηρούν υψηλή συχνότητα εμφάνισης και στις δύο εξεταζόμενες περιόδους. Η κατηγορία «Είδη Συσκευασίας προς Πώληση» εμφανίζει το υψηλότερο ποσοστό υποστήριξης, συμμετέχοντας σε περισσότερο από το 43% των συναλλαγών. Ωστόσο, επειδή αφορά κυρίως σακούλες μεταφοράς, η παρουσία της συνδέεται περισσότερο με τη λειτουργική ολοκλήρωση της αγοράς και δεν αντιμετωπίζεται ως αυτόνομη καταναλωτική προτίμηση. Για τον λόγο αυτό, δεν λαμβάνεται υπόψη στην ουσιαστική ερμηνεία των αγοραστικών προτύπων.

Αν δεν ληφθεί υπόψη η συγκεκριμένη κατηγορία, το «Φρέσκο Γάλα» παρουσιάζει το υψηλότερο ποσοστό υποστήριξης μεταξύ των πραγματικών κατηγοριών προϊόντων και στις δύο περιόδους, με 28,57% κατά την περίοδο COVID-19 και 24,05% κατά τη μεταπανδημική περίοδο. Υψηλή συχνότητα εμφάνισης διατηρούν επίσης τα «Λαχανικά Πράσινα Ποικιλία», με ποσοστά 22,11% και 20,27% αντίστοιχα, καθώς και τα «Γιαούρτια», με 20,37% και 17,38%. Η σταθερή παρουσία των συγκεκριμένων κατηγοριών στις υψηλότερες θέσεις δείχνει ότι βασικά είδη καθημερινής διατροφής εξακολουθούν να κατέχουν κεντρικό ρόλο στα καλάθια αγορών και στις δύο εξεταζόμενες περιόδους.

3.4.2 Σύγκριση των ποσοστών υποστήριξης

Για την καλύτερη αποτύπωση των διαφορών μεταξύ των δύο περιόδων δημιουργήθηκε το Σχήμα 3.1, στο οποίο παρουσιάζονται συγκριτικά τα ποσοστά υποστήριξης των σημαντικότερων κατηγοριών προϊόντων.



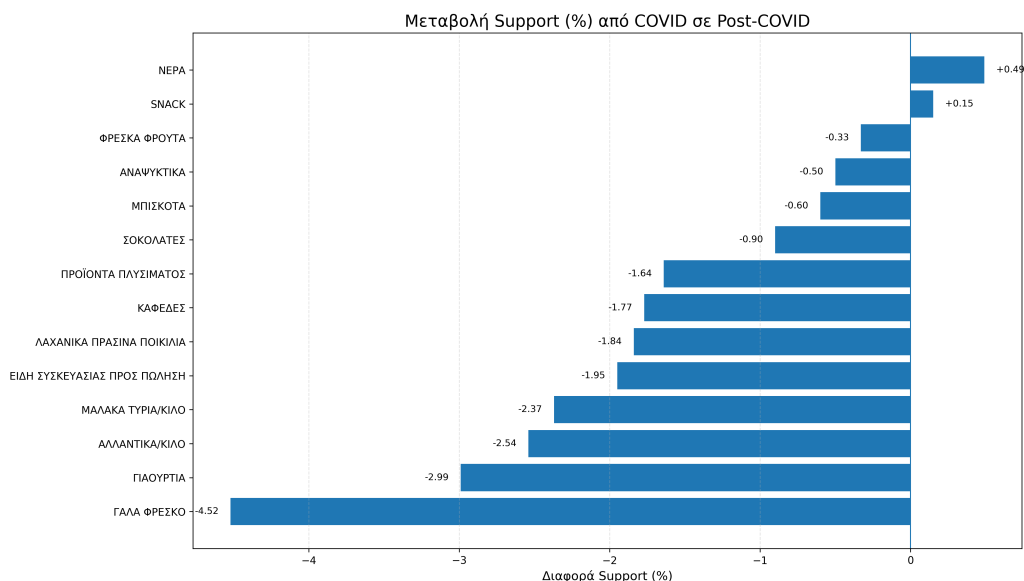
Σχήμα 3.1: Σύγκριση ποσοστών υποστήριξης (Support %) μεταξύ της περιόδου COVID-19 και της μεταπανδημικής περιόδου

Όπως φαίνεται στο Σχήμα 3.1, η γενική εικόνα των κατηγοριών με τα υψηλότερα ποσοστά υποστήριξης παραμένει σε μεγάλο βαθμό κοινή στις δύο περιόδους, αν και στις περισσότερες περιπτώσεις οι τιμές είναι χαμηλότερες κατά τη μεταπανδημική περίοδο. Η κατηγορία «Είδη Συσκευασίας προς Πώληση» εμφανίζει το υψηλότερο ποσοστό υποστήριξης, αλλά, για τους λόγους που αναφέρθηκαν προηγουμένως, δεν λαμβάνεται υπόψη στην ουσιαστική σύγκριση των καταναλωτικών προτύπων.

Μεταξύ των κατηγοριών που παρουσιάζονται στο Σχήμα 3.1 οι μεγαλύτερες μειώσεις καταγράφηκαν στο «Φρέσκο Γάλα» (-4,52 ποσοστιαίες μονάδες), στα «Γιαούρτια» (-2,99 ποσοστιαίες μονάδες) και στα «Αλλαντικά/Κιλό» (-2,54 ποσοστιαίες μονάδες). Παρότι οι κατηγορίες αυτές εξακολουθούν να συγκαταλέγονται στις σημαντικότερες και στις δύο περιόδους, η συχνότητα εμφάνισής τους στις συναλλαγές παρουσιάζει αισθητή υποχώρηση κατά τη μεταπανδημική περίοδο. Το εύρημα αυτό αποτυπώνει μια διαφοροποίηση στη συμμετοχή των συγκεκριμένων κατηγοριών στα καλάθια αγορών, χωρίς να επιτρέπει από μόνο του τον προσδιορισμό των παραγόντων που την προκάλεσαν.

3.4.3 Μεταβολή του ποσοστού υποστήριξης

Οι μεταβολές που παρουσιάστηκαν αριθμητικά στην προηγούμενη υποενότητα αποδίδονται οπτικά στο Σχήμα 3.2, ώστε να γίνει σαφέστερη η κατεύθυνση και το μέγεθός τους. Οι τιμές εκφράζονται σε ποσοστιαίες μονάδες και υπολογίζονται ως η διαφορά του support της μεταπανδημικής περιόδου από το support της περιόδου COVID-19.



Σχήμα 3.2: Μεταβολή του support των σημαντικότερων κατηγοριών προϊόντων μεταξύ των δύο περιόδων. Οι τιμές εκφράζονται σε ποσοστιαίες μονάδες και υπολογίζονται ως Support Post-COVID μείον Support COVID-19.

Η οπτική απεικόνιση δείχνει ότι η πλειονότητα των σημαντικότερων κατηγοριών παρουσίασε χαμηλότερο support κατά τη μεταπανδημική περίοδο, ενώ θετική μεταβολή καταγράφηκε σε περιορισμένο αριθμό κατηγοριών. Επειδή το σχήμα αφορά μόνο τις κατηγορίες με την υψηλότερη συχνότητα εμφάνισης, στην επόμενη ενότητα η σύγκριση επεκτείνεται στο σύνολο των κοινών κατηγοριών των δύο περιόδων.

3.4.4 Συμπληρωματική σύγκριση των κοινών 1-itemsets

Για να διευρυνθεί η εικόνα που προέκυψε από την αρχική εξέταση των συχνότερων κατηγοριών, πραγματοποιήθηκε συμπληρωματική σύγκριση του ποσοστού υποστήριξης για τις κατηγορίες που εμφανίζονται και στις δύο χρονικές περιόδους. Μετά τις εξαιρέσεις που αναφέρθηκαν προηγουμένως, εντοπίστηκαν 315 κοινές κατηγορίες προϊόντων, οι οποίες αποτέλεσαν τη βάση της συγκεκριμένης σύγκρισης.

Οι κατηγορίες που εμφανίζονται μόνο στη μία περίοδο καταγράφηκαν χωριστά και δεν συμπεριλήφθηκαν στην κατάταξη των μεγαλύτερων αυξήσεων και μειώσεων. Για τις κατηγορίες που εμφανίζονται μόνο στη μεταπανδημική περίοδο, η σχετική μεταβολή δεν μπορεί να υπολογιστεί, καθώς το support της περιόδου COVID-19 είναι μηδενικό. Αντίστοιχα, στις κατηγορίες που εμφανίζονται μόνο κατά την περίοδο COVID-19, η αριθμητική μεταβολή ισούται με -100%, χωρίς όμως αυτό να τεκμηριώνει απαραίτητα την πλήρη εξαφάνιση μιας καταναλωτικής επιλογής. Η παρουσία μιας κατηγορίας μόνο στη μία περίοδο μπορεί να συνδέεται και με αλλαγές στην ταξινόμηση ή στη σύνθεση του διαθέσιμου καταλόγου προϊόντων. Για τον λόγο αυτό, η κύρια σύγκριση περιορίστηκε στις ίδιες κατηγορίες προϊόντων μεταξύ των δύο περιόδων.

Για κάθε κοινή κατηγορία υπολογίστηκαν τόσο η απόλυτη διαφορά του support όσο και η σχετική ποσοστιαία μεταβολή του. Η απόλυτη διαφορά εκφράζεται σε ποσοστιαίες μονάδες και υπολογίζεται ως:

$$\Delta Support = Support_{Post-COVID} - Support_{COVID}$$

Η σχετική ποσοστιαία μεταβολή υπολογίζεται σε σχέση με την περίοδο COVID-19 ως εξής:

$$\% \Delta Support = \frac{Support_{Post-COVID} - Support_{COVID}}{Support_{COVID}} \times 100$$

Θετική τιμή της διαφοράς δηλώνει ότι η κατηγορία εμφανίζεται σε μεγαλύτερο ποσοστό συναλλαγών κατά τη μεταπανδημική περίοδο, ενώ αρνητική τιμή δηλώνει μείωση της συμμετοχής της. Οι δέκα μεγαλύτερες αυξήσεις παρουσιάζονται στον Πίνακα 3.3.

Πίνακας 3.3: Κατηγορίες με τις δέκα μεγαλύτερες αυξήσεις του support μεταξύ των δύο περιόδων

Κατηγορία	Support COVID (%)	Support Post-COVID (%)	Δ Support (π.μ.)	Σχετική μεταβολή (%)
ΕΤΟΙΜΟ ΦΑΓΗΤΟ CATERING	2,379	3,448	+1,069	+44,93
ΚΡΕΜΜΥΔΙΑ ΚΙΛΟΥ	3,700	4,495	+0,796	+21,50
ΕΝΕΡΓΕΙΑΚΑ ΠΟΤΑ	1,703	2,498	+0,795	+46,68
ΕΠΙΔΟΡΠΙΑ	1,841	2,593	+0,753	+40,89
ΑΠΕΡΙΤΙΦ	1,859	2,413	+0,554	+29,78
ΧΟΙΡΙΝΟ ΝΩΠΟ	6,772	7,266	+0,494	+7,30
ΝΕΡΑ	14,592	15,077	+0,485	+3,32
ΕΛΙΕΣ ΚΙΛΟΥ	2,145	2,615	+0,470	+21,89
ΠΡΟΪΟΝΤΑ ΖΥΜΗΣ ΓΛΥΚΑ	7,047	7,441	+0,394	+5,60
ΒΟΕΙΟ ΝΩΠΟ	7,881	8,220	+0,339	+4,30

Η μεγαλύτερη αύξηση σε απόλυτους όρους καταγράφηκε στην κατηγορία «Έτοιμο Φαγητό Catering», της οποίας το support αυξήθηκε από 2,379% σε 3,448%, δηλαδή κατά 1,069 ποσοστιαίες μονάδες. Ακολούθησαν τα «Κρεμμύδια Κιλού» και τα «Ενεργειακά Ποτά», με αύξηση περίπου 0,80 ποσοστιαίων μονάδων. Η σχετική μεταβολή των ενεργειακών ποτών ήταν υψηλότερη, καθώς έφθασε το 46,68%, λόγω του χαμηλότερου αρχικού ποσοστού υποστήριξης. Το αποτέλεσμα αυτό δείχνει ότι η απόλυτη και η σχετική μεταβολή πρέπει να εξετάζονται συνδυαστικά.

Οι δέκα μεγαλύτερες μειώσεις του support μεταξύ των κοινών κατηγοριών παρουσιάζονται στον Πίνακα 3.4.

Πίνακας 3.4: Κατηγορίες με τις δέκα μεγαλύτερες μειώσεις του support μεταξύ των δύο περιόδων

Κατηγορία	Support COVID (%)	Support Post-COVID (%)	Δ Support (π.μ.)	Σχετική μεταβολή (%)
ΓΑΛΑ ΦΡΕΣΚΟ	28,570	24,049	-4,521	-15,83
ΓΙΑΟΥΡΤΙΑ	20,366	17,377	-2,989	-14,68
ΑΦΡΟΛΟΥΤΡΑ	6,047	3,288	-2,759	-45,63
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	13,629	11,089	-2,541	-18,64
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	16,603	14,231	-2,372	-14,28
ΠΑΡΑΦΑΡΜΑΚΕΥΤΙΚΑ ΕΙΔΗ	3,247	0,968	-2,278	-70,18
ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	22,110	20,271	-1,839	-8,32
ΕΙΔΗ ΖΑΧΑΡΟΠΛΑΣΤΙΚΗΣ	6,963	5,191	-1,772	-25,45
ΚΑΦΕΔΕΣ	12,588	10,821	-1,767	-14,04
ΨΩΜΙΑ ΣΥΣΚΕΤΑΣΜΕΝΑ	12,329	10,620	-1,709	-13,86

Η μεγαλύτερη απόλυτη μείωση καταγράφηκε στο «Φρέσκο Γάλα», του οποίου το support μειώθηκε από 28,570% σε 24,049%, παρουσιάζοντας διαφορά -4,521 ποσοστιαίων μονάδων και σχετική μεταβολή -15,83%. Ακολούθησαν τα «Γιαούρτια» και τα «Αφρόλουτρα», με μειώσεις 2,989 και 2,759 ποσοστιαίων μονάδων αντίστοιχα. Σημαντική ήταν και η μεταβολή στα «Παραφαρμακευτικά Είδη», των οποίων το support μειώθηκε από 3,247% σε 0,968%. Η απόλυτη μείωση ήταν 2,278 ποσοστιαίες μονάδες, ενώ, σε σχέση με το αρχικό support της κατηγορίας, αντιστοιχεί σε σχετική μείωση 70,18%. Η τιμή αυτή εκφράζει την αναλογική μεταβολή και δεν πρέπει να συγχέεται με την απόλυτη διαφορά του support. Η σημαντικά υψηλότερη παρουσία των «Παραφαρμακευτικών Ειδών» κατά την περίοδο COVID-19 θα μπορούσε να συνδέεται με το αυξημένο ενδιαφέρον των καταναλωτών για προϊόντα υγείας και προστασίας κατά τη συγκεκριμένη περίοδο. Ωστόσο, η διαθέσιμη ανάλυση δεν επιτρέπει την αιτιώδη απόδοση της μεταβολής στην πανδημία, καθώς η συγκεκριμένη κατηγορία περιλαμβάνει διαφορετικά είδη προϊόντων και ενδέχεται να επηρεάστηκε και από μεταβολές στη διαθεσιμότητα, στην ταξινόμηση ή στις γενικότερες συνθήκες της αγοράς.

Συνολικά, τα αποτελέσματα δείχνουν ότι οι μεταβολές δεν ήταν ομοιόμορφες για όλες τις κατηγορίες. Ορισμένες κατηγορίες αύξησαν τη συμμετοχή τους στις συναλλαγές κατά τη μεταπανδημική περίοδο, ενώ αρκετές κατηγορίες με υψηλό αρχικό support παρουσίασαν μείωση. Οι διαφοροποιήσεις αυτές αποτυπώνουν αλλαγές στη σύνθεση των αγορών, χωρίς όμως να μπορούν να αποδοθούν αποκλειστικά στην πανδημία, καθώς μεταξύ των δύο περιόδων μεταβλήθηκαν και άλλες οικονομικές και κοινωνικές συνθήκες. Για να διερευνηθεί εάν η γενικότερη εικόνα συνδέεται και με διαφορά στο πλήθος των κατηγοριών που περιλαμβάνονται ανά συναλλαγή, στην επόμενη υποενότητα εξετάζεται το μέσο μέγεθος του καλαθιού.

3.4.5 Σύγκριση του μέσου μεγέθους καλαθιού

Μετά τη σύγκριση των ποσοστών υποστήριξης, εξετάστηκε το μέσο μέγεθος του καλαθιού στις δύο χρονικές περιόδους. Σκοπός της ανάλυσης ήταν να διερευνηθεί εάν ο μεγαλύτερος αριθμός συχνών στοιχειοσυνόλων που εντοπίστηκε κατά την περίοδο COVID-19 μπορεί να σχετίζεται και με μια γενικότερη διαφορά στο πλήθος των κατηγοριών που περιλαμβάνονται στις συναλλαγές.

Όπως αναφέρθηκε προηγουμένως, ως μέγεθος καλαθιού ορίστηκε το πλήθος των διαφορετικών κατηγοριών προϊόντων ανά συναλλαγή και όχι το πλήθος των τεμαχίων. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 3.5.

Πίνακας 3.5: Σύγκριση του μεγέθους καλαθιού μεταξύ των δύο περιόδων

Περίοδος	Πλήθος συναλλαγών	Μέσο μέγεθος καλαθιού	Διάμεσο μέγεθος καλαθιού
COVID-19	4.408.281	7,1085	5
Μεταπανδημική	5.101.823	6,4014	5

Κατά την περίοδο COVID-19, κάθε συναλλαγή περιλάμβανε κατά μέσο όρο 7,1085 διαφορετικές κατηγορίες προϊόντων, ενώ κατά τη μεταπανδημική περίοδο ο αντίστοιχος μέσος όρος μειώθηκε σε 6,4014 κατηγορίες. Η διαφορά ανέρχεται σε περίπου 0,71 κατηγορίες ανά συναλλαγή και αντιστοιχεί σε σχετική μείωση 9,95%.¹

¹Η απόλυτη διαφορά και η σχετική μεταβολή υπολογίστηκαν με βάση τις μη στρογγυλοποιημένες τιμές των μέσων όρων.

Παρότι το διάμεσο μέγεθος του καλαθιού παρέμεινε στις πέντε κατηγορίες και στις δύο περιόδους, ο μέσος όρος παρουσίασε μείωση. Το αποτέλεσμα δείχνει ότι η κεντρική θέση της κατανομής δεν μεταβλήθηκε, ενώ η συνολική κατανομή των μεγεθών των καλαθιών διαφοροποιήθηκε. Η υψηλότερη μέση τιμή κατά την περίοδο COVID-19 είναι συμβατή με συχνότερη παρουσία καλαθιών που περιλαμβάνουν περισσότερες κατηγορίες, χωρίς όμως η ακριβής μορφή της κατανομής να μπορεί να προσδιοριστεί αποκλειστικά από τον μέσο όρο και τη διάμεσο.

Το εύρημα αυτό μπορεί να εξηγήσει εν μέρει τον μεγαλύτερο αριθμό συχνών στοιχειοσυνόλων που εντοπίστηκε κατά την περίοδο COVID-19. Όταν ένα καλάθι περιλαμβάνει περισσότερες διαφορετικές κατηγορίες, δημιουργούνται περισσότεροι πιθανοί συνδυασμοί προϊόντων και αυξάνεται η πιθανότητα κάποιοι από αυτούς να υπερβούν το ελάχιστο όριο υποστήριξης. Συνεπώς, τα περισσότερα frequent itemsets της περιόδου COVID-19 δεν αποτελούν από μόνα τους ένδειξη ισχυρότερων συσχετίσεων, αλλά ενδέχεται να συνδέονται εν μέρει και με το μεγαλύτερο μέσο μέγεθος των καλαθιών.

3.4.6 Σύνδεση με την εξόρυξη κανόνων συσχέτισης

Η διερευνητική ανάλυση έδειξε ότι αρκετές από τις βασικές κατηγορίες προϊόντων διατηρούν υψηλά ποσοστά υποστήριξης και στις δύο χρονικές περιόδους. Ωστόσο, η συνολική σύγκριση των 315 κοινών 1-itemsets ανέδειξε μια γενικότερη υποχώρηση των ποσοστών υποστήριξης κατά τη μεταπανδημική περίοδο. Συγκεκριμένα, 229 κατηγορίες (72,70%) παρουσίασαν χαμηλότερο support, ενώ 86 κατηγορίες (27,30%) παρουσίασαν υψηλότερο support. Η τάση ήταν ακόμη σαφέστερη στις 124 κοινές κατηγορίες με support τουλάχιστον 1% σε μία από τις δύο περιόδους, καθώς 102 κατηγορίες (82,26%) παρουσίασαν μείωση και 22 κατηγορίες (17,74%) αύξηση. Παρά τη γενικότερη αυτή υποχώρηση, οι μεταβολές δεν ήταν ομοιόμορφες για όλες τις κατηγορίες.

Παράλληλα, το μεγαλύτερο μέσο μέγεθος του καλαθιού κατά την περίοδο COVID-19 μπορεί να εξηγήσει εν μέρει τον μεγαλύτερο αριθμό συχνών στοιχειοσυνόλων που εντοπίστηκε στη συγκεκριμένη περίοδο. Οι επιμέρους αυξήσεις και μειώσεις του support υποδεικνύουν επίσης διαφοροποιήσεις στη σύνθεση των αγορών. Τα ευρήματα αυτά δεν συνεπάγονται ισχυρότερες συσχετίσεις ούτε μπορούν να αποδοθούν αποκλειστικά στην πανδημία, καθώς ενδέχεται να σχετίζονται και με άλλους οικονομικούς, κοινωνικούς και εμπορικούς παράγοντες.

Παρά τη χρησιμότητά τους, το support των επιμέρους κατηγοριών και το μέγεθος του καλαθιού δεν μπορούν να αποκαλύψουν ποια προϊόντα αγοράζονται μαζί ούτε να αξιολογήσουν την ισχύ της μεταξύ τους σχέσης. Η ανάλυση των 1-itemsets απαντά κυρίως στο ερώτημα «ποιες κατηγορίες εμφανίζονται συχνότερα», αλλά όχι στο ερώτημα «ποιοι συνδυασμοί κατηγοριών εμφανίζονται συστηματικά στις ίδιες συναλλαγές».

Για τον λόγο αυτό, στο επόμενο μέρος εφαρμόζονται τεχνικές εξόρυξης συχνών στοιχειοσυνόλων και κανόνων συσχέτισης. Αρχικά διερευνάται η κατάλληλη τιμή της παραμέτρου minimum support με τη χρήση του FP-Growth, ο οποίος επιλέχθηκε για τη συγκεκριμένη διαδικασία λόγω της υπολογιστικής του αποδοτικότητας. Στη συνέχεια παρουσιάζονται η παραμετροποίηση και η εφαρμογή του Apriori, καθώς και η αξιολόγηση των παραγόμενων κανόνων με βάση τις μετρικές support, confidence, lift, leverage και conviction.

3.5 Διερεύνηση της Παραμέτρου Minimum Support

Η επιλογή κατάλληλης τιμής για τον προσδιορισμό της παραμέτρου minimum support επηρεάζει άμεσα τον αριθμό των παραγόμενων frequent itemsets αλλά και το υπολογιστικό κόστος της διαδικασίας εξόρυξης.

Πριν από την εκτέλεση της κύριας πειραματικής διαδικασίας πραγματοποιήθηκε ανάλυση της ευαισθησίας της συγκεκριμένης παραμέτρου προκειμένου να εντοπιστεί μια τιμή που μπορεί να προσφέρει ισορροπία μεταξύ πληρότητας αποτελεσμάτων και υπολογιστικής αποδοτικότητας. Ο πλήρης κώδικας που χρησιμοποιήθηκε για την ανάλυση ευαισθησίας της παραμέτρου minimum support παρουσιάζεται στην Ενότητα Α'2 του Παραρτήματος Α.

Σε αυτό το πείραμα χρησιμοποιήθηκε ο αλγόριθμος FP-Growth. Η επιλογή του συγκεκριμένου αλγορίθμου βασίστηκε στην υψηλή υπολογιστική του αποδοτικότητα σε μεγάλα σύνολα δεδομένων, επιτρέποντας την εκτέλεση πολλαπλών επαναληπτικών πειραμάτων σε σημαντικά μικρότερο χρόνο σε σχέση με άλλους αλγορίθμους εξόρυξης συχνών προτύπων. Σκοπός της διαδικασίας δεν ήταν η τελική εξαγωγή κανόνων συσχέτισης αλλά η διερεύνηση του κατάλληλου εύρους τιμών της παραμέτρου minimum support. Πραγματοποιήθηκαν διαδοχικές εκτελέσεις για διαφορετικές τιμές της παραμέτρου και για κάθε εκτέλεση καταγράφηκαν ο αριθμός των παραγόμενων frequent itemsets και ο απαιτούμενος χρόνος εκτέλεσης. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 3.6.

Πίνακας 3.6: Επίδραση της παραμέτρου minimum support στον αριθμό frequent itemsets και στον χρόνο εκτέλεσης του FP-Growth

Minimum Support	Frequent Itemsets	Χρόνος Εκτέλεσης (sec)
0,001	183.916	1.754
0,002	42.613	964
0,005	6.611	510
0,010	1.687	268
0,015	761	211
0,020	440	195
0,050	81	102

Από τα αποτελέσματα του Πίνακα 3.6 παρατηρείται ότι οι πολύ χαμηλές τιμές support οδηγούν σε πολύ έντονη αύξηση του αριθμού των παραγόμενων frequent itemsets. Ενδεικτικά, η μείωση του support από 0,002 σε 0,001 είχε ως αποτέλεσμα την αύξηση των συχνών συνόλων από 42.613 σε 183.916, δηλαδή περισσότερο από τετραπλάσιο αριθμό αποτελεσμάτων. Ταυτόχρονα, παρατηρείται σημαντική αύξηση του απαιτούμενου χρόνου εκτέλεσης.

Αντίθετα, η αύξηση της τιμής του support οδηγεί σε σημαντική μείωση του αριθμού των εξαγόμενων προτύπων. Για τιμές μεγαλύτερες του 0,02 ο αριθμός των frequent itemsets περιορίζεται σημαντικά. Αυτό αυξάνει την πιθανότητα αποκλεισμού ενδιαφερουσών συσχετίσεων και καταναλωτικών προτύπων.

Η ανάλυση των αποτελεσμάτων ανέδειξε την ύπαρξη ενός σημείου ισορροπίας (elbow point) στην περιοχή μεταξύ 0,01 και 0,015. Στην περιοχή αυτή παρατηρείται σημαντική μείωση του υπολογιστικού κόστους σε σχέση με μικρότερες τιμές support, χωρίς όμως να χάνεται μεγάλο μέρος της πληροφορίας που περιέχουν τα δεδομένα. Για τιμές μικρότερες του 0,01 ο αριθμός των frequent itemsets αυξάνεται δυσανάλογα, ενώ για τιμές μεγαλύτερες του 0,015 η πληροφορία που εξάγεται

από τα δεδομένα αρχίζει να περιορίζεται αισθητά.

Με βάση τα παραπάνω επιλέχθηκε η τιμή $\text{minimum support} = 0,01$ για την κύρια πειραματική διαδικασία. Η συγκεκριμένη τιμή παρήγαγε 1.687 frequent itemsets, αριθμό επαρκή για την εξαγωγή ουσιαστικών κανόνων συσχέτισης, ενώ διατήρησε το υπολογιστικό κόστος σε αποδεκτά επίπεδα. Επιπρόσθετα, η τιμή αυτή βρίσκεται εντός της περιοχής ισορροπίας που προέκυψε από την ανάλυση ευαισθησίας, αποτελώντας τον βέλτιστο συμβιβασμό μεταξύ πληρότητας αποτελεσμάτων και αποδοτικότητας της διαδικασίας εξόρυξης.

3.6 Αρχική Παραμετροποίηση του Αλγορίθμου Apriori

Το επόμενο στάδιο της έρευνας περιελάμβανε την πρώτη εφαρμογή του αλγορίθμου Apriori με σκοπό την εξαγωγή κανόνων συσχέτισης από το σύνολο δεδομένων της περιόδου COVID-19. Η παραγωγή των κανόνων βασίστηκε αρχικά στις δύο πιο διαδεδομένες μετρικές αξιολόγησης κανόνων συσχέτισης, το confidence και το lift. Η μετρική confidence χρησιμοποιήθηκε για την εκτίμηση της πιθανότητας εμφάνισης του consequent δεδομένης της παρουσίας του antecedent, ενώ η μετρική lift για την αξιολόγηση της ισχύος της συσχέτισης σε σχέση με την ανεξάρτητη εμφάνιση των προϊόντων. Για την πρώτη πειραματική προσέγγιση χρησιμοποιήθηκαν οι παράμετροι: Minimum Support = 0,01 - Minimum Confidence = 0,30 - Minimum Lift = 1,20 Τα αποτελέσματα παρουσιάζονται στον Πίνακα 3.7.

Πίνακας 3.7: Αντιπροσωπευτικοί κανόνες συσχέτισης με υψηλές τιμές lift κατά τη διερεύνηση των παραμέτρων του αλγορίθμου Apriori

Antecedent	Consequent	Support	Confidence	Lift
KENA	ΚΡΑΣΙΑ	0,0107	0,4053	7,67
KENA	ΜΠΥΡΕΣ	0,0159	0,6060	6,82
ΠΡΩΙΝΑ, ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΓΑΛΑ ΦΡΕΣΚΟ	0,0107	0,4290	7,25
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΓΙΑΟΥΡΤΙΑ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΓΑΛΑ ΦΡΕΣΚΟ	0,0178	0,4189	5,86
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΖΥΜΑΡΙΚΑ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΓΑΛΑ ΦΡΕΣΚΟ	0,0102	0,4165	5,82
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, SNACK	ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ, ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	0,0106	0,4203	5,30
ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ, ΤΥΡΙΑ ΔΙΑΦΟΡΑ ΣΥΣΚ	ΑΛΛΑΝΤΙΚΑ ΣΥΣΚΕΥΑΣΜΕΝΑ ΨΥΓΕΙΟΥ	0,0126	0,4485	5,10
ΔΙΑΦΟΡΑ ΑΡΤΟΣΚΕΥΑΣΜΑΤΑ, ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	0,0108	0,6973	5,12
ΑΛΛΑΝΤΙΚΑ ΣΥΣΚΕΥΑΣΜΕΝΑ ΨΥΓΕΙΟΥ, ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ	ΤΥΡΙΑ ΔΙΑΦΟΡΑ ΣΥΣΚ	0,0126	0,4824	4,82
ΠΡΩΙΝΑ, ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	0,0161	0,6489	4,76

Από τους κανόνες που προέκυψαν κατά την πρώτη εκτέλεση παρατηρήθηκε ότι αρκετές κατηγορίες προϊόντων εμφανίζονταν επανειλημμένα σε διαφορετικούς συνδυασμούς. Ιδιαίτερα συχνή ήταν η παρουσία των κατηγοριών «Αλλαντικά/Κιλό», «Μαλακά Τυριά/Κιλό», «Γάλα Φρέσκο» και «Ψωμιά Συσκευασμένα». Στα αποτελέσματα εμφανιζόταν συχνά και η κατηγορία «Είδη Συσκευασίας προς Πώληση», λόγω του υψηλού συνολικού ποσοστού υποστήριξής της. Ωστόσο, επειδή αφορά κυρίως σακούλες μεταφοράς, η παρουσία της δεν ερμηνεύεται ως ένδειξη ουσιαστικής καταναλωτικής προτίμησης.

Για λόγους συντομίας, στον Πίνακα 3.7 παρουσιάζονται δέκα αντιπροσωπευτικοί κανόνες με υψηλές τιμές support, confidence και lift. Η εξέταση του συνόλου των κανόνων έδειξε ότι οι παραπάνω κατηγορίες τροφίμων συμμετείχαν σε μεγάλο αριθμό συσχετίσεων, αναδεικνύοντας επαναλαμβανόμενα πρότυπα συνδυαστικής αγοράς.

Επίσης, διαπιστώθηκε ότι μεγάλο μέρος των αποτελεσμάτων αφορούσε παραλλαγές των ίδιων βασικών σχέσεων. Για παράδειγμα, εμφανίζονταν πολλαπλοί κανόνες που περιλάμβαναν τις κατηγορίες «Αλλαντικά/Κιλό», «Μαλακά Τυριά/Κιλό» και «Γάλα Φρέσκο», με μικρές διαφοροποιήσεις ως προς την προσθήκη επιπλέον προϊόντων στο antecedent ή στο consequent. Το γεγονός αυτό υποδηλώνει ότι η τιμή του confidence ήταν σχετικά χαμηλή και επέτρεπε την παραγωγή μεγάλου αριθμού παρόμοιων κανόνων.

Για τον λόγο αυτό πραγματοποιήθηκε δεύτερη εκτέλεση αυξάνοντας την τιμή της παραμέτρου confidence από 0,30 σε 0,40, διατηρώντας σταθερή την τιμή του lift. Πιο συγκεκριμένα, οι παράμετροι της εκτέλεσης ήταν:

- Minimum Support = 0,01
- Minimum Confidence = 0,40
- Minimum Lift = 1,20

Τα αποτελέσματα των δύο πρώτων πειραμάτων παρουσιάζονται στον Πίνακα 3.8. Σκοπός της δεύτερης εκτέλεσης ήταν η διερεύνηση της επίδρασης της παραμέτρου confidence στον αριθμό και στην ποιότητα των παραγόμενων κανόνων συσχέτισης.

Πίνακας 3.8: Επίδραση της παραμέτρου confidence στον αριθμό των παραγόμενων κανόνων

Πείραμα	Support	Confidence	Lift	Κανόνες
P1	0,01	0,30	1,20	895
P2	0,01	0,40	1,20	745

Όπως φαίνεται στον Πίνακα 3.8, η αύξηση της τιμής του confidence από 0,30 σε 0,40 οδήγησε σε μείωση του αριθμού των παραγόμενων κανόνων από 895 σε 745, δηλαδή κατά περίπου 16,8%. Η μείωση αυτή φανερώνει ότι ένα μέρος των κανόνων που προέκυπταν στην πρώτη εκτέλεση δεν πληρούσε το αυστηρότερο κριτήριο αξιοπιστίας που επιβλήθηκε με τη χρήση υψηλότερης τιμής για confidence.

Παρότι η αύξηση αυτή οδήγησε σε μείωση του συνολικού αριθμού των παραγόμενων κανόνων, η εξέταση των αποτελεσμάτων έδειξε ότι οι κανόνες με τις υψηλότερες τιμές lift παρέμειναν σε μεγάλο βαθμό οι ίδιοι με εκείνους που προέκυψαν στο προηγούμενο πείραμα. Οι κυρίαρχες κατηγορίες προϊόντων, όπως τα «Αλλαντικά/Κιλό», τα «Μαλακά Τυριά/Κιλό», το «Γάλα Φρέσκο» και τα «Ψωμιά Συσκευασμένα», συνέχισαν να συμμετέχουν στους ισχυρότερους κανόνες συσχέτισης.

Για τον λόγο αυτό δεν παρουσιάζεται εκ νέου πίνακας αντιπροσωπευτικών κανόνων, καθώς τα αποτελέσματα δεν διαφοροποιούνται ουσιαστικά από εκείνα του προηγούμενου πειράματος. Αντίθετα, η ανάλυση επικεντρώνεται στη μεταβολή του πλήθους και της ποιότητας των κανόνων, η οποία αποτυπώνει αποτελεσματικότερα την επίδραση της παραμέτρου confidence.

Η παρατήρηση αυτή υποδηλώνει ότι η αύξηση της τιμής του confidence, αν και περιορίσε τον συνολικό αριθμό κανόνων, δεν ήταν επαρκής για την εξάλειψη των επαναλαμβανόμενων μοτίβων που κυριαρχούσαν στα αποτελέσματα. Ως εκ τούτου, κρίθηκε απαραίτητη η περαιτέρω αυστηροποίηση των κριτηρίων αξιολόγησης μέσω της αύξησης της παραμέτρου lift και της μετέπειτα ενσωμάτωσης των μετρικών leverage και conviction.

3.7 Βελτίωση της Διαδικασίας Φιλτραρίσματος Κανόνων

Η ανάλυση των αρχικών αποτελεσμάτων ανέδειξε ότι η αποκλειστική χρήση των μετρικών confidence και lift δεν ήταν επαρκής για την αποτελεσματική αξιολόγηση και ιεράρχηση των παραγόμενων κανόνων συσχέτισης. Παρότι οι συγκεκριμένες μετρικές επέτρεπαν τον αποκλεισμό κανόνων χαμηλής αξιοπιστίας, εξακολουθούσε να παρατηρείται μεγάλος αριθμός επαναλαμβανόμενων και ιδιαίτερα παρόμοιων κανόνων, δυσχεραίνοντας την ερμηνεία των αποτελεσμάτων.

Για την αντιμετώπιση του προβλήματος αυτού πραγματοποιήθηκε αναθεώρηση της διαδικασίας του φιλτραρίσματος και αξιολόγησης των κανόνων. Αρχικά ενσωματώθηκαν δύο επιπλέον μετρικές ποιότητας, η leverage και η conviction. Η μετρική leverage χρησιμοποιήθηκε για την αξιολόγηση του βαθμού απόκλισης της πραγματικής συν-εμφάνισης δύο προϊόντων από την αναμενόμενη συν-εμφάνισή τους υπό συνθήκες στατιστικής ανεξαρτησίας. Έτσι δίνεται έμφαση σε συσχετίσεις οι οποίες παρουσιάζουν ουσιαστική διαφοροποίηση από την τυχαία συμπεριφορά. Η μετρική conviction χρησιμοποιήθηκε για να γίνει εκτίμηση της κατευθυντικότητας και της αξιοπιστίας ενός κανόνα, λαμβάνοντας υπόψη τις περιπτώσεις στις οποίες ο κανόνας δεν επαληθεύεται.

Με την εισαγωγή των νέων μετρικών αξιολόγησης, κρίθηκε απαραίτητη και η αναθεώρηση της διαδικασίας παραγωγής των κανόνων συσχέτισης. Η διερεύνηση των αρχικών αποτελεσμάτων έδειξε ότι σημαντικό ποσοστό των παραγόμενων κανόνων αποτελούσε ουσιαστικά διαφορετικές εκδοχές της ίδιας συσχέτισης μεταξύ προϊόντων. Ως αποτέλεσμα, ο συνολικός αριθμός των κανόνων αυξανόταν σημαντικά χωρίς να προστίθεται αντίστοιχη αναλυτική αξία.

Έτσι δόθηκε έμφαση στην ανάδειξη των απλούστερων και περισσότερο ερμηνεύσιμων σχέσεων μεταξύ των προϊόντων. Στόχος δεν ήταν η παραγωγή όσο το δυνατόν περισσότερων κανόνων, αλλά ο εντοπισμός εκείνων των συσχετίσεων που μπορούν να περιγράψουν με σαφήνεια τις αγοραστικές συνήθειες των καταναλωτών και να υποστηρίξουν τη συγκριτική ανάλυση των δύο χρονικών περιόδων. Κατά τη διαδικασία αξιολόγησης διαπιστώθηκε ότι πολλοί σύνθετοι κανόνες δεν παρείχαν ουσιαστικά νέα γνώση, αλλά αποτελούσαν επεκτάσεις των ήδη γνωστών συσχετίσεων με την προσθήκη επιπλέον προϊόντων. Για παράδειγμα, όταν μία ισχυρή σχέση μεταξύ δύο προϊόντων είχε ήδη εντοπιστεί, η εμφάνιση πολλών παραλλαγών της ίδιας σχέσης με τη συμμετοχή τρίτων προϊόντων οδηγούσε κυρίως σε αύξηση της πολυπλοκότητας των αποτελεσμάτων και όχι σε βαθύτερη κατανόηση της καταναλωτικής συμπεριφοράς.

Με βάση τα παραπάνω, η διαδικασία φιλτραρίσματος προσαρμόστηκε έτσι ώστε να δίνεται προτεραιότητα στους πλέον αντιπροσωπευτικούς και ουσιαστικούς κανόνες, περιορίζοντας την παρουσία πλεοναζόντων συσχετίσεων. Η προσέγγιση αυτή οδήγησε σε ένα περισσότερο συμπαγές και ευκολότερα ερμηνεύσιμο σύνολο αποτελεσμάτων, το οποίο κρίθηκε καταλληλότερο για την εξαγωγή συμπερασμάτων σχετικά με τις μεταβολές της καταναλωτικής συμπεριφοράς μεταξύ της περιόδου COVID-19 και της μεταπανδημικής περιόδου.

Τέλος, οι κανόνες ταξινομήθηκαν με βάση τις μετρικές leverage και lift, δίνοντας προτεραιότητα σε συσχετίσεις που παρουσιάζουν τόσο υψηλή ισχύ όσο και ουσιαστική στατιστική σημασία. Η προσέγγιση αυτή περιορίσε σημαντικά την παρουσία πλεοναζόντων κανόνων και διαμόρφωσε ένα σύνολο αποτελεσμάτων περισσότερο κατάλληλο για την αξιολόγηση και ερμηνεία των προτύπων καταναλωτικής συμπεριφοράς που εξετάζονται.

3.8 Διερεύνηση και Αξιολόγηση Παραμέτρων του Αλγορίθμου Apriori

Μετά το πέρας της αναθεωρημένης διαδικασίας φιλτραρίσματος και αξιολόγησης των κανόνων συσχέτισης, κρίθηκε απαραίτητη η συστηματική διερεύνηση των παραμέτρων του αλγορίθμου Apriori, προκειμένου να προσδιοριστεί ο κατάλληλος συνδυασμός παραμέτρων για την εξαγωγή των τελικών αποτελεσμάτων.

Η διαδικασία αυτή πραγματοποιήθηκε ξεχωριστά για τις δύο εξεταζόμενες χρονικές περιόδους, COVID-19 και μεταπανδημική περίοδο, με σκοπό να αξιολογηθεί η επίδραση των παραμέτρων στην ποσότητα και την ποιότητα των παραγόμενων κανόνων συσχέτισης. Για κάθε μία από τις δύο περιόδους εξετάστηκαν διαφορετικά πειραματικά σενάρια και συγκρίθηκαν τα αποτελέσματά τους με βάση τον αριθμό των συχνών συνόλων αντικειμένων, το πλήθος των κανόνων, τις μετρικές αξιολόγησης και τα υπολογιστικά χαρακτηριστικά της εκτέλεσης.

Στις επόμενες υποενότητες παρουσιάζονται αναλυτικά τα πειράματα που πραγματοποιήθηκαν και η διαδικασία επιλογής των τελικών παραμέτρων που χρησιμοποιήθηκαν στην κύρια ανάλυση.

3.8.1 Διερεύνηση και Αξιολόγηση Παραμέτρων του Αλγορίθμου Apriori για την Περίοδο COVID-19

Η διαδικασία εξόρυξης κανόνων συσχέτισης επηρεάζεται σημαντικά από τις τιμές των παραμέτρων που χρησιμοποιούνται για την εξαγωγή των κανόνων. Παρότι η τιμή της παραμέτρου minimum support είχε ήδη προσδιοριστεί μέσω της προηγούμενης ανάλυσης ευαισθησίας και ορίστηκε στο 0.01, παρέμενε αναγκαία η διερεύνηση της επίδρασης των μετρικών confidence, lift και leverage στην ποιότητα των παραγόμενων αποτελεσμάτων.

Για τον σκοπό αυτό πραγματοποιήθηκαν έξι διαφορετικά πειράματα με τον αλγόριθμο Apriori. Οι συνδυασμοί των παραμέτρων επιλέχθηκαν έτσι ώστε να καλύπτουν τόσο σχετικά χαλαρές όσο και αυστηρότερες συνθήκες εξαγωγής κανόνων προκειμένου να εντοπιστεί ένας συνδυασμός που να προσφέρει ισορροπία μεταξύ ποσότητας και ποιότητας των κανόνων συσχέτισης.

Η τελική φάση της διερεύνησης των παραμέτρων του αλγορίθμου Apriori πραγματοποιήθηκε μέσω έξι πειραμάτων, στα οποία εξετάστηκαν διαφορετικοί συνδυασμοί των μετρικών confidence, lift και leverage, διατηρώντας σταθερές τις τιμές των παραμέτρων minimum support και conviction. Στόχος της διαδικασίας ήταν η αξιολόγηση της επίδρασης κάθε παραμέτρου στο πλήθος και στην ποιότητα των παραγόμενων κανόνων συσχέτισης, καθώς και η επιλογή του καταλληλότερου συνδυασμού για την κύρια ανάλυση των δεδομένων.

Οι τιμές των παραμέτρων που εξετάστηκαν παρουσιάζονται στον Πίνακα 3.9.

Πίνακας 3.9: Συνδυασμοί παραμέτρων που χρησιμοποιήθηκαν στα πειράματα Apriori για την περίοδο COVID-19

Πείραμα	Support	Confidence	Lift	Leverage	Conviction
P1	0,01	0,20	1,20	0,005	1,0
P2	0,01	0,30	1,20	0,005	1,0
P3	0,01	0,20	1,30	0,005	1,0
P4	0,01	0,20	1,20	0,010	1,0
P5	0,01	0,30	1,30	0,005	1,0
P6	0,01	0,30	1,30	0,010	1,0

Σε όλες τις εκτελέσεις χρησιμοποιήθηκε το ίδιο σύνολο δεδομένων της περιόδου COVID-19, το οποίο περιλάμβανε 4.408.323 συναλλαγές και 323 μοναδικές κατηγορίες προϊόντων. Επιπλέον, εφαρμόστηκε η ίδια διαδικασία παραγωγής και επεξεργασίας των κανόνων. Επομένως, οι διαφοροποιήσεις των αποτελεσμάτων δεν προέρχονται από αλλαγές στα δεδομένα εισόδου, αλλά αντανακλούν την επίδραση των διαφορετικών κατωφλίων σε συνδυασμό με την κοινή διαδικασία μεταγενέστερου φιλτραρίσματος.

Μετά την εφαρμογή των καθορισμένων κατωφλίων πραγματοποιήθηκε συμπληρωματική επεξεργασία των παραγόμενων κανόνων. Αρχικά αφαιρέθηκαν οι αντίστροφοι κανόνες, διατηρώντας τον κανόνα που προηγούνταν βάσει των τιμών leverage, lift και confidence. Στη συνέχεια αφαιρέθηκαν σύνθετοι κανόνες όταν η σχέση μεταξύ κατηγοριών του antecedent και του consequent είχε ήδη αποτυπωθεί μέσω απλού κανόνα $1 \rightarrow 1$. Συνεπώς, το τελικό πλήθος των κανόνων διαμορφώθηκε τόσο από τα επιλεγμένα κατώφλια όσο και από τη συγκεκριμένη διαδικασία περιορισμού των αντίστροφων και πλεοναζόντων κανόνων.

Για κάθε πείραμα καταγράφηκαν το πλήθος των τελικών κανόνων συσχέτισης, το πλήθος των συχνών συνόλων αντικειμένων (frequent itemsets), ο χρόνος εκτέλεσης και η μέγιστη καταγεγραμμένη κατανάλωση μνήμης. Τα συγκεντρωτικά αποτελέσματα παρουσιάζονται στον Πίνακα 3.10.

Πίνακας 3.10: Συγκεντρωτικά αποτελέσματα των πειραμάτων Apriori για την περίοδο COVID-19

Πείραμα	Rules	Itemsets	Runtime (sec)	Memory (MB)
P1	648	1687	337,69	7101,98
P2	362	1687	338,53	7099,03
P3	654	1687	337,24	7101,06
P4	160	1687	341,37	7099,88
P5	354	1687	347,09	7099,48
P6	116	1687	339,24	7101,45

Αυτό που παρατηρείται από τον πίνακα είναι ότι το πλήθος των συχνών συνόλων αντικειμένων παρέμεινε σταθερό και ίσο με 1.687 itemsets σε όλα τα πειράματα. Το αποτέλεσμα αυτό ήταν αναμενόμενο, καθώς η παράμετρος minimum support διατηρήθηκε σταθερή σε όλες τις εκτελέσεις. Συνεπώς, οι μεταβολές που παρατηρούνται αφορούν αποκλειστικά τη διαδικασία εξαγωγής των κανόνων συσχέτισης επιβεβαιώνοντας τον προηγούμενο ισχυρισμό.

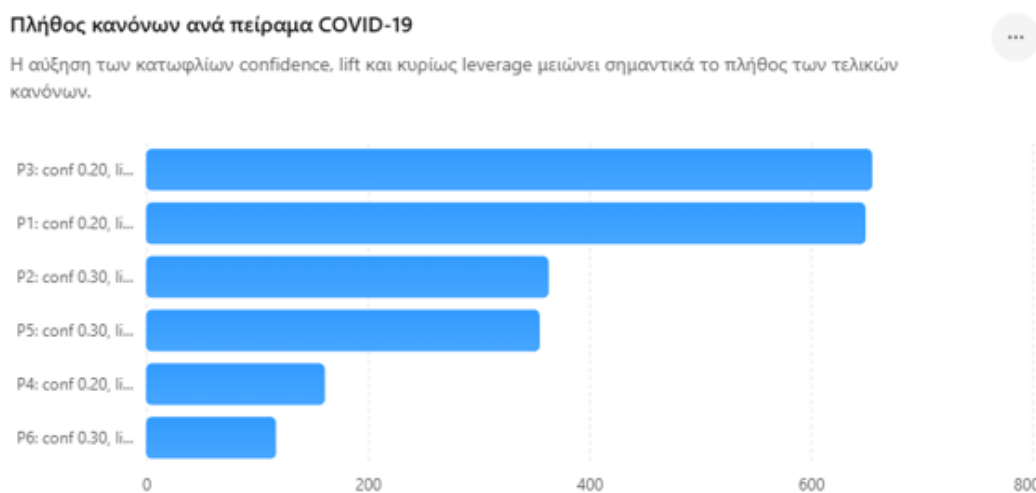
Αντίστοιχα, τόσο ο χρόνος εκτέλεσης όσο και η κατανάλωση μνήμης παρουσίασαν εξαιρετικά μικρές αποκλίσεις. Ο χρόνος εκτέλεσης κυμάνθηκε μεταξύ 337 και 347 δευτερολέπτων, ενώ η απαιτούμενη μνήμη παρέμεινε περίπου στα 7,1 GB. Το εύρημα αυτό δείχνει ότι οι συγκεκριμένες παράμετροι επηρεάζουν κυρίως το πλήθος των τελικών κανόνων και όχι την υπολογιστική πολυπλοκότητα της διαδικασίας.

Η σημαντικότερη διαφοροποίηση που προέκυψε μετά την εκτέλεση των πειραμάτων αφορά τον αριθμό των παραγόμενων κανόνων. Η αύξηση του confidence από 0.20 σε 0.30 οδήγησε σε σημαντική μείωση των κανόνων, το οποίο υποδεικνύει ότι μεγάλο ποσοστό των συσχετίσεων εμφανίζει πιθανότητα πραγματοποίησης μεταξύ 20% και 30%.

Η αύξηση του κατωφλίου του lift από 1,20 σε 1,30 περιορίσε, όπως αναμενόταν, το αρχικό σύνολο των κανόνων που πληρούσαν τα κριτήρια. Μετά, όμως, την αφαίρεση των αντίστροφων και πλεοναζόντων σύνθετων κανόνων, τα πειράματα P1 και P3 κατέληξαν σε 648 και 654 κανόνες αντί-

στοιχα. Η μικρή αυτή αντιστροφή δεν σημαίνει ότι το αυστηρότερο όριο του lift δημιούργησε περισσότερους κανόνες. Στο P1, το χαμηλότερο κατώφλι διατήρησε περισσότερους απλούς κανόνες, με αποτέλεσμα περισσότεροι σύνθετοι κανόνες να χαρακτηριστούν στη συνέχεια ως πλεονάζοντες και να αφαιρεθούν. Συνεπώς, τα τελικά πλήθη πρέπει να ερμηνεύονται σε συνδυασμό με τη διαδικασία μεταγενέστερου φιλτραρίσματος.

Η μεγαλύτερη επίδραση παρατηρήθηκε κατά την αύξηση της παραμέτρου leverage από 0.005 σε 0.010. Στις περιπτώσεις αυτές ο αριθμός των κανόνων μειώθηκε δραστικά. Το εύρημα αυτό οδηγεί στο συμπέρασμα ότι σημαντικό ποσοστό των κανόνων διαθέτει θετική συσχέτιση αλλά περιορισμένη πρακτική απόκλιση από την αναμενόμενη συχνότητα εμφάνισης.

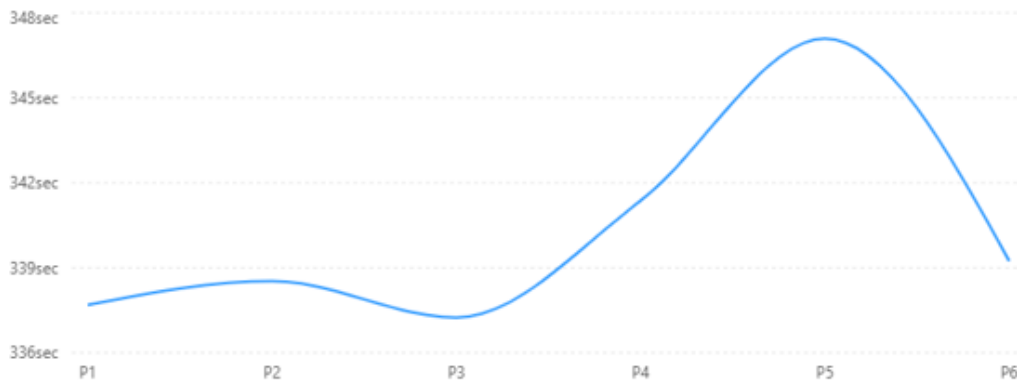


Σχήμα 3.3: Πλήθος κανόνων ανά πείραμα για την περίοδο COVID-19

Παρατηρείται ότι η αυστηροποίηση του leverage επιφέρει τη μεγαλύτερη μείωση του πλήθους των κανόνων. Τα πειράματα P4 και P6 παρουσιάζουν δραστικό περιορισμό των αποτελεσμάτων, αποδεικνύοντας ότι η συγκεκριμένη μετρική λειτουργεί ως ιδιαίτερα ισχυρό φίλτρο επιλογής.

Χρόνος εκτέλεσης ανά πείραμα

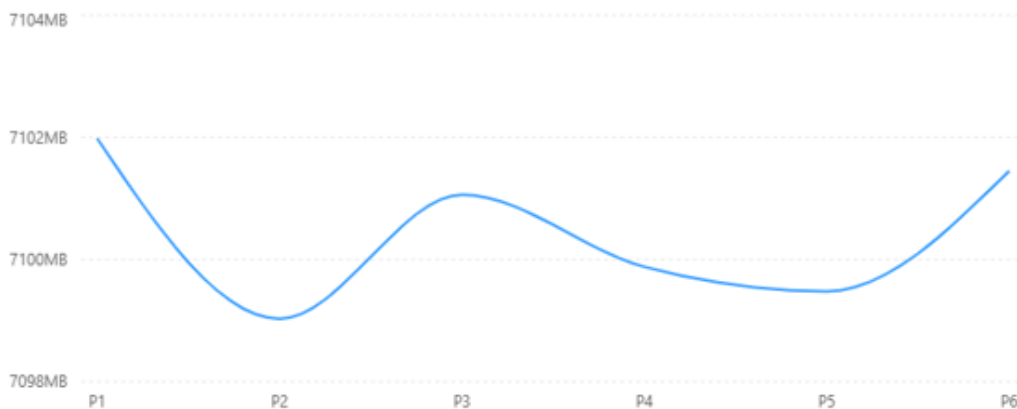
Οι χρόνοι εκτέλεσης παραμένουν σχεδόν σταθεροί, γεγονός που δείχνει ότι οι παράμετροι επηρεάζουν κυρίως το φιλτράρισμα των κανόνων.



Σχήμα 3.4: Χρόνος εκτέλεσης ανά πείραμα για την περίοδο COVID-19

Κατανάλωση μνήμης ανά πείραμα

Η κατανάλωση μνήμης παραμένει πρακτικά αμετάβλητη σε όλες τις εκτελέσεις.



Σχήμα 3.5: Κατανάλωση μνήμης ανά πείραμα για την περίοδο COVID-19

Σ'Όπως προκύπτει από τα δύο προηγούμενα σχήματα, οι διαφορές μεταξύ των πειραμάτων είναι περιορισμένες. Η διαδικασία δημιουργίας των frequent itemsets αποτελεί το κυρίαρχο υπολογιστικό στάδιο του Apriori και ως εκ τούτου οι διαφοροποιήσεις των κατωφλίων confidence, lift και leverage δεν επηρεάζουν σε σημαντικό βαθμό ούτε τον συνολικό χρόνο εκτέλεσης αλλά ούτε και την συνολική κατανάλωση μνήμης.

Συμπερασματικά, τα πειράματα P1 και P3 παρήγαγαν τα υψηλότερα και μεταξύ τους παρόμοια πλήθη τελικών κανόνων, 648 και 654 αντίστοιχα, ενώ το P6 παρήγαγε μόλις 116 κανόνες, περιορίζοντας σημαντικά τον όγκο της διαθέσιμης γνώσης.

Μεταξύ των δύο ακραίων περιπτώσεων, το πείραμα P5 παρουσιάζει την καλύτερη ισορροπία μεταξύ αυστηρότητας και πληρότητας αποτελεσμάτων. Συγκεκριμένα, διατηρεί υψηλά κατώφλια confidence και lift, απορρίπτοντας ασθενέστερες συσχετίσεις, ενώ ταυτόχρονα διατηρεί επαρκή

αριθμό κανόνων για ουσιαστική ανάλυση της καταναλωτικής συμπεριφοράς.

Εξαιτίας όσων προαναφέρθηκαν, το πείραμα P5 επιλέγεται ως η τελική διαμόρφωση του αλγορίθμου Apriori για την περίοδο COVID-19 και χρησιμοποιείται στην επόμενη ενότητα για τη λεπτομερή ανάλυση των κανόνων.

3.8.2 Διερεύνηση και Αξιολόγηση Παραμέτρων του Αλγορίθμου Apriori για τη Μεταπανδημική Περίοδο

Μετά την ολοκλήρωση της διαδικασίας επιλογής παραμέτρων για την περίοδο COVID-19, ακολουθήθηκε η ίδια πειραματική διαδικασία και για το σύνολο δεδομένων της μεταπανδημικής περιόδου (POST COVID).

Η τιμή της παραμέτρου minimum support διατηρήθηκε και σε αυτήν την περίπτωση σταθερή στο 0,01, σύμφωνα με τα αποτελέσματα της προηγούμενης ανάλυσης ευαισθησίας. Με τον τρόπο αυτό διασφαλίστηκε ότι οι διαφορές που παρατηρούνται μεταξύ των πειραμάτων οφείλονται αποκλειστικά στις παραμέτρους αξιολόγησης των κανόνων και όχι στη διαδικασία εξαγωγής των frequent itemsets.

Για τον σκοπό αυτό πραγματοποιήθηκαν και πάλι έξι διαφορετικά πειράματα με τον αλγόριθμο Apriori, χρησιμοποιώντας διαφορετικούς συνδυασμούς των παραμέτρων confidence, lift και leverage. Οι συνδυασμοί αυτοί παρουσιάζονται στον Πίνακα 3.11.

Πίνακας 3.11: Συνδυασμοί παραμέτρων που χρησιμοποιήθηκαν στα πειράματα Apriori για τη μεταπανδημική περίοδο

Πείραμα	Support	Confidence	Lift	Leverage	Conviction
P1	0,01	0,20	1,20	0,005	1,0
P2	0,01	0,30	1,20	0,005	1,0
P3	0,01	0,20	1,30	0,005	1,0
P4	0,01	0,20	1,20	0,010	1,0
P5	0,01	0,30	1,30	0,005	1,0
P6	0,01	0,30	1,30	0,010	1,0

Το σύνολο δεδομένων της μεταπανδημικής περιόδου περιλαμβάνει 5.101.828 συναλλαγές και 322 μοναδικές κατηγορίες προϊόντων. Η χρήση του ίδιου συνόλου δεδομένων σε όλα τα πειράματα επιτρέπει την αντικειμενική αξιολόγηση της επίδρασης των παραμέτρων στα τελικά αποτελέσματα.

Τα αποτελέσματα των έξι εκτελέσεων παρουσιάζονται στον Πίνακα 3.12.

Πίνακας 3.12: Συγκεντρωτικά αποτελέσματα των πειραμάτων Apriori για τη μεταπανδημική περίοδο

Πείραμα	Rules	Itemsets	Runtime (sec)	Memory (MB)
P1	367	1104	276,90	7762,70
P2	229	1104	278,14	7762,13
P3	354	1104	279,98	7763,21
P4	98	1104	307,38	7758,34
P5	216	1104	342,40	7762,17
P6	73	1104	302,76	7764,85

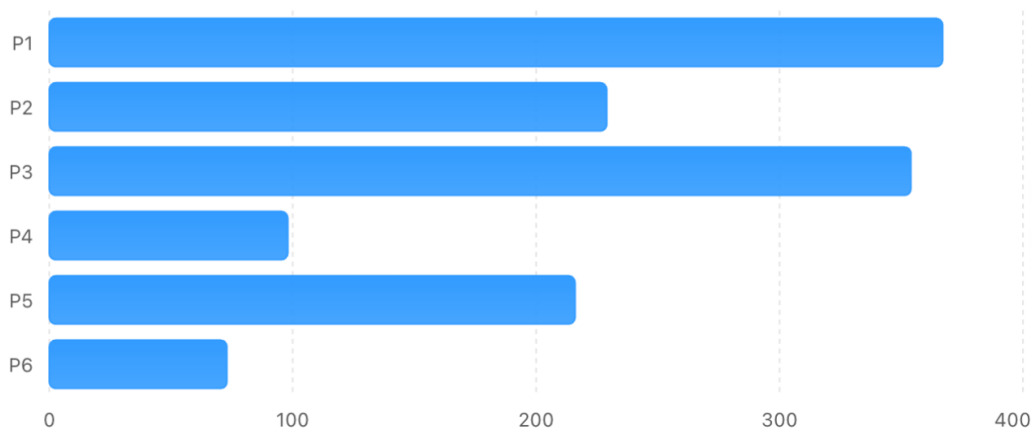
Παρατηρείται ότι το πλήθος των frequent itemsets παραμένει σταθερό και ίσο με 1.104 σε όλες τις εκτελέσεις. Όπως προαναφέραμε, το αποτέλεσμα αυτό ήταν αναμενόμενο, καθώς η παράμετρος minimum support παρέμεινε σταθερή καθιστώντας τις όποιες διαφοροποιήσεις που παρατηρούνται να αφορούν αποκλειστικά το στάδιο παραγωγής και φιλτραρίσματος των κανόνων συσχέτισης.

Αντίστοιχα, η κατανάλωση μνήμης παραμένει ουσιαστικά αμετάβλητη μεταξύ των πειραμάτων και κυμαίνεται περίπου στα 7,76 GB δείχνοντας ότι οι μεταβολές των παραμέτρων confidence, lift και leverage επηρεάζουν κυρίως το πλήθος των τελικών κανόνων και όχι τις απαιτήσεις μνήμης της διαδικασίας.

Παρόμοια εικόνα παρατηρείται και στον χρόνο εκτέλεσης ο οποίος παραμένει σχετικά σταθερός. Αυτό υποδηλώνει ότι το κυριότερο υπολογιστικό κόστος του αλγορίθμου συνδέεται με την εξαγωγή των frequent itemsets και όχι με το τελικό φιλτράρισμα των κανόνων.

Πλήθος κανόνων ανά πείραμα

Αριθμός κανόνων συσχέτισης που παρήχθησαν σε κάθε πείραμα της μεταπανδημικής περιόδου.



Σχήμα 3.6: Πλήθος κανόνων ανά πείραμα για τη μεταπανδημική περίοδο

Από το Σχήμα 3.6 προκύπτει ότι η αυστηροποίηση των παραμέτρων οδηγεί σε σημαντική μείωση του πλήθους των παραγόμενων κανόνων. Το πείραμα P1 παρήγαγε τον μεγαλύτερο αριθμό κανόνων (367), ενώ το πείραμα P6 τον μικρότερο (73).

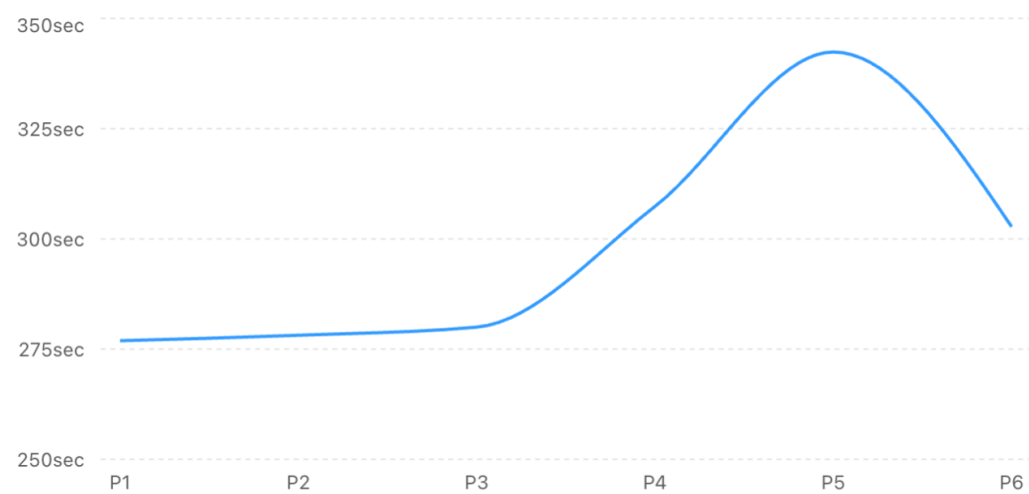
Η αύξηση της παραμέτρου confidence από 0.20 σε 0.30 οδηγεί σε αισθητή μείωση του πλήθους των κανόνων. Συγκεκριμένα, το P1 παρήγαγε 367 κανόνες ενώ το P2 229, που υποδηλώνει ότι σημαντικός αριθμός κανόνων παρουσιάζει confidence μεταξύ 20% και 30%.

Παράλληλα, η αύξηση της παραμέτρου lift από 1.20 σε 1.30 προκαλεί σχετικά περιορισμένη μείωση των κανόνων. Η μετάβαση από το P1 στο P3 μειώνει τους κανόνες από 367 σε 354, το οποίο δείχνει ότι οι περισσότερες συσχετίσεις διαθέτουν ήδη σχετικά υψηλή ισχύ.

Η μεγαλύτερη επίδραση παρατηρείται στην παράμετρο leverage. Όταν το κατώφλι αυξάνεται από 0.005 σε 0.010, το πλήθος των κανόνων μειώνεται δραστικά. Το αποτέλεσμα αυτό δείχνει ότι μεγάλο ποσοστό κανόνων διαθέτει θετική συσχέτιση, αλλά μικρότερη πρακτική σημαντικότητα ως προς την απόκλιση από την αναμενόμενη τυχαία συνύπαρξη.

Χρόνος εκτέλεσης ανά πείραμα

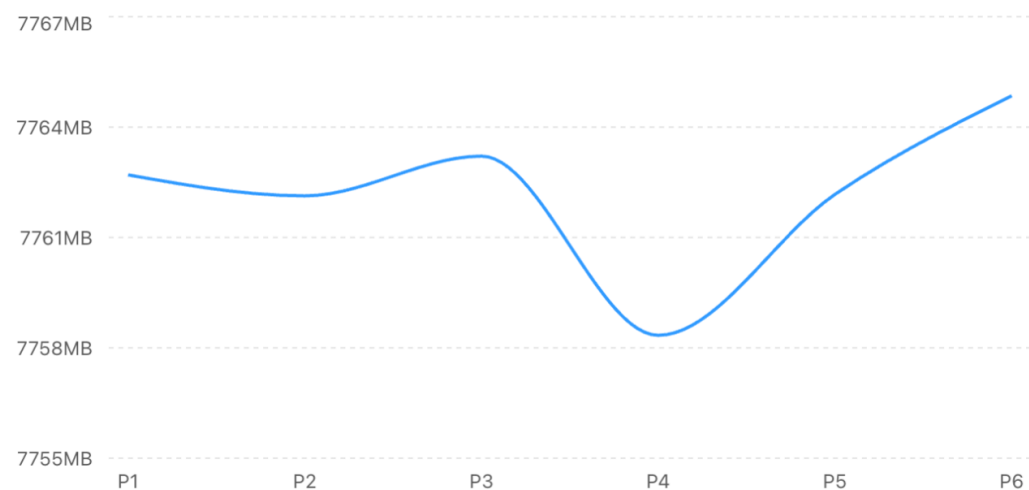
Οι χρόνοι εκτέλεσης ανά πείραμα για τη μεταπανδημική περίοδο.



Σχήμα 3.7: Χρόνος εκτέλεσης ανά πείραμα για τη μεταπανδημική περίοδο

Κατανάλωση μνήμης ανά πείραμα

Η κατανάλωση μνήμης παραμένει σχεδόν σταθερή σε όλα τα πειράματα.



Σχήμα 3.8: Κατανάλωση μνήμης ανά πείραμα για τη μεταπανδημική περίοδο

Η κατανάλωση μνήμης παραμένει πρακτικά σταθερή σε όλα τα πειράματα, δείχνοντας ότι οι διαφορετικές τιμές confidence, lift και leverage δεν επηρεάζουν ουσιαστικά τις απαιτήσεις μνήμης της διαδικασίας εξόρυξης.

Συμπέρασμα Επιλογής Παραμέτρων

Η αξιολόγηση των έξι πειραμάτων δείχνει ότι τα πειράματα P1 και P3 παράγουν μεγαλύτερο

αριθμό κανόνων, ωστόσο περιλαμβάνουν και μεγαλύτερο ποσοστό κανόνων χαμηλότερης αξιοπιστίας. Αντίθετα, τα πειράματα P4 και P6 εφαρμόζουν ιδιαίτερα αυστηρά κριτήρια leverage, οδηγώντας σε σημαντική μείωση του πλήθους των κανόνων και ενδεχομένως στον αποκλεισμό χρήσιμων συσχετίσεων.

Το πείραμα P5 παρουσιάζει την καλύτερη ισορροπία μεταξύ της αυστηρότητας και πληρότητας των αποτελεσμάτων. Συγκεκριμένα, συνδυάζει υψηλότερα κατώφλια confidence και lift, διατηρώντας επαρκή αριθμό κανόνων για ουσιαστική ανάλυση της καταναλωτικής συμπεριφοράς.

Για τον λόγο αυτό το πείραμα P5 επιλέγεται ως η τελική διαμόρφωση του αλγορίθμου Apriori για τη μεταπανδημική περίοδο και χρησιμοποιείται στην επόμενη ενότητα για την ανάλυση και σύγκριση των κανόνων συσχέτισης μεταξύ των δύο χρονικών περιόδων.

3.8.3 Τελική Επιλογή Μοντέλου Apriori

Μετά την ολοκλήρωση της διαδικασίας αξιολόγησης των διαφορετικών συνδυασμών παραμέτρων για τις δύο εξεταζόμενες χρονικές περιόδους, πραγματοποιήθηκε η επιλογή του τελικού μοντέλου Apriori που χρησιμοποιήθηκε για την εξαγωγή και σύγκριση των καταναλωτικών προτύπων.

Η επιλογή αυτή βασίστηκε σε συνδυασμό ποσοτικών και ποιοτικών κριτηρίων. Συγκεκριμένα αξιολογήθηκαν:

- το πλήθος των παραγόμενων κανόνων συσχέτισης,
- το πλήθος των frequent itemsets,
- η ισχύς των συσχετίσεων μέσω της μετρικής lift,
- η πρακτική σημαντικότητα των κανόνων μέσω της μετρικής leverage,
- η δυνατότητα ερμηνείας των αποτελεσμάτων,
- καθώς και η συγκρισιμότητα μεταξύ των δύο χρονικών περιόδων.

Στην περίοδο COVID-19, το πείραμα P5 παρήγαγε 354 κανόνες συσχέτισης και 1.687 frequent itemsets, ενώ στη μεταπανδημική περίοδο παρήγαγε 216 κανόνες και 1.104 frequent itemsets. Παρότι ορισμένα πειράματα παρήγαγαν μεγαλύτερο αριθμό κανόνων, οι κανόνες αυτοί συνοδεύονταν από χαμηλότερα κατώφλια confidence ή lift, αυξάνοντας την πιθανότητα εμφάνισης λιγότερο ουσιαστικών συσχετίσεων.

Αντίθετα, τα αυστηρότερα πειράματα με τιμή leverage ίση με 0,010 περιόρισαν σημαντικά τον αριθμό των παραγόμενων κανόνων, το οποίο ενδέχεται να οδηγεί σε απώλεια χρήσιμης πληροφορίας και σε περιορισμό της δυνατότητας ερμηνείας των καταναλωτικών προτύπων.

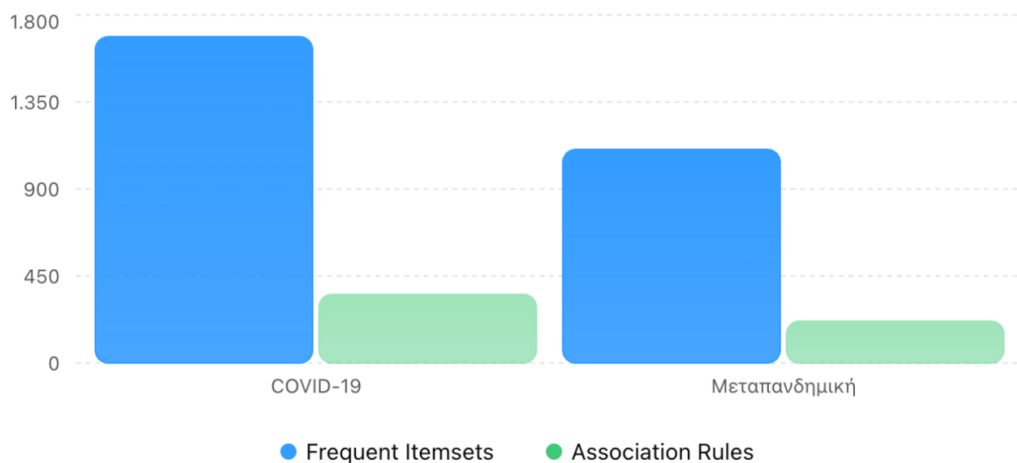
Ο Πίνακας 3.13 παρουσιάζει συγκριτικά τα βασικά αποτελέσματα του επιλεγμένου πειράματος P5 για τις δύο χρονικές περιόδους.

Πίνακας 3.13: Σύγκριση του τελικού μοντέλου P5 μεταξύ της περιόδου COVID-19 και της μεταπανδημικής περιόδου

Μετρική	COVID-19	Μεταπανδημική Περίοδος
Transactions	4.408.323	5.101.828
Frequent Itemsets	1.687	1.104
Association Rules	354	216
Runtime (sec)	347,09	342,40

Σύγκριση αποτελεσμάτων Apriori ανά περίοδο

Συγκριτική απεικόνιση του πλήθους Frequent Itemsets και Association Rules για την περίοδο COVID-19 και τη μεταπανδημική περίοδο.



Σχήμα 3.9: Σύγκριση αποτελεσμάτων του τελικού μοντέλου P5

Όπως φαίνεται στο Σχήμα 3.9, η περίοδος COVID-19 παρουσίασε μεγαλύτερο αριθμό συχνών στοιχειοσυνόλων (1.687 frequent itemsets) και μεγαλύτερο αριθμό κανόνων συσχέτισης (354 association rules) σε σύγκριση με τη μεταπανδημική περίοδο, όπου εξήχθησαν 1.104 frequent itemsets και 216 κανόνες.

Το εύρημα αυτό δείχνει ότι κατά την περίοδο COVID-19 εντοπίστηκε μεγαλύτερος αριθμός συχνών στοιχειοσυνόλων και κανόνων συσχέτισης. Δεν τεκμηριώνει, ωστόσο, ότι οι συσχετίσεις ήταν ισχυρότερες, καθώς το πλήθος των εξαγόμενων προτύπων δεν αποτελεί μέτρο της ισχύος τους. Επιπλέον, το μέσο lift των κανόνων ήταν 2,5289 κατά την περίοδο COVID-19 και 2,9215 κατά τη μεταπανδημική περίοδο. Ο μεγαλύτερος αριθμός frequent itemsets της περιόδου COVID-19 μπορεί να συνδέεται εν μέρει με τα γενικά υψηλότερα ποσοστά υποστήριξης και με το μεγαλύτερο μέγεθος των καλαθιών, χωρίς να συνεπάγεται ισχυρότερες ή περισσότερες προβλέψιμες συσχετίσεις.

Από τα αποτελέσματα προκύπτει ότι το πείραμα P5 προσφέρει την καλύτερη ισορροπία μεταξύ ποιότητας και πληρότητας αποτελεσμάτων. Η επιλογή του ίδιου συνδυασμού παραμέτρων και για τις δύο περιόδους διασφαλίζει τη συγκρισιμότητα των αποτελεσμάτων και επιτρέπει την αξιόπιστη διερεύνηση των μεταβολών της καταναλωτικής συμπεριφοράς.

Κατά συνέπεια, ως τελικό μοντέλο της παρούσας έρευνας επιλέγεται ο ακόλουθος συνδυασμός παραμέτρων:

Πίνακας 3.14: Τελικές παράμετροι του αλγορίθμου Apriori

Παράμετρος	Τιμή
Minimum Support	0,01
Minimum Confidence	0,30
Minimum Lift	1,30
Minimum Leverage	0,005
Minimum Conviction	1,0

Το μοντέλο αυτό χρησιμοποιείται στην επόμενη ενότητα για την ανάλυση των κανόνων συσχέτισης και τη συγκριτική αξιολόγηση των καταναλωτικών προτύπων μεταξύ της περιόδου COVID-19 και της μεταπανδημικής περιόδου.

Κεφάλαιο 4

Συγκριτική Διερεύνηση της Καταναλωτικής Συμπεριφοράς κατά την Περίοδο COVID-19 και τη Μεταπανδημική Περίοδο

4.1 Αποτελέσματα Apriori για την Περίοδο COVID-19

Μετά την ολοκλήρωση της διαδικασίας διερεύνησης των παραμέτρων του αλγορίθμου Apriori, επιλέχθηκε το πείραμα P5, το οποίο χρησιμοποίησε τις τιμές $\text{min_support} = 0,01$, $\text{min_confidence} = 0,30$, $\text{min_lift} = 1,30$ και $\text{min_leverage} = 0,005$. Η συγκεκριμένη παραμετροποίηση παρήγαγε ένα ικανοποιητικό πλήθος κανόνων συσχέτισης, κρατώντας υψηλό επίπεδο αξιοπιστίας και ερμηνευσιμότητας των αποτελεσμάτων.

Η ανάλυση πραγματοποιήθηκε σε σύνολο 4.408.323 συναλλαγών και 323 μοναδικών κατηγοριών προϊόντων.

Πίνακας 4.1: Συνοπτικά αποτελέσματα του επιλεγμένου πειραματικού σεναρίου P5 για την περίοδο COVID-19

Μετρική	Τιμή
Transactions	4.408.323
Unique Items	323
Frequent Itemsets	1.687
Association Rules	354
Runtime (sec)	347,09
Memory Usage (MB)	7.099,48

Πίνακας 4.2: Στατιστικά χαρακτηριστικά των παραγόμενων κανόνων συσχέτισης στο επιλεγμένο πειραματικό σενάριο P5 για την περίοδο COVID-19

Μετρική	Τιμή
Αλγόριθμος	Apriori
Σύνολο Δεδομένων	COVID-19
Minimum Support	0,01
Minimum Confidence	0,30
Minimum Lift	1,30
Minimum Leverage	0,005
Minimum Conviction	1,00
Πλήθος Συναλλαγών	4.408.323
Πλήθος Διαφορετικών Κατηγοριών	323
Frequent Itemsets	1.687
Association Rules	354
Μέσο Support	0,0195
Μέσο Confidence	0,4028
Μέσο Lift	2,5289
Μέσο Leverage	0,0100
Μέσο Conviction	1,4812
Χρόνος Εκτέλεσης (sec)	347,09
Μέγιστη Καταγεγραμμένη Μνήμη (MB)	7.099,48

4.1.1 Συχνά Στοιχειοσύνολα (Frequent Itemsets) COVID-19

Η εφαρμογή του Apriori οδήγησε στην εξαγωγή 1.687 συχνών συνόλων αντικειμένων (frequent itemsets). Ο αριθμός αυτός δείχνει ότι παρά το μεγάλο μέγεθος του συνόλου δεδομένων, σημαντικός αριθμός κατηγοριών προϊόντων εμφανίζεται συστηματικά στις ίδιες συναλλαγές.

Τα frequent itemsets αποτελούν τη βάση για την παραγωγή των κανόνων συσχέτισης και αποτυπώνουν επαναλαμβανόμενα πρότυπα αγορών των καταναλωτών κατά την περίοδο της πανδημίας.

Πίνακας 4.3: Δεκαπέντε συχνότερα πολυ-στοιχειακά frequent itemsets κατά την περίοδο COVID-19

#	Frequent Itemset	Support (%)
1	ΓΑΛΑ ΦΡΕΣΚΟ + ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	13,78
2	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ + ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	10,77
3	ΓΙΑΟΥΡΤΙΑ + ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	10,09
4	ΑΝΑΨΥΚΤΙΚΑ + ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	10,05
5	ΓΑΛΑ ΦΡΕΣΚΟ + ΓΙΑΟΥΡΤΙΑ	9,29
6	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ + ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	9,23
7	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ + ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	7,93
8	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ + ΓΑΛΑ ΦΡΕΣΚΟ	7,67
9	SNACK + ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	7,64
10	ΣΟΚΟΛΑΤΕΣ + ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	7,37
11	ΝΕΡΑ + ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	7,35
12	ΦΡΕΣΚΑ ΦΡΟΥΤΑ + ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	7,16
13	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ + ΓΑΛΑ ΦΡΕΣΚΟ	7,15
14	ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ + ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	6,94
15	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ + ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	6,89

Ο Πίνακας 4.3 παρουσιάζει τα δεκαπέντε συχνότερα frequent itemsets που εξήχθησαν από το σύνολο δεδομένων της περιόδου COVID-19. Παρατηρείται ότι τα συχνότερα itemsets περιλαμβάνουν κυρίως βασικές κατηγορίες τροφίμων και προϊόντων καθημερινής κατανάλωσης, αντανακλώντας τις συνήθειες αγοραστικές συνήθειες των καταναλωτών κατά την εξεταζόμενη περίοδο.

Η κατηγορία «Είδη Συσκευασίας προς Πώληση» συμμετέχει σε αρκετά από τα συχνότερα στοιχειοσύνολα, λόγω της υψηλής συχνότητας εμφάνισής της στις συναλλαγές. Επειδή, όμως, αφορά κυρίως σακούλες μεταφοράς, οι συνδυασμοί στους οποίους συμμετέχει δεν ερμηνεύονται ως ουσιαστικές καταναλωτικές προτιμήσεις. Για την ερμηνεία των αποτελεσμάτων δίνεται μεγαλύτερη έμφαση στους συνδυασμούς μεταξύ κατηγοριών προϊόντων διατροφής και καθημερινής χρήσης.

Ταυτόχρονα, παρατηρούνται ισχυροί συνδυασμοί μεταξύ των κατηγοριών βασικών τροφίμων. Ενδεικτικά, το itemset «Αλλαντικά/Κιλό – Μαλακά Τυριά/Κιλό» εμφανίζεται στο 9,23% των συναλλαγών, το οποίο υποδηλώνει συχνή συνύπαρξη μεταξύ των δύο κατηγοριών. Αντίστοιχα, οι συνδυασμοί «Γάλα Φρέσκο – Γιαούρτια» και «Λαχανικά Πράσινα Ποικιλία – Γάλα Φρέσκο» υποδηλώνουν κοινά πρότυπα κατανάλωσης που σχετίζονται με βασικά είδη διατροφής.

Τα αποτελέσματα αυτά αποτελούν την πρώτη ένδειξη ύπαρξης ουσιαστικών συσχετίσεων μεταξύ συγκεκριμένων κατηγοριών προϊόντων και δικαιολογούν τη συνέχεια της ανάλυσης μέσω κανόνων συσχέτισης, οι οποίοι παρουσιάζονται στην επόμενη ενότητα.

4.1.2 Κανόνες Συσχέτισης (Association Rules)

Με βάση τα παραγόμενα frequent itemsets εξήχθησαν συνολικά 354 κανόνες συσχέτισης. Για λόγους παρουσίασης και ανάλυσης επιλέχθηκαν οι είκοσι σημαντικότεροι κανόνες με βάση το leverage και το lift.

Πίνακας 4.4: Κορυφαίοι κανόνες συσχέτισης που εξήχθησαν από το επιλεγμένο πειραματικό σενάριο P5 για την περίοδο COVID-19

Antecedent	Consequent	Support (%)	Confidence (%)	Lift	Leverage	Conviction
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	9,23	55,62	4,0807	0,0697	1,9460
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΕΙΔΗ ΣΥΣΚΕΤΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	4,82	35,36	4,4586	0,0374	1,4243
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΕΙΔΗ ΣΥΣΚΕΤΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	4,82	69,91	4,2109	0,0367	2,7717
ΦΡΕΣΚΑ ΦΡΟΥΤΑ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	6,75	46,51	2,1037	0,0354	1,4563
ΓΙΑΟΥΡΤΙΑ	ΓΑΛΑ ΦΡΕΣΚΟ	9,29	45,62	1,5968	0,0347	1,3135
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΓΑΛΑ ΦΡΕΣΚΟ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	4,25	71,91	4,3315	0,0327	2,9694
ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ	ΦΡΕΣΚΑ ΦΡΟΥΤΑ	4,32	40,65	2,8002	0,0278	1,4403
ΨΩΜΙΑ ΣΥΣΚΕΤΑΣΜΕΝΑ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	4,64	37,67	2,2688	0,0260	1,3380
ΚΑΘΑΡΙΣΤΙΚΑ	ΠΡΟΪΟΝΤΑ ΠΛΥΣΙΜΑΤΟΣ	3,67	43,79	3,2526	0,0254	1,5395
ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	4,79	44,98	2,0344	0,0243	1,4157
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	ΓΑΛΑ ΦΡΕΣΚΟ	7,15	43,09	1,5083	0,0241	1,2552
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΓΙΑΟΥΡΤΙΑ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	3,08	72,53	4,3684	0,0237	3,0356
ΑΛΛΑΝΤΙΚΑ ΣΥΣΚΕΤΑΣΜΕΝΑ ΨΥΓΕΙΟΥ	ΤΥΡΙΑ ΔΙΑΦΟΡΑ ΣΥΣΚ	3,22	36,66	3,6597	0,0234	1,4206
ΨΩΜΙΑ ΣΥΣΚΕΤΑΣΜΕΝΑ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	4,02	32,60	2,3922	0,0234	1,2815
ΠΡΩΙΝΑ	ΓΑΛΑ ΦΡΕΣΚΟ	4,95	50,72	1,7751	0,0216	1,4493
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	2,74	68,26	4,1113	0,0207	2,6275
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	ΨΩΜΙΑ ΣΥΣΚΕΤΑΣΜΕΝΑ	3,19	34,58	2,8045	0,0205	1,3400
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	ΓΙΑΟΥΡΤΙΑ	5,43	32,68	1,6044	0,0204	1,1828
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	ΓΑΛΑ ΦΡΕΣΚΟ	5,92	43,41	1,5194	0,0202	1,2622
ΛΑΧΑΝΙΚΑ ΣΕ ΚΟΝΣΕΡΒΑΧΑΡΤΙΝΗ ΣΥΣΚΕΤΑΣΙΑ	ΖΥΜΑΡΙΚΑ	2,81	36,99	3,3748	0,0198	1,4131

Οι σημαντικότεροι κανόνες παρουσιάζουν ισχυρές σχέσεις μεταξύ βασικών κατηγοριών προϊόντων καθημερινής κατανάλωσης. Ο ισχυρότερος κανόνας που προέκυψε ήταν:

ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ → ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ

με support 9,23%, confidence 55,62% και lift 4,08.

Μια συναλλαγή που περιλαμβάνει μαλακά τυριά έχει περίπου 4,08 φορές μεγαλύτερη πιθανότητα να περιλαμβάνει και αλλαντικά από όση θα αναμενόταν με βάση τη συνολική συχνότητα εμφάνισης των αλλαντικών.

Παρόμοια ισχυρή σχέση παρατηρείται μεταξύ των κατηγοριών:

- **ΦΡΕΣΚΑ ΦΡΟΥΤΑ → ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ**
- **ΓΙΑΟΥΡΤΙΑ → ΓΑΛΑ ΦΡΕΣΚΟ**
- **ΚΑΘΑΡΙΣΤΙΚΑ → ΠΡΟΪΟΝΤΑ ΠΛΥΣΙΜΑΤΟΣ**
- **ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ → ΦΡΕΣΚΑ ΦΡΟΥΤΑ**

Οι κανόνες αυτοί αναδεικνύουν πρότυπα συμπληρωματικής κατανάλωσης, καθώς οι αντίστοιχες κατηγορίες προϊόντων χρησιμοποιούνται συχνά από κοινού.

4.1.3 Ερμηνεία Αποτελεσμάτων COVID-19

Τα αποτελέσματα της περιόδου COVID-19 δείχνουν υψηλή παρουσία βασικών ειδών διατροφής και οικιακής χρήσης. Η εικόνα αυτή είναι συμβατή με ευρήματα της βιβλιογραφίας σχετικά με τις αγοραστικές συνήθειες κατά την πανδημία. Ωστόσο, η παρούσα ανάλυση δεν επιτρέπει την αιτιώδη απόδοση των συγκεκριμένων προτύπων αποκλειστικά στις συνθήκες της πανδημίας.

4.2 Αποτελέσματα Apriori για τη Μεταπανδημική Περίοδο

Για λόγους συγκρισιμότητας εφαρμόστηκε η ίδια ακριβώς παραμετροποίηση και στο σύνολο δεδομένων της μεταπανδημικής περιόδου. Η ανάλυση πραγματοποιήθηκε σε 5.101.828 συναλλαγές και 322 μοναδικές κατηγορίες προϊόντων.

Πίνακας 4.5: Συνοπτικά αποτελέσματα του επιλεγμένου πειραματικού σεναρίου P5 για τη μεταπανδημική περίοδο

Μετρική	Τιμή
Transactions	5.101.828
Unique Items	322
Frequent Itemsets	1.104
Association Rules	216
Runtime (sec)	342,40
Memory Usage (MB)	7.762,17

Πίνακας 4.6: Στατιστικά χαρακτηριστικά των παραγόμενων κανόνων συσχέτισης στο επιλεγμένο πειραματικό σενάριο P5 για τη μεταπανδημική περίοδο

Μετρική	Τιμή
Αλγόριθμος	Apriori
Σύνολο Δεδομένων	Μεταπανδημική Περίοδος
Minimum Support	0,01
Minimum Confidence	0,30
Minimum Lift	1,30
Minimum Leverage	0,005
Minimum Conviction	1,00
Πλήθος Συναλλαγών	5.101.828
Πλήθος Διαφορετικών Κατηγοριών	322
Frequent Itemsets	1.104
Association Rules	216
Μέσο Support	0,0186
Μέσο Confidence	0,3975
Μέσο Lift	2,9215
Μέσο Leverage	0,0099
Μέσο Conviction	1,6153
Χρόνος Εκτέλεσης (sec)	342,40
Μέγιστη Χρήση Μνήμης (MB)	7.762,17

4.2.1 Συχνά Στοιχειοσύνολα (Frequent Itemsets) Μεταπανδημικής Περιόδου

Η εφαρμογή του Apriori οδήγησε στην εξαγωγή 1.104 frequent itemsets, αριθμός μικρότερος σε σχέση με την περίοδο COVID-19. Το αποτέλεσμα αυτό δείχνει ότι λιγότεροι συνδυασμοί κατηγοριών υπερέβησαν το καθορισμένο ελάχιστο όριο υποστήριξης, χωρίς να συνεπάγεται από μόνο του μικρότερη ποικιλία ή μεγαλύτερη ομοιογένεια των αγορών.

Πίνακας 4.7: Δεκαπέντε συχνότερα πολυ-στοιχειακά frequent itemsets κατά τη μεταπανδημική περίοδο

#	Frequent Itemset	Support (%)
1	ΓΑΛΑ ΦΡΕΣΚΟ + ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	11,22
2	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ + ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	9,59
3	ΑΝΑΨΥΚΤΙΚΑ + ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	9,35
4	ΓΙΑΟΥΡΤΙΑ + ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	8,49
5	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ + ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	7,46
6	SNACK + ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	7,34
7	ΝΕΡΑ + ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	7,16
8	ΓΙΑΟΥΡΤΙΑ + ΓΑΛΑ ΦΡΕΣΚΟ	7,01
9	ΦΡΕΣΚΑ ΦΡΟΥΤΑ + ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	6,87
10	ΣΟΚΟΛΑΤΕΣ + ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	6,56
11	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ + ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	6,46
12	ΦΡΕΣΚΑ ΦΡΟΥΤΑ + ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	6,14
13	ΜΠΙΣΚΟΤΑ + ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	6,14
14	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ + ΓΑΛΑ ΦΡΕΣΚΟ	6,04
15	ΠΡΟΪΟΝΤΑ ΠΛΥΣΙΜΑΤΟΣ + ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	5,96

Ο Πίνακας 4.7 παρουσιάζει τα δεκαπέντε συχνότερα frequent itemsets που εξήχθησαν από το σύνολο δεδομένων της μεταπανδημικής περιόδου. Τα αποτελέσματα αναδεικνύουν συνδυασμούς προϊόντων που εμφανίζονται συστηματικά στα καταναλωτικά καλάθια και αποτελούν τη βάση για την εξαγωγή των κανόνων συσχέτισης που παρουσιάζονται στη συνέχεια.

Αρκετά από τα συχνότερα στοιχειοσύνολα περιλαμβάνουν την κατηγορία «Είδη Συσκευασίας προς Πώληση», γεγονός που οφείλεται στην υψηλή συχνότητα εμφάνισής της στις συναλλαγές. Καθώς η συγκεκριμένη κατηγορία αφορά κυρίως σακούλες μεταφοράς, οι συνδυασμοί στους οποίους συμμετέχει δεν ερμηνεύονται ως ουσιαστικές καταναλωτικές προτιμήσεις.

Το πρώτο στοιχειοσύνολο του πίνακα που δεν περιλαμβάνει τη συγκεκριμένη κατηγορία είναι ο συνδυασμός «Αλλαντικά/Κιλό – Μαλακά Τυριά/Κιλό», ο οποίος εμφανίζεται στο 7,46% των συναλλαγών. Η συχνή συνύπαρξη των δύο κατηγοριών δείχνει ότι τα συγκεκριμένα προϊόντα επιλέγονται συχνά στο ίδιο καλάθι. Ακολουθεί ο συνδυασμός «Γιαούρτια – Γάλα Φρέσκο» με support 7,01%, ενώ οι συνδυασμοί «Φρέσκα Φρούτα – Λαχανικά Πράσινα Ποικιλία» και «Λαχανικά Πράσινα Ποικιλία – Γάλα Φρέσκο» εμφανίζονται στο 6,14% και στο 6,04% των συναλλαγών αντίστοιχα. Τα στοιχειοσύνολα αυτά δείχνουν ότι οι συχνότεροι ουσιαστικοί συνδυασμοί της μεταπανδημικής περιόδου αφορούν κυρίως βασικά είδη καθημερινής διατροφής.

Συνολικά, τα frequent itemsets της μεταπανδημικής περιόδου δείχνουν ότι οι συχνότεροι συνδυασμοί προϊόντων αφορούν κυρίως βασικές κατηγορίες τροφίμων και είδη καθημερινής χρήσης, γεγονός που αντανακλά τη σταθερότητα των καταναλωτικών προτύπων στο εξεταζόμενο χρονικό

διάστημα.

4.2.2 Κανόνες Συσχέτισης (Association Rules) Μεταπανδημικής Περιόδου

Συνολικά εξήχθησαν 216 κανόνες συσχέτισης.

Πίνακας 4.8: Κορυφαίοι κανόνες συσχέτισης που εξήχθησαν από το επιλεγμένο πειραματικό σενάριο P5 για τη μεταπανδημική περίοδο

Antecedent	Consequent	Support(%)	Confidence(%)	Lift	Leverage	Conviction
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	7,46	67,30	4,73	0,059	2,62
ΦΡΕΣΚΑ ΦΡΟΥΤΑ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	6,14	43,30	2,14	0,033	1,41
ΕΙΔΗ ΣΥΣΚΕΤΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ, ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	3,72	57,55	5,19	0,030	2,09
ΓΙΑΟΥΡΤΙΑ	ΓΑΛΑ ΦΡΕΣΚΟ	7,01	40,36	1,68	0,028	1,27
ΕΤΟΙΜΟ ΦΑΓΗΤΟ CATERING	ΑΠΑΡΑΔΕΚΤΑ ΕΜΠΟΡΕΥΜΑΤΑ	2,87	83,31	28,28	0,028	5,82
ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ	ΦΡΕΣΚΑ ΦΡΟΥΤΑ	4,08	39,19	2,76	0,026	1,41
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΓΑΛΑ ΦΡΕΣΚΟ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	3,04	55,95	5,05	0,024	2,02
ΚΑΘΑΡΙΣΤΙΚΑ	ΠΡΟΪΟΝΤΑ ΠΛΥΣΙΜΑΤΟΣ	2,95	43,17	3,65	0,021	1,55
ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	4,25	40,80	2,01	0,021	1,35
ΨΩΜΙΑ ΣΥΣΚΕΤΑΣΜΕΝΑ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	3,63	34,15	2,40	0,021	1,30
ΑΛΛΑΝΤΙΚΑ ΣΥΣΚΕΤΑΣΜΕΝΑ ΨΥΓΕΙΟΥ	ΤΥΡΙΑ ΔΙΑΦΟΡΑ ΣΥΣΚ	2,82	36,64	4,00	0,021	1,43
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	ΓΑΛΑ ΦΡΕΣΚΟ	5,43	38,18	1,59	0,020	1,23
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΨΩΜΙΑ ΣΥΣΚΕΤΑΣΜΕΝΑ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	2,37	65,46	5,90	0,020	2,57
ΜΑΛΑΚΤΙΚΑ	ΠΡΟΪΟΝΤΑ ΠΛΥΣΙΜΑΤΟΣ	2,47	55,25	4,67	0,019	1,97
Κρεμμύδια κιλού	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	2,83	62,88	3,10	0,019	2,15
ΠΡΩΙΝΑ	ΓΑΛΑ ΦΡΕΣΚΟ	3,71	45,65	1,90	0,018	1,40
SNACK	ΑΝΑΨΥΚΤΙΚΑ	4,17	30,21	1,72	0,017	1,18
ΓΙΑΟΥΡΤΙΑ, ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	2,19	53,26	4,80	0,017	1,90
ΛΑΧΑΝΙΚΑ ΣΕ ΚΟΝΣΕΡΒΑΧΑΡΤΙΝΗ ΣΥΣΚΕΤΑΣΙΑ	ΖΥΜΑΡΙΚΑ	2,35	36,72	3,80	0,017	1,43
ΕΙΔΗ ΣΥΣΚΕΤΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ, ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	ΦΡΕΣΚΑ ΦΡΟΥΤΑ	3,09	32,19	2,27	0,017	1,27

Ο ισχυρότερος κανόνας που προέκυψε ήταν:

ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ → ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ

με confidence 67,30% και lift 4,73.

Παρατηρείται ότι η σχέση μεταξύ των δύο κατηγοριών παραμένει ιδιαίτερα ισχυρή και μετά την πανδημία. Ισχυροί κανόνες εντοπίστηκαν επίσης μεταξύ των κατηγοριών: • ΦΡΕΣΚΑ ΦΡΟΥΤΑ → ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ • ΓΙΑΟΥΡΤΙΑ → ΓΑΛΑ ΦΡΕΣΚΟ • ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ → ΦΡΕΣΚΑ ΦΡΟΥΤΑ • ΚΑΘΑΡΙΣΤΙΚΑ → ΠΡΟΪΟΝΤΑ ΠΛΥΣΙΜΑΤΟΣ

4.2.3 Ερμηνεία Αποτελεσμάτων Μεταπανδημικής Περιόδου

Τα αποτελέσματα δείχνουν ότι οι βασικές αγοραστικές συνήθειες των καταναλωτών διατηρήθηκαν σε μεγάλο βαθμό και μετά το τέλος της πανδημίας. Οι κυρίαρχες συσχετίσεις εξακολουθούν να αφορούν προϊόντα καθημερινής κατανάλωσης και κατηγορίες που χρησιμοποιούνται συμπληρωματικά.

Παράλληλα, κατά τη μεταπανδημική περίοδο εντοπίστηκε μικρότερος αριθμός frequent itemsets και παραγόμενων κανόνων. Ωστόσο, η διαφορά αυτή δεν επαρκεί από μόνη της για την εξαγωγή συμπερασμάτων σχετικά με τη σταθερότητα ή την πολυπλοκότητα της καταναλωτικής συμπεριφοράς, καθώς μπορεί να συνδέεται και με τη γενικότερη υποχώρηση του support και το μικρότερο μέσο μέγεθος των καλαθιών.

4.3 Συγκριτική Ανάλυση COVID-19 και Μεταπανδημικής Περιόδου

Σε αυτή την ενότητα πραγματοποιείται συγκριτική αξιολόγηση των αποτελεσμάτων που προέκυψαν από την εφαρμογή του αλγορίθμου Apriori στις δύο εξεταζόμενες χρονικές περιόδους. Στόχος είναι να διερευνηθούν οι διαφοροποιήσεις των αγοραστικών προτύπων μεταξύ της περιόδου COVID-19 και της μεταπανδημικής περιόδου και να εντοπιστούν πιθανές μεταβολές στις συσχετίσεις μεταξύ των κατηγοριών προϊόντων.

Η σύγκριση πραγματοποιείται τόσο σε επίπεδο συνολικών μετρικών του αλγορίθμου όσο και σε επίπεδο συχνών συνόλων αντικειμένων και κανόνων συσχέτισης.

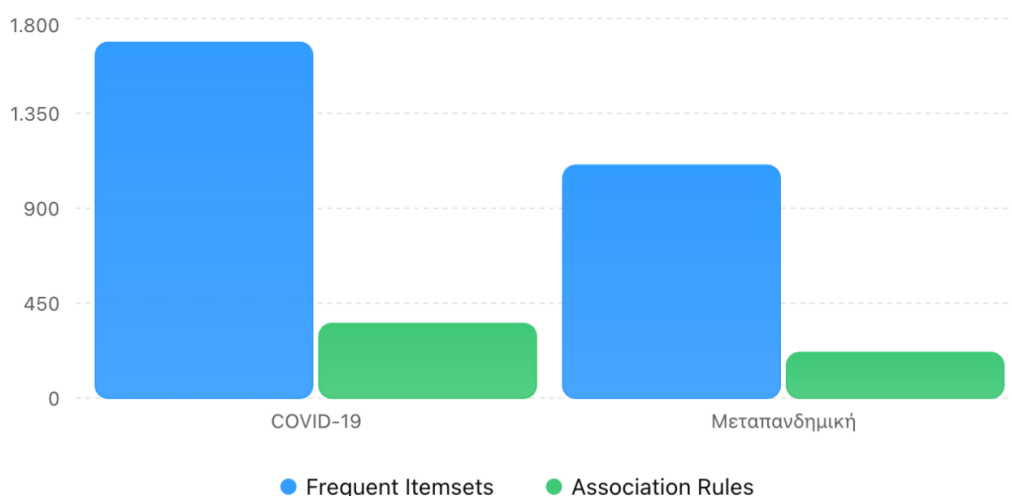
4.3.1 Σύγκριση Συνολικών Αποτελεσμάτων

Πίνακας 4.9: Συγκριτικά αποτελέσματα εφαρμογής του αλγορίθμου Apriori στις δύο εξεταζόμενες χρονικές περιόδους

Μετρική	COVID-19	Μεταπανδημική Περίοδος
Transactions	4.408.323	5.101.828
Frequent Itemsets	1.687	1.104
Association Rules	354	216
Runtime (sec)	347,09	342,40
Memory (MB)	7.099,48	7.762,17

Σύγκριση Frequent Itemsets και Association Rules

Συγκριτική παρουσίαση των αποτελεσμάτων Apriori στις δύο εξεταζόμενες περιόδους.



Σχήμα 4.1: Σύγκριση του πλήθους των frequent itemsets και των κανόνων συσχέτισης που εξήχθησαν από τον αλγόριθμο Apriori για την περίοδο COVID-19 και τη μεταπανδημική περίοδο

Όπως φαίνεται στο Σχήμα 4.1, κατά την περίοδο COVID-19 εξήχθησαν 1.687 frequent itemsets και 354 κανόνες συσχέτισης, ενώ κατά τη μεταπανδημική περίοδο εξήχθησαν 1.104 frequent itemsets και 216 κανόνες. Η διαφορά αυτή δείχνει ότι στην πρώτη περίοδο περισσότερα πρότυπα υπερέβησαν τα καθορισμένα όρια υποστήριξης και αξιολόγησης, χωρίς όμως να τεκμηριώνει μεγαλύτερη ισχύ των συσχετίσεων ή μεγαλύτερη συγκέντρωση των αγοραστικών επιλογών. Όπως έδειξε η διερευνητική ανάλυση, ο μεγαλύτερος αριθμός συχνών στοιχειοσυνόλων κατά την περίοδο COVID-19 μπορεί να συνδέεται εν μέρει με τα γενικά υψηλότερα ποσοστά υποστήριξης και με το μεγαλύτερο μέσο μέγεθος των καλαθιών.

Η υπολογιστική απόδοση του Apriori παρέμεινε σχετικά σταθερή στις δύο περιόδους, καθώς ο χρόνος εκτέλεσης κυμάνθηκε περίπου στα 345 δευτερόλεπτα και η μέγιστη κατανάλωση μνήμης μεταξύ 7 και 8 GB. Επομένως, οι διαφορές στον αριθμό των εξαγόμενων προτύπων δεν συνοδεύτηκαν από αντίστοιχα μεγάλη διαφοροποίηση του χρόνου εκτέλεσης.

4.3.2 Σύγκριση Frequent Itemsets

Για την πληρέστερη σύγκριση των αγοραστικών προτύπων εξετάστηκαν τα frequent itemsets που εμφανίζονται και στις δύο χρονικές περιόδους. Αρχικά εντοπίστηκαν 1.088 κοινά itemsets. Μετά την εξαίρεση των κατηγοριών «Είδη Συσκευασίας προς Πώληση» και «Απαράδεκτα Εμπορεύματα», παρέμειναν 818 κοινά itemsets. Διευκρινίζεται ότι οι εξαιρέσεις εφαρμόστηκαν κατά τη συγκριτική επεξεργασία των ήδη παραχθέντων αποτελεσμάτων και δεν πραγματοποιήθηκε εκ νέου εκτέλεση του αλγορίθμου Apriori. Η ίδια διαδικασία ακολουθήθηκε και στη σύγκριση των κοινών κανόνων συσχέτισης που παρουσιάζεται στην επόμενη ενότητα.

Από τα 818 κοινά itemsets, τα 113 ήταν 1-itemsets, ενώ τα υπόλοιπα 705 περιλάμβαναν τουλάχιστον δύο διαφορετικές κατηγορίες προϊόντων. Οι 113 κοινές κατηγορίες αντιστοιχούν στα 1-itemsets

που υπερέβησαν το καθορισμένο ελάχιστο όριο support και στις δύο περιόδους. Αντίθετα, οι 315 κοινές κατηγορίες που εξετάστηκαν στη συμπληρωματική διερευνητική ανάλυση περιλαμβάνουν όλες τις κατηγορίες που εμφανίστηκαν και στις δύο περιόδους, ανεξάρτητα από το αν υπερέβησαν το συγκεκριμένο όριο. Η παρούσα σύγκριση επικεντρώθηκε στα 705 πολυστοιχειακά frequent itemsets, καθώς τα 1-itemsets έχουν ήδη εξεταστεί χωριστά στο πλαίσιο της διερευνητικής ανάλυσης.

Για κάθε κοινό frequent itemset υπολογίστηκε η απόλυτη διαφορά του support μεταξύ των δύο περιόδων:

$$\Delta Support = Support_{Post-COVID} - Support_{COVID}$$

Η διαφορά αυτή εκφράζεται σε ποσοστιαίες μονάδες. Παράλληλα, υπολογίστηκε και η σχετική ποσοστιαία μεταβολή ως προς την περίοδο COVID-19:

$$\text{Σχετική μεταβολή} = \frac{Support_{Post-COVID} - Support_{COVID}}{Support_{COVID}} \times 100$$

Η κατάταξη των frequent itemsets πραγματοποιήθηκε με βάση την απόλυτη διαφορά του support. Η επιλογή αυτή επιτρέπει να αναδειχθούν τα patterns με τη μεγαλύτερη απόλυτη μεταβολή στη συμμετοχή τους στις συναλλαγές, χωρίς να δίνεται δυσανάλογη βαρύτητα σε itemsets με πολύ χαμηλό αρχικό support. Η σχετική μεταβολή παρουσιάζεται συμπληρωματικά, ώστε να αποτυπώνεται και το μέγεθος της αλλαγής σε σχέση με την αρχική τιμή.

Τα δέκα κοινά frequent itemsets με τις μεγαλύτερες αυξήσεις του support παρουσιάζονται στον Πίνακα 4.10.

Πίνακας 4.10: Κοινά frequent itemsets με τις δέκα μεγαλύτερες αυξήσεις του support

Frequent itemset	Support COVID (%)	Support Post-COVID (%)	Δ Support (π.μ.)	Σχετική μεταβολή (%)
Κρεμμύδια κιλού + Λαχανικά Πράσινα Ποικιλία	2,358	2,826	+0,469	+19,87
Κρεμμύδια κιλού + Φρέσκα Φρούτα	1,180	1,373	+0,194	+16,44
Βόειο Νωπό + Χοιρινό Νωπό	1,938	2,122	+0,184	+9,49
Ζαχαρώδη + Νερά	1,608	1,766	+0,157	+9,78
Κοτόπουλα + Χοιρινό Νωπό	1,439	1,590	+0,151	+10,51
Νερά + Προϊόντα Ζύμης Γλυκά	1,497	1,635	+0,138	+9,21
Snack + Νερά	2,823	2,952	+0,128	+4,54
Snack + Προϊόντα Ζύμης Γλυκά	2,030	2,132	+0,101	+4,99
Βόειο Νωπό + Κοτόπουλα	1,387	1,484	+0,097	+7,02
Νερά + Ψωμί	1,895	1,968	+0,073	+3,84

Η μεγαλύτερη αύξηση καταγράφηκε στον συνδυασμό «Κρεμμύδια κιλού – Λαχανικά Πράσινα Ποικιλία», του οποίου το support αυξήθηκε από 2,358% σε 2,826%. Η απόλυτη διαφορά ανέρχεται σε 0,469 ποσοστιαίες μονάδες και αντιστοιχεί σε σχετική αύξηση 19,87%. Ακολουθεί ο συνδυασμός «Κρεμμύδια κιλού – Φρέσκα Φρούτα», με αύξηση 0,194 ποσοστιαίων μονάδων και σχετική μεταβολή 16,44%.

Στις μεγαλύτερες αυξήσεις περιλαμβάνονται επίσης συνδυασμοί νωπού κρέατος, νερού, σνακ, προϊόντων ζύμης και φρέσκων προϊόντων. Ωστόσο, οι απόλυτες αυξήσεις είναι σχετικά περιορισμένες, καθώς μόνο ο πρώτος συνδυασμός υπερβαίνει τις 0,4 ποσοστιαίες μονάδες.

Τα δέκα κοινά frequent itemsets με τις μεγαλύτερες μειώσεις του support παρουσιάζονται στον Πίνακα 4.11.

Πίνακας 4.11: Κοινά frequent itemsets με τις δέκα μεγαλύτερες μειώσεις του support

Frequent itemset	Support COVID (%)	Support Post-COVID (%)	Δ Support (π.μ.)	Σχετική μεταβολή (%)
Γάλα Φρέσκο + Γιαούρτια	9,291	7,013	-2,278	-24,52
Αλλαντικά/Κιλό + Μαλακά Τυριά/Κιλό	9,234	7,462	-1,772	-19,19
Γάλα Φρέσκο + Μαλακά Τυριά/Κιλό	7,154	5,433	-1,721	-24,06
Αλλαντικά/Κιλό + Γάλα Φρέσκο	5,916	4,215	-1,701	-28,76
Γάλα Φρέσκο + Λαχανικά Πράσινα Ποικιλία	7,675	6,044	-1,631	-21,25
Γάλα Φρέσκο + Ψωμιά Συσκευασμένα	5,435	4,079	-1,356	-24,95
Γιαούρτια + Μαλακά Τυριά/Κιλό	5,425	4,109	-1,316	-24,25
Γάλα Φρέσκο + Πρωινά	4,946	3,707	-1,239	-25,05
Αλλαντικά/Κιλό + Γάλα Φρέσκο + Μαλακά Τυριά/Κιλό	4,255	3,040	-1,215	-28,56
Γάλα Φρέσκο + Προϊόντα Πλυσίματος	4,620	3,416	-1,204	-26,07

Η μεγαλύτερη μείωση καταγράφηκε στον συνδυασμό «Γάλα Φρέσκο – Γιαούρτια», του οποίου το support μειώθηκε από 9,291% σε 7,013%. Η μεταβολή αντιστοιχεί σε μείωση 2,278 ποσοστιαίων μονάδων ή 24,52% σε σχέση με την περίοδο COVID-19. Ακολουθούν οι συνδυασμοί «Αλλαντικά/Κιλό – Μαλακά Τυριά/Κιλό» και «Γάλα Φρέσκο – Μαλακά Τυριά/Κιλό», με μειώσεις 1,772 και 1,721 ποσοστιαίων μονάδων αντίστοιχα.

Στις μεγαλύτερες μειώσεις παρατηρείται έντονη παρουσία του φρέσκου γάλακτος και άλλων γαλακτοκομικών προϊόντων, καθώς και συνδυασμών αλλαντικών και τυριών. Το φρέσκο γάλα συμμετέχει σε επτά από τα δέκα patterns με τη μεγαλύτερη μείωση, γεγονός που συμφωνεί με τη μείωση του support της συγκεκριμένης κατηγορίας που καταγράφηκε κατά τη σύγκριση των κοινών 1-itemsets.

Συνολικά, από τα 705 κοινά πολυστοιχειακά frequent itemsets που εξετάστηκαν, τα 34 παρουσίασαν αύξηση και τα 671 μείωση του support κατά τη μεταπανδημική περίοδο. Επομένως, η υποχώρηση του support δεν περιορίζεται σε μεμονωμένους συνδυασμούς, αλλά χαρακτηρίζει τη μεγάλη πλειονότητα των κοινών patterns.

Η γενικότερη αυτή τάση είναι συνεπής με τα αποτελέσματα της διερευνητικής ανάλυσης των 1-itemsets και με το μικρότερο μέσο μέγεθος καλαθιού που καταγράφηκε στη μεταπανδημική περίοδο. Κατά την περίοδο COVID-19, κάθε συναλλαγή περιλάμβανε κατά μέσο όρο 7,1085 διαφορετικές κατηγορίες προϊόντων, έναντι 6,4014 στη μεταπανδημική περίοδο. Όταν ένα καλάθι περιλαμβάνει περισσότερες διαφορετικές κατηγορίες, δημιουργούνται περισσότεροι πιθανοί συνδυασμοί και αυξάνεται η πιθανότητα κάποιοι από αυτούς να υπερβούν το καθορισμένο ελάχιστο όριο υποστήριξης.

Συνεπώς, ο μεγαλύτερος αριθμός frequent itemsets και οι υψηλότερες τιμές support που παρατηρούνται κατά την περίοδο COVID-19 μπορεί να συνδέονται εν μέρει με το μεγαλύτερο μέσο μέγεθος των καλαθιών. Το εύρημα αυτό δεν επαρκεί για να χαρακτηριστούν τα καλάθια της μίας περιόδου περισσότερο ομοιογενή ή περισσότερο διαφοροποιημένα από εκείνα της άλλης, ούτε επιτρέπει την απόδοση των μεταβολών αποκλειστικά στην πανδημία.

Επιπλέον, η μείωση του support δεν συνεπάγεται ότι οι σχέσεις μεταξύ των κατηγοριών έγιναν απαραίτητα ασθενέστερες. Το support εκφράζει τη συχνότητα εμφάνισης ενός συνδυασμού, αλλά

δεν λαμβάνει υπόψη τη γενικότερη συχνότητα των επιμέρους κατηγοριών. Για τον λόγο αυτό, στην επόμενη ενότητα εξετάζονται συμπληρωματικά οι μεταβολές του confidence και του lift στους κοινούς κανόνες συσχέτισης.

4.3.3 Σύγκριση Κανόνων Συσχέτισης

Η σύγκριση των κανόνων συσχέτισης ανέδειξε ένα σύνολο κανόνων που εμφανίζονται τόσο κατά την περίοδο COVID-19 όσο και κατά τη μεταπανδημική περίοδο. Στον Πίνακα 4.12 παρουσιάζονται ενδεικτικά ορισμένοι από τους ισχυρούς κοινούς κανόνες, με βάση τις τιμές του lift στις δύο χρονικές περιόδους.

Πίνακας 4.12: Ενδεικτικοί ισχυροί κανόνες που εμφανίζονται και στις δύο περιόδους

Κανόνας	Lift COVID	Lift Μεταπανδημικής Περιόδου
ΑΛΛΑΝΤΙΚΑ ΣΥΣΚΕΥΑΣΜΕΝΑ ΨΥΓΕΙΟΥ → ΤΥΡΙΑ ΔΙΑΦΟΡΑ ΣΥΣΚ	3,660	4,003
ΓΙΑΟΥΡΤΙΑ → ΓΑΛΑ ΦΡΕΣΚΟ	1,597	1,678
ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ → ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	2,034	2,013
ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ → ΦΡΕΣΚΑ ΦΡΟΥΤΑ	2,800	2,763
ΚΑΘΑΡΙΣΤΙΚΑ → ΠΡΟΪΟΝΤΑ ΠΛΥΣΙΜΑΤΟΣ	3,253	3,651
ΛΑΧΑΝΙΚΑ ΣΕ ΚΟΝΣΕΡΒΑ/ΧΑΡΤΙΝΗ ΣΥΣΚΕΥΑΣΙΑ → ΖΥΜΑΡΙΚΑ	3,375	3,799
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ → ΓΑΛΑ ΦΡΕΣΚΟ	1,508	1,588
ΠΡΩΙΝΑ → ΓΑΛΑ ΦΡΕΣΚΟ	1,775	1,898
ΦΡΕΣΚΑ ΦΡΟΥΤΑ → ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	2,104	2,136
ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ → ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	2,269	2,400

Ιδιαίτερα υψηλές τιμές lift εμφανίζουν οι κανόνες **Αλλαντικά Συσκευασμένα Ψυγείου → Τυριά Διάφορα Συσκ, Καθαριστικά → Προϊόντα Πλυσίματος** και **Λαχανικά σε Κονσέρβα/Χάρτινη Συσκευασία → Ζυμαρικά**. Οι κανόνες αυτοί εμφανίζονται και στις δύο περιόδους και αποτυπώνουν συστηματική συνύπαρξη κατηγοριών που εξυπηρετούν συναφείς καταναλωτικές ανάγκες.

Παράλληλα, οι κανόνες **Γιαούρτια → Γάλα Φρέσκο, Μαλακά Τυριά/Κιλό → Γάλα Φρέσκο** και **Πρωινά → Γάλα Φρέσκο** παραμένουν κοινοί στις δύο περιόδους. Αντίστοιχα, η παρουσία κανόνων που περιλαμβάνουν φρέσκα φρούτα και λαχανικά αποτελεί ένδειξη συνέχειας ορισμένων προτύπων συνδυαστικής αγοράς.

Για την πληρέστερη αποτύπωση των μεταβολών δεν εξετάστηκαν μόνο οι ενδεικτικοί κανόνες του Πίνακα 4.12, αλλά πραγματοποιήθηκε συστηματική σύγκριση όλων των κανόνων που εμφανίζονται και στις δύο περιόδους. Ένας κανόνας θεωρήθηκε κοινός μόνο όταν παρουσίαζε το ίδιο antecedent και το ίδιο consequent και στις δύο περιόδους. Η κατεύθυνση διατηρήθηκε, επομένως ο κανόνας $A \rightarrow B$ αντιμετωπίστηκε ως διαφορετικός από τον $B \rightarrow A$.

Με αυτή τη διαδικασία εντοπίστηκαν αρχικά 147 κοινοί κατευθυνόμενοι κανόνες. Μετά την εξαίρεση των κανόνων που περιείχαν τις κατηγορίες «Είδη Συσκευασίας προς Πώληση» και «Απαράδεκτα Εμπορεύματα», παρέμειναν 124 κοινοί κανόνες. Από αυτούς, οι 72 είχαν μία κατηγορία στο antecedent και μία στο consequent, οι 50 είχαν δύο κατηγορίες στο antecedent και μία στο consequent, ενώ δύο κανόνες είχαν τρεις κατηγορίες στο antecedent και μία στο consequent.

Για κάθε κοινό κανόνα υπολογίστηκαν η απόλυτη και η σχετική μεταβολή του confidence και του lift. Η απόλυτη διαφορά του confidence εκφράστηκε σε ποσοστιαίες μονάδες και υπολογίστηκε ως:

$$\Delta Confidence = Confidence_{Post-COVID} - Confidence_{COVID}$$

Αντίστοιχα, η μεταβολή του lift υπολογίστηκε ως:

$$\Delta Lift = Lift_{Post-COVID} - Lift_{COVID}$$

Η συνολική εικόνα των μεταβολών των δύο μετρικών παρουσιάζεται στον Πίνακα 4.13.

Πίνακας 4.13: Συνολική σύγκριση confidence και lift στους 124 κοινούς κανόνες

Μετρική	Μέση τιμή COVID	Μέση τιμή Post-COVID	Μέση μεταβολή	Κανόνες με αύξηση	Κανόνες με μείωση
Confidence	38,525%	36,398%	-2,128 π.μ.	20	104
Lift	2,256	2,379	+0,123	112	12

Όπως προκύπτει από τον Πίνακα 4.13, το confidence μειώθηκε στους 104 από τους 124 κοινούς κανόνες και αυξήθηκε στους 20. Η μέση τιμή του υποχώρησε από 38,525% σε 36,398%, παρουσιάζοντας μέση διαφορά -2,128 ποσοστιαίων μονάδων. Η διάμεση μεταβολή ήταν επίσης αρνητική και διαμορφώθηκε στις -2,110 ποσοστιαίες μονάδες, γεγονός που δείχνει ότι η μείωση δεν περιορίζεται σε έναν μικρό αριθμό ακραίων περιπτώσεων.

Η μεγαλύτερη αύξηση του confidence καταγράφηκε στον κανόνα **Γιαούρτια + Νερά → Φρέσκα Φρούτα**, του οποίου το confidence αυξήθηκε από 30,999% σε 33,009%, δηλαδή κατά 2,010 ποσοστιαίες μονάδες. Αντίθετα, η μεγαλύτερη μείωση εμφανίστηκε στον κανόνα **Χυμοί Μακράς Διαρκείας → Γάλα Φρέσκο**, με το confidence να υποχωρεί από 39,735% σε 34,009%, παρουσιάζοντας διαφορά -5,727 ποσοστιαίων μονάδων.

Αξιοσημείωτο είναι ότι και οι δέκα κανόνες με τη μεγαλύτερη μείωση του confidence έχουν ως consequent το φρέσκο γάλα. Το αποτέλεσμα αυτό συμφωνεί με τη μείωση της συχνότητας εμφάνισης της συγκεκριμένης κατηγορίας στη μεταπανδημική περίοδο. Δείχνει, επομένως, ότι όταν αγοράζονται άλλες σχετικές κατηγορίες, η πιθανότητα να περιλαμβάνεται στο ίδιο καλάθι και φρέσκο γάλα είναι χαμηλότερη σε σύγκριση με την περίοδο COVID-19.

Διαφορετική εικόνα παρουσιάζει το lift, το οποίο αυξήθηκε στους 112 από τους 124 κοινούς κανόνες και μειώθηκε μόνο στους 12. Η μέση τιμή του αυξήθηκε από 2,256 σε 2,379, με μέση μεταβολή +0,123. Θετική ήταν και η διάμεση μεταβολή, η οποία διαμορφώθηκε σε +0,091. Επομένως, η αύξηση του lift χαρακτηρίζει τη μεγάλη πλειονότητα των κοινών κανόνων και δεν οφείλεται αποκλειστικά σε λίγες ακραίες τιμές. Οι συγκεκριμένες μέσες τιμές αφορούν αποκλειστικά τους 124 κοινούς κανόνες και όχι το σύνολο των κανόνων που παρήχθησαν χωριστά σε κάθε περίοδο.

Η μεγαλύτερη αύξηση του lift καταγράφηκε στον κανόνα **Αλεύρι-Σιμιγδάλι → Είδη Ζαχαροπλαστικής**, όπου η τιμή του αυξήθηκε από 4,947 σε 6,138, παρουσιάζοντας διαφορά +1,191. Αντίθετα, η μεγαλύτερη μείωση εμφανίστηκε στον κανόνα **Λαχανικά Πράσινα Ποικιλία + Φρέσκα Φρούτα → Εξωτικά Φρούτα**. Στην περίπτωση αυτή το lift μειώθηκε από 3,266 σε 3,203, δηλαδή μόνο κατά 0,063.

Η ταυτόχρονη μείωση του confidence και αύξηση του lift στους περισσότερους κοινούς κανόνες δεν αποτελεί αντίφαση. Το lift υπολογίζεται σε σχέση με τη γενικότερη συχνότητα εμφάνισης του consequent:

$$Lift(A \rightarrow B) = \frac{Confidence(A \rightarrow B)}{Support(B)}$$

Επομένως, όταν το support του consequent μειώνεται αναλογικά περισσότερο από το confidence του κανόνα, το lift μπορεί να αυξάνεται ακόμη και αν η πιθανότητα εμφάνισης του consequent παρουσία του antecedent υποχωρεί. Χαρακτηριστικό παράδειγμα αποτελεί ο κανόνας **Χυμοί Μακράς Διαρκείας → Γάλα Φρέσκο**: παρότι το confidence μειώθηκε κατά 5,727 ποσοστιαίες μονάδες, το lift αυξήθηκε ελαφρά από 1,391 σε 1,414.

Το αποτέλεσμα αυτό εξηγεί γιατί πρέπει να διακρίνεται η συχνότητα εμφάνισης ενός συνδυασμού από τη σχετική ισχύ της συσχέτισης. Παρότι τα κοινά frequent itemsets εμφανίζουν γενικά χαμηλότερο support και οι περισσότεροι κοινόι κανόνες χαμηλότερο confidence κατά τη μεταπανδημική περίοδο, η πλειονότητα των κανόνων παρουσιάζει υψηλότερο lift. Συνεπώς, ο μεγαλύτερος αριθμός frequent itemsets και κανόνων που εντοπίστηκε κατά την περίοδο COVID-19 δεν αποτελεί από μόνος του ένδειξη ισχυρότερων συσχετίσεων.

Η επανεμφάνιση των συγκεκριμένων κανόνων και στις δύο περιόδους αποτελεί ένδειξη συνέχειας ορισμένων προτύπων συνδυαστικής αγοράς. Παράλληλα, οι μεταβολές του confidence και του lift αναδεικνύουν διαφοροποιήσεις στον τρόπο με τον οποίο συνδέονται οι κατηγορίες προϊόντων μεταξύ των δύο χρονικών περιόδων. Οι διαφοροποιήσεις αυτές δεν μπορούν να αποδοθούν αποκλειστικά στην πανδημία, καθώς ενδέχεται να συνδέονται και με άλλους οικονομικούς, κοινωνικούς και εμπορικούς παράγοντες.

Οι αναλυτικοί πίνακες με τις δέκα μεγαλύτερες αυξήσεις και μειώσεις του confidence και του lift παρουσιάζονται στο Παράρτημα Β (Πίνακες Β'3– Β'6).

4.3.4 Συμπέρασμα Συγκριτικής Ανάλυσης Περιόδων με χρήση Apriori

Η συγκριτική ανάλυση των δύο περιόδων ανέδειξε τόσο στοιχεία συνέχειας όσο και επιμέρους μεταβολές στην καταναλωτική συμπεριφορά. Κατά την περίοδο COVID-19 εντοπίστηκε μεγαλύτερος συνολικός αριθμός frequent itemsets και κανόνων συσχέτισης, καθώς προέκυψαν 1.687 frequent itemsets και 354 κανόνες, έναντι 1.104 frequent itemsets και 216 κανόνων κατά τη μεταπανδημική περίοδο. Παράλληλα, η σύγκριση των κοινών frequent itemsets έδειξε γενικότερη υποχώρηση της συχνότητας εμφάνισής τους. Από τα 705 κοινά itemsets που περιείχαν τουλάχιστον δύο κατηγορίες προϊόντων, μετά τις εξαιρέσεις που αναφέρθηκαν προηγουμένως, τα 671 παρουσίασαν μείωση του support και τα 34 αύξηση.

Παρά τις μεταβολές αυτές, αρκετοί από τους ισχυρούς κανόνες συσχέτισης εμφανίζονται και στις δύο χρονικές περιόδους, γεγονός που αποτελεί ένδειξη συνέχειας ορισμένων προτύπων συνδυαστικής αγοράς. Μεταξύ αυτών περιλαμβάνονται συσχετίσεις που αφορούν γαλακτοκομικά προϊό-

ντα, φρέσκα τρόφιμα, προϊόντα πρωινού και είδη οικιακής χρήσης. Επομένως, παρότι η συχνότητα εμφάνισης αρκετών συνδυασμών μεταβλήθηκε, ορισμένες βασικές σχέσεις μεταξύ συμπληρωματικών κατηγοριών εξακολούθησαν να εμφανίζονται στα καλάθια αγορών και στις δύο περιόδους.

Η σύγκριση των 124 κοινών κανόνων αποτύπωσε διαφορετική κατεύθυνση στις μεταβολές του confidence και του lift. Το confidence μειώθηκε στους 104 κανόνες και αυξήθηκε στους 20, με τη μέση τιμή του να υποχωρεί από 38,525% σε 36,398%. Αντίθετα, το lift αυξήθηκε στους 112 κανόνες και μειώθηκε μόνο στους 12, ενώ η μέση τιμή του αυξήθηκε από 2,256 σε 2,379. Το αποτέλεσμα αυτό δείχνει ότι, μολονότι σε αρκετές περιπτώσεις μειώθηκε η πιθανότητα εμφάνισης του consequent παρουσία του antecedent, η σχετική ισχύς της συσχέτισης, λαμβάνοντας υπόψη τη γενικότερη συχνότητα εμφάνισης του consequent, δεν παρουσίασε αντίστοιχη εξασθένηση. Συνεπώς, ο μεγαλύτερος αριθμός frequent itemsets και κανόνων κατά την περίοδο COVID-19 δεν αποτελεί από μόνος του ένδειξη ισχυρότερων συσχετίσεων.

Ο μικρότερος αριθμός frequent itemsets κατά τη μεταπανδημική περίοδο μπορεί να συνδέεται εν μέρει με τη γενικότερη υποχώρηση των τιμών support και με το μικρότερο μέσο μέγεθος των καλάθιων. Όταν μία συναλλαγή περιλαμβάνει λιγότερες διαφορετικές κατηγορίες προϊόντων, περιορίζεται και το πλήθος των πιθανών συνδυασμών που μπορούν να υπερβούν το ελάχιστο όριο υποστήριξης. Το εύρημα αυτό δεν επαρκεί, ωστόσο, για να χαρακτηριστούν τα καλάθια της μίας περιόδου περισσότερο ομοιογενή ή περισσότερο διαφοροποιημένα από εκείνα της άλλης.

Συνολικά, τα αποτελέσματα δείχνουν ότι ορισμένα βασικά πρότυπα συνδυαστικής αγοράς διατηρήθηκαν, ενώ ταυτόχρονα καταγράφηκαν διαφοροποιήσεις στη συχνότητα εμφάνισης επιμέρους κατηγοριών, itemsets και κανόνων συσχέτισης. Οι μεταβολές αυτές δεν μπορούν να αποδοθούν αποκλειστικά στην πανδημία COVID-19, καθώς μεταξύ των δύο περιόδων ενδέχεται να διαφοροποιήθηκαν και άλλοι οικονομικοί, κοινωνικοί και εμπορικοί παράγοντες. Επομένως, τα ευρήματα ερμηνεύονται ως διαφοροποιήσεις μεταξύ των δύο εξεταζόμενων χρονικών περιόδων και όχι ως αποδεδειγμένες επιδράσεις της πανδημίας.

4.4 Συγκριτική Αξιολόγηση των Αλγορίθμων Apriori, FP-Growth και Eclat

Μετά την ολοκλήρωση της ανάλυσης της καταναλωτικής συμπεριφοράς για την περίοδο της πανδημίας COVID-19 και τη μεταπανδημική περίοδο μέσω του αλγορίθμου Apriori, κρίθηκε σκόπιμη η διερεύνηση της αποδοτικότητας διαφορετικών αλγορίθμων εξόρυξης κανόνων συσχέτισης στο σύνολο των διαθέσιμων δεδομένων.

Προκειμένου να επιτευχθεί αυτό δημιουργήθηκε ένα ενοποιημένο σύνολο δεδομένων (merged dataset), το οποίο περιελάμβανε το σύνολο των συναλλαγών και από τις δύο χρονικές περιόδους. Η χρήση του ενοποιημένου συνόλου επιτρέπει την αξιολόγηση των αλγορίθμων σε ένα ιδιαίτερα μεγάλο περιβάλλον δεδομένων, προσεγγίζοντας ρεαλιστικές συνθήκες εφαρμογής σε μεγάλης κλίμακας καταστήματα λιανικού εμπορίου.

Η θεωρητική σύγκριση παρουσιάστηκε στην Ενότητα 2.5.5, ενώ στην παρούσα ενότητα παρουσιάζεται η πειραματική αξιολόγηση των αλγορίθμων Apriori, FP-Growth και Eclat με βάση κοινές παραμέτρους εξόρυξης, ώστε τα αποτελέσματα να είναι άμεσα συγκρίσιμα. Συγκεκριμένα, και στις τρεις περιπτώσεις χρησιμοποιήθηκαν οι ακόλουθες τιμές παραμέτρων:

- Minimum Support = 0,01
- Minimum Confidence = 0,30
- Minimum Lift = 1,30
- Minimum Leverage = 0,005
- Minimum Conviction = 1,0

Για λόγους συγκρισιμότητας και ερμηνευσιμότητας, μετά την εφαρμογή των κοινών κριτηρίων φιλτραρίσματος διατηρήθηκαν και στους τρεις αλγορίθμους μόνο κανόνες με μία κατηγορία στο antecedent και μία κατηγορία στο consequent, δηλαδή κανόνες $1 \rightarrow 1$.

Η επιλογή των συγκεκριμένων τιμών βασίστηκε στα αποτελέσματα της προηγούμενης πειραματικής διαδικασίας, κατά την οποία οι παράμετροι αυτοί παρήγαγαν κανόνες με υψηλή ερμηνευτική αξία και ικανοποιητική ισορροπία μεταξύ πληρότητας και ποιότητας αποτελεσμάτων.

Η σύγκριση των αλγορίθμων πραγματοποιείται τόσο ως προς την αποτελεσματικότητα των αποτελεσμάτων που παράγουν όσο και ως προς την υπολογιστική τους απόδοση. Ειδικότερα, αξιολογείται το πλήθος των συχνών συνόλων αντικειμένων (frequent itemsets), το πλήθος των παραγόμενων κανόνων συσχέτισης, ο συνολικός χρόνος εκτέλεσης καθώς και η μέγιστη κατανάλωση μνήμης κατά τη διάρκεια της επεξεργασίας.

Μέσω της ανάλυσης αυτής επιδιώκεται να προσδιοριστεί ποιος αλγόριθμος είναι πιο κατάλληλος για την εξόρυξη κανόνων συσχέτισης σε μεγάλα σύνολα δεδομένων, λαμβάνοντας υπόψη την ικανότητά του να παράγει τα επιθυμητά αποτελέσματα με το μικρότερο δυνατό υπολογιστικό κόστος σε χρόνο και μνήμη.

4.4.1 Αποτελέσματα Apriori

Ο αλγόριθμος Apriori εφαρμόστηκε στο ενοποιημένο σύνολο δεδομένων, το οποίο περιλάμβανε συνολικά 9.510.151 συναλλαγές και 329 μοναδικές κατηγορίες προϊόντων. Για την εξόρυξη των κανόνων συσχέτισης χρησιμοποιήθηκαν οι παράμετροι που επιλέχθηκαν από την προηγούμενη πειραματική διαδικασία (min_support = 0.01, min_confidence = 0.30, min_lift = 1.30, min_leverage = 0.005 και min_conviction = 1.0). Ο πλήρης κώδικας που χρησιμοποιήθηκε για την εκτέλεση του αλγορίθμου Apriori παρουσιάζεται στην Ενότητα Α'3 του Παραρτήματος Α.

Η εκτέλεση του αλγορίθμου οδήγησε στην εξαγωγή 1.329 συχνών στοιχειοσυνόλων (frequent itemsets) και 101 κανόνων συσχέτισης που ικανοποιούσαν τα καθορισμένα κριτήρια φιλτραρίσματος. Παράλληλα, καταγράφηκαν ο συνολικός χρόνος εκτέλεσης και η μέγιστη κατανάλωση μνήμης, προκειμένου να αξιολογηθεί η υπολογιστική αποδοτικότητα του αλγορίθμου.

Ο Πίνακας 4.14 παρουσιάζει τα συνοπτικά αποτελέσματα της εκτέλεσης.

Πίνακας 4.14: Αποτελέσματα εκτέλεσης Apriori

Μετρική	Τιμή
Transactions	9.510.151
Unique Items	329
Frequent Itemsets	1.329
Association Rules	101
Runtime (sec)	741,98
Peak Memory (MB)	14.634,14

Από τα αποτελέσματα προκύπτει ότι ο Apriori κατάφερε να επεξεργαστεί επιτυχώς ένα ιδιαίτερα μεγάλο σύνολο δεδομένων, εξάγοντας σημαντικό αριθμό συχνών συνόλων και ισχυρών κανόνων συσχέτισης. Παρόλα αυτά, η διαδικασία απαίτησε σημαντικούς υπολογιστικούς πόρους, καθώς ο χρόνος εκτέλεσης ξεπέρασε τα 12 λεπτά, ενώ η μέγιστη κατανάλωση μνήμης προσέγγισε τα 14,6 GB.

Χάρη στην ανάλυση των κανόνων συσχέτισης παρουσιάστηκαν ορισμένα ιδιαίτερα ισχυρά και επαναλαμβανόμενα πρότυπα αγορών. Οι ισχυρότεροι κανόνες αφορούν κυρίως τις κατηγορίες γαλακτοκομικών προϊόντων, αλλαντικών, φρέσκων τροφίμων και ειδών καθημερινής κατανάλωσης.

Ενδεικτικά, ο ισχυρότερος κανόνας που προέκυψε ήταν:

ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ → ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ

με support 8,28%, confidence 54,03%, lift 4,40 και leverage 0,064. Η συγκεκριμένη συσχέτιση υποδηλώνει ότι οι καταναλωτές που αγοράζουν μαλακά τυριά κοπής έχουν σημαντικά αυξημένη πιθανότητα να αγοράσουν και αλλαντικά χύμα κατά την ίδια συναλλαγή.

Αντίστοιχα, ιδιαίτερα ισχυρή παρουσιάζεται και η αντίστροφη σχέση:

ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ → ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ

με confidence 67,53% και τον ίδιο υψηλό δείκτη lift (4,40), επιβεβαιώνοντας τη στενή συμπληρωματικότητα των δύο κατηγοριών.

Σημαντικοί κανόνες εντοπίστηκαν επίσης μεταξύ κατηγοριών φρέσκων τροφίμων, όπως:

• ΦΡΕΣΚΑ ΦΡΟΥΤΑ → ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ

• ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ → ΦΡΕΣΚΑ ΦΡΟΥΤΑ

• ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ → ΦΡΕΣΚΑ ΦΡΟΥΤΑ

οι οποίοι υποδεικνύουν την τάση των καταναλωτών να συνδυάζουν διαφορετικές κατηγορίες φρέσκων προϊόντων στο ίδιο καλάθι αγορών.

Ταυτόχρονα, εμφανίζονται ισχυρές σχέσεις μεταξύ κατηγοριών που σχετίζονται με το πρωινό και την καθημερινή κατανάλωση, όπως οι συσχετίσεις μεταξύ γάλακτος, γιαουρτιών και προϊόντων πρωινού. Ενδεικτικά, ο κανόνας:

ΓΙΑΟΥΡΤΙΑ → ΓΑΛΑ ΦΡΕΣΚΟ

παρουσιάζει support 8,07%, confidence 43,01% και lift 1,64, γεγονός που υποδηλώνει τη συχνή συνδυαστική αγορά των δύο κατηγοριών.

Ταυτόχρονα, παρατηρήθηκαν ισχυρές συσχετίσεις σε κατηγορίες ειδών οικιακής χρήσης, όπως:

ΚΑΘΑΡΙΣΤΙΚΑ → ΠΡΟΪΟΝΤΑ ΠΛΥΣΙΜΑΤΟΣ

με lift 3,46, που επιβεβαιώνει ότι οι καταναλωτές τείνουν να προμηθεύονται ταυτόχρονα προϊόντα που εξυπηρετούν παρόμοιες ανάγκες καθαριότητας.

Τα αποτελέσματα του Apriori δείχνουν ότι παρά το ιδιαίτερα μεγάλο μέγεθος του συνόλου δεδομένων, οι σημαντικότεροι κανόνες συσχέτισης αφορούν κυρίως συμπληρωματικές κατηγορίες προϊόντων οι οποίες εμφανίζονται διαχρονικά στα καλάθια αγορών των καταναλωτών. Τα ευρήματα αυτά επιβεβαιώνουν τη δυνατότητα του αλγορίθμου να αναδεικνύει ουσιαστικά πρότυπα αγοραστικής συμπεριφοράς, αν και με αυξημένες απαιτήσεις σε χρόνο εκτέλεσης και μνήμη.

4.4.2 Αποτελέσματα FP-Growth

Ο αλγόριθμος FP-Growth εφαρμόστηκε επίσης στο ενοποιημένο σύνολο δεδομένων. Για τη διασφάλιση της συγκρισιμότητας των αποτελεσμάτων χρησιμοποιήθηκαν οι ίδιες παράμετροι εξόρυξης που εφαρμόστηκαν σε όλα τα πειράματα, δηλαδή $\text{min_support} = 0.01$, $\text{min_confidence} = 0.30$, $\text{min_lift} = 1.30$, $\text{min_leverage} = 0.005$ και $\text{min_conviction} = 1.0$. Ο πλήρης κώδικας που χρησιμοποιήθηκε για την εκτέλεση του αλγορίθμου FP-Growth παρουσιάζεται στην Ενότητα Α'4 του Παραρτήματος Α.

Η εκτέλεση του αλγορίθμου οδήγησε στην εξαγωγή 1.329 συχνών συνόλων αντικειμένων (frequent itemsets) και 101 κανόνων συσχέτισης που ικανοποιούσαν τα προκαθορισμένα κριτήρια φιλτραρίσματος. Τα αποτελέσματα αυτά αποτυπώνονται συνοπτικά στον Πίνακα 4.15.

Πίνακας 4.15: Αποτελέσματα εκτέλεσης FP-Growth

Μετρική	Τιμή
Transactions	9.510.151
Unique Items	329
Frequent Itemsets	1.329
Association Rules	101
Runtime (sec)	651,77
Peak Memory (MB)	22.660,63

Από την ανάλυση των συχνών στοιχειοσυνόλων προκύπτει ότι βασικές κατηγορίες προϊόντων καθημερινής κατανάλωσης, όπως το φρέσκο γάλα, τα γιαούρτια, τα φρέσκα φρούτα και τα λαχανικά, εμφανίζουν υψηλά ποσοστά υποστήριξης στο ενοποιημένο σύνολο δεδομένων. Στα συχνότερα αποτελέσματα περιλαμβάνεται επίσης η κατηγορία «Είδη Συσκευασίας προς Πώληση», λόγω της παρουσίας της σε μεγάλο αριθμό συναλλαγών. Ωστόσο, καθώς αφορά κυρίως σακούλες μεταφοράς, δεν λαμβάνεται υπόψη στην ουσιαστική ερμηνεία των αγοραστικών προτύπων.

Οι σημαντικότεροι κανόνες συσχέτισης που προέκυψαν αφορούν κυρίως συμπληρωματικές κατηγορίες προϊόντων. Ο ισχυρότερος κανόνας που εντοπίστηκε ήταν ο:

ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ → ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ

με support 8,28%, confidence 54,03%, lift 4,40 και leverage 0,064. Το αποτέλεσμα αυτό υποδηλώνει ότι περισσότεροι από τους μισούς καταναλωτές που αγοράζουν μαλακά τυριά από τον πάγκο επιλέγουν ταυτόχρονα και αλλαντικά χύμα.

Αντίστοιχα, ιδιαίτερα ισχυρός εμφανίζεται και ο κανόνας:

ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ → ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ

με confidence 67,53% και lift 4,40, επιβεβαιώνοντας τη στενή συνάφεια των δύο κατηγοριών προϊόντων. Σημαντικοί κανόνες προέκυψαν επίσης μεταξύ κατηγοριών φρέσκων τροφίμων, όπως:

- **ΦΡΕΣΚΑ ΦΡΟΥΤΑ → ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ**
- **ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ → ΦΡΕΣΚΑ ΦΡΟΥΤΑ**
- **ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ → ΦΡΕΣΚΑ ΦΡΟΥΤΑ**

Οι συγκεκριμένοι κανόνες υποδεικνύουν ότι οι καταναλωτές συνηθίζουν να προμηθεύονται διαφορετικές κατηγορίες φρέσκων προϊόντων στο πλαίσιο της ίδιας επίσκεψης στο κατάστημα. Ακόμη, παρατηρούνται ισχυρές συσχετίσεις μεταξύ προϊόντων πρωινού και γαλακτοκομικών ειδών, όπως ο κανόνας:

ΓΙΑΟΥΡΤΙΑ → ΓΑΛΑ ΦΡΕΣΚΟ

ο οποίος παρουσιάζει υψηλή τιμή confidence και lift, επιβεβαιώνοντας ότι οι δύο κατηγορίες εμφανίζονται συχνά μαζί στο καλάθι αγορών. Επίσης και εδώ εντοπίστηκαν σημαντικές συσχετίσεις μεταξύ των κατηγοριών ειδών καθαριότητας και οικιακής χρήσης.

Συνολικά και τα αποτελέσματα του FP-Growth ανέδειξαν ένα σύνολο κανόνων με υψηλή ερμηνευτική αξία, οι οποίοι αποτυπώνουν σταθερά πρότυπα αγοραστικής συμπεριφοράς στο σύνολο των διαθέσιμων δεδομένων και επιβεβαιώνουν την ύπαρξη ισχυρών σχέσεων μεταξύ συμπληρωματικών κατηγοριών προϊόντων.

4.4.3 Αποτελέσματα Eclat

Τέλος ο αλγόριθμος Eclat εφαρμόστηκε στο ενοποιημένο σύνολο δεδομένων των δύο εξεταζόμενων περιόδων με τις ίδιες παραμέτρους που εφαρμόστηκαν και στα προηγούμενα πειράματα, δηλαδή min_support = 0.01, min_confidence = 0.30, min_lift = 1.30, min_leverage = 0.005 και min_conviction = 1.0. Ο πλήρης κώδικας που χρησιμοποιήθηκε για την εκτέλεση του αλγορίθμου Eclat παρουσιάζεται στην Ενότητα Α΄.5 του Παραρτήματος Α.

Η εκτέλεση του αλγορίθμου οδήγησε στην εξαγωγή 1.329 συχνών συνόλων αντικειμένων (frequent itemsets) και 101 κανόνων συσχέτισης που πληρούσαν τα καθορισμένα κριτήρια ποιότητας. Τα συνοπτικά αποτελέσματα παρουσιάζονται στον Πίνακα 4.16.

Πίνακας 4.16: Αποτελέσματα εκτέλεσης Eclat

Μετρική	Τιμή
Transactions	9.510.151
Unique Items	329
Frequent Itemsets	1.329
Association Rules	101
Runtime (sec)	102,53
Peak Memory (MB)	9.170,88

Από την ανάλυση των συχνών στοιχειοσυνόλων προκύπτει ότι βασικές κατηγορίες προϊόντων

καθημερινής κατανάλωσης, όπως το φρέσκο γάλα, τα γιαούρτια, τα φρέσκα φρούτα και τα λαχανικά, εμφανίζουν υψηλά ποσοστά υποστήριξης στο ενοποιημένο σύνολο δεδομένων. Στα συχνότερα αποτελέσματα περιλαμβάνεται επίσης η κατηγορία «Είδη Συσκευασίας προς Πώληση», λόγω της παρουσίας της σε μεγάλο αριθμό συναλλαγών. Ωστόσο, καθώς αφορά κυρίως σακούλες μεταφοράς, δεν λαμβάνεται υπόψη στην ουσιαστική ερμηνεία των αγοραστικών προτύπων.

Οι σημαντικότεροι κανόνες συσχέτισης που προέκυψαν αφορούν κυρίως συμπληρωματικές κατηγορίες προϊόντων, οι οποίες εμφανίζονται συστηματικά μαζί στις ίδιες συναλλαγές. Ο ισχυρότερος κανόνας που εντοπίστηκε και σε αυτόν τον αλγόριθμο ήταν ο:

ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ → ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ

με support 8,28%, confidence 54,03%, lift 4,40 και leverage 0,064. Ο κανόνας αυτός υποδηλώνει ότι οι καταναλωτές που αγοράζουν μαλακά τυριά από τον πάγκο έχουν αυξημένη πιθανότητα να επιλέξουν ταυτόχρονα και αλλαντικά χύμα.

Αντίστοιχα, ιδιαίτερα ισχυρή συσχέτιση εμφανίζεται και η αντίστροφη:

ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ → ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ

η οποία παρουσιάζει confidence 67,53% και lift 4,40, επιβεβαιώνοντας τη στενή συμπληρωματική σχέση των δύο κατηγοριών προϊόντων.

Σημαντικοί κανόνες προέκυψαν επίσης μεταξύ κατηγοριών φρέσκων τροφίμων. Ενδεικτικά, οι κανόνες:

- **ΦΡΕΣΚΑ ΦΡΟΥΤΑ → ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ**
- **ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ → ΦΡΕΣΚΑ ΦΡΟΥΤΑ**
- **ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ → ΦΡΕΣΚΑ ΦΡΟΥΤΑ**

καταδεικνύουν την τάση των καταναλωτών να συνδυάζουν διαφορετικές κατηγορίες φρέσκων προϊόντων κατά την ίδια επίσκεψη στο κατάστημα. Παράλληλα, εντοπίστηκαν ισχυρές συσχετίσεις μεταξύ γαλακτοκομικών προϊόντων και ειδών πρωινού. Χαρακτηριστικό παράδειγμα αποτελεί ο κανόνας:

ΓΙΑΟΥΡΤΙΑ → ΓΑΛΑ ΦΡΕΣΚΟ

ο οποίος παρουσιάζει υψηλές τιμές confidence και lift, γεγονός που επιβεβαιώνει τη συχνή συνδυαστική αγορά των δύο κατηγοριών. Και σε αυτό το πείραμα παρουσιάστηκαν σημαντικές σχέσεις μεταξύ προϊόντων οικιακής χρήσης και καθαριότητας.

Τέλος, τα αποτελέσματα του Eclat ανέδειξαν ένα σύνολο κανόνων συσχέτισης με υψηλή ερμηνευτική αξία, οι οποίοι αποτυπώνουν σταθερά πρότυπα αγοραστικής συμπεριφοράς στο σύνολο των διαθέσιμων δεδομένων και επιβεβαιώνουν την ύπαρξη ισχυρών σχέσεων μεταξύ συμπληρωματικών κατηγοριών προϊόντων.

4.4.4 Συγκριτική Αξιολόγηση των Αλγορίθμων Apriori, FP-Growth και Eclat

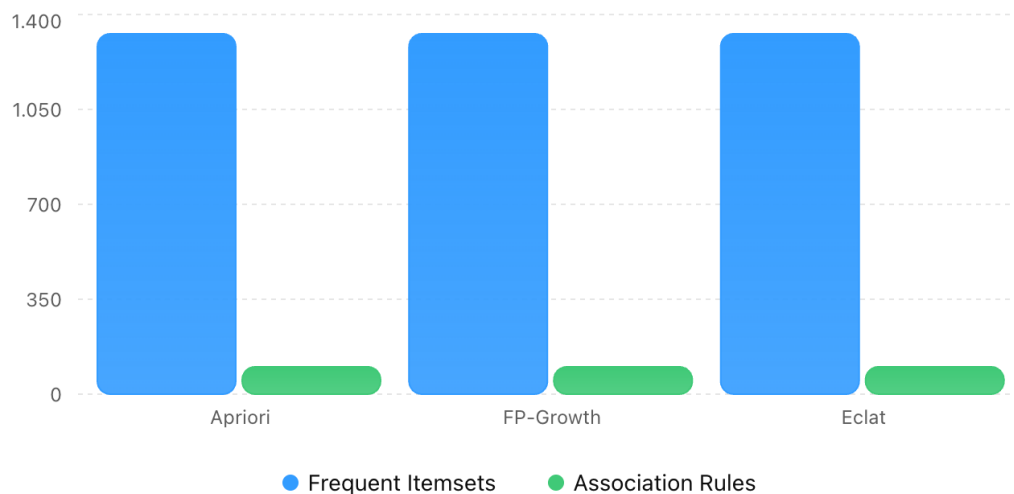
Πίνακας 4.17: Συγκριτική Αξιολόγηση των Αλγορίθμων Apriori, FP-Growth και Eclat

Αλγόριθμος	Frequent Itemsets	Association Rules	Runtime (sec)	Peak Memory (MB)
Apriori	1.329	101	741,98	14.634,14
FP-Growth	1.329	101	651,77	22.660,63
Eclat	1.329	101	102,53	9.170,88

Μετά την ολοκλήρωση των πειραμάτων πραγματοποιήθηκε συγκριτική αξιολόγηση των αλγορίθμων Apriori, FP-Growth και Eclat με στόχο τη διερεύνηση της αποτελεσματικότητας και της υπολογιστικής τους αποδοτικότητας στο ενοποιημένο σύνολο δεδομένων. Η αξιολόγηση βασίστηκε τόσο στην ποιότητα των παραγόμενων αποτελεσμάτων όσο και στις απαιτήσεις των αλγορίθμων σε χρόνο εκτέλεσης και κατανάλωση μνήμης.

Frequent itemsets και association rules

Πλήθος αποτελεσμάτων ανά αλγόριθμο.



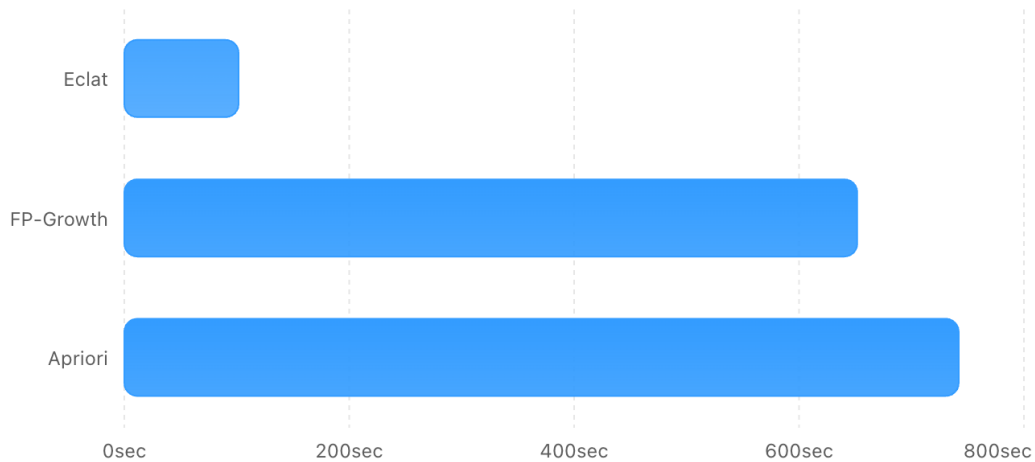
Σχήμα 4.2: Σύγκριση του πλήθους των frequent itemsets και των κανόνων συσχέτισης που εξήχθησαν από τους αλγορίθμους Apriori, FP-Growth και Eclat.

Αρχικά, όπως φαίνεται στο Σχήμα 4.2, και οι τρεις αλγόριθμοι παράγααν τον ίδιο αριθμό συχνών στοιχειοσυνόλων, δηλαδή 1.329 frequent itemsets, καθώς και τον ίδιο αριθμό κανόνων συσχέτισης, δηλαδή 101 association rules. Η αριθμητική αυτή συμφωνία δείχνει ότι, με τις ίδιες παραμέτρους εξόρυξης, οι τρεις αλγόριθμοι παράγααν ισάριθμα αποτελέσματα. Ωστόσο, το ίδιο πλήθος δεν αρκεί από μόνο του για να τεκμηριώσει ότι τα παραγόμενα itemsets και οι κανόνες ήταν απολύτως ταυτόσημοι ως προς το περιεχόμενό τους.

Κατά συνέπεια, η συγκριτική αξιολόγηση επικεντρώνεται κυρίως στις διαφορές της υπολογιστικής απόδοσης των αλγορίθμων και ειδικότερα στον χρόνο εκτέλεσης και στη μέγιστη καταγεγραμμένη κατανάλωση μνήμης.

Σύγκριση χρόνου εκτέλεσης

Μικρότερη τιμή σημαίνει ταχύτερη εκτέλεση.



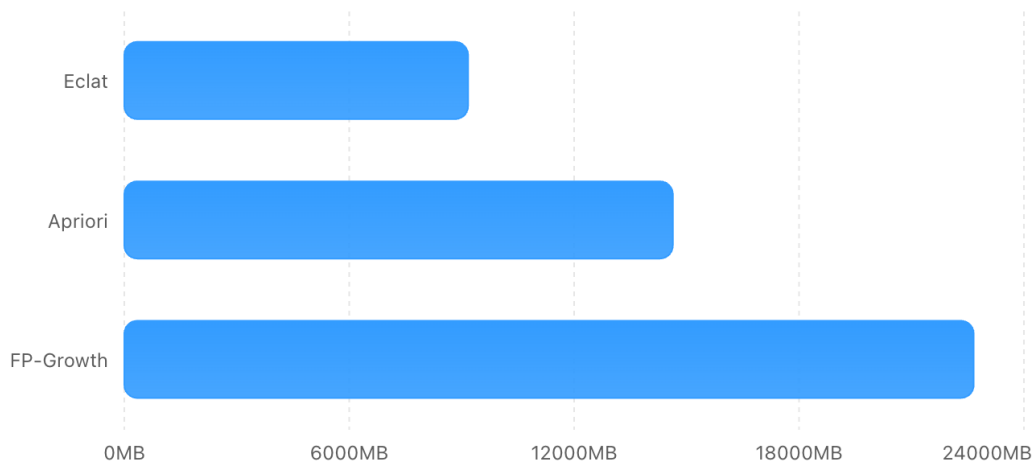
Σχήμα 4.3: Συγκριτική απεικόνιση χρόνου εκτέλεσης των αλγορίθμων Apriori, FP-Growth και Eclat.

Όπως φαίνεται στο Σχήμα 4.3, ο αλγόριθμος Eclat παρουσίασε τον μικρότερο χρόνο εκτέλεσης, ολοκληρώνοντας την επεξεργασία του συνόλου δεδομένων σε μόλις 102,53 δευτερόλεπτα. Αντίθετα, ο FP-Growth χρειάστηκε 651,77 δευτερόλεπτα, ενώ ο Apriori 741,98 δευτερόλεπτα. Η διαφορά είναι ιδιαίτερα έντονη, καθώς ο Eclat ολοκλήρωσε την επεξεργασία περίπου επτά φορές ταχύτερα από τον FP-Growth και περισσότερο από επτά φορές ταχύτερα από τον Apriori.

Η καλύτερη επίδοση του Eclat στη συγκεκριμένη εκτέλεση μπορεί να συνδέεται με τον κάθετο τρόπο αναπαράστασης των δεδομένων (vertical data layout), ο οποίος επιτρέπει την αποτελεσματικότερη διαχείριση των συνόλων αντικειμένων και μειώνει σημαντικά τον αριθμό των απαιτούμενων σαρώσεων του συνόλου δεδομένων. Αντίθετα, ο Apriori βασίζεται σε επαναλαμβανόμενη παραγωγή υποψήφιων συνόλων αντικειμένων (candidate generation), διαδικασία που αυξάνει σημαντικά το υπολογιστικό κόστος όταν το πλήθος των συναλλαγών είναι ιδιαίτερα μεγάλο.

Σύγκριση μέγιστης κατανάλωσης μνήμης

Μικρότερη τιμή σημαίνει χαμηλότερη απαίτηση σε μνήμη.



Σχήμα 4.4: Συγκριτική απεικόνιση μέγιστης κατανάλωσης μνήμης των αλγορίθμων Apriori, FP-Growth και Eclat.

Παρόμοια εικόνα παρατηρείται και ως προς τη χρήση μνήμης όπως φαίνεται στο Σχήμα 4.4. Ο Eclat παρουσίασε τη χαμηλότερη κατανάλωση μνήμης, απαιτώντας μόλις 9.170,88 MB κατά τη διάρκεια της εκτέλεσης. Ο Apriori ακολούθησε με 14.634,14 MB, ενώ ο FP-Growth εμφάνισε τη μεγαλύτερη απαίτηση σε μνήμη, παρουσιάζοντας μέγιστη κατανάλωση μνήμης συνολικά 22.660,63 MB.

Το αποτέλεσμα αυτό παρουσιάζει ιδιαίτερο ενδιαφέρον, καθώς στη διεθνή βιβλιογραφία ο FP-Growth θεωρείται συνήθως αποδοτικότερος από τον Apriori. Σε αυτή την περίπτωση όμως, το μεγάλο μέγεθος του συνόλου δεδομένων και η απαίτηση διατήρησης της δομής FP-Tree στη μνήμη οδήγησαν σε σημαντικά αυξημένη κατανάλωση πόρων. Αντίθετα, ο Eclat αξιοποίησε αποτελεσματικότερα τους διαθέσιμους πόρους, επιτυγχάνοντας ταυτόχρονα χαμηλότερο χρόνο εκτέλεσης και μικρότερη χρήση μνήμης.

Από τα αποτελέσματα της συγκριτικής αξιολόγησης προκύπτει ότι οι τρεις αλγόριθμοι παράγαν τον ίδιο αριθμό frequent itemsets και κανόνων συσχέτισης. Παρουσίασαν, ωστόσο, σημαντικές διαφορές ως προς την υπολογιστική τους απόδοση. Ο Eclat παρουσίασε την καλύτερη συνολική επίδοση, συνδυάζοντας τον μικρότερο χρόνο εκτέλεσης με τη χαμηλότερη κατανάλωση μνήμης. Αυτό τον καθιστά ιδιαίτερα κατάλληλο στο συγκεκριμένο σύνολο δεδομένων και πειραματικό περιβάλλον, για εφαρμογές εξόρυξης κανόνων συσχέτισης σε μεγάλα σύνολα δεδομένων λιανικού εμπορίου.

Κεφάλαιο 5

Συμπεράσματα και Μελλοντική Έρευνα

5.1 Συμπεράσματα της Έρευνας

Στο πλαίσιο της παρούσας μελέτης πραγματοποιήθηκε ανάλυση της καταναλωτικής συμπεριφοράς κατά την περίοδο COVID-19 και κατά τη μεταπανδημική περίοδο μέσω τεχνικών εξόρυξης κανόνων συσχέτισης (Association Rule Mining). Η ανάλυση βασίστηκε σε περισσότερες από 9,5 εκατομμύρια πραγματικές συναλλαγές από δεκαεπτά καταστήματα αλυσίδας λιανικού εμπορίου στον νομό Καβάλας. Ο μεγάλος όγκος και η προέλευση των δεδομένων από πραγματικές συναλλαγές επέτρεψαν τη διερεύνηση των αγοραστικών προτύπων σε συνθήκες λειτουργίας του λιανεμπορίου.

Η αρχική διερευνητική ανάλυση έδειξε ότι αρκετές βασικές κατηγορίες προϊόντων, όπως το φρέσκο γάλα, τα γιαούρτια, τα φρέσκα φρούτα και τα λαχανικά, διατήρησαν υψηλή συμμετοχή στα καλάθια αγορών και στις δύο χρονικές περιόδους. Παράλληλα, μετά τις εξαιρέσεις που αναφέρθηκαν στη μεθοδολογία, η συμπληρωματική σύγκριση των 315 κοινών κατηγοριών που εμφανίστηκαν και στις δύο περιόδους, ανεξάρτητα από το αν υπερέβησαν το καθορισμένο ελάχιστο όριο support, ανέδειξε μια γενικότερη υποχώρηση των ποσοστών υποστήριξης κατά τη μεταπανδημική περίοδο. Συγκεκριμένα, 229 κατηγορίες παρουσίασαν χαμηλότερο support και 86 υψηλότερο. Από αυτές, οι 113 υπερέβησαν το καθορισμένο ελάχιστο όριο support και στις δύο περιόδους και αποτέλεσαν τα κοινά frequent 1-itemsets. Οι μεγαλύτερες μειώσεις καταγράφηκαν στις κατηγορίες «Γάλα Φρέσκο» (-4,521 ποσοστιαίες μονάδες), «Γιαούρτια» (-2,989 ποσοστιαίες μονάδες) και «Αφρόλουτρα» (-2,759 ποσοστιαίες μονάδες). Αντίθετα, οι μεγαλύτερες αυξήσεις εμφανίστηκαν στις κατηγορίες «Έτοιμο Φαγητό Catering» (+1,069 ποσοστιαίες μονάδες), «Κρεμμύδια Κιλού» (+0,796 ποσοστιαίες μονάδες) και «Ενεργειακά Ποτά» (+0,795 ποσοστιαίες μονάδες). Οι μεταβολές, επομένως, δεν ήταν ομοιόμορφες, καθώς σε ορισμένες κατηγορίες καταγράφηκαν αυξήσεις και σε άλλες αισθητές μειώσεις. Το αποτέλεσμα αυτό παρέχει ένδειξη ότι, εκτός από τη γενικότερη υποχώρηση της συχνότητας εμφάνισης αρκετών κατηγοριών, σημειώθηκαν και επιμέρους διαφοροποιήσεις στη σύνθεση των αγορών.

Η σύγκριση του μεγέθους των καλαθιών συμπλήρωσε την παραπάνω εικόνα. Κατά την περίοδο COVID-19, κάθε συναλλαγή περιλάμβανε κατά μέσο όρο 7,1085 διαφορετικές κατηγορίες προϊόντων, ενώ κατά τη μεταπανδημική περίοδο ο αντίστοιχος μέσος όρος μειώθηκε σε 6,4014 κατηγορίες, παρουσιάζοντας σχετική μείωση 9,95%. Αντίθετα, η διάμεσος παρέμεινε ίση με πέντε κατηγορίες και στις δύο περιόδους. Η διαφορά του μέσου μεγέθους δείχνει ότι, συνολικά, οι συναλλαγές της περιόδου COVID-19 περιλάμβαναν περισσότερες διαφορετικές κατηγορίες προϊόντων, χωρίς όμως να μεταβάλλεται η κεντρική θέση της κατανομής, όπως αυτή αποτυπώνεται από τη διάμεσο.

Η εφαρμογή του αλγορίθμου Apriori οδήγησε στην εξαγωγή 1.687 συχνών στοιχειοσυνόλων και 354 κανόνων συσχέτισης κατά την περίοδο COVID-19, έναντι 1.104 στοιχειοσυνόλων και 216 κανόνων κατά τη μεταπανδημική περίοδο. Ο μεγαλύτερος αριθμός frequent itemsets και κανόνων κατά την περίοδο COVID-19 μπορεί να συνδέεται εν μέρει με τα γενικά υψηλότερα ποσοστά υποστήριξης και με το μεγαλύτερο μέσο μέγεθος των καλαθιών, καθώς η παρουσία περισσότερων κατηγοριών στην ίδια συναλλαγή δημιουργεί περισσότερους πιθανούς συνδυασμούς προϊόντων.

Το μεγαλύτερο πλήθος των παραγόμενων προτύπων δεν συνεπάγεται, ωστόσο, ότι οι συσχετίσεις κατά την περίοδο COVID-19 ήταν ισχυρότερες. Στο σύνολο των κανόνων που παρήχθησαν χωριστά σε κάθε περίοδο, το μέσο lift ήταν 2,5289 κατά την περίοδο COVID-19 και 2,9215 κατά τη μεταπανδημική περίοδο. Επομένως, το πλήθος των στοιχειοσυνόλων και των κανόνων πρέπει να διακρίνεται από την ισχύ των συσχετίσεων και να εξετάζεται σε συνδυασμό με τις αντίστοιχες μετρικές αξιολόγησης.

Η συστηματική σύγκριση των κοινών frequent itemsets και των κανόνων συσχέτισης ανέδειξε τόσο μεταβολές όσο και στοιχεία συνέχειας μεταξύ των δύο περιόδων. Από τα 705 κοινά itemsets που περιείχαν τουλάχιστον δύο κατηγορίες προϊόντων, μετά τις εξαιρέσεις που αναφέρθηκαν προηγουμένως, τα 671 παρουσίασαν χαμηλότερο support κατά τη μεταπανδημική περίοδο και τα 34 υψηλότερο. Ενδεικτικά, η μεγαλύτερη μείωση καταγράφηκε στον συνδυασμό «Γάλα Φρέσκο – Γιαούρτια», του οποίου το support μειώθηκε από 9,291% σε 7,013%, δηλαδή κατά 2,278 ποσοστιαίες μονάδες. Αντίθετα, η μεγαλύτερη αύξηση εμφανίστηκε στον συνδυασμό «Κρεμμύδια Κιλού – Λαχανικά Πράσινα Ποικιλία», με αύξηση του support από 2,358% σε 2,826%, δηλαδή κατά 0,469 ποσοστιαίες μονάδες. Στους 124 κοινούς κανόνες, το confidence μειώθηκε στους 104 και αυξήθηκε στους 20, με τη μέση τιμή του να υποχωρεί από 38,525% σε 36,398%. Αντίθετα, το lift αυξήθηκε στους 112 κανόνες και μειώθηκε στους 12, ενώ η μέση τιμή του αυξήθηκε από 2,256 σε 2,379. Τα αποτελέσματα δείχνουν ότι η συχνότητα εμφάνισης των συνδυασμών και η σχετική ισχύς των συσχετίσεων δεν μεταβλήθηκαν απαραίτητα προς την ίδια κατεύθυνση.

Παράλληλα, αρκετοί από τους κυριότερους συνδυασμούς και κανόνες εμφανίστηκαν και στις δύο χρονικές περιόδους. Μεταξύ αυτών περιλαμβάνονται συσχετίσεις που αφορούν γαλακτοκομικά προϊόντα, φρέσκα τρόφιμα, προϊόντα πρωινού και είδη οικιακής χρήσης. Η επανεμφάνιση αυτών των σχέσεων αποτελεί ένδειξη συνέχειας σε συγκεκριμένα πρότυπα συνδυαστικής αγοράς, χωρίς να σημαίνει ότι η συνολική καταναλωτική συμπεριφορά παρέμεινε αμετάβλητη.

Συνολικά, από τη σύγκριση των δύο περιόδων προκύπτουν τόσο γενικότερες όσο και επιμέρους διαφοροποιήσεις. Από τη μία πλευρά, οι χαμηλότερες τιμές support πολλών κατηγοριών και το μικρότερο μέσο μέγεθος του καλαθιού κατά τη μεταπανδημική περίοδο δείχνουν μια συνολικότερη μεταβολή στη δομή των συναλλαγών. Από την άλλη πλευρά, οι διαφορετικές αυξήσεις και μειώσεις ανά κατηγορία, καθώς και οι μεταβολές στις μετρικές των κοινών κανόνων, υποδεικνύουν ότι άλλαξε και η σχετική συμμετοχή συγκεκριμένων προϊόντων και συνδυασμών στα καλάθια αγορών.

Οι παρατηρούμενες διαφοροποιήσεις δεν μπορούν να αποδοθούν αποκλειστικά στην πανδημία COVID-19. Οι δύο εξεταζόμενες περίοδοι διαφέρουν και ως προς άλλους οικονομικούς, κοινωνικούς και εμπορικούς παράγοντες, όπως ο πληθωρισμός, οι μεταβολές των τιμών, η διαθεσιμότητα των προϊόντων και οι γενικότερες συνθήκες της αγοράς. Για τον λόγο αυτό, τα αποτελέσματα ερμηνεύονται ως διαφοροποιήσεις μεταξύ των δύο χρονικών περιόδων και όχι ως αποδεδειγμένες επιδράσεις της πανδημίας.

Εκτός από τη μελέτη των αγοραστικών προτύπων, πραγματοποιήθηκε συγκριτική αξιολόγηση

των αλγορίθμων Apriori, FP-Growth και Eclat στο ενοποιημένο σύνολο των 9.510.151 συναλλαγών. Με τις κοινές παραμέτρους που εφαρμόστηκαν, και οι τρεις αλγόριθμοι παράγαγαν τον ίδιο αριθμό αποτελεσμάτων, δηλαδή 1.329 συχνά στοιχειοσύνολα και 101 κανόνες συσχέτισης. Παρουσίασαν, ωστόσο, σημαντικές διαφορές ως προς τον χρόνο εκτέλεσης και τη μέγιστη καταγεγραμμένη χρήση μνήμης.

Στη συγκεκριμένη πειραματική διαδικασία, ο Eclat παρουσίασε τον μικρότερο χρόνο εκτέλεσης, ολοκληρώνοντας την ανάλυση σε 102,53 δευτερόλεπτα. Η μέγιστη καταγεγραμμένη χρήση μνήμης του ήταν επίσης η χαμηλότερη και ανήλθε σε 9.170,88 MB. Ο FP-Growth ολοκλήρωσε την εκτέλεση σε 651,77 δευτερόλεπτα, ενώ η μέγιστη καταγεγραμμένη χρήση μνήμης του ήταν η υψηλότερη και ανήλθε σε 22.660,63 MB. Ο Apriori κατέγραψε τον μεγαλύτερο χρόνο εκτέλεσης, 741,98 δευτερόλεπτα, και μέγιστη καταγεγραμμένη χρήση μνήμης 14.634,14 MB.

Τα αποτελέσματα αυτά αφορούν το συγκεκριμένο σύνολο δεδομένων, τις υλοποιήσεις και το υπολογιστικό περιβάλλον που χρησιμοποιήθηκαν και δεν πρέπει να ερμηνεύονται ως απόλυτη εγγενής υπεροχή ενός αλγορίθμου σε κάθε πιθανό σύνολο δεδομένων. Αναδεικνύουν, ωστόσο, τη σημασία επιλογής της κατάλληλης αλγοριθμικής προσέγγισης κατά την επεξεργασία μεγάλων συνόλων συναλλαγών.

Ο συνδυασμός της επιχειρησιακής ερμηνείας των αγοραστικών προτύπων με τη συγκριτική αξιολόγηση της υπολογιστικής αποδοτικότητας των αλγορίθμων επέτρεψε την εξαγωγή συμπερασμάτων τόσο για τη συμπεριφορά των καταναλωτών όσο και για την καταλληλότητα των μεθόδων που εφαρμόστηκαν. Παράλληλα, επιβεβαιώθηκε η χρησιμότητα του Market Basket Analysis ως εργαλείου για την αναγνώριση συχνών συνδυασμών προϊόντων, τη μελέτη των διαφοροποιήσεων μεταξύ χρονικών περιόδων και την υποστήριξη επιχειρηματικών αποφάσεων.

Συμπερασματικά, η παρούσα εργασία αναδεικνύει τη δυνατότητα αξιοποίησης των τεχνικών εξόρυξης κανόνων συσχέτισης για τη μελέτη πραγματικών δεδομένων λιανεμπορίου μεγάλης κλίμακας. Τα αποτελέσματα μπορούν να χρησιμοποιηθούν ως υποστηρικτικό εργαλείο για την καλύτερη κατανόηση των αγοραστικών συνηθειών, την οργάνωση των προϊόντων και τον σχεδιασμό προωθητικών ενεργειών, με την προϋπόθεση ότι ερμηνεύονται λαμβάνοντας υπόψη τη δομή των δεδομένων και τους περιορισμούς της ανάλυσης.

5.2 Συνεισφορά της Έρευνας

Η παρούσα εργασία συνεισφέρει τόσο σε ερευνητικό όσο και σε πρακτικό επίπεδο. Αρχικά, αξιοποιήθηκε ένα ιδιαίτερα μεγάλο και ρεαλιστικό σύνολο δεδομένων λιανικού εμπορίου, το οποίο περιελάμβανε περισσότερες από 9,5 εκατομμύρια συναλλαγές που προέρχονται από δεκαεπτά καταστήματα της ίδιας αλυσίδας λιανικής.

Ταυτόχρονα, πραγματοποιήθηκε συστηματική σύγκριση δύο διακριτών ιστορικών περιόδων, της περιόδου COVID-19 και της μεταπανδημικής περιόδου, επιτρέποντας την αποτύπωση των διαφοροποιήσεων που προέκυψαν στην καταναλωτική συμπεριφορά.

Η σύγκριση των δύο περιόδων ενισχύθηκε με την εξέταση των ποσοστών υποστήριξης των κοινών κατηγοριών προϊόντων και του μέσου μεγέθους του καλαθιού. Επιπλέον, υπολογίστηκαν οι μεταβολές του support στα κοινά συχνά στοιχειοσύνολα, καθώς και οι μεταβολές του confidence και του lift στους κοινούς κανόνες συσχέτισης. Με τον τρόπο αυτό κατέστη δυνατή η πληρέστερη αποτύπωση των διαφοροποιήσεων μεταξύ των δύο περιόδων και η διάκριση μεταξύ της συχνότητας

εμφάνισης των αγοραστικών προτύπων και της ισχύος των αντίστοιχων συσχετίσεων.

Ιδιαίτερη συνεισφορά αποτελεί επίσης ο συνδυασμός επιχειρησιακής ανάλυσης και υπολογιστικής αξιολόγησης. Εκτός από τη διερεύνηση των καταναλωτικών προτύπων, πραγματοποιήθηκε συγκριτική αξιολόγηση των αλγορίθμων Apriori, FP-Growth και Eclat σε πραγματικές συνθήκες λιανικού εμπορίου μεγάλης κλίμακας.

Τέλος, τα αποτελέσματα της εργασίας μπορούν να αξιοποιηθούν από επιχειρήσεις λιανικής για τη λήψη αποφάσεων που σχετίζονται με την οργάνωση προϊόντων, τη διαμόρφωση προωθητικών ενεργειών και τη βελτιστοποίηση της εμπειρίας αγορών των καταναλωτών.

5.3 Περιορισμοί της Μελέτης

Παρά τη μεγάλη έκταση του συνόλου δεδομένων και τη σημαντική ερμηνευτική αξία των αποτελεσμάτων, η παρούσα μελέτη παρουσιάζει ορισμένους περιορισμούς.

Αρχικά, τα δεδομένα προέρχονται από έναν μόνο νομό της Ελλάδας και συνεπώς τα αποτελέσματα ενδέχεται να μην είναι πλήρως γενικεύσιμα σε άλλες γεωγραφικές περιοχές ή σε διαφορετικά καταναλωτικά περιβάλλοντα.

Αξίζει επίσης να σημειωθεί ότι η ανάλυση πραγματοποιήθηκε σε επίπεδο κατηγορίας προϊόντων και όχι σε επίπεδο μεμονωμένων προϊόντων (SKU). Ως αποτέλεσμα, ενδέχεται να αποκρύπτονται πιο εξειδικευμένα πρότυπα αγορών που εμφανίζονται μεταξύ συγκεκριμένων προϊόντων. Επιπλέον, τα διαθέσιμα δεδομένα δεν περιλάμβαναν πληροφορία για την ποσότητα ή το πλήθος των τεμαχίων ανά κατηγορία και συναλλαγή. Για τον λόγο αυτό, κάθε κατηγορία προσμετρήθηκε μία φορά ανά συναλλαγή και το μέγεθος του καλαθιού ορίστηκε ως το πλήθος των διαφορετικών κατηγοριών προϊόντων και όχι ως το πλήθος των τεμαχίων. Συνεπώς, η ανάλυση αποτυπώνει την παρουσία των κατηγοριών στα καλάθια, αλλά δεν επιτρέπει τη διερεύνηση μεταβολών στις αγοραζόμενες ποσότητες.

Επιπλέον, δεδομένου ότι η ανάλυση βασίστηκε σε πραγματικά δεδομένα συναλλαγών προερχόμενα από λειτουργικό πληροφοριακό σύστημα λιανεμπορίου, δεν μπορεί να αποκλειστεί η ύπαρξη μεμονωμένων διαφοροποιήσεων στην ονοματολογία ορισμένων κατηγοριών προϊόντων. Παρότι κατά το στάδιο της προεπεξεργασίας πραγματοποιήθηκε καθαρισμός και έλεγχος των δεδομένων, είναι πιθανό να παρέμειναν επιμέρους περιπτώσεις διαφορετικής λεκτικής αναπαράστασης της ίδιας κατηγορίας. Οι περιπτώσεις αυτές ενδέχεται να επηρεάζουν τοπικά τη συχνότητα εμφάνισης συγκεκριμένων κατηγοριών και, κατ' επέκταση, τα αντίστοιχα συχνά στοιχειοσύνολα και τους κανόνες συσχέτισης. Σε μελλοντική επέκταση της μελέτης θα μπορούσε να εφαρμοστεί αυστηρότερη διαδικασία κανονικοποίησης της ονοματολογίας των κατηγοριών πριν από την εφαρμογή των αλγορίθμων εξόρυξης.

Ως προς τη συγκριτική αξιολόγηση της υπολογιστικής αποδοτικότητας, οι τιμές μνήμης προέκυψαν από καταγραφή της χρήσης RSS μέσω της βιβλιοθήκης `psutil` σε προκαθορισμένα στάδια της εκτέλεσης. Συνεπώς, εκφράζουν τη μέγιστη καταγεγραμμένη τιμή στα συγκεκριμένα σημεία μέτρησης και όχι κατ' ανάγκη το απόλυτο μέγιστο κατά τη διάρκεια ολόκληρης της εκτέλεσης. Επιπλέον, κάθε πείραμα εκτελέστηκε μία φορά και, επομένως, δεν αξιολογήθηκε η πιθανή μεταβλητότητα του χρόνου εκτέλεσης και της χρήσης μνήμης μεταξύ επαναλαμβανόμενων εκτελέσεων. Τα αποτελέσματα της υπολογιστικής σύγκρισης αφορούν τις συγκεκριμένες υλοποιήσεις, το εξεταζόμενο σύνολο δεδομένων και το συγκεκριμένο υπολογιστικό περιβάλλον και δεν πρέπει να ερμηνεύονται

ως απόλυτη εγγενής υπεροχή ενός αλγορίθμου σε κάθε πιθανό σύνολο δεδομένων.

Τέλος, η μελέτη επικεντρώθηκε στη συνολική σύγκριση δύο ευρύτερων χρονικών περιόδων και δεν εξέτασε πιθανές εποχικές μεταβολές ή μεταβολές που σχετίζονται με συγκεκριμένα κύματα της πανδημίας και επιμέρους χρονικά διαστήματα.

5.4 Μελλοντική Έρευνα

Η παρούσα εργασία δημιουργεί ένα ισχυρό υπόβαθρο για περαιτέρω έρευνα στον τομέα της ανάλυσης καταναλωτικής συμπεριφοράς μέσω τεχνικών εξόρυξης γνώσης.

Μία πρώτη κατεύθυνση αφορά την εφαρμογή της ανάλυσης σε επίπεδο μεμονωμένων προϊόντων (SKU), γεγονός που θα μπορούσε να αποκαλύπτει λεπτομερέστερα πρότυπα αγορών και πιο εξειδικευμένες σχέσεις μεταξύ προϊόντων.

Επιπλέον, ιδιαίτερο ενδιαφέρον παρουσιάζει η χρονική ανάλυση της καταναλωτικής συμπεριφοράς ανά έτος, μήνα ή κύμα πανδημίας, ώστε να καταγραφούν με μεγαλύτερη ακρίβεια οι δυναμικές μεταβολές που σημειώθηκαν κατά τη διάρκεια της εξεταζόμενης περιόδου.

Μια ακόμη προοπτική αποτελεί η εφαρμογή πιο προηγμένων τεχνικών εξόρυξης γνώσης, όπως το Sequential Pattern Mining, το οποίο επιτρέπει τη μελέτη της χρονικής ακολουθίας των αγορών, καθώς και το Utility Mining, το οποίο λαμβάνει υπόψη την οικονομική αξία των προϊόντων και όχι μόνο τη συχνότητα εμφάνισής τους.

Επιπροσθέτως, η επέκταση της μελέτης σε δεδομένα από περισσότερες γεωγραφικές περιοχές ή από διαφορετικές αλυσίδες λιανικής θα μπορούσε να προσφέρει μια πιο ολοκληρωμένη εικόνα της καταναλωτικής συμπεριφοράς σε εθνικό επίπεδο.

Τέλος, ιδιαίτερο ενδιαφέρον παρουσιάζει η ανάπτυξη συστημάτων εξόρυξης κανόνων συσχέτισης σε πραγματικό χρόνο, τα οποία θα επιτρέπουν την άμεση αξιοποίηση των συναλλακτικών δεδομένων για την υποστήριξη επιχειρησιακών αποφάσεων και την προσαρμογή των στρατηγικών μάρκετινγκ στις τρέχουσες ανάγκες των καταναλωτών.

Βιβλιογραφία

- [1] S. Rana and M. Mondal, “A Seasonal and Multilevel Association Based Approach for Market Basket Analysis in Retail Supermarket,” *European Journal of Information Technologies and Computer Science*, vol. 1, no. 4, pp. 9–15, Oct. 2021.
- [2] M. Alawadh and A. Barnawi, “A Survey on Methods and Applications of Intelligent Market Basket Analysis Based on Association Rule,” *Journal on Big Data*, vol. 4, pp. 1–25, Oct. 2022.
- [3] Y. Ünvan, “Market basket analysis with association rules,” *Communications in Statistics—Theory and Methods*, vol. 50, no. 7, pp. 1615–1628, 2021.
- [4] M. Martinez, B. Escobar, M. García Díaz, and D. Pinto-Roa, “Market basket analysis with association rules in the retail sector using Orange. Case Study: Appliances Sales Company,” *CLEI Electronic Journal*, vol. 24, no. 2, Aug. 2021, art. no. 12.
- [5] J. Sheth, “Impact of Covid-19 on consumer behavior: Will the old habits return or die?” *Journal of Business Research*, vol. 117, pp. 280–283, Sep. 2020.
- [6] S. Billore and T. Anisimova, “Panic buying research: A systematic literature review and future research agenda,” *International Journal of Consumer Studies*, vol. 45, no. 4, pp. 777–804, 2021.
- [7] R. Jazemi, S. Farahani, W. Otieno, and J. Jang, “Review on Panic Buying Behavior during Pandemics: Influencing Factors, Stockpiling, and Intervention Strategies,” *Behavioral Sciences*, vol. 14, no. 3, Mar. 2024, art. no. 222.
- [8] G. Chua, K. F. Yuen, X. Wang, and Y. D. Wong, “The Determinants of Panic Buying during COVID-19,” *International Journal of Environmental Research and Public Health*, vol. 18, no. 6, p. 3247, Mar. 2021.
- [9] N. A. Omar, M. A. Nazri, M. H. Ali, and S. S. Alam, “The panic buying behavior of consumers during the COVID-19 pandemic: Examining the influences of uncertainty, perceptions of severity, perceptions of scarcity, and anxiety,” *Journal of Retailing and Consumer Services*, vol. 62, p. 102600, Sep. 2021.
- [10] P. I. N. Fernando, K. J. T. Perera, U. D. I. C. Udawaththa, R. M. C. S. Ratnayake, and D. G. M. P. Madushani, “Consumer stockpiling behaviour within the face of covid-19 pandemic: Systematic review,” *Sri Lanka Journal of Management Studies*, vol. 3, no. 1, pp. 66–86, 2021.
- [11] A. Schumacher and L. Micheli, “Anticipated scarcity and stockpiling during the COVID-19 pandemic: The role of perceived threat, childhood SES and materialism,” *PLOS ONE*, vol. 19, no. 3, p. e0294497, 2024.

- [12] J. Żurek and M. Rudy, “Impact of the COVID-19 Pandemic on Changes in Consumer Purchasing Behavior in the Food Market with a Focus on Meat and Meat Products—A Comprehensive Literature Review,” *Foods*, vol. 13, no. 6, p. 933, Mar. 2024.
- [13] O. Tyrväinen and H. Karjaluo, “Online grocery shopping before and during the COVID-19 pandemic: A meta-analytical review,” *Telematics and Informatics*, vol. 71, p. 101839, Jul. 2022.
- [14] V. Galushko and A. Riabchyk, “The demand for online grocery shopping: COVID-induced changes in grocery shopping behavior of Canadian consumers,” *PLOS ONE*, vol. 19, no. 2, p. e0295538, Feb. 2024.
- [15] J. M. Diaz-Gutierrez, H. Mohammadi-Mavi, and A. Ranjbari, “COVID-19 Impacts on Online and In-Store Shopping Behaviors: Why they Happened and Whether they Will Last Post Pandemic,” *Transportation Research Record*, vol. 2678, no. 12, pp. 580–598, Dec. 2024.
- [16] R. Agrawal, T. Imieliński, and A. Swami, “Mining association rules between sets of items in large databases,” *ACM SIGMOD Record*, vol. 22, no. 2, pp. 207–216, Jun. 1993.
- [17] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed., ser. Morgan Kaufmann Series in Data Management Systems. Amsterdam Boston: Elsevier/Morgan Kaufmann, 2012.
- [18] R. Agrawal and R. Srikant, “Fast Algorithms for Mining Association Rules in Large Databases,” in *Proceedings of the 20th International Conference on Very Large Data Bases*, ser. VLDB ’94. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., Jun. 1994, pp. 487–499.
- [19] K. Malliaridis, S. Ougiaroglou, and D. A. Dervos, “WebApriori: A web application for association rules mining,” in *Intelligent Tutoring Systems*, ser. Lecture Notes in Computer Science, vol. 12149. Cham: Springer, 2020, pp. 371–377.
- [20] K. Malliaridis and S. Ougiaroglou, “Efficient techniques for retrieving top-K frequent itemsets,” *Expert Systems with Applications*, vol. 311, p. 131250, May 2026.
- [21] J. Han, J. Pei, and Y. Yin, “Mining frequent patterns without candidate generation,” *ACM SIGMOD Record*, vol. 29, no. 2, pp. 1–12, Jun. 2000.
- [22] M. J. Zaki, “Scalable algorithms for association mining,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 12, no. 3, pp. 372–390, May/Jun. 2000.
- [23] M. Kholod, A. Celani, and G. Ciaramella, “The Analysis of Customers’ Transactions Based on POS and RFID Data Using Big Data Analytics Tools in the Retail Space of the Future,” *Applied Sciences*, vol. 14, p. 11567, Dec. 2024.
- [24] B. B. Gupta, A. Gaurav, and P. K. Panigrahi, “Analysis of retail sector research evolution and trends during covid-19,” *Technological Forecasting and Social Change*, vol. 194, p. 122671, Sep. 2023.
- [25] L. Szász, C. Bálint, O. Csíki, B. Z. Nagy, B. G. Rácz, D. Csala, and L. C. Harris, “The impact of covid-19 on the evolution of online retail: The pandemic as a window of opportunity,” *Journal of Retailing and Consumer Services*, vol. 69, p. 103089, Nov. 2022.

Παράρτημα Α΄

Κώδικες Υλοποίησης

Α΄.1 Μετατροπή του συνόλου δεδομένων σε μορφή καλαθιού

Ο ακόλουθος κώδικας χρησιμοποιήθηκε για τη μετατροπή των αρχικών δεδομένων συναλλαγών σε μορφή καλαθιού. Οι εγγραφές ομαδοποιήθηκαν με βάση το αναγνωριστικό της συναλλαγής και οι διαφορετικές κατηγορίες προϊόντων κάθε συναλλαγής αποθηκεύτηκαν σε μία γραμμή, διαχωρισμένες με ελληνικό ερωτηματικό (;).

```
1 import numpy as np
2 import pandas as pd
3 import os
4
5 # Declaration Section
6 filename = 'AfterCovid.csv'
7 groupColumn = 'TransId'
8 itemColumn = 'Description'
9 separator = ";"
10 filepath = filename
11
12 # Import to a DataFrame
13 dataset = pd.read_csv(filepath, sep=separator, encoding='utf-8')
14
15 # Isolate only the groupColumn and itemColumn
16 dataset = dataset[[groupColumn, itemColumn]]
17
18 # Sort dataset by groupColumn
19 datasetSorted = dataset.sort_values(by=groupColumn)
20
21 TempInv = ''
22 records = []
23 setRec = set()
24
25 for index, row in datasetSorted.iterrows():
26     if TempInv != row[groupColumn]:
27         if len(setRec) >= 1:
28             records.append(sorted(setRec))
29             setRec = set()
30             setRec.add(str(row[itemColumn]).strip())
31             TempInv = row[groupColumn]
32         else:
33             setRec.add(str(row[itemColumn]).strip())
34
35 if len(setRec) >= 1:
```

```

36     records.append(sorted(setRec))
37
38 os.makedirs('public', exist_ok=True)
39
40 output_path = os.path.join('public', '1' + filename[1:])
41
42 with open(output_path, 'w', encoding='utf-8-sig') as f:
43     for sublist in records:
44         line = ';'.join(sublist)
45         f.write(line + '\n')
46
47 print("Αρχείο:")
48 print(os.path.abspath(output_path))
49 print("Υπάρχει;", os.path.exists(output_path))
50 print("Πλήθος εγγραφών:", len(records))

```

Α.2 Ανάλυση ευαισθησίας της παραμέτρου minimum support

Ο ακόλουθος κώδικας χρησιμοποιήθηκε για τη διερεύνηση της κατάλληλης τιμής της παραμέτρου minimum support. Ο αλγόριθμος FP-Growth εφαρμόστηκε επαναληπτικά για διαφορετικές τιμές της παραμέτρου και, για κάθε εκτέλεση, καταγράφηκαν το πλήθος των frequent itemsets και ο αντίστοιχος χρόνος εκτέλεσης.

```

1
2 import os
3 import time
4 import pandas as pd
5 from mlxtend.preprocessing import TransactionEncoder
6 from mlxtend.frequent_patterns import fpgrowth
7
8 # =====
9 # CONFIGURATION
10 # =====
11
12 #basket_file = os.path.join('public', '1Covid.csv')
13 #basket_file = "/Users/nikievangeliidou/python-projects/public/1Covid.csv"
14
15 basket_file = "/home/stoug/evangelidou/public/1Covid.csv"
16
17
18
19 separator = ';'
20
21 # Εύρος τιμών για διερεύνηση του min_support
22 MIN_SUPPORT_RANGE = [
23     0.001,
24     0.002,
25     0.005,
26     0.01,
27     0.015,
28     0.02,
29     0.05
30 ]
31
32 OUTPUT_FILE = "support_sensitivity_fp_growth_Covid.csv"
33
34 # =====

```

```

35 # LOAD DATA (BASKET FORMAT)
36 # =====
37
38 transactions = []
39
40 with open(basket_file, "r", encoding="utf-8") as f:
41     for line in f:
42         line = line.strip()
43
44         # αγνόηση κενών γραμμών
45         if not line:
46             continue
47
48         # κάθε γραμμή = 1 transaction
49         items = [x.strip() for x in line.split(separator) if x.strip()]
50
51         if len(items) > 0:
52             transactions.append(items)
53
54 # Μετατροπή σε one-hot encoded matrix
55 te = TransactionEncoder()
56 df = pd.DataFrame(
57     te.fit(transactions).transform(transactions),
58     columns=te.columns_
59 )
60
61 # =====
62 # EXPERIMENT LOOP
63 # =====
64
65 results = []
66
67 for min_sup in MIN_SUPPORT_RANGE:
68
69     print(f"Running FP-Growth with min_support = {min_sup}")
70
71     start_time = time.time()
72
73     # Frequent itemsets extraction
74     itemsets = fpgrowth(
75         df,
76         min_support=min_sup,
77         use_colnames=True
78     )
79
80     runtime = time.time() - start_time
81
82     # Καταγραφή αποτελεσμάτων
83     results.append({
84         "min_support": min_sup,
85         "num_itemsets": len(itemsets),
86         "runtime_seconds": runtime
87     })
88
89 # =====
90 # SAVE RESULTS
91 # =====
92
93 results_df = pd.DataFrame(results)
94 results_df.to_csv(OUTPUT_FILE, index=False)
95
96 print("\nPhase 1 completed successfully.")

```

```
97 print("Results saved to:", OUTPUT_FILE)
```

Α.3 Υλοποίηση του αλγορίθμου Apriori

Ο ακόλουθος κώδικας αποτελεί την τελική υλοποίηση του αλγορίθμου Apriori στο ενοποιημένο σύνολο δεδομένων. Περιλαμβάνει τη φόρτωση και κωδικοποίηση των συναλλαγών, την εξαγωγή των frequent itemsets, τη δημιουργία και το φιλτράρισμα των κανόνων συσχέτισης, καθώς και την καταγραφή του χρόνου εκτέλεσης και της μέγιστης κατανάλωσης μνήμης.

```
1
2 import pandas as pd
3 import time
4 import psutil
5 import os
6 from mlxtend.preprocessing import TransactionEncoder
7 from mlxtend.frequent_patterns import apriori, association_rules
8
9 # =====
10 # FINAL APRIORI EXPERIMENT - MERGED DATASET
11 # =====
12
13 # -----
14 # PARAMETERS
15 # -----
16 basket_file = "/home/stoug/evangelidou/public/mergedcsv.csv"
17 separator = ";"
18
19 MIN_SUPPORT = 0.01
20 MIN_CONFIDENCE = 0.30
21 MIN_LIFT = 1.30
22 MIN_LEVERAGE = 0.005
23 MIN_CONVICTON = 1.0
24
25 # Output files
26 metrics_file = "FinalMetricsApriori.csv"
27 itemsets_file = "FinalApriori_FrequentItemsets.csv"
28 rules_file = "FinalApriori_AssociationRules.csv"
29
30 # -----
31 # MEMORY TRACKING
32 # -----
33 process = psutil.Process(os.getpid())
34 peak_memory = process.memory_info().rss / 1024 / 1024
35
36 def memory_mb():
37     return process.memory_info().rss / 1024 / 1024
38
39 def update_peak():
40     global peak_memory
41     peak_memory = max(peak_memory, memory_mb())
42
43 # -----
44 # START
45 # -----
46 start_time = time.time()
47
48 print("Loading transactions...")
```

```

49
50 transactions = []
51
52 with open(basket_file, "r", encoding="utf-8-sig") as f:
53     for line in f:
54         items = [item.strip() for item in line.strip().split(separator) if item.strip()]
55
56         # Αφαίρεση διπλότυπων προϊόντων μέσα στην ίδια συναλλαγή
57         items = list(set(items))
58
59         if len(items) > 0:
60             transactions.append(items)
61
62 update_peak()
63
64 n_transactions = len(transactions)
65 unique_items = sorted(set(item for transaction in transactions for item in transaction))
66 n_unique_items = len(unique_items)
67
68 print(f"Transactions: {n_transactions}")
69 print(f"Unique items: {n_unique_items}")
70 print(f"Memory after loading: {memory_mb():.2f} MB")
71
72 # -----
73 # TRANSACTION ENCODING
74 # -----
75 print("Encoding transactions...")
76
77 te = TransactionEncoder()
78 te_array = te.fit(transactions).transform(transactions)
79
80 df = pd.DataFrame(te_array, columns=te.columns_)
81
82 update_peak()
83
84 print("Encoding completed.")
85 print(f"Memory after encoding: {memory_mb():.2f} MB")
86
87 # -----
88 # APRIORI
89 # -----
90 print("Running Apriori...")
91
92 frequent_itemsets = apriori(
93     df,
94     min_support=MIN_SUPPORT,
95     use_colnames=True,
96     low_memory=True
97 )
98
99 update_peak()
100
101 print(f"Frequent Itemsets: {len(frequent_itemsets)}")
102
103 # -----
104 # ASSOCIATION RULES
105 # -----
106 print("Generating association rules...")
107
108 rules = association_rules(
109     frequent_itemsets,
110     metric="confidence",
111     min_threshold=MIN_CONFIDENCE

```

```

112 )
113
114 rules = rules[
115     (rules["lift"] >= MIN_LIFT) &
116     (rules["leverage"] >= MIN_LEVERAGE) &
117     (rules["conviction"] >= MIN_CONVICTON)
118 ]
119
120 # Κρατάμε μόνο κανόνες 1 → 1
121 rules = rules[
122     (rules["antecedents"].apply(lambda x: len(x)) == 1) &
123     (rules["consequents"].apply(lambda x: len(x)) == 1)
124 ]
125
126 rules = rules.sort_values(
127     by=["leverage", "lift", "confidence"],
128     ascending=False
129 )
130
131 update_peak()
132
133 print(f"Association Rules: {len(rules)}")
134
135 # -----
136 # SAVE DETAILED RESULTS
137 # -----
138 frequent_itemsets.to_csv(itemsets_file, index=False, encoding="utf-8-sig")
139 rules.to_csv(rules_file, index=False, encoding="utf-8-sig")
140
141 # -----
142 # FINAL METRICS
143 # -----
144 runtime = time.time() - start_time
145
146 metrics = pd.DataFrame({
147     "Metric": [
148         "Transactions",
149         "Unique Items",
150         "Frequent Itemsets",
151         "Association Rules",
152         "Runtime (sec)",
153         "Peak Memory (MB)"
154     ],
155     "Value": [
156         n_transactions,
157         n_unique_items,
158         len(frequent_itemsets),
159         len(rules),
160         round(runtime, 2),
161         round(peak_memory, 2)
162     ]
163 })
164
165 metrics.to_csv(metrics_file, index=False, encoding="utf-8-sig")
166
167 # -----
168 # PRINT SUMMARY
169 # -----
170 print("\n===== FINAL APRIORI METRICS =====")
171 print(metrics)
172 print("=====")
173
174 print("\nSaved files:")

```

```

175 print(f"- {metrics_file}")
176 print(f"- {itemsets_file}")
177 print(f"- {rules_file}")
178

```

Α'.4 Υλοποίηση του αλγορίθμου FP-Growth

Ο ακόλουθος κώδικας αποτελεί την τελική υλοποίηση του αλγορίθμου FP-Growth στο ενοποιημένο σύνολο δεδομένων. Για να εξασφαλιστεί η συγκρισιμότητα των αποτελεσμάτων, χρησιμοποιήθηκαν οι ίδιες παράμετροι, η ίδια διαδικασία παραγωγής κανόνων και οι ίδιες μετρικές υπολογιστικής αποδοτικότητας με την υλοποίηση του Apriori.

```

1
2
3 import pandas as pd
4 import time
5 import psutil
6 import os
7 from mlxtend.preprocessing import TransactionEncoder
8 from mlxtend.frequent_patterns import fpgrowth, association_rules
9
10 # =====
11 # FINAL FP-GROWTH EXPERIMENT - MERGED DATASET
12 # =====
13
14 # -----
15 # PARAMETERS
16 # -----
17 basket_file = "/home/stoug/evangelidou/public/mergedcsv.csv"
18 separator = ";"
19
20 MIN_SUPPORT = 0.01
21 MIN_CONFIDENCE = 0.30
22 MIN_LIFT = 1.30
23 MIN_LEVERAGE = 0.005
24 MIN_CONVICTON = 1.0
25
26 # Output files
27 metrics_file = "FinalMetricsFPGrowth.csv"
28 itemsets_file = "FinalFPGrowth_FrequentItemsets.csv"
29 rules_file = "FinalFPGrowth_AssociationRules.csv"
30
31 # -----
32 # MEMORY TRACKING
33 # -----
34 process = psutil.Process(os.getpid())
35 peak_memory = process.memory_info().rss / 1024 / 1024
36
37 def memory_mb():
38     return process.memory_info().rss / 1024 / 1024
39
40 def update_peak():
41     global peak_memory
42     peak_memory = max(peak_memory, memory_mb())
43
44 # -----
45 # START

```

```

46 # -----
47 start_time = time.time()
48
49 print("Loading transactions...")
50
51 transactions = []
52
53 with open(basket_file, "r", encoding="utf-8-sig") as f:
54     for line in f:
55         items = [item.strip() for item in line.strip().split(separator) if item.strip()]
56
57         # Αφαίρεση διπλότυπων προϊόντων μέσα στην ίδια συναλλαγή
58         items = list(set(items))
59
60         if len(items) > 0:
61             transactions.append(items)
62
63 update_peak()
64
65 n_transactions = len(transactions)
66 unique_items = sorted(set(item for transaction in transactions for item in transaction))
67 n_unique_items = len(unique_items)
68
69 print(f"Transactions: {n_transactions}")
70 print(f"Unique items: {n_unique_items}")
71 print(f"Memory after loading: {memory_mb():.2f} MB")
72
73 # -----
74 # TRANSACTION ENCODING
75 # -----
76 print("Encoding transactions...")
77
78 te = TransactionEncoder()
79 te_array = te.fit(transactions).transform(transactions)
80
81 df = pd.DataFrame(te_array, columns=te.columns_)
82
83 update_peak()
84
85 print("Encoding completed.")
86 print(f"Memory after encoding: {memory_mb():.2f} MB")
87
88 # -----
89 # FP-GROWTH
90 # -----
91 print("Running FP-Growth...")
92
93 frequent_itemsets = fpgrowth(
94     df,
95     min_support=MIN_SUPPORT,
96     use_colnames=True
97 )
98
99 update_peak()
100
101 print(f"Frequent Itemsets: {len(frequent_itemsets)}")
102
103 # -----
104 # ASSOCIATION RULES
105 # -----
106 print("Generating association rules...")
107
108 rules = association_rules(

```

```

109     frequent_itemsets,
110     metric="confidence",
111     min_threshold=MIN_CONFIDENCE
112 )
113
114 rules = rules[
115     (rules["lift"] >= MIN_LIFT) &
116     (rules["leverage"] >= MIN_LEVERAGE) &
117     (rules["conviction"] >= MIN_CONVICTON)
118 ]
119
120 # Κρατάμε μόνο κανόνες 1 → 1
121 rules = rules[
122     (rules["antecedents"].apply(lambda x: len(x)) == 1) &
123     (rules["consequents"].apply(lambda x: len(x)) == 1)
124 ]
125
126 rules = rules.sort_values(
127     by=["leverage", "lift", "confidence"],
128     ascending=False
129 )
130
131 update_peak()
132
133 print(f"Association Rules: {len(rules)}")
134
135 # -----
136 # SAVE DETAILED RESULTS
137 # -----
138 frequent_itemsets.to_csv(itemsets_file, index=False, encoding="utf-8-sig")
139 rules.to_csv(rules_file, index=False, encoding="utf-8-sig")
140
141 # -----
142 # FINAL METRICS
143 # -----
144 runtime = time.time() - start_time
145
146 metrics = pd.DataFrame({
147     "Metric": [
148         "Transactions",
149         "Unique Items",
150         "Frequent Itemsets",
151         "Association Rules",
152         "Runtime (sec)",
153         "Peak Memory (MB)"
154     ],
155     "Value": [
156         n_transactions,
157         n_unique_items,
158         len(frequent_itemsets),
159         len(rules),
160         round(runtime, 2),
161         round(peak_memory, 2)
162     ]
163 })
164
165 metrics.to_csv(metrics_file, index=False, encoding="utf-8-sig")
166
167 # -----
168 # PRINT SUMMARY
169 # -----
170 print("\n===== FINAL FP-GROWTH METRICS =====")
171 print(metrics)

```

```

172 print("=====")
173
174 print("\nSaved files:")
175 print(f"- {metrics_file}")
176 print(f"- {itemsets_file}")
177 print(f"- {rules_file}")
178

```

A'.5 Υλοποίηση του αλγορίθμου Eclat

Ο ακόλουθος κώδικας αποτελεί την τελική υλοποίηση του αλγορίθμου Eclat στο ενοποιημένο σύνολο δεδομένων. Τα αποτελέσματα του αλγορίθμου μετατράπηκαν στην κατάλληλη μορφή για την παραγωγή κανόνων συσχέτισης με κοινά κατώφλια και για τη σύγκριση του χρόνου εκτέλεσης και της μέγιστης κατανάλωσης μνήμης με τους άλλους δύο αλγορίθμους.

```

1
2 import pandas as pd
3 import time
4 import psutil
5 import os
6 from fim import eclat
7 from mlxtend.frequent_patterns import association_rules
8
9 # =====
10 # FINAL ECLAT EXPERIMENT - MERGED DATASET
11 # =====
12
13 # -----
14 # PARAMETERS
15 # -----
16 basket_file = "/home/stoug/evangelidou/public/mergedcsv.csv"
17 separator = ";"
18
19 MIN_SUPPORT = 0.01
20 MIN_CONFIDENCE = 0.30
21 MIN_LIFT = 1.30
22 MIN_LEVERAGE = 0.005
23 MIN_CONVICTION = 1.0
24
25 # Output files
26 metrics_file = "FinalMetricsEclat.csv"
27 itemsets_file = "FinalEclat_FrequentItemsets.csv"
28 rules_file = "FinalEclat_AssociationRules.csv"
29
30 # -----
31 # MEMORY TRACKING
32 # -----
33 process = psutil.Process(os.getpid())
34 peak_memory = process.memory_info().rss / 1024 / 1024
35
36 def memory_mb():
37     return process.memory_info().rss / 1024 / 1024
38
39 def update_peak():
40     global peak_memory
41     peak_memory = max(peak_memory, memory_mb())
42

```

```

43 # -----
44 # START
45 # -----
46 start_time = time.time()
47
48 print("Loading transactions...")
49
50 transactions = []
51
52 with open(basket_file, "r", encoding="utf-8-sig") as f:
53     for line in f:
54         items = [item.strip() for item in line.strip().split(separator) if item.strip()]
55
56         # Αφαίρεση διπλότυπων προϊόντων μέσα στην ίδια συναλλαγή
57         items = list(set(items))
58
59         if len(items) > 0:
60             transactions.append(items)
61
62 update_peak()
63
64 n_transactions = len(transactions)
65 unique_items = sorted(set(item for transaction in transactions for item in transaction))
66 n_unique_items = len(unique_items)
67
68 print(f"Transactions: {n_transactions}")
69 print(f"Unique items: {n_unique_items}")
70 print(f"Memory after loading: {memory_mb():.2f} MB")
71
72 # -----
73 # ECLAT
74 # -----
75 print("Running ECLAT...")
76
77 # To fim.eclat δέχεται το supp ως ποσοστό επί τοις εκατό.
78 # Άρα για MIN_SUPPORT = 0.01 δίνουμε supp = 1.0
79 min_support_percent = MIN_SUPPORT * 100
80
81 eclat_results = eclat(
82     transactions,
83     supp=min_support_percent,
84     target="s",
85     report="s"
86 )
87
88 update_peak()
89
90 print(f"ECLAT raw itemsets: {len(eclat_results)}")
91
92 # -----
93 # CONVERT ECLAT RESULTS TO DATAFRAME
94 # -----
95
96 frequent_itemsets = pd.DataFrame(
97     [
98         {
99             "support": support,
100             "itemsets": frozenset(itemset)
101         }
102         for itemset, support in eclat_results
103     ]
104 )
105

```

```

106 frequent_itemsets = frequent_itemsets[["support", "itemsets"]]
107
108 # Ταξινόμηση για πιο καθαρό αρχείο εξόδου
109 frequent_itemsets = frequent_itemsets.sort_values(
110     by="support",
111     ascending=False
112 ).reset_index(drop=True)
113
114 update_peak()
115
116 print(f"Frequent Itemsets: {len(frequent_itemsets)}")
117
118 # -----
119 # ASSOCIATION RULES
120 # -----
121 print("Generating association rules...")
122
123 rules = association_rules(
124     frequent_itemsets,
125     metric="confidence",
126     min_threshold=MIN_CONFIDENCE
127 )
128
129 rules = rules[
130     (rules["lift"] >= MIN_LIFT) &
131     (rules["leverage"] >= MIN_LEVERAGE) &
132     (rules["conviction"] >= MIN_CONVICTON)
133 ]
134
135 # Κρατάμε μόνο κανόνες 1 → 1
136 rules = rules[
137     (rules["antecedents"].apply(lambda x: len(x)) == 1) &
138     (rules["consequents"].apply(lambda x: len(x)) == 1)
139 ]
140
141 rules = rules.sort_values(
142     by=["leverage", "lift", "confidence"],
143     ascending=False
144 ).reset_index(drop=True)
145
146 update_peak()
147
148 print(f"Association Rules: {len(rules)}")
149
150 # -----
151 # SAVE DETAILED RESULTS
152 # -----
153 frequent_itemsets.to_csv(itemsets_file, index=False, encoding="utf-8-sig")
154 rules.to_csv(rules_file, index=False, encoding="utf-8-sig")
155
156 # -----
157 # FINAL METRICS
158 # -----
159 runtime = time.time() - start_time
160
161 metrics = pd.DataFrame({
162     "Metric": [
163         "Transactions",
164         "Unique Items",
165         "Frequent Itemsets",
166         "Association Rules",
167         "Runtime (sec)",
168         "Peak Memory (MB)"

```

```

169     ],
170     "Value": [
171         n_transactions,
172         n_unique_items,
173         len(frequent_itemsets),
174         len(rules),
175         round(runtime, 2),
176         round(peak_memory, 2)
177     ]
178 })
179
180 metrics.to_csv(metrics_file, index=False, encoding="utf-8-sig")
181
182 # -----
183 # PRINT SUMMARY
184 # -----
185 print("\n===== FINAL ECLAT METRICS =====")
186 print(metrics)
187 print("=====")
188
189 print("\nSaved files:")
190 print(f"- {metrics_file}")
191 print(f"- {itemsets_file}")
192 print(f"- {rules_file}")
193

```

Α.6 Σύγκριση του support των 1-itemsets και του μεγέθους καλαθιού

Ο ακόλουθος κώδικας χρησιμοποιήθηκε για τη συμπληρωματική σύγκριση του support των 1-itemsets και του μεγέθους καλαθιού στις δύο χρονικές περιόδους. Κάθε κατηγορία προσμετρήθηκε μία φορά ανά συναλλαγή, ενώ ως μέγεθος καλαθιού ορίστηκε το πλήθος των διαφορετικών κατηγοριών προϊόντων που περιλαμβάνονταν σε αυτή.

Από την ανάλυση εξαιρέθηκαν οι μη καταναλωτικές κατηγορίες που αναφέρθηκαν στη μεθοδολογία. Για κάθε κοινή κατηγορία υπολογίστηκαν η διαφορά του support σε ποσοστιαίες μονάδες και η σχετική ποσοστιαία μεταβολή. Παράλληλα, υπολογίστηκαν ο μέσος και ο διάμεσος αριθμός διαφορετικών κατηγοριών ανά συναλλαγή.

```

1 import csv
2 import io
3 import zipfile
4 from collections import Counter
5 from pathlib import Path
6
7
8 # =====
9 # PYΘΜΙΣΕΙΣ
10 # =====
11
12
13 BASE_DIR = Path(__file__).resolve().parent
14
15 COVID_ZIP = BASE_DIR / "1Covid.csv.zip"
16 AFTER_COVID_ZIP = BASE_DIR / "1AfterCovid.csv.zip"
17
18 COVID_CSV = "1Covid.csv"
19 AFTER_COVID_CSV = "1AfterCovid.csv"

```

```

20
21 # Εδώ θα αποθηκευτούν τα αποτελέσματα.
22 OUTPUT_DIR = BASE_DIR / "results_support_basket"
23
24 # Η κατηγορία αυτή αφορά κυρίως σακούλες μεταφοράς
25 # και δεν χρησιμοποιείται στην ουσιαστική ανάλυση.
26 EXCLUDED_CATEGORIES = {
27     "ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ",
28     "ΑΠΑΡΑΔΕΚΤΑ ΕΜΠΟΡΕΥΜΑΤΑ"
29 }
30
31
32
33 # =====
34 # ΒΟΗΘΗΤΙΚΕΣ ΣΥΝΑΡΤΗΣΕΙΣ
35 # =====
36
37 def format_decimal(value, decimals=4):
38     """
39     Μετατρέπει έναν αριθμό σε κείμενο με ελληνικό
40     δεκαδικό κόμμα, ώστε να εμφανίζεται σωστά στα CSV.
41     """
42
43     if value is None:
44         return ""
45
46     return f"{value:.{decimals}f}".replace(".", ",")
47
48
49 def calculate_median_from_counts(
50     size_counts,
51     total_transactions
52 ):
53     """
54     Υπολογίζει τη διάμεσο του μεγέθους καλαθιού
55     χωρίς να αποθηκεύει όλα τα μεγέθη στη μνήμη.
56     """
57
58     if total_transactions == 0:
59         raise ValueError(
60             "Δεν βρέθηκαν συναλλαγές για τον "
61             "υπολογισμό της διαμέσου."
62         )
63
64     # Θέσεις των δύο κεντρικών παρατηρήσεων.
65     # Για μονό πλήθος συναλλαγών είναι η ίδια θέση.
66     middle_position_1 = (total_transactions + 1) // 2
67     middle_position_2 = (total_transactions + 2) // 2
68
69     cumulative_count = 0
70     middle_value_1 = None
71     middle_value_2 = None
72
73     for basket_size in sorted(size_counts):
74
75         cumulative_count += size_counts[basket_size]
76
77         if (
78             middle_value_1 is None
79             and cumulative_count >= middle_position_1
80         ):
81             middle_value_1 = basket_size
82

```

```

83         if cumulative_count >= middle_position_2:
84             middle_value_2 = basket_size
85             break
86
87     return (
88         middle_value_1 + middle_value_2
89     ) / 2
90
91
92 # =====
93 # ΑΝΑΛΥΣΗ ΜΙΑΣ ΧΡΟΝΙΚΗΣ ΠΕΡΙΟΔΟΥ
94 # =====
95
96 def analyze_period(
97     zip_path,
98     csv_filename,
99     period_name
100 ):
101
102     # Πλήθος συναλλαγών στις οποίες εμφανίζεται
103     # κάθε κατηγορία.
104     item_counts = Counter()
105
106     # Κατανομή του μεγέθους των καλαθιών.
107     # Παράδειγμα: {1: 1000, 2: 3000, 3: 5000}
108     basket_size_counts = Counter()
109
110     total_transactions = 0
111     total_categories_in_baskets = 0
112     empty_baskets_after_exclusion = 0
113
114     if not zip_path.exists():
115         raise FileNotFoundError(
116             f"Δεν βρέθηκε το αρχείο "
117             f"{zip_path.name} στον φάκελο:\n"
118             f"{BASE_DIR}"
119         )
120
121     with zipfile.ZipFile(zip_path, "r") as zip_file:
122
123         if csv_filename not in zip_file.namelist():
124             raise FileNotFoundError(
125                 f"Το {csv_filename} δεν βρέθηκε "
126                 f"μέσα στο {zip_path.name}."
127             )
128
129         with zip_file.open(
130             csv_filename,
131             "r"
132         ) as raw_file:
133
134             text_file = io.TextIOWrapper(
135                 raw_file,
136                 encoding="utf-8-sig",
137                 newline=""
138             )
139
140             reader = csv.reader(
141                 text_file,
142                 delimiter=";"
143             )
144
145             for row in reader:

```

```

146
147     items = [
148         item.strip()
149         for item in row
150         if item.strip()
151     ]
152
153     # Παραλείπουμε μόνο τις πραγματικά
154     # κενές γραμμές.
155     if not items:
156         continue
157
158     total_transactions += 1
159
160     # Το set διασφαλίζει ότι κάθε κατηγορία
161     # μετρά μόνο μία φορά ανά συναλλαγή.
162     unique_items = set(items)
163
164     # Αφαιρούμε την κατηγορία των σακουλών.
165     unique_items.difference_update(
166         EXCLUDED_CATEGORIES
167     )
168
169     # Μέγεθος καλαθιού =
170     # πλήθος διαφορετικών κατηγοριών.
171     basket_size = len(unique_items)
172
173     if basket_size == 0:
174         empty_baskets_after_exclusion += 1
175         total_transactions -= 1
176         continue
177
178     basket_size_counts[basket_size] += 1
179
180     total_categories_in_baskets += (
181         basket_size
182     )
183
184     # Κάθε κατηγορία αυξάνεται κατά μία
185     # για τη συγκεκριμένη συναλλαγή.
186     item_counts.update(unique_items)
187
188     # Εμφάνιση προόδου ανά 500.000
189     # συναλλαγές.
190     if (
191         total_transactions
192         % 500_000 == 0
193     ):
194         print(
195             f"{period_name}: "
196             f"διαβάστηκαν "
197             f"{total_transactions:,} "
198             f"συναλλαγές"
199         )
200
201     if total_transactions == 0:
202         raise ValueError(
203             f"Δεν βρέθηκαν συναλλαγές "
204             f"για την περίοδο {period_name}."
205         )
206
207     mean_basket_size = (
208         total_categories_in_baskets

```

```

209         / total_transactions
210     )
211
212     median_basket_size = (
213         calculate_median_from_counts(
214             basket_size_counts,
215             total_transactions
216         )
217     )
218
219     return {
220         "period": period_name,
221         "transactions": total_transactions,
222         "item_counts": item_counts,
223         "basket_size_counts": basket_size_counts,
224         "total_categories_in_baskets":
225             total_categories_in_baskets,
226         "mean_basket_size": mean_basket_size,
227         "median_basket_size": median_basket_size,
228         "empty_baskets_after_exclusion":
229             empty_baskets_after_exclusion
230     }
231
232
233 # =====
234 # ΣΥΓΚΡΙΣΗ SUPPORT ΤΩΝ 1-ITEMSETS
235 # =====
236
237 def build_comparison(
238     covid_results,
239     after_results
240 ):
241
242     covid_categories = set(
243         covid_results["item_counts"]
244     )
245
246     after_categories = set(
247         after_results["item_counts"]
248     )
249
250     # Περιλαμβάνει όλες τις κατηγορίες που εμφανίζονται
251     # σε τουλάχιστον μία από τις δύο περιόδους.
252     all_categories = (
253         covid_categories
254         | after_categories
255     )
256
257     comparison_rows = []
258
259     for category in all_categories:
260
261         covid_count = (
262             covid_results["item_counts"]
263             .get(category, 0)
264         )
265
266         after_count = (
267             after_results["item_counts"]
268             .get(category, 0)
269         )
270
271         covid_support = (

```

```

272         covid_count
273         / covid_results["transactions"]
274     )
275
276     after_support = (
277         after_count
278         / after_results["transactions"]
279     )
280
281     # Απόλυτη διαφορά αναλογιών.
282     support_difference = (
283         after_support - covid_support
284     )
285
286     # Μετατροπή της διαφοράς σε
287     # ποσοστιαίες μονάδες.
288     support_difference_pp = (
289         support_difference * 100
290     )
291
292     # Σχετική ποσοστιαία μεταβολή
293     # με βάση την περίοδο COVID.
294     #
295     # Δεν υπολογίζεται όταν το COVID support
296     # είναι μηδέν.
297     if covid_support > 0:
298
299         relative_change_pct = (
300             support_difference
301             / covid_support
302             * 100
303         )
304
305     else:
306         relative_change_pct = None
307
308     # Καταγραφή του αν η κατηγορία είναι κοινή
309     # ή εμφανίζεται μόνο στη μία περίοδο.
310     if (
311         covid_count > 0
312         and after_count > 0
313     ):
314         status = "Κοινή κατηγορία"
315
316     elif covid_count > 0:
317         status = "Μόνο COVID"
318
319     else:
320         status = "Μόνο Post-COVID"
321
322     comparison_rows.append({
323         "category": category,
324         "status": status,
325         "covid_count": covid_count,
326         "after_count": after_count,
327         "covid_support_pct":
328             covid_support * 100,
329         "after_support_pct":
330             after_support * 100,
331         "support_difference_pp":
332             support_difference_pp,
333         "relative_change_pct":
334             relative_change_pct

```

```

335     })
336
337     # Ταξινόμηση από τη μεγαλύτερη αύξηση
338     # προς τη μεγαλύτερη μείωση.
339     comparison_rows.sort(
340         key=lambda row:
341             row["support_difference_pp"],
342         reverse=True
343     )
344
345     return comparison_rows
346
347
348 # =====
349 # ΑΠΟΘΗΚΕΥΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΣΕ CSV
350 # =====
351
352 COMPARISON_HEADER = [
353     "Κατηγορία",
354     "Κατάσταση",
355     "Πλήθος COVID",
356     "Πλήθος Post-COVID",
357     "Support COVID (%)",
358     "Support Post-COVID (%)",
359     "Δ Support (ποσοστιαίες μονάδες)",
360     "Σχετική μεταβολή Support (%)"
361 ]
362
363
364 def write_comparison_csv(
365     output_path,
366     rows
367 ):
368
369     with output_path.open(
370         "w",
371         encoding="utf-8-sig",
372         newline=""
373     ) as output_file:
374
375         writer = csv.writer(
376             output_file,
377             delimiter=";"
378         )
379
380         writer.writerow(
381             COMPARISON_HEADER
382         )
383
384         for row in rows:
385
386             writer.writerow([
387                 row["category"],
388                 row["status"],
389                 row["covid_count"],
390                 row["after_count"],
391                 format_decimal(
392                     row["covid_support_pct"],
393                     4
394                 ),
395                 format_decimal(
396                     row["after_support_pct"],
397                     4

```

```

398         ),
399         format_decimal(
400             row["support_difference_pp"],
401             4
402         ),
403         format_decimal(
404             row["relative_change_pct"],
405             2
406         )
407     ])
408
409
410 def write_basket_summary(
411     output_path,
412     covid_results,
413     after_results
414 ):
415
416     mean_difference = (
417         after_results["mean_basket_size"]
418         - covid_results["mean_basket_size"]
419     )
420
421     mean_relative_change = (
422         mean_difference
423         / covid_results["mean_basket_size"]
424         * 100
425     )
426
427     with output_path.open(
428         "w",
429         encoding="utf-8-sig",
430         newline=""
431     ) as output_file:
432
433         writer = csv.writer(
434             output_file,
435             delimiter=";"
436         )
437
438         writer.writerow([
439             "Περίοδος",
440             "Συναλλαγές",
441             "Διαφορετικές κατηγορίες",
442             "Μέσο μέγεθος καλαθιού",
443             "Median μεγέθους καλαθιού",
444             "Κενά καλάθια μετά τις εξαιρέσεις"
445         ])
446
447     for results in (
448         covid_results,
449         after_results
450     ):
451
452         writer.writerow([
453             results["period"],
454             results["transactions"],
455             len(results["item_counts"]),
456             format_decimal(
457                 results["mean_basket_size"],
458                 4
459             ),
460             format_decimal(

```

```

461         results["median_basket_size"],
462         2
463     ),
464     results[
465         "empty_baskets_after_exclusion"
466     ]
467 ])
468
469 writer.writerow([])
470
471 writer.writerow([
472     "Διαφορά μέσου "
473     "(Post-COVID - COVID)",
474     format_decimal(
475         mean_difference,
476         4
477     )
478 ])
479
480 writer.writerow([
481     "Σχετική μεταβολή μέσου (%)",
482     format_decimal(
483         mean_relative_change,
484         2
485     )
486 ])
487
488
489 # =====
490 # ΕΜΦΑΝΙΣΗ TOP-10
491 # =====
492
493 def print_top_rows(
494     title,
495     rows
496 ):
497
498     print("\n" + "=" * 70)
499     print(title)
500     print("=" * 70)
501
502     for position, row in enumerate(
503         rows,
504         start=1
505     ):
506
507         print(
508             f"{position}. "
509             f"{row['category']}\n"
510
511             f"    COVID: "
512             f"{row['covid_support_pct']:.4f}% | "
513
514             f"Post-COVID: "
515             f"{row['after_support_pct']:.4f}% | "
516
517             f"Δ Support: "
518             f"{row['support_difference_pp']:+.4f} "
519             f"π.μ. | "
520
521             f"Σχετική μεταβολή: "
522             f"{row['relative_change_pct']:+.2f}%"
523         )

```

```

524
525
526 # =====
527 # ΚΥΡΙΑ ΕΚΤΕΛΕΣΗ
528 # =====
529
530 def main():
531
532     OUTPUT_DIR.mkdir(
533         exist_ok=True
534     )
535
536     covid_results = analyze_period(
537         COVID_ZIP,
538         COVID_CSV,
539         "COVID"
540     )
541
542     after_results = analyze_period(
543         AFTER_COVID_ZIP,
544         AFTER_COVID_CSV,
545         "Post-COVID"
546     )
547
548     comparison_rows = build_comparison(
549         covid_results,
550         after_results
551     )
552
553     # Για τα Top-10 και την κύρια σύγκριση
554     # κρατάμε μόνο κατηγορίες που εμφανίζονται
555     # και στις δύο περιόδους.
556     common_rows = [
557         row
558         for row in comparison_rows
559         if row["status"] == "Κοινή κατηγορία"
560     ]
561
562     # Οι κατηγορίες που εμφανίζονται μόνο
563     # στη μία περίοδο καταγράφονται χωριστά.
564     categories_only_one_period = [
565         row
566         for row in comparison_rows
567         if row["status"] != "Κοινή κατηγορία"
568     ]
569
570     top_increases = sorted(
571         [
572             row
573             for row in common_rows
574             if (
575                 row["support_difference_pp"]
576                 > 0
577             )
578         ],
579         key=lambda row:
580             row["support_difference_pp"],
581         reverse=True
582     )[:10]
583
584     top_decreases = sorted(
585         [
586             row

```

```

587         for row in common_rows
588         if (
589             row["support_difference_pp"]
590             < 0
591         )
592     ],
593     key=lambda row:
594         row["support_difference_pp"]
595 )[:10]
596
597 # Όλες οι κατηγορίες.
598 write_comparison_csv(
599     OUTPUT_DIR
600     / "comparison_1_itemsets_all.csv",
601     comparison_rows
602 )
603
604 # Μόνο οι κοινές κατηγορίες.
605 write_comparison_csv(
606     OUTPUT_DIR
607     / "comparison_1_itemsets_common.csv",
608     common_rows
609 )
610
611 # Top-10 αυξήσεις.
612 write_comparison_csv(
613     OUTPUT_DIR
614     / "top10_support_increases.csv",
615     top_increases
616 )
617
618 # Top-10 μειώσεις.
619 write_comparison_csv(
620     OUTPUT_DIR
621     / "top10_support_decreases.csv",
622     top_decreases
623 )
624
625 # Κατηγορίες που εμφανίζονται μόνο
626 # στη μία από τις δύο περιόδους.
627 write_comparison_csv(
628     OUTPUT_DIR
629     / "categories_only_one_period.csv",
630     categories_only_one_period
631 )
632
633 # Σύνοψη μεγέθους καλαθιού.
634 write_basket_summary(
635     OUTPUT_DIR
636     / "basket_size_summary.csv",
637     covid_results,
638     after_results
639 )
640
641 # =====
642 # ΕΜΦΑΝΙΣΗ ΒΑΣΙΚΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ
643 # =====
644
645 print("\n" + "=" * 70)
646 print("ΠΛΗΘΟΣ ΣΥΝΑΛΛΑΓΩΝ")
647 print("=" * 70)
648
649 print(

```

```

650         f"COVID:      "
651         f"{covid_results['transactions']:,}"
652     )
653
654     print(
655         f"Post-COVID: "
656         f"{after_results['transactions']:,}"
657     )
658
659     print("\n" + "=" * 70)
660     print("ΜΕΓΕΘΟΣ ΚΑΛΑΘΙΟΥ")
661     print("=" * 70)
662
663     print(
664         "COVID - Μέσο πλήθος "
665         "διαφορετικών κατηγοριών: "
666         f"{covid_results['mean_basket_size']:.4f}"
667     )
668
669     print(
670         "COVID - Median: "
671         f"{covid_results['median_basket_size']:.2f}"
672     )
673
674     print(
675         "Post-COVID - Μέσο πλήθος "
676         "διαφορετικών κατηγοριών: "
677         f"{after_results['mean_basket_size']:.4f}"
678     )
679
680     print(
681         "Post-COVID - Median: "
682         f"{after_results['median_basket_size']:.2f}"
683     )
684
685     mean_difference = (
686         after_results["mean_basket_size"]
687         - covid_results["mean_basket_size"]
688     )
689
690     mean_relative_change = (
691         mean_difference
692         / covid_results["mean_basket_size"]
693         * 100
694     )
695
696     print(
697         "Διαφορά μέσου "
698         "(Post-COVID - COVID): "
699         f"{mean_difference:+.4f} κατηγορίες"
700     )
701
702     print(
703         "Σχετική μεταβολή μέσου: "
704         f"{mean_relative_change:+.2f}%"
705     )
706
707     print_top_rows(
708         "TOP-10 ΜΕΓΑΛΥΤΕΡΕΣ ΑΥΞΗΣΕΙΣ SUPPORT "
709         "(ΜΟΝΟ ΚΟΙΝΕΣ ΚΑΤΗΓΟΡΙΕΣ)",
710         top_increases
711     )
712

```

```

713     print_top_rows(
714         "TOP-10 ΜΕΓΑΛΥΤΕΡΕΣ ΜΕΙΩΣΕΙΣ SUPPORT "
715         "(ΜΟΝΟ ΚΟΙΝΕΣ ΚΑΤΗΓΟΡΙΕΣ)",
716         top_decreases
717     )
718
719     print("\n" + "=" * 70)
720     print(
721         "ΚΑΤΗΓΟΡΙΕΣ ΠΟΥ ΕΜΦΑΝΙΖΟΝΤΑΙ "
722         "ΜΟΝΟ ΣΤΗ ΜΙΑ ΠΕΡΙΟΔΟ"
723     )
724     print("=" * 70)
725
726     print(
727         "Συνολικό πλήθος: "
728         f"{len(categories_only_one_period)}"
729     )
730
731     print(
732         "Καταγράφηκαν χωριστά στο αρχείο: "
733         "categories_only_one_period.csv"
734     )
735
736     print(
737         "\nΗ ανάλυση ολοκληρώθηκε επιτυχώς."
738     )
739
740     print(
741         "Τα αποτελέσματα αποθηκεύτηκαν "
742         f"στον φάκελο:\n{OUTPUT_DIR}"
743     )
744
745
746 if __name__ == "__main__":
747     main()

```

A'.7 Σύγκριση του support των κοινών frequent itemsets

Ο ακόλουθος κώδικας χρησιμοποιήθηκε για την αριθμητική σύγκριση του support των frequent itemsets που εμφανίστηκαν και στις δύο χρονικές περιόδους. Τα itemsets αντιστοιχίστηκαν ανεξάρτητα από τη σειρά των κατηγοριών τους και εφαρμόστηκαν οι ίδιες εξαιρέσεις μη καταναλωτικών κατηγοριών.

Για κάθε κοινό itemset υπολογίστηκαν η διαφορά του support σε ποσοστιαίες μονάδες και η σχετική ποσοστιαία μεταβολή. Για την κατάταξη των μεγαλύτερων αυξήσεων και μειώσεων χρησιμοποιήθηκαν τα patterns που περιείχαν τουλάχιστον δύο κατηγορίες προϊόντων.

```

1  #!/usr/bin/env python3
2  # -*- coding: utf-8 -*-
3
4  """Compare the support of common frequent itemsets in two periods.
5
6  The script is intended for the final Apriori P5 outputs of the COVID-19
7  and Post-COVID periods. It:
8
9  1. reads the itemsets CSV directly or from a ZIP archive;
10 2. matches itemsets independently of the order of their categories;

```

```

11 3. keeps only itemsets that occur in both periods;
12 4. removes itemsets containing operational/non-product categories;
13 5. calculates the support difference and relative percentage change;
14 6. creates Top-10 increases and decreases for patterns of length >= 2.
15
16 Top-10 rankings are based on the absolute support difference, expressed
17 in percentage points. Relative percentage change is also reported.
18 """
19
20 from __future__ import annotations
21
22 import argparse
23 import ast
24 import csv
25 import io
26 import zipfile
27 from pathlib import Path
28 from typing import Iterable, TextIO
29
30
31 DEFAULT_COVID_FILE = Path("P5_conf30_lift13_lev005.zip")
32 DEFAULT_POSTCOVID_FILE = Path("P5AC_conf30_lift13_lev005.zip")
33 DEFAULT_OUTPUT_DIR = Path("Support_Comparison_Results_FINAL")
34
35 EXCLUDED_CATEGORIES = {
36     "ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ",
37     "ΑΠΑΡΑΔΕΚΤΑ ΕΜΠΟΡΕΥΜΑΤΑ",
38 }
39
40
41 def parse_itemset(text: str) -> tuple[str, ...]:
42     """Convert a textual set/frozenset into a sorted tuple of categories."""
43
44     value = text.strip()
45     tree = ast.parse(value, mode="eval")
46     node = tree.body
47
48     if isinstance(node, ast.Call):
49         if not (
50             isinstance(node.func, ast.Name)
51             and node.func.id in {"frozenset", "set"}
52             and len(node.args) == 1
53         ):
54             raise ValueError(f"Unexpected itemset format: {text}")
55         node = node.args[0]
56
57     if not isinstance(node, (ast.Set, ast.List, ast.Tuple)):
58         raise ValueError(f"Unexpected itemset contents: {text}")
59
60     items: list[str] = []
61     for element in node.elts:
62         item = ast.literal_eval(element)
63         if not isinstance(item, str):
64             raise ValueError(f"Non-text value found in itemset: {text}")
65         cleaned_item = item.strip()
66         if cleaned_item:
67             items.append(cleaned_item)
68
69     if not items:
70         raise ValueError(f"Empty itemset: {text}")
71
72     return tuple(sorted(set(items)))
73

```

```

74
75 def select_itemsets_member(zip_file: zipfile.ZipFile) -> str:
76     """Select the itemsets CSV contained in a ZIP archive."""
77
78     candidates = [
79         name
80         for name in zip_file.namelist()
81         if not name.endswith("/")
82         and name.lower().endswith(".csv")
83         and "itemset" in Path(name).name.lower()
84     ]
85
86     if len(candidates) != 1:
87         raise ValueError(
88             "The ZIP must contain exactly one itemsets CSV. "
89             f"Candidates found: {candidates}"
90         )
91
92     return candidates[0]
93
94
95 def read_rows(text_file: TextIO, source_name: str) -> dict[tuple[str, ...], float]:
96     """Read support and itemsets columns from an open CSV text stream."""
97
98     reader = csv.DictReader(text_file, delimiter=";")
99     required_columns = {"support", "itemsets"}
100
101     if not required_columns.issubset(reader.fieldnames or []):
102         raise ValueError(
103             f"{source_name} must contain the columns 'support' and 'itemsets'. "
104             f"Columns found: {reader.fieldnames}"
105         )
106
107     itemsets: dict[tuple[str, ...], float] = {}
108
109     for line_number, row in enumerate(reader, start=2):
110         key = parse_itemset(row["itemsets"])
111         support = float(row["support"])
112
113         if key in itemsets and itemsets[key] != support:
114             raise ValueError(
115                 f"Duplicate itemset with different support in {source_name}, "
116                 f"line {line_number}: {key}"
117             )
118
119         itemsets[key] = support
120
121     return itemsets
122
123
124 def read_itemsets(path: Path) -> dict[tuple[str, ...], float]:
125     """Read an itemsets CSV supplied directly or inside a ZIP archive."""
126
127     if not path.exists():
128         raise FileNotFoundError(f"Input file not found: {path}")
129
130     if path.suffix.lower() == ".zip":
131         with zipfile.ZipFile(path, "r") as zip_file:
132             member = select_itemsets_member(zip_file)
133             with zip_file.open(member, "r") as binary_file:
134                 with io.TextIOWrapper(
135                     binary_file,
136                     encoding="utf-8-sig",

```

```

137         newline="",
138     ) as text_file:
139         return read_rows(text_file, f"{path.name}:{member}")
140
141     with path.open("r", encoding="utf-8-sig", newline="") as text_file:
142         return read_rows(text_file, path.name)
143
144
145 def format_itemset(items: tuple[str, ...]) -> str:
146     return "{" + ", ".join(items) + "}"
147
148
149 def build_row(
150     itemset: tuple[str, ...],
151     support_covid: float,
152     support_postcovid: float,
153 ) -> dict[str, object]:
154     difference = support_postcovid - support_covid
155     relative_change = difference / support_covid * 100
156
157     return {
158         "Itemset": format_itemset(itemset),
159         "Itemset_Length": len(itemset),
160         "Support_COVID_Percent": support_covid * 100,
161         "Support_PostCOVID_Percent": support_postcovid * 100,
162         "Delta_Support_Percentage_Points": difference * 100,
163         "Relative_Change_Percent": relative_change,
164     }
165
166
167 OUTPUT_COLUMNS = [
168     "Itemset",
169     "Itemset_Length",
170     "Support_COVID_Percent",
171     "Support_PostCOVID_Percent",
172     "Delta_Support_Percentage_Points",
173     "Relative_Change_Percent",
174 ]
175
176
177 def write_comparison_csv(path: Path, rows: Iterable[dict[str, object]]) -> None:
178     with path.open("w", encoding="utf-8-sig", newline="") as output_file:
179         writer = csv.DictWriter(
180             output_file,
181             fieldnames=OUTPUT_COLUMNS,
182             delimiter=";",
183         )
184         writer.writeheader()
185
186         for row in rows:
187             writer.writerow(
188                 {
189                     "Itemset": row["Itemset"],
190                     "Itemset_Length": row["Itemset_Length"],
191                     "Support_COVID_Percent": (
192                         f"{row['Support_COVID_Percent']:.8f}"
193                     ),
194                     "Support_PostCOVID_Percent": (
195                         f"{row['Support_PostCOVID_Percent']:.8f}"
196                     ),
197                     "Delta_Support_Percentage_Points": (
198                         f"{row['Delta_Support_Percentage_Points']:.8f}"
199                     ),

```

```

200         "Relative_Change_Percent": (
201             f"{row['Relative_Change_Percent']:.8f}"
202         ),
203     }
204 )
205
206
207 def write_summary_csv(path: Path, summary_rows: list[tuple[str, object]]) -> None:
208     with path.open("w", encoding="utf-8-sig", newline="") as output_file:
209         writer = csv.writer(output_file, delimiter=",")
210         writer.writerow(["Metric", "Value"])
211         writer.writerows(summary_rows)
212
213
214 def print_top_rows(title: str, rows: list[dict[str, object]]) -> None:
215     print("\n" + "=" * 90)
216     print(title)
217     print("=" * 90)
218
219     for position, row in enumerate(rows, start=1):
220         print(
221             f"{position}. {row['Itemset']}\n"
222             f"COVID: {row['Support_COVID_Percent']:.3f}% | "
223             f"Post-COVID: {row['Support_PostCOVID_Percent']:.3f}% | "
224             f"Delta: {row['Delta_Support_Percentage_Points']:+.3f} p.p. | "
225             f"Relative change: {row['Relative_Change_Percent']:+.2f}%"
226         )
227
228
229 def create_results_zip(output_dir: Path) -> Path:
230     zip_path = output_dir.with_suffix(".zip")
231
232     with zipfile.ZipFile(zip_path, "w", compression=zipfile.ZIP_DEFLATED) as zip_file:
233         for result_file in sorted(output_dir.iterdir()):
234             if result_file.is_file():
235                 zip_file.write(result_file, arcname=result_file.name)
236
237     return zip_path
238
239
240 def main() -> None:
241     parser = argparse.ArgumentParser(
242         description=(
243             "Compare support of common P5 frequent itemsets between "
244             "COVID and Post-COVID."
245         )
246     )
247     parser.add_argument(
248         "covid_file",
249         nargs="?",
250         type=Path,
251         default=DEFAULT_COVID_FILE,
252         help="COVID itemsets CSV or ZIP",
253     )
254     parser.add_argument(
255         "postcovid_file",
256         nargs="?",
257         type=Path,
258         default=DEFAULT_POSTCOVID_FILE,
259         help="Post-COVID itemsets CSV or ZIP",
260     )
261     parser.add_argument(
262         "--output-dir",

```

```

263     type=Path,
264     default=DEFAULT_OUTPUT_DIR,
265     help="Directory in which result files will be saved",
266 )
267 parser.add_argument("--top-n", type=int, default=10)
268 parser.add_argument("--min-pattern-length", type=int, default=2)
269 args = parser.parse_args()
270
271 covid_itemsets = read_itemsets(args.covid_file)
272 postcovid_itemsets = read_itemsets(args.postcovid_file)
273
274 common_before_exclusion = set(covid_itemsets) & set(postcovid_itemsets)
275
276 common_after_exclusion = {
277     itemset
278     for itemset in common_before_exclusion
279     if not any(category in EXCLUDED_CATEGORIES for category in itemset)
280 }
281
282 all_common_rows = [
283     build_row(
284         itemset,
285         covid_itemsets[itemset],
286         postcovid_itemsets[itemset],
287     )
288     for itemset in common_after_exclusion
289 ]
290 all_common_rows.sort(
291     key=lambda row: (row["Itemset_Length"], row["Itemset"])
292 )
293
294 pattern_rows = [
295     row
296     for row in all_common_rows
297     if row["Itemset_Length"] >= args.min_pattern_length
298 ]
299
300 increasing_rows = [
301     row
302     for row in pattern_rows
303     if row["Delta_Support_Percentage_Points"] > 0
304 ]
305 decreasing_rows = [
306     row
307     for row in pattern_rows
308     if row["Delta_Support_Percentage_Points"] < 0
309 ]
310 unchanged_rows = [
311     row
312     for row in pattern_rows
313     if row["Delta_Support_Percentage_Points"] == 0
314 ]
315
316 top_increases = sorted(
317     increasing_rows,
318     key=lambda row: (
319         -row["Delta_Support_Percentage_Points"],
320         row["Itemset"],
321     ),
322 )[: args.top_n]
323
324 top_decreases = sorted(
325     decreasing_rows,

```

```

326         key=lambda row: (
327             row["Delta_Support_Percentage_Points"],
328             row["Itemset"],
329         ),
330    )[: args.top_n]
331
332     args.output_dir.mkdir(parents=True, exist_ok=True)
333
334     write_comparison_csv(
335         args.output_dir / "Common_Itemsets_Support_Comparison.csv",
336         all_common_rows,
337     )
338     write_comparison_csv(
339         args.output_dir / "Common_Patterns_Length_2_or_More.csv",
340         pattern_rows,
341     )
342     write_comparison_csv(
343         args.output_dir / "Top10_Support_Increases.csv",
344         top_increases,
345     )
346     write_comparison_csv(
347         args.output_dir / "Top10_Support_Decreases.csv",
348         top_decreases,
349     )
350
351     summary_rows = [
352         ("COVID frequent itemsets", len(covid_itemsets)),
353         ("Post-COVID frequent itemsets", len(postcovid_itemsets)),
354         ("Common itemsets before exclusions", len(common_before_exclusion)),
355         ("Common itemsets after exclusions", len(all_common_rows)),
356         (
357             f"Common patterns with length >= {args.min_pattern_length}",
358             len(pattern_rows),
359         ),
360         ("Patterns with increased support", len(increasing_rows)),
361         ("Patterns with decreased support", len(decreasing_rows)),
362         ("Patterns with unchanged support", len(unchanged_rows)),
363         ("Top-N", args.top_n),
364         ("Excluded categories", " | ".join(sorted(EXCLUDED_CATEGORIES))),
365         (
366             "Top-N ranking criterion",
367             "Delta Support in percentage points",
368         ),
369     ]
370     write_summary_csv(args.output_dir / "Analysis_Summary.csv", summary_rows)
371
372     print("\nAnalysis completed successfully.")
373     print(f"COVID frequent itemsets: {len(covid_itemsets):,}")
374     print(f"Post-COVID frequent itemsets: {len(postcovid_itemsets):,}")
375     print(
376         "Common itemsets before exclusions: "
377         f"{len(common_before_exclusion):,}"
378     )
379     print(
380         "Common itemsets after exclusions: "
381         f"{len(all_common_rows):,}"
382     )
383     print(
384         f"Common patterns with length >= {args.min_pattern_length}: "
385         f"{len(pattern_rows):,}"
386     )
387     print(f"Patterns with increased support: {len(increasing_rows):,}")
388     print(f"Patterns with decreased support: {len(decreasing_rows):,}")

```

```

389     print(f"Patterns with unchanged support: {len(unchanged_rows):,}")
390
391     print_top_rows("TOP SUPPORT INCREASES", top_increases)
392     print_top_rows("TOP SUPPORT DECREASES", top_decreases)
393
394     zip_path = create_results_zip(args.output_dir)
395     print(f"\nResults directory: {args.output_dir.resolve()}")
396     print(f"Results ZIP: {zip_path.resolve()}")
397
398
399 if __name__ == "__main__":
400     main()

```

A'8 Σύγκριση του confidence και του lift των κοινών κανόνων

Ο ακόλουθος κώδικας χρησιμοποιήθηκε για τη σύγκριση των κοινών κατευθυνόμενων κανόνων συσχέτισης μεταξύ των δύο περιόδων. Ως κοινός θεωρήθηκε ένας κανόνας όταν είχε το ίδιο antecedent και το ίδιο consequent και στις δύο περιόδους, διατηρώντας την κατεύθυνσή του.

Για κάθε κοινό κανόνα υπολογίστηκαν οι μεταβολές του support, του confidence και του lift. Επιπλέον, δημιουργήθηκαν οι κατατάξεις με τις δέκα μεγαλύτερες αυξήσεις και μειώσεις του confidence και του lift.

```

1  #!/usr/bin/env python3
2  # -*- coding: utf-8 -*-
3
4  """Σύγκριση confidence και lift κοινών κανόνων COVID / Post-COVID.
5
6  Το πρόγραμμα:
7  1. διαβάζει τα CSV των association rules απευθείας από τα δύο ZIP,
8  2. αντιστοιχίζει μόνο κανόνες με ίδιο antecedent και ίδιο consequent,
9  3. διατηρεί την κατεύθυνση του κανόνα (A -> B != B -> A),
10 4. εξαιρεί κανόνες που περιέχουν λειτουργικές/μη προϊόντικές κατηγορίες,
115. υπολογίζει τις μεταβολές support, confidence και lift,
126. δημιουργεί Top-10 αυξήσεις και μειώσεις για confidence και lift.
13
14Η κατάταξη γίνεται:
15- για confidence: με βάση τη διαφορά σε ποσοστιαίες μονάδες,
16- για lift: με βάση την απόλυτη αριθμητική διαφορά του lift.
17
18Δεν απαιτούνται pandas, openpyxl ή άλλες εξωτερικές βιβλιοθήκες.
19"""
20
21from __future__ import annotations
22
23import argparse
24import ast
25import csv
26import io
27import zipfile
28from pathlib import Path
29from typing import Iterable
30
31
32# =====
33# PYΘΜΙΣΕΙΣ
34# =====
35

```

```

36 DEFAULT_COVID_FILE = Path("P5_conf30_lift13_lev005.zip")
37 DEFAULT_POSTCOVID_FILE = Path("P5AC_conf30_lift13_lev005.zip")
38 DEFAULT_OUTPUT_DIR = Path("Common_Rules_Metrics_Results_FINAL")
39
40 EXCLUDED_CATEGORIES = {
41     "ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ",
42     "ΑΠΑΡΑΔΕΚΤΑ ΕΜΠΟΡΕΥΜΑΤΑ",
43 }
44
45 TOP_N = 10
46
47
48 # =====
49 # ΒΟΗΘΗΤΙΚΕΣ ΣΥΝΑΡΤΗΣΕΙΣ
50 # =====
51
52 def normalize_header(value: str) -> str:
53     """Κανονικοποιεί ένα όνομα στήλης για ασφαλή αντιστοίχιση."""
54
55     return " ".join(
56         value.strip().lower().replace("_", " ").split()
57     )
58
59
60 def parse_number(value: str, field_name: str, row_number: int) -> float:
61     """Μετατρέπει αριθμητικό κείμενο σε float."""
62
63     text = str(value).strip()
64
65     if not text:
66         raise ValueError(
67             f"Κενή τιμή στη στήλη {field_name}, γραμμή {row_number}."
68         )
69
70     # Υποστήριξη τόσο δεκαδικής τελείας όσο και δεκαδικού κόμματος.
71     if "," in text and "." not in text:
72         text = text.replace(",", ".")
73
74     try:
75         return float(text)
76     except ValueError as exc:
77         raise ValueError(
78             f"Μη έγκυρη τιμή στη στήλη {field_name}, "
79             f"γραμμή {row_number}: {value}"
80         ) from exc
81
82
83 def parse_itemset(value: str) -> tuple[str, ...]:
84     """Μετατρέπει μία πλευρά κανόνα σε ταξινομημένο tuple.
85
86     Υποστηρίζονται τόσο δομημένες τιμές, όπως
87     ``frozenset({'ΓΑΛΛΑ', 'ΓΙΑΟΥΡΤΙΑ'})``, όσο και απλό κείμενο,
88     όπως ``ΓΑΛΛΑ ΦΡΕΣΚΟ``. Στη δεύτερη περίπτωση η τιμή θεωρείται
89     μία μοναδική κατηγορία.
90     """
91
92     text = str(value).strip()
93
94     if not text:
95         raise ValueError("Βρέθηκε κενή πλευρά κανόνα.")
96
97     structured_prefixes = ("frozenset(", "set(", "{", "[", "(")
98

```

```

99  # Στα συγκεκριμένα rules CSV τα antecedents/consequents μπορεί να
100  # αποθηκεύονται ως απλό κείμενο. Όταν υπάρχουν περισσότερες κατηγορίες,
101  # χωρίζονται με κόμμα (π.χ. "SNACK, ΜΠΙΣΚΟΤΑ"). Τις διαχωρίζουμε και
102  # τις ταξινομούμε ώστε οι μορφές "A, B" και "B, A" να ταυτίζονται.
103  if not text.startswith(structured_prefixes):
104      plain_items = {
105          item.strip()
106          for item in text.split(",")
107          if item.strip()
108      }
109
110      if not plain_items:
111          raise ValueError("Βρέθηκε κενή πλευρά κανόνα.")
112
113      return tuple(sorted(plain_items))
114
115  try:
116      tree = ast.parse(text, mode="eval")
117      node = tree.body
118
119      if isinstance(node, ast.Constant) and isinstance(node.value, str):
120          item = node.value.strip()
121
122          if not item:
123              raise ValueError
124
125          return (item,)
126
127      if isinstance(node, ast.Call):
128          if not (
129              isinstance(node.func, ast.Name)
130              and node.func.id in {"frozenset", "set"}
131              and len(node.args) == 1
132          ):
133              raise ValueError
134
135          node = node.args[0]
136
137      if not isinstance(node, (ast.Set, ast.List, ast.Tuple)):
138          raise ValueError
139
140      items: list[str] = []
141
142      for element in node.elts:
143          item = ast.literal_eval(element)
144
145          if not isinstance(item, str):
146              raise ValueError
147
148          item = item.strip()
149
150          if item:
151              items.append(item)
152
153      if not items:
154          raise ValueError
155
156      return tuple(sorted(set(items)))
157
158  except Exception as exc:
159      raise ValueError(
160          f"Δεν ήταν δυνατή η ανάγνωση του itemset: {text}"
161      ) from exc

```

```

162
163
164 def display_itemset(itemset: tuple[str, ...]) -> str:
165     """Επιστρέφει αναγνώσιμη μορφή ενός itemset."""
166
167     return "{" + ", ".join(itemset) + "}"
168
169
170 def display_rule(
171     antecedent: tuple[str, ...],
172     consequent: tuple[str, ...],
173 ) -> str:
174     """Επιστρέφει αναγνώσιμη μορφή ενός κατευθυνόμενου κανόνα."""
175
176     return f"{display_itemset(antecedent)} -> {display_itemset(consequent)}"
177
178
179 def find_rules_csv(zip_file: zipfile.ZipFile, zip_name: str) -> str:
180     """Εντοπίζει το CSV των association rules μέσα σε ένα ZIP."""
181
182     candidates = [
183         name
184         for name in zip_file.namelist()
185         if not name.endswith("/")
186         and name.lower().endswith(".csv")
187         and "rule" in Path(name).name.lower()
188         and "itemset" not in Path(name).name.lower()
189     ]
190
191     if not candidates:
192         raise FileNotFoundError(
193             f"Δεν βρέθηκε rules CSV μέσα στο {zip_name}.\n"
194             "Περιεχόμενα ZIP:\n"
195             + "\n".join(zip_file.namelist())
196         )
197
198     if len(candidates) == 1:
199         return candidates[0]
200
201     preferred = [
202         name
203         for name in candidates
204         if Path(name).name.lower().endswith("_rules.csv")
205     ]
206
207     if len(preferred) == 1:
208         return preferred[0]
209
210     raise ValueError(
211         f"Βρέθηκαν περισσότερα από ένα πιθανά rules CSV στο {zip_name}:\n"
212         + "\n".join(candidates)
213     )
214
215
216 def detect_delimiter(text: str) -> str:
217     """Εντοπίζει αν το CSV χρησιμοποιεί ; ή , ως διαχωριστικό."""
218
219     try:
220         dialect = csv.Sniffer().sniff(text[:8192], delimiters=";,")
221         return dialect.delimiter
222     except csv.Error:
223         return ";"
224

```

```

225
226 def find_column(
227     normalized_headers: dict[str, str],
228     possible_names: set[str],
229     source_name: str,
230 ) -> str:
231     """Επιστρέφει το πραγματικό όνομα μιας απαιτούμενης στήλης."""
232
233     for name in possible_names:
234         if name in normalized_headers:
235             return normalized_headers[name]
236
237     raise ValueError(
238         f"Δεν βρέθηκε μία από τις στήλες {sorted(possible_names)} "
239         f"στο {source_name}.\n"
240         f"Διαθέσιμες στήλες: {list(normalized_headers.values())}"
241     )
242
243
244 # =====
245 # ΑΝΑΓΝΩΣΗ ΚΑΝΟΝΩΝ
246 # =====
247
248 RuleKey = tuple[tuple[str, ...], tuple[str, ...]]
249
250
251 def read_rules_text(
252     text: str,
253     source_name: str,
254 ) -> dict[RuleKey, dict[str, float]]:
255     """Διαβάζει τους κανόνες από το κείμενο ενός CSV."""
256
257     delimiter = detect_delimiter(text)
258     reader = csv.DictReader(io.StringIO(text), delimiter=delimiter)
259
260     if reader.fieldnames is None:
261         raise ValueError(f"Δεν υπάρχουν headers στο {source_name}.")
262
263     normalized_headers = {
264         normalize_header(header): header
265         for header in reader.fieldnames
266         if header is not None
267     }
268
269     antecedent_col = find_column(
270         normalized_headers,
271         {"antecedent", "antecedents"},
272         source_name,
273     )
274     consequent_col = find_column(
275         normalized_headers,
276         {"consequent", "consequents"},
277         source_name,
278     )
279     support_col = find_column(
280         normalized_headers,
281         {"support"},
282         source_name,
283     )
284     confidence_col = find_column(
285         normalized_headers,
286         {"confidence"},
287         source_name,

```

```

288     )
289     lift_col = find_column(
290         normalized_headers,
291         {"lift"},
292         source_name,
293     )
294
295     rules: dict[RuleKey, dict[str, float]] = {}
296
297     for row_number, row in enumerate(reader, start=2):
298         antecedent_text = row.get(antecedent_col)
299         consequent_text = row.get(consequent_col)
300
301         if antecedent_text is None or consequent_text is None:
302             continue
303
304         if not antecedent_text.strip() or not consequent_text.strip():
305             continue
306
307         antecedent = parse_itemset(antecedent_text)
308         consequent = parse_itemset(consequent_text)
309         rule_key: RuleKey = (antecedent, consequent)
310
311         metrics = {
312             "support": parse_number(
313                 row[support_col], "support", row_number
314             ),
315             "confidence": parse_number(
316                 row[confidence_col], "confidence", row_number
317             ),
318             "lift": parse_number(
319                 row[lift_col], "lift", row_number
320             ),
321         }
322
323         if rule_key in rules:
324             previous = rules[rule_key]
325
326             if any(
327                 abs(previous[name] - metrics[name]) > 1e-12
328                 for name in metrics
329             ):
330                 raise ValueError(
331                     f"Ο κανόνας {display_rule(*rule_key)} εμφανίζεται "
332                     f"περισσότερες από μία φορές με διαφορετικές τιμές "
333                     f"στο {source_name}."
334                 )
335
336             rules[rule_key] = metrics
337
338     return rules
339
340
341 def read_rules(path: Path) -> dict[RuleKey, dict[str, float]]:
342     """Διαβάζει rules CSV είτε απευθείας είτε μέσα από ZIP."""
343
344     if not path.exists():
345         raise FileNotFoundError(
346             f"Δεν βρέθηκε το αρχείο εισόδου: {path.resolve()}"
347         )
348
349     if path.suffix.lower() == ".zip":
350         with zipfile.ZipFile(path, "r") as zip_file:

```

```

351         rules_member = find_rules_csv(zip_file, path.name)
352
353         print(f"\nZIP: {path.name}")
354         print(f"Rules file: {rules_member}")
355
356         raw_text = zip_file.read(rules_member).decode("utf-8-sig")
357         return read_rules_text(
358             raw_text,
359             f"{path.name}:{rules_member}",
360         )
361
362     raw_text = path.read_text(encoding="utf-8-sig")
363     return read_rules_text(raw_text, path.name)
364
365
366 # =====
367 # ΦΙΛΤΡΑΡΙΣΜΑ ΚΑΙ ΥΠΟΛΟΓΙΣΜΟΙ
368 # =====
369
370 def contains_excluded_category(rule_key: RuleKey) -> bool:
371     """Ελέγχει και τις δύο πλευρές του κανόνα."""
372
373     antecedent, consequent = rule_key
374     return any(
375         category in EXCLUDED_CATEGORIES
376         for category in (*antecedent, *consequent)
377     )
378
379
380 def relative_change(new_value: float, old_value: float) -> float | None:
381     """Υπολογίζει σχετική μεταβολή επί τοις εκατό."""
382
383     if old_value == 0:
384         return None
385
386     return (new_value - old_value) / old_value * 100
387
388
389 def build_comparison_row(
390     rule_key: RuleKey,
391     covid_metrics: dict[str, float],
392     postcovid_metrics: dict[str, float],
393 ) -> dict[str, object]:
394     """Δημιουργεί μία γραμμή σύγκρισης για κοινό κανόνα."""
395
396     antecedent, consequent = rule_key
397
398     support_covid = covid_metrics["support"]
399     support_post = postcovid_metrics["support"]
400     confidence_covid = covid_metrics["confidence"]
401     confidence_post = postcovid_metrics["confidence"]
402     lift_covid = covid_metrics["lift"]
403     lift_post = postcovid_metrics["lift"]
404
405     return {
406         "Antecedent": display_itemset(antecedent),
407         "Consequent": display_itemset(consequent),
408         "Rule": display_rule(antecedent, consequent),
409         "Antecedent_Length": len(antecedent),
410         "Consequent_Length": len(consequent),
411         "Support_COVID_Percent": support_covid * 100,
412         "Support_PostCOVID_Percent": support_post * 100,
413         "Delta_Support_Percentage_Points": (

```

```

414         support_post - support_covid
415     ) * 100,
416     "Confidence_COVID_Percent": confidence_covid * 100,
417     "Confidence_PostCOVID_Percent": confidence_post * 100,
418     "Delta_Confidence_Percentage_Points": (
419         confidence_post - confidence_covid
420     ) * 100,
421     "Confidence_Relative_Change_Percent": relative_change(
422         confidence_post,
423         confidence_covid,
424     ),
425     "Lift_COVID": lift_covid,
426     "Lift_PostCOVID": lift_post,
427     "Delta_Lift": lift_post - lift_covid,
428     "Lift_Relative_Change_Percent": relative_change(
429         lift_post,
430         lift_covid,
431     ),
432 }
433
434
435 # =====
436 # ΑΠΟΘΗΚΕΥΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ
437 # =====
438
439 OUTPUT_COLUMNS = [
440     "Antecedent",
441     "Consequent",
442     "Rule",
443     "Antecedent_Length",
444     "Consequent_Length",
445     "Support_COVID_Percent",
446     "Support_PostCOVID_Percent",
447     "Delta_Support_Percentage_Points",
448     "Confidence_COVID_Percent",
449     "Confidence_PostCOVID_Percent",
450     "Delta_Confidence_Percentage_Points",
451     "Confidence_Relative_Change_Percent",
452     "Lift_COVID",
453     "Lift_PostCOVID",
454     "Delta_Lift",
455     "Lift_Relative_Change_Percent",
456 ]
457
458
459 DECIMAL_COLUMNS = set(OUTPUT_COLUMNS) - {
460     "Antecedent",
461     "Consequent",
462     "Rule",
463     "Antecedent_Length",
464     "Consequent_Length",
465 }
466
467
468 def format_output_value(column: str, value: object) -> object:
469     """Μορφοποιεί με αρκετή ακρίβεια χωρίς να αλλάζει τη μονάδα."""
470
471     if value is None:
472         return ""
473
474     if column in DECIMAL_COLUMNS:
475         return f"{float(value):.8f}"
476

```

```

477     return value
478
479
480 def write_comparison_csv(
481     file_path: Path,
482     rows: Iterable[dict[str, object]],
483 ) -> None:
484     """Αποθηκεύει πίνακα αποτελεσμάτων σε CSV με UTF-8-SIG."""
485
486     with file_path.open(
487         "w",
488         encoding="utf-8-sig",
489         newline="",
490     ) as output_file:
491         writer = csv.DictWriter(
492             output_file,
493             fieldnames=OUTPUT_COLUMNS,
494             delimiter=";",
495         )
496         writer.writeheader()
497
498         for row in rows:
499             writer.writerow(
500                 {
501                     column: format_output_value(column, row[column])
502                     for column in OUTPUT_COLUMNS
503                 }
504             )
505
506
507 def write_summary_csv(
508     file_path: Path,
509     rows: list[tuple[str, object]],
510 ) -> None:
511     """Αποθηκεύει τη σύνοψη της ανάλυσης."""
512
513     with file_path.open(
514         "w",
515         encoding="utf-8-sig",
516         newline="",
517     ) as output_file:
518         writer = csv.writer(output_file, delimiter=";")
519         writer.writerow(["Metric", "Value"])
520         writer.writerows(rows)
521
522
523 def print_top_rows(
524     title: str,
525     rows: list[dict[str, object]],
526     metric: str,
527 ) -> None:
528     """Εμφανίζει συνοπτικά ένα Top-N στο terminal."""
529
530     print("\n" + "=" * 120)
531     print(title)
532     print("=" * 120)
533
534     for position, row in enumerate(rows, start=1):
535         print(f"{{position}}. {row['Rule']}")
536
537         if metric == "confidence":
538             print(
539                 "    Confidence COVID: "

```

```

540         f"{row['Confidence_COVID_Percent']:.3f}% | "
541         "Post-COVID: "
542         f"{row['Confidence_PostCOVID_Percent']:.3f}% | "
543         "Δ: "
544         f"{row['Delta_Confidence_Percentage_Points']:+.3f} π.μ. | "
545         "Σχετική μεταβολή: "
546         f"{row['Confidence_Relative_Change_Percent']:+.2f}%"
547     )
548     else:
549         print(
550             f"    Lift COVID: {row['Lift_COVID']:.4f} | "
551             f"Post-COVID: {row['Lift_PostCOVID']:.4f} | "
552             f"Δ: {row['Delta_Lift']:+.4f} | "
553             "Σχετική μεταβολή: "
554             f"{row['Lift_Relative_Change_Percent']:+.2f}%"
555         )
556
557
558 def create_results_zip(output_dir: Path) -> Path:
559     """Συμπιέζει όλα τα παραγόμενα CSV σε ένα ZIP."""
560
561     zip_path = output_dir.with_suffix(".zip")
562
563     with zipfile.ZipFile(
564         zip_path,
565         "w",
566         compression=zipfile.ZIP_DEFLATED,
567     ) as zip_file:
568         for result_file in sorted(output_dir.iterdir()):
569             if result_file.is_file():
570                 zip_file.write(result_file, arcname=result_file.name)
571
572     return zip_path
573
574
575 # =====
576 # ΚΥΡΙΑ ΕΚΤΕΛΕΣΗ
577 # =====
578
579 def main() -> None:
580     parser = argparse.ArgumentParser(
581         description=(
582             "Σύγκριση confidence και lift κοινών association rules "
583             "μεταξύ COVID και Post-COVID."
584         )
585     )
586     parser.add_argument(
587         "covid_file",
588         nargs="?",
589         type=Path,
590         default=DEFAULT_COVID_FILE,
591         help="COVID rules CSV ή ZIP",
592     )
593     parser.add_argument(
594         "postcovid_file",
595         nargs="?",
596         type=Path,
597         default=DEFAULT_POSTCOVID_FILE,
598         help="Post-COVID rules CSV ή ZIP",
599     )
600     parser.add_argument(
601         "--output-dir",
602         type=Path,

```

```

603     default=DEFAULT_OUTPUT_DIR,
604     help="Φάκελος αποθήκευσης αποτελεσμάτων",
605 )
606 parser.add_argument(
607     "--top-n",
608     type=int,
609     default=TOP_N,
610     help="Πλήθος κανόνων στους πίνακες Top-N",
611 )
612 args = parser.parse_args()
613
614 if args.top_n <= 0:
615     raise ValueError("To --top-n πρέπει να είναι θετικός ακέραιος.")
616
617 print("\n" + "=" * 80)
618 print("COVID vs POST-COVID")
619 print("ΣΥΓΚΡΙΣΗ CONFIDENCE ΚΑΙ LIFT ΚΟΙΝΩΝ ΚΑΝΟΝΩΝ")
620 print("=" * 80)
621
622 covid_rules = read_rules(args.covid_file)
623 postcovid_rules = read_rules(args.postcovid_file)
624
625 common_before_exclusion = set(covid_rules) & set(postcovid_rules)
626
627 common_after_exclusion = {
628     rule_key
629     for rule_key in common_before_exclusion
630     if not contains_excluded_category(rule_key)
631 }
632
633 comparison_rows = [
634     build_comparison_row(
635         rule_key,
636         covid_rules[rule_key],
637         postcovid_rules[rule_key],
638     )
639     for rule_key in common_after_exclusion
640 ]
641
642 comparison_rows.sort(key=lambda row: row["Rule"])
643
644 confidence_increases = [
645     row
646     for row in comparison_rows
647     if row["Delta_Confidence_Percentage_Points"] > 0
648 ]
649 confidence_decreases = [
650     row
651     for row in comparison_rows
652     if row["Delta_Confidence_Percentage_Points"] < 0
653 ]
654 confidence_unchanged = [
655     row
656     for row in comparison_rows
657     if row["Delta_Confidence_Percentage_Points"] == 0
658 ]
659
660 lift_increases = [
661     row
662     for row in comparison_rows
663     if row["Delta_Lift"] > 0
664 ]
665 lift_decreases = [

```

```

666         row
667         for row in comparison_rows
668         if row["Delta_Lift"] < 0
669     ]
670     lift_unchanged = [
671         row
672         for row in comparison_rows
673         if row["Delta_Lift"] == 0
674     ]
675
676     top_confidence_increases = sorted(
677         confidence_increases,
678         key=lambda row: (
679             -row["Delta_Confidence_Percentage_Points"],
680             row["Rule"],
681         ),
682     )[: args.top_n]
683
684     top_confidence_decreases = sorted(
685         confidence_decreases,
686         key=lambda row: (
687             row["Delta_Confidence_Percentage_Points"],
688             row["Rule"],
689         ),
690     )[: args.top_n]
691
692     top_lift_increases = sorted(
693         lift_increases,
694         key=lambda row: (-row["Delta_Lift"], row["Rule"]),
695     )[: args.top_n]
696
697     top_lift_decreases = sorted(
698         lift_decreases,
699         key=lambda row: (row["Delta_Lift"], row["Rule"]),
700     )[: args.top_n]
701
702     args.output_dir.mkdir(parents=True, exist_ok=True)
703
704     write_comparison_csv(
705         args.output_dir / "Common_Rules_Metrics_Comparison.csv",
706         comparison_rows,
707     )
708     write_comparison_csv(
709         args.output_dir / "Top10_Confidence_Increases.csv",
710         top_confidence_increases,
711     )
712     write_comparison_csv(
713         args.output_dir / "Top10_Confidence_Decreases.csv",
714         top_confidence_decreases,
715     )
716     write_comparison_csv(
717         args.output_dir / "Top10_Lift_Increases.csv",
718         top_lift_increases,
719     )
720     write_comparison_csv(
721         args.output_dir / "Top10_Lift_Decreases.csv",
722         top_lift_decreases,
723     )
724
725     summary_rows = [
726         ("COVID rules", len(covid_rules)),
727         ("Post-COVID rules", len(postcovid_rules)),
728         (

```

```

729         "Common directed rules before exclusions",
730         len(common_before_exclusion),
731     ),
732     (
733         "Common directed rules after exclusions",
734         len(comparison_rows),
735     ),
736     ("Rules with increased confidence", len(confidence_increases)),
737     ("Rules with decreased confidence", len(confidence_decreases)),
738     ("Rules with unchanged confidence", len(confidence_unchanged)),
739     ("Rules with increased lift", len(lift_increases)),
740     ("Rules with decreased lift", len(lift_decreases)),
741     ("Rules with unchanged lift", len(lift_unchanged)),
742     ("Top-N", args.top_n),
743     (
744         "Excluded categories",
745         " | ".join(sorted(EXCLUDED_CATEGORIES)),
746     ),
747     (
748         "Rule matching criterion",
749         "Same antecedent and same consequent; direction preserved",
750     ),
751     (
752         "Confidence ranking criterion",
753         "Delta Confidence in percentage points",
754     ),
755     ("Lift ranking criterion", "Absolute Delta Lift"),
756 ]
757
758 write_summary_csv(
759     args.output_dir / "Analysis_Summary.csv",
760     summary_rows,
761 )
762
763 print("\n" + "-" * 80)
764 print("ΣΥΝΟΨΗ")
765 print("-" * 80)
766 print(f"COVID rules: {len(covid_rules):,}")
767 print(f"Post-COVID rules: {len(postcovid_rules):,}")
768 print(
769     "Κοινοί κατευθυνόμενοι κανόνες πριν τις εξαιρέσεις: "
770     f"{len(common_before_exclusion):,}"
771 )
772 print(
773     "Κοινοί κατευθυνόμενοι κανόνες μετά τις εξαιρέσεις: "
774     f"{len(comparison_rows):,}"
775 )
776 print(f"Αύξηση confidence: {len(confidence_increases):,}")
777 print(f"Μείωση confidence: {len(confidence_decreases):,}")
778 print(f"Αμετάβλητο confidence: {len(confidence_unchanged):,}")
779 print(f"Αύξηση lift: {len(lift_increases):,}")
780 print(f"Μείωση lift: {len(lift_decreases):,}")
781 print(f"Αμετάβλητο lift: {len(lift_unchanged):,}")
782
783 print_top_rows(
784     f"TOP-{args.top_n} ΑΥΞΗΣΕΙΣ CONFIDENCE",
785     top_confidence_increases,
786     "confidence",
787 )
788 print_top_rows(
789     f"TOP-{args.top_n} ΜΕΙΩΣΕΙΣ CONFIDENCE",
790     top_confidence_decreases,
791     "confidence",

```

```

792     )
793     print_top_rows(
794         f"TOP-{args.top_n} ΑΥΞΗΣΕΙΣ LIFT",
795         top_lift_increases,
796         "lift",
797     )
798     print_top_rows(
799         f"TOP-{args.top_n} ΜΕΙΩΣΕΙΣ LIFT",
800         top_lift_decreases,
801         "lift",
802     )
803
804     zip_path = create_results_zip(args.output_dir)
805
806     print("\n" + "=" * 80)
807     print("Η ΑΝΑΛΥΣΗ ΟΛΟΚΛΗΡΩΘΗΚΕ")
808     print("=" * 80)
809     print(f"Φάκελος αποτελεσμάτων: {args.output_dir.resolve()}")
810     print(f"ZIP αποτελεσμάτων: {zip_path.resolve()}")
811
812
813 if __name__ == "__main__":
814     main()

```

Παράρτημα Β΄

Συμπληρωματικοί Πίνακες Κανόνων Συσχέτισης

Β΄.1 Top-100 Κανόνες Συσχέτισης Περιόδου COVID-19

Στην παρούσα ενότητα παρουσιάζονται οι 100 σημαντικότεροι κανόνες συσχέτισης που εξήχθησαν από το επιλεγμένο πειραματικό σενάριο P5 για την περίοδο COVID-19.

Πίνακας Β΄.1: Top-100 κανόνες συσχέτισης περιόδου COVID-19

Antecedent	Consequent	Support (%)	Confidence (%)	Lift	Leverage	Conviction
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	9,23	55,62	4,08	0,07	1,95
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	4,82	35,36	4,46	0,037	1,42
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	4,82	69,91	4,21	0,037	2,77
ΦΡΕΣΚΑ ΦΡΟΥΤΑ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	6,75	46,51	2,1	0,035	1,46
ΓΙΑΟΥΡΤΙΑ	ΓΑΛΑ ΦΡΕΣΚΟ	9,29	45,62	1,6	0,035	1,31
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΓΑΛΑ ΦΡΕΣΚΟ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	4,25	71,91	4,33	0,033	2,97
ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ	ΦΡΕΣΚΑ ΦΡΟΥΤΑ	4,32	40,65	2,8	0,028	1,44
ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	4,64	37,67	2,27	0,026	1,34
ΚΑΘΑΡΙΣΤΙΚΑ	ΠΡΟΙΟΝΤΑ ΠΛΥΣΙΜΑΤΟΣ	3,67	43,79	3,25	0,025	1,54
ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	4,79	44,98	2,03	0,024	1,42
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	ΓΑΛΑ ΦΡΕΣΚΟ	7,15	43,09	1,51	0,024	1,26
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΓΙΑΟΥΡΤΙΑ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	3,08	72,53	4,37	0,024	3,04
ΑΛΛΑΝΤΙΚΑ ΣΥΣΚΕΥΑΣΜΕΝΑ ΨΥΓΕΙΟΥ	ΤΥΡΙΑ ΔΙΑΦΟΡΑ ΣΥΣΚ	3,22	36,66	3,66	0,023	1,42
ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	4,02	32,6	2,39	0,023	1,28
ΠΡΩΙΝΑ	ΓΑΛΑ ΦΡΕΣΚΟ	4,95	50,72	1,78	0,022	1,45
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	2,74	68,26	4,11	0,021	2,63
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ	3,19	34,58	2,8	0,021	1,34
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	ΓΙΑΟΥΡΤΙΑ	5,42	32,68	1,6	0,02	1,18
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	ΓΑΛΑ ΦΡΕΣΚΟ	5,92	43,41	1,52	0,02	1,26

Συνέχεια του Πίνακα Β'.1

Antecedent	Consequent	Support (%)	Confidence (%)	Lift	Leverage	Conviction
ΛΑΧΑΝΙΚΑ ΣΕ ΚΟΝΣΕΡΒΑ- ΧΑΡΤΙΝΗ ΣΥΣΚΕΥΑΣΙΑ	ΖΥΜΑΡΙΚΑ	2,81	36,99	3,37	0,02	1,41
ΜΑΛΑΚΤΙΚΑ	ΠΡΟΙΟΝΤΑ ΠΛΥΣΙΜΑΤΟΣ	2,55	54,8	4,07	0,019	1,91
ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ	ΓΑΛΑ ΦΡΕΣΚΟ	5,44	44,08	1,54	0,019	1,28
ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙ- ΚΙΛΙΑ, ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	ΦΡΕΣΚΑ ΦΡΟΥΤΑ	3,46	32,13	2,21	0,019	1,26
ΓΑΛΑ ΦΡΕΣΚΟ, ΕΙΔΗ ΣΥ- ΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	ΓΙΑΟΥΡΤΙΑ	4,66	33,85	1,66	0,019	1,2
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΓΑΛΑ ΦΡΕΣΚΟ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	2,28	38,44	4,85	0,018	1,5
SNACK	ΑΝΑΨΥΚΤΙΚΑ	4,26	31,21	1,73	0,018	1,19
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΓΑΛΑ ΦΡΕΣΚΟ, ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	2,28	73,73	4,44	0,018	3,17
ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙ- ΛΙΑ, ΓΑΛΑ ΦΡΕΣΚΟ	ΦΡΕΣΚΑ ΦΡΟΥΤΑ	2,83	36,89	2,54	0,017	1,35
ΠΑΤΑΤΕΣ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙ- ΛΙΑ	2,71	59,35	2,68	0,017	1,92
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΑΝΑΨΥ- ΚΤΙΚΑ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	2,23	70,29	4,23	0,017	2,81
ΓΙΑΟΥΡΤΙΑ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙ- ΛΙΑ	6,19	30,4	1,37	0,017	1,12
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ	2,63	33,13	2,69	0,016	1,31
ΜΠΥΡΕΣ	ΑΝΑΨΥΚΤΙΚΑ	3,26	36,63	2,03	0,016	1,29
ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙ- ΛΙΑ, ΦΡΕΣΚΑ ΦΡΟΥΤΑ	ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ	2,35	34,74	3,27	0,016	1,37
ΦΡΕΣΚΑ ΦΡΟΥΤΑ	ΓΙΑΟΥΡΤΙΑ	4,56	31,41	1,54	0,016	1,16
ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙ- ΛΙΑ, ΓΙΑΟΥΡΤΙΑ	ΦΡΕΣΚΑ ΦΡΟΥΤΑ	2,49	40,23	2,77	0,016	1,43
ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ	ΓΙΑΟΥΡΤΙΑ	4,08	33,13	1,63	0,016	1,19
ΑΛΕΥΡΙ-ΣΙΜΙΓΔΑΛΙ	ΕΙΔΗ ΖΑΧΑΡΟΠΛΑΣΤΙΚΗΣ	1,97	34,44	4,95	0,016	1,42
ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ, ΦΡΕΣΚΑ ΦΡΟΥΤΑ	ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ	2,32	32,41	3,05	0,016	1,32
ΤΥΡΙΑ ΔΙΑΦΟΡΑ ΣΥΣΚ	ΓΙΑΟΥΡΤΙΑ	3,6	35,92	1,76	0,016	1,24
Κρεμμύδια κλού.	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙ- ΛΙΑ	2,36	63,73	2,88	0,015	2,15
ΕΙΔΗ ΜΙΑΣ ΧΡΗΣΕΩΣ	ΠΡΟΙΟΝΤΑ ΠΛΥΣΙΜΑΤΟΣ	2,52	33,48	2,49	0,015	1,3
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ	2,36	34,19	2,77	0,015	1,33
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΓΑΛΑ ΦΡΕΣΚΟ	ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ	2,36	32,93	2,67	0,015	1,31
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	ΓΙΑΟΥΡΤΙΑ	4,25	31,16	1,53	0,015	1,16
ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙ- ΛΙΑ, ΓΑΛΑ ΦΡΕΣΚΟ	ΓΙΑΟΥΡΤΙΑ	3,03	39,47	1,94	0,015	1,32
ΤΥΡΙΑ ΔΙΑΦΟΡΑ ΣΥΣΚ, ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	ΑΛΛΑΝΤΙΚΑ ΣΥΣΚΕΥΑ- ΣΜΕΝΑ ΨΥΓΕΙΟΥ	1,95	35,24	4,01	0,015	1,41
ΦΕΤΑ/ΚΙΛΟ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	2,47	40,47	2,44	0,015	1,4
ΡΥΖΙΑ-ΟΣΠΡΙΑ	ΖΥΜΑΡΙΚΑ	2,04	36,33	3,31	0,014	1,4
ΧΥΜΟΙ ΜΑΚΡΑΣ ΔΙΑΡΚΕΙΑΣ	ΑΝΑΨΥΚΤΙΚΑ	3,09	33,24	1,84	0,014	1,23
ΓΑΛΑ ΦΡΕΣΚΟ, ΓΙΑΟΥΡΤΙΑ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	2,95	31,78	1,91	0,014	1,22
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΖΥΜΑ- ΡΙΚΑ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	1,8	73,73	4,44	0,014	3,17

Συνέχεια του Πίνακα Β'.1

Antecedent	Consequent	Support (%)	Confidence (%)	Lift	Leverage	Conviction
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΠΡΟΙΟΝΤΑ ΠΛΥΣΙΜΑΤΟΣ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	1,8	73,48	4,43	0,014	3,14
ΑΥΓΑ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	2,9	42,5	1,92	0,014	1,35
ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ	ΓΙΑΟΥΡΤΙΑ	3,56	33,45	1,64	0,014	1,2
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΓΑΛΛΑ ΦΡΕΣΚΟ, ΓΙΑΟΥΡΤΙΑ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	1,78	75,39	4,54	0,014	3,39
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, SNACK	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	1,79	71,32	4,3	0,014	2,91
ΣΟΚΟΛΑΤΕΣ, ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	1,79	71,08	4,28	0,014	2,88
ΤΥΡΙΑ ΔΙΑΦΟΡΑ ΣΥΣΚ	ΓΑΛΛΑ ΦΡΕΣΚΟ	4,23	42,21	1,48	0,014	1,24
ΚΕΝΑ	ΜΠΥΡΕΣ	1,59	60,6	6,82	0,014	2,31
ΖΥΜΑΡΙΚΑ	ΓΑΛΛΑ ΦΡΕΣΚΟ	4,48	40,85	1,43	0,013	1,21
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΚΑΦΕΔΕΣ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	1,73	73,51	4,43	0,013	3,15
ΠΡΩΙΝΑ	ΓΙΑΟΥΡΤΙΑ	3,32	34,08	1,67	0,013	1,21
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΓΙΑΟΥΡΤΙΑ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	1,67	39,39	4,97	0,013	1,52
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΓΑΛΛΑ ΦΡΕΣΚΟ	ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ	2,05	34,72	2,82	0,013	1,34
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΦΡΕΣΚΑ ΦΡΟΥΤΑ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	1,72	68,64	4,13	0,013	2,66
ΓΑΛΛΑ ΦΡΕΣΚΟ, ΦΡΕΣΚΑ ΦΡΟΥΤΑ	ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ	1,87	34,61	3,25	0,013	1,37
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ, ΓΙΑΟΥΡΤΙΑ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	1,67	73,56	4,43	0,013	3,15
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ	1,89	39,18	3,18	0,013	1,44
ΦΡΕΣΚΑ ΦΡΟΥΤΑ	ΓΑΛΛΑ ΦΡΕΣΚΟ	5,4	37,23	1,3	0,013	1,14
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΠΡΩΙΝΑ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	1,61	75,09	4,52	0,013	3,35
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΜΠΙΣΚΟΤΑ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	1,61	72,74	4,38	0,012	3,06
ΓΑΛΑΤΑ	ΚΑΦΕΔΕΣ	2,1	30,78	2,45	0,012	1,26
ΖΥΜΑΡΙΚΑ	ΓΙΑΟΥΡΤΙΑ	3,47	31,66	1,55	0,012	1,17
ΣΑΛΑΤΕΣ ΦΡΕΣΚΙΕΣ ΣΥΣΚΕΥΑΣΜΕΝΕΣ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	2,23	48,95	2,21	0,012	1,53
ΧΑΡΤΙΑ ΚΟΥΖΙΝΑΣ	ΠΡΟΙΟΝΤΑ ΠΛΥΣΙΜΑΤΟΣ	1,96	34,95	2,6	0,012	1,33
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	1,52	37,9	4,78	0,012	1,48
ΓΙΑΟΥΡΤΙΑ, ΦΡΕΣΚΑ ΦΡΟΥΤΑ	ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ	1,68	36,94	3,47	0,012	1,42
ΛΑΧΑΝΙΚΑ ΣΕ ΚΟΝΣΕΡΒΑ-ΧΑΡΤΙΝΗ ΣΥΣΚΕΥΑΣΙΑ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	2,88	37,88	1,71	0,012	1,25
ΣΑΜΠΟΥΤΑΝ	Αφρόλουτρα.	1,49	30,46	5,04	0,012	1,35
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	1,54	72,13	4,34	0,012	2,99
ΜΑΡΓΑΡΙΝΕΣ	ΓΑΛΛΑ ΦΡΕΣΚΟ	3,17	45,65	1,6	0,012	1,31
ΜΠΙΣΚΟΤΑ, ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	SNACK	2,09	31,45	2,3	0,012	1,26
ΨΩΜΙ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	3,62	32,72	1,48	0,012	1,16
ΤΥΡΙΑ ΔΙΑΦΟΡΑ ΣΥΣΚ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	3,39	33,82	1,53	0,012	1,18
ΧΑΡΤΙΑ ΥΓΕΙΑΣ	ΠΡΟΙΟΝΤΑ ΠΛΥΣΙΜΑΤΟΣ	1,94	33,63	2,5	0,012	1,3

Συνέχεια του Πίνακα Β'.1

Antecedent	Consequent	Support (%)	Confidence (%)	Lift	Leverage	Conviction
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ, ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	1,52	69,82	4,21	0,012	2,76
ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ	ΓΑΛΑ ΦΡΕΣΚΟ	4,19	39,39	1,38	0,012	1,18
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΓΑΛΑ ΦΡΕΣΚΟ	ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ	1,67	39,16	3,18	0,011	1,44
ΑΛΛΑΝΤΙΚΑ ΣΥΣΚΕΥΑΣΜΕΝΑ ΨΥΓΕΙΟΥ	ΓΙΑΟΥΡΤΙΑ	2,92	33,19	1,63	0,011	1,19
ΛΕΥΚΟ ΤΥΡΙ ΚΙΛΟΥ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	1,92	39,96	2,41	0,011	1,39
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΓΙΑΟΥΡΤΙΑ	ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ	1,79	33,05	2,68	0,011	1,31
ΑΥΓΑ	ΓΑΛΑ ΦΡΕΣΚΟ	3,07	44,94	1,57	0,011	1,3
ΝΕΡΑ, ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	ΑΝΑΨΥΚΤΙΚΑ	2,44	33,17	1,83	0,011	1,23
ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	3,83	31,09	1,41	0,011	1,13
ΑΛΛΑΝΤΙΚΑ ΣΥΣΚΕΥΑΣΜΕΝΑ ΨΥΓΕΙΟΥ	ΓΑΛΑ ΦΡΕΣΚΟ	3,61	41,1	1,44	0,011	1,21
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΝΕΡΑ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	1,45	68,54	4,13	0,011	2,65
ΜΑΡΓΑΡΙΝΕΣ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	2,24	32,33	1,95	0,011	1,23
ΖΥΜΑΡΙΚΑ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	3,51	32,01	1,45	0,011	1,15
ΤΥΡΙΑ ΔΙΑΦΟΡΑ ΣΥΣΚ, ΓΑΛΑ ΦΡΕΣΚΟ	ΑΛΛΑΝΤΙΚΑ ΣΥΣΚΕΥΑΣΜΕΝΑ ΨΥΓΕΙΟΥ	1,45	34,31	3,9	0,011	1,39

Β'.2 Top-100 Κανόνες Συσχέτισης Μεταπανδημικής Περιόδου

Στην παρούσα ενότητα παρουσιάζονται οι 100 σημαντικότεροι κανόνες συσχέτισης που εξήχθησαν από το επιλεγμένο πειραματικό σενάριο P5 για τη μεταπανδημική περίοδο.

Πίνακας Β'.2: Top-100 κανόνες συσχέτισης μεταπανδημικής περιόδου

Antecedent	Consequent	Support (%)	Confidence (%)	Lift	Leverage	Conviction
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	7,46	67,3	4,73	0,059	2,62
ΦΡΕΣΚΑ ΦΡΟΥΤΑ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	6,14	43,3	2,14	0,033	1,41
ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ, ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	3,72	57,55	5,19	0,03	2,09
ΓΙΑΟΥΡΤΙΑ	ΓΑΛΑ ΦΡΕΣΚΟ	7,01	40,36	1,68	0,028	1,27
ΕΤΟΙΜΟ ΦΑΓΗΤΟ CATERING	ΑΠΑΡΑΔΕΚΤΑ ΕΜΠΟΡΕΥΜΑΤΑ	2,87	83,31	28,28	0,028	5,82
ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ	ΦΡΕΣΚΑ ΦΡΟΥΤΑ	4,08	39,19	2,76	0,026	1,41
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΓΑΛΑ ΦΡΕΣΚΟ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	3,04	55,94	5,05	0,024	2,02
ΚΑΘΑΡΙΣΤΙΚΑ	ΠΡΟΙΟΝΤΑ ΠΛΥΣΙΜΑΤΟΣ	2,95	43,17	3,65	0,021	1,55
ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	4,25	40,8	2,01	0,021	1,35
ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	3,63	34,15	2,4	0,021	1,3
ΑΛΛΑΝΤΙΚΑ ΣΥΣΚΕΥΑΣΜΕΝΑ ΨΥΓΕΙΟΥ	ΤΥΡΙΑ ΔΙΑΦΟΡΑ ΣΥΣΚ	2,82	36,64	4	0,021	1,43

Συνέχεια του Πίνακα Β'2

Antecedent	Consequent	Support (%)	Confidence (%)	Lift	Leverage	Conviction
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	ΓΑΛΑ ΦΡΕΣΚΟ	5,43	38,18	1,59	0,02	1,23
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΨΩ- ΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	2,37	65,46	5,9	0,02	2,57
ΜΑΛΑΚΤΙΚΑ	ΠΡΟΙΟΝΤΑ ΠΛΥΣΙΜΑΤΟΣ	2,47	55,25	4,67	0,019	1,97
Κρεμμύδια κιλού.	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙ- ΛΙΑ	2,83	62,88	3,1	0,019	2,15
ΠΡΩΙΝΑ	ΓΑΛΑ ΦΡΕΣΚΟ	3,71	45,65	1,9	0,018	1,4
SNACK	ΑΝΑΨΥΚΤΙΚΑ	4,17	30,21	1,72	0,017	1,18
ΓΙΑΟΥΡΤΙΑ, ΜΑΛΑΚΑ ΤΥΡΙΑ- /ΚΙΛΟ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	2,19	53,26	4,8	0,017	1,9
ΛΑΧΑΝΙΚΑ ΣΕ ΚΟΝΣΕΡΒΑ- ΧΑΡΤΙΝΗ ΣΥΣΚΕΥΑΣΙΑ	ΖΥΜΑΡΙΚΑ	2,35	36,72	3,8	0,017	1,43
ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ, ΛΑΧΑΝΙΚΑ ΠΡΑ- ΣΙΝΑ ΠΟΙΚΙΛΙΑ	ΦΡΕΣΚΑ ΦΡΟΥΤΑ	3,09	32,19	2,27	0,017	1,27
ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙ- ΛΙΑ, ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	2,1	54,08	4,88	0,017	1,94
ΠΑΤΑΤΕΣ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙ- ΛΙΑ	2,53	56,54	2,79	0,016	1,83
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ	2,37	31,82	3	0,016	1,31
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	ΓΑΛΑ ΦΡΕΣΚΟ	4,21	38,01	1,58	0,015	1,23
ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ, ΓΑΛΑ ΦΡΕΣΚΟ	ΓΙΑΟΥΡΤΙΑ	3,48	31,02	1,78	0,015	1,2
ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ	ΓΑΛΑ ΦΡΕΣΚΟ	4,08	38,41	1,6	0,015	1,23
ΜΠΥΡΕΣ	ΑΝΑΨΥΚΤΙΚΑ	3,03	34,9	1,99	0,015	1,27
ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ, ΑΠΑΡΑΔΕΚΤΑ ΕΜΠΟΡΕΥΜΑΤΑ	ΕΤΟΙΜΟ ΦΑΓΗΤΟ CATERING	1,5	97,53	28,29	0,014	39,02
ΦΡΕΣΚΑ ΦΡΟΥΤΑ, ΕΙΔΗ ΣΥ- ΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ	2,14	31,21	3	0,014	1,3
ΤΥΡΙΑ ΔΙΑΦΟΡΑ ΣΥΣΚ	ΓΙΑΟΥΡΤΙΑ	3	32,82	1,89	0,014	1,23
ΦΡΕΣΚΑ ΦΡΟΥΤΑ, ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ	2,05	33,35	3,2	0,014	1,34
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΑΝΑ- ΨΥΚΤΙΚΑ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	1,72	59,38	5,36	0,014	2,19
ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙ- ΛΙΑ, ΓΑΛΑ ΦΡΕΣΚΟ	ΦΡΕΣΚΑ ΦΡΟΥΤΑ	2,25	37,16	2,62	0,014	1,37
ΓΙΑΟΥΡΤΙΑ, ΛΑΧΑΝΙΚΑ ΠΡΑ- ΣΙΝΑ ΠΟΙΚΙΛΙΑ	ΦΡΕΣΚΑ ΦΡΟΥΤΑ	2,1	41,37	2,92	0,014	1,46
ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ	ΓΙΑΟΥΡΤΙΑ	3,13	30,08	1,73	0,013	1,18
ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ, ΤΥΡΙΑ ΔΙΑΦΟΡΑ ΣΥΣΚ	ΑΛΛΑΝΤΙΚΑ ΣΥΣΚΕΥΑ- ΣΜΕΝΑ ΨΥΓΕΙΟΥ	1,69	33,89	4,41	0,013	1,4
ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ, ΜΑΛΑΚΑ ΤΥΡΙΑ- /ΚΙΛΟ, ΓΑΛΑ ΦΡΕΣΚΟ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	1,57	60,91	5,49	0,013	2,27
ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ, ΜΑΛΑΚΑ ΤΥΡΙΑ- /ΚΙΛΟ	ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ	1,97	30,47	2,87	0,013	1,29
ΦΕΤΑ/ΚΙΛΟ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	2,01	39,07	2,75	0,013	1,41
ΑΥΓΑ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙ- ΛΙΑ	2,47	40,94	2,02	0,012	1,35
ΕΙΔΗ ΜΙΑΣ ΧΡΗΣΕΩΣ	ΠΡΟΙΟΝΤΑ ΠΛΥΣΙΜΑΤΟΣ	1,95	31,51	2,67	0,012	1,29
ΧΥΜΟΙ ΜΑΚΡΑΣ ΔΙΑΡΚΕΙΑΣ	ΑΝΑΨΥΚΤΙΚΑ	2,61	32,48	1,85	0,012	1,22

Συνέχεια του Πίνακα Β'2

Antecedent	Consequent	Support (%)	Confidence (%)	Lift	Leverage	Conviction
ΡΥΖΙΑ-ΟΣΠΡΙΑ	ΖΥΜΑΡΙΚΑ	1,67	33,73	3,49	0,012	1,36
ΑΛΕΥΡΙ-ΣΙΜΙΓΔΑΛΙ	ΕΙΔΗ ΖΑΧΑΡΟΠΛΑΣΤΙΚΗΣ	1,42	31,86	6,14	0,012	1,39
ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙ- ΛΙΑ, ΓΑΛΑ ΦΡΕΣΚΟ	ΓΙΑΟΥΡΤΙΑ	2,23	36,94	2,13	0,012	1,31
ΚΕΝΑ	ΜΠΥΡΕΣ	1,37	60,72	6,99	0,012	2,32
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, SNACK	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	1,42	60,46	5,45	0,012	2,25
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΨΩ- ΜΙΑ ΣΥΣΚΕΤΑΣΜΕΝΑ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	1,36	37,42	6,96	0,012	1,51
ΤΥΡΙΑ ΔΙΑΦΟΡΑ ΣΥΣΚ	ΓΑΛΑ ΦΡΕΣΚΟ	3,35	36,61	1,52	0,011	1,2
ΨΩΜΙ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙ- ΛΙΑ	3,43	30,42	1,5	0,011	1,15
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	ΨΩΜΙΑ ΣΥΣΚΕΤΑΣΜΕΝΑ	1,71	31,87	3	0,011	1,31
ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ, ΜΠΙΣΚΟΤΑ	SNACK	1,99	32,38	2,35	0,011	1,27
ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ, ΜΑΛΑΚΑ ΤΥΡΙΑ- /ΚΙΛΟ, ΨΩΜΙΑ ΣΥΣΚΕΥΑ- ΣΜΕΝΑ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	1,36	68,98	6,22	0,011	2,87
ΖΥΜΑΡΙΚΑ	ΓΑΛΑ ΦΡΕΣΚΟ	3,46	35,76	1,49	0,011	1,18
ΤΥΡΙΑ ΔΙΑΦΟΡΑ ΣΥΣΚ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙ- ΛΙΑ	2,96	32,32	1,59	0,011	1,18
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΓΑΛΑ ΦΡΕΣΚΟ	ΓΙΑΟΥΡΤΙΑ	2,04	37,51	2,16	0,011	1,32
ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ, ΝΕΡΑ	ΑΝΑΨΥΚΤΙΚΑ	2,35	32,78	1,87	0,011	1,23
ΦΡΕΣΚΑ ΦΡΟΥΤΑ	ΓΑΛΑ ΦΡΕΣΚΟ	4,5	31,69	1,32	0,011	1,11
ΖΥΜΑΡΙΚΑ, ΜΑΛΑΚΑ ΤΥΡΙΑ- /ΚΙΛΟ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	1,34	56,22	5,07	0,011	2,03
ΧΑΡΤΙΑ ΚΟΥΖΙΝΑΣ	ΠΡΟΙΟΝΤΑ ΠΛΥΣΙΜΑΤΟΣ	1,73	31,34	2,65	0,011	1,28
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΓΑΛΑ ΦΡΕΣΚΟ	ΨΩΜΙΑ ΣΥΣΚΕΤΑΣΜΕΝΑ	1,65	30,39	2,86	0,011	1,28
ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ	ΓΑΛΑ ΦΡΕΣΚΟ	3,58	34,34	1,43	0,011	1,16
ΦΡΕΣΚΑ ΦΡΟΥΤΑ, ΓΑΛΑ ΦΡΕΣΚΟ	ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ	1,54	34,15	3,28	0,011	1,36
ΠΡΩΙΝΑ	ΓΙΑΟΥΡΤΙΑ	2,47	30,41	1,75	0,011	1,19
ΦΡΕΣΚΑ ΦΡΟΥΤΑ, ΓΙΑΟΥΡΤΙΑ	ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ	1,46	36,44	3,5	0,01	1,41
ΛΑΧΑΝΙΚΑ ΣΕ ΚΟΝΣΕΡΒΑ- ΧΑΡΤΙΝΗ ΣΥΣΚΕΥΑΣΙΑ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙ- ΛΙΑ	2,34	36,56	1,8	0,01	1,26
ΣΟΚΟΛΑΤΕΣ, ΜΑΛΑΚΑ ΤΥ- ΡΙΑ/ΚΙΛΟ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	1,3	56,69	5,11	0,01	2,05
ΦΡΕΣΚΑ ΦΡΟΥΤΑ, ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	1,32	52,68	4,75	0,01	1,88
ΦΡΕΣΚΑ ΦΡΟΥΤΑ, ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ	ΓΙΑΟΥΡΤΙΑ	2,1	34,13	1,96	0,01	1,25
ΦΡΕΣΚΑ ΦΡΟΥΤΑ, ΓΑΛΑ ΦΡΕΣΚΟ	ΓΙΑΟΥΡΤΙΑ	1,81	40,26	2,32	0,01	1,38
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΠΡΟΙΟΝΤΑ ΠΛΥΣΙΜΑΤΟΣ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	1,28	54,66	4,93	0,01	1,96
ΜΑΡΓΑΡΙΝΕΣ	ΓΑΛΑ ΦΡΕΣΚΟ	2,46	40,81	1,7	0,01	1,28
ΣΑΛΑΤΕΣ ΦΡΕΣΚΙΕΣ ΣΥ- ΣΚΕΥΑΣΜΕΝΕΣ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙ- ΛΙΑ	1,82	45,41	2,24	0,01	1,46
ΧΑΡΤΙΑ ΥΓΕΙΑΣ	ΠΡΟΙΟΝΤΑ ΠΛΥΣΙΜΑΤΟΣ	1,62	30,55	2,58	0,01	1,27
ΜΑΡΓΑΡΙΝΕΣ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	1,85	30,76	2,16	0,01	1,24

Συνέχεια του Πίνακα Β'2

Antecedent	Consequent	Support (%)	Confidence (%)	Lift	Leverage	Conviction
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΕΞΩ-ΤΙΚΑ ΦΡΟΥΤΑ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	1,23	55,06	4,97	0,01	1,98
ΛΕΥΚΟ ΤΥΡΙ ΚΙΛΟΥ	ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	1,54	38,28	2,69	0,01	1,39
ΖΥΜΑΡΙΚΑ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙ-ΛΙΑ	2,93	30,29	1,49	0,01	1,14
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ	1,36	36,52	3,44	0,01	1,41
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΚΑΦΕ-ΔΕΣ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	1,2	55,28	4,99	0,01	1,99
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΜΠΙ-ΣΚΟΤΑ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	1,17	58,22	5,25	0,009	2,13
ΦΡΕΣΚΑ ΦΡΟΥΤΑ, ΕΙΔΗ ΣΥ-ΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	ΓΙΑΟΥΡΤΙΑ	2,13	31,07	1,79	0,009	1,2
ΓΙΑΟΥΡΤΙΑ, ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ, ΜΑ-ΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	1,16	58,07	5,24	0,009	2,12
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΓΑΛΑ ΦΡΕΣΚΟ, ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	1,12	67,64	6,1	0,009	2,75
ΑΥΓΑ	ΓΑΛΑ ΦΡΕΣΚΟ	2,38	39,46	1,64	0,009	1,25
ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ, ΓΑΛΑ ΦΡΕΣΚΟ	ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ	1,38	32,68	3,08	0,009	1,33
ΓΙΑΟΥΡΤΙΑ, ΜΑΛΑΚΑ ΤΥΡΙΑ-/ΚΙΛΟ, ΓΑΛΑ ΦΡΕΣΚΟ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	1,15	56,58	5,1	0,009	2,05
ΓΑΛΑ ΦΡΕΣΚΟ, ΨΩΜΙΑ ΣΥ-ΣΚΕΥΑΣΜΕΝΑ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	1,38	33,77	3,05	0,009	1,34
ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ, ΛΑΧΑΝΙΚΑ ΠΡΑ-ΣΙΝΑ ΠΟΙΚΙΛΙΑ, ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	1,12	58,71	5,3	0,009	2,15
ΑΛΛΑΝΤΙΚΑ ΣΥΣΚΕΥΑΣΜΕΝΑ ΨΥΓΕΙΟΥ	ΓΑΛΑ ΦΡΕΣΚΟ	2,75	35,8	1,49	0,009	1,18
ΠΡΩΙΝΑ, ΜΑΛΑΚΑ ΤΥΡΙΑ-/ΚΙΛΟ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	1,09	61,55	5,55	0,009	2,31
ΑΥΓΑ	ΓΙΑΟΥΡΤΙΑ	1,94	32,17	1,85	0,009	1,22
ΨΩΜΙ, ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	1,11	55,85	5,04	0,009	2,01
ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙ-ΛΙΑ, ΝΕΡΑ	ΦΡΕΣΚΑ ΦΡΟΥΤΑ	1,36	40,79	2,88	0,009	1,45
ΓΙΑΟΥΡΤΙΑ, ΕΙΔΗ ΣΥΣΚΕΥΑΣΙΑΣ ΠΡΟΣ ΠΩΛΗΣΗ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙ-ΛΙΑ	2,61	30,72	1,52	0,009	1,15
ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ, ΝΕΡΑ	ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ	1,1	55,76	5,03	0,009	2,01
ΜΥΡΩΔΙΚΑ ΔΙΑΦΟΡΑ	ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙ-ΛΙΑ	1,26	66,48	3,28	0,009	2,38
ΤΥΡΙΑ ΔΙΑΦΟΡΑ ΣΥΣΚ, ΨΩ-ΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ	ΑΛΛΑΝΤΙΚΑ ΣΥΣΚΕΥΑΣΜΕΝΑ ΨΥΓΕΙΟΥ	1,05	45,27	5,88	0,009	1,69
ΜΑΡΓΑΡΙΝΕΣ	ΓΙΑΟΥΡΤΙΑ	1,92	31,86	1,83	0,009	1,21
ΤΥΡΙΑ ΔΙΑΦΟΡΑ ΣΥΣΚ, ΓΑΛΑ ΦΡΕΣΚΟ	ΑΛΛΑΝΤΙΚΑ ΣΥΣΚΕΥΑΣΜΕΝΑ ΨΥΓΕΙΟΥ	1,13	33,62	4,37	0,009	1,39

Β'.3 Κοινοί κανόνες: αυξήσεις του confidence

Ο Πίνακας Β'.3 παρουσιάζει τους δέκα κοινούς κανόνες με τη μεγαλύτερη αύξηση του confidence. Η μεταβολή υπολογίστηκε από την περίοδο COVID-19 προς τη μεταπανδημική περίοδο και εκφράζεται σε ποσοστιαίες μονάδες.

Πίνακας Β'.3: Κοινοί κανόνες με τις δέκα μεγαλύτερες αυξήσεις του confidence

Κανόνας	Confidence COVID (%)	Confidence Post-COVID (%)	Δ Confidence (π.μ.)	Σχετική μεταβολή (%)
{ΓΙΑΟΥΡΤΙΑ, ΝΕΡΑ} → {ΦΡΕΣΚΑ ΦΡΟΥΤΑ}	30,999	33,009	+2,010	+6,48
{ΑΝΑΨΥΚΤΙΚΑ, ΜΠΥΡΕΣ} → {ΝΕΡΑ}	33,326	34,834	+1,508	+4,52
{ΓΑΛΑ ΦΡΕΣΚΟ, ΓΙΑΟΥΡΤΙΑ, ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ} → {ΦΡΕΣΚΑ ΦΡΟΥΤΑ}	44,299	45,804	+1,505	+3,40
{ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ, ΝΕΡΑ} → {ΦΡΕΣΚΑ ΦΡΟΥΤΑ}	39,496	40,792	+1,296	+3,28
{ΓΙΑΟΥΡΤΙΑ, ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ} → {ΦΡΕΣΚΑ ΦΡΟΥΤΑ}	40,234	41,370	+1,136	+2,82
{ΓΑΛΑ ΦΡΕΣΚΟ, ΜΠΙΣΚΟΤΑ} → {SNACK}	30,746	31,561	+0,815	+2,65
{ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ, ΜΠΙΣΚΟΤΑ} → {ΦΡΕΣΚΑ ΦΡΟΥΤΑ}	38,964	39,724	+0,760	+1,95
{SNACK, ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ} → {ΦΡΕΣΚΑ ΦΡΟΥΤΑ}	37,019	37,668	+0,649	+1,75
{ΜΠΙΣΚΟΤΑ, ΣΟΚΟΛΑΤΕΣ} → {SNACK}	37,125	37,739	+0,614	+1,65
{ΑΝΑΨΥΚΤΙΚΑ, ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ} → {ΦΡΕΣΚΑ ΦΡΟΥΤΑ}	33,862	34,458	+0,596	+1,76

Β'.4 Κοινοί κανόνες: μειώσεις του confidence

Ο Πίνακας Β'.4 παρουσιάζει τους δέκα κοινούς κανόνες με τη μεγαλύτερη μείωση του confidence. Η μεταβολή υπολογίστηκε από την περίοδο COVID-19 προς τη μεταπανδημική περίοδο και εκφράζεται σε ποσοστιαίες μονάδες.

Πίνακας Β'.4: Κοινοί κανόνες με τις δέκα μεγαλύτερες μειώσεις του confidence

Κανόνας	Confidence COVID (%)	Confidence Post-COVID (%)	Δ Confidence (π.μ.)	Σχετική μεταβολή (%)
{ΧΥΜΟΙ ΜΑΚΡΑΣ ΔΙΑΡΚΕΙΑΣ} → {ΓΑΛΑ ΦΡΕΣΚΟ}	39,735	34,009	-5,727	-14,41
{ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ} → {ΓΑΛΑ ΦΡΕΣΚΟ}	44,083	38,407	-5,675	-12,87
{ΤΥΡΙΑ ΔΙΑΦΟΡΑ ΣΥΣΚ} → {ΓΑΛΑ ΦΡΕΣΚΟ}	42,208	36,609	-5,598	-13,26
{ΦΡΕΣΚΑ ΦΡΟΥΤΑ} → {ΓΑΛΑ ΦΡΕΣΚΟ}	37,234	31,694	-5,540	-14,88
{ΑΥΓΑ} → {ΓΑΛΑ ΦΡΕΣΚΟ}	44,936	39,462	-5,473	-12,18
{ΧΥΜΟΙ ΦΡΟΥΤΩΝ ΦΡΕΣΚΙΟΙ-ΤΣΑΙ} → {ΓΑΛΑ ΦΡΕΣΚΟ}	39,691	34,275	-5,416	-13,65
{ΑΛΛΑΝΤΙΚΑ/ΚΙΛΟ} → {ΓΑΛΑ ΦΡΕΣΚΟ}	43,410	38,012	-5,398	-12,44
{ΑΛΛΑΝΤΙΚΑ ΣΥΣΚΕΥΑΣΜΕΝΑ ΨΥΓΕΙΟΥ} → {ΓΑΛΑ ΦΡΕΣΚΟ}	41,103	35,800	-5,303	-12,90
{ΒΟΥΤΥΡΟ} → {ΓΑΛΑ ΦΡΕΣΚΟ}	47,785	42,484	-5,301	-11,09
{ΓΙΑΟΥΡΤΙΑ} → {ΓΑΛΑ ΦΡΕΣΚΟ}	45,620	40,359	-5,261	-11,53

Β'.5 Κοινοί κανόνες: αυξήσεις του lift

Ο Πίνακας Β'.5 παρουσιάζει τους δέκα κοινούς κανόνες με τη μεγαλύτερη αύξηση του lift από την περίοδο COVID-19 προς τη μεταπανδημική περίοδο. Η κατάταξη βασίστηκε στην αριθμητική διαφορά των δύο τιμών.

Πίνακας Β'.5: Κοινοί κανόνες με τις δέκα μεγαλύτερες αυξήσεις του lift

Κανόνας	Lift COVID	Lift Post-COVID	Δ Lift	Σχετική μεταβολή (%)
{ΑΛΕΥΡΙ-ΣΙΜΙΓΔΑΛΙ} → {ΕΙΔΗ ΖΑΧΑΡΟΠΛΑΣΤΙΚΗΣ}	4,947	6,138	+1,191	+24,08
{ΤΥΡΙΑ ΔΙΑΦΟΡΑ ΣΥΣΚ, ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ} → {ΑΛΛΑΝΤΙΚΑ ΣΥΣΚΕΥΑΣΜΕΝΑ ΨΥΓΕΙΟΥ}	5,102	5,884	+0,782	+15,32
{ΜΑΛΑΚΤΙΚΑ} → {ΠΡΟΙΟΝΤΑ ΠΛΥΣΙΜΑΤΟΣ}	4,071	4,673	+0,602	+14,79
{ΓΑΛΑ ΦΡΕΣΚΟ, ΤΥΡΙΑ ΔΙΑΦΟΡΑ ΣΥΣΚ} → {ΑΛΛΑΝΤΙΚΑ ΣΥΣΚΕΥΑΣΜΕΝΑ ΨΥΓΕΙΟΥ}	3,902	4,370	+0,467	+11,98
{ΛΑΧΑΝΙΚΑ ΣΕ ΚΟΝΣΕΡΒΑ-ΧΑΡΤΙΝΗ ΣΥΣΚΕΥΑΣΙΑ} → {ΖΥΜΑΡΙΚΑ}	3,375	3,799	+0,424	+12,57
{ΚΑΘΑΡΙΣΤΙΚΑ} → {ΠΡΟΙΟΝΤΑ ΠΛΥΣΙΜΑΤΟΣ}	3,253	3,651	+0,399	+12,26
{ΑΛΛΑΝΤΙΚΑ ΣΥΣΚΕΥΑΣΜΕΝΑ ΨΥΓΕΙΟΥ, ΤΥΡΙΑ ΔΙΑΦΟΡΑ ΣΥΣΚ} → {ΨΩΜΙΑ ΣΥΣΚΕΥΑΣΜΕΝΑ}	3,171	3,519	+0,348	+10,99
{ΑΛΛΑΝΤΙΚΑ ΣΥΣΚΕΥΑΣΜΕΝΑ ΨΥΓΕΙΟΥ} → {ΤΥΡΙΑ ΔΙΑΦΟΡΑ ΣΥΣΚ}	3,660	4,003	+0,343	+9,38
{ΦΕΤΑ/ΚΙΛΟ} → {ΜΑΛΑΚΑ ΤΥΡΙΑ/ΚΙΛΟ}	2,438	2,746	+0,308	+12,63
{ΣΑΠΟΥΝΙΑ} → {ΠΡΟΙΟΝΤΑ ΠΛΥΣΙΜΑΤΟΣ}	2,773	3,073	+0,299	+10,79

Β'.6 Κοινοί κανόνες: μειώσεις του lift

Ο Πίνακας Β'.6 παρουσιάζει τους δέκα κοινούς κανόνες με τη μεγαλύτερη μείωση του lift από την περίοδο COVID-19 προς τη μεταπανδημική περίοδο. Η κατάταξη βασίστηκε στην αριθμητική διαφορά των δύο τιμών.

Πίνακας Β'.6: Κοινοί κανόνες με τις δέκα μεγαλύτερες μειώσεις του lift

Κανόνας	Lift COVID	Lift Post-COVID	Δ Lift	Σχετική μεταβολή (%)
{ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ, ΦΡΕΣΚΑ ΦΡΟΥΤΑ} → {ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ}	3,266	3,203	-0,063	-1,93
{ΑΝΑΨΥΚΤΙΚΑ, ΣΟΚΟΛΑΤΕΣ} → {SNACK}	2,716	2,666	-0,050	-1,85
{ΚΡΑΣΙΑ} → {ΑΝΑΨΥΚΤΙΚΑ}	1,894	1,845	-0,049	-2,61
{SNACK, ΜΠΙΣΚΟΤΑ} → {ΣΟΚΟΛΑΤΕΣ}	2,569	2,519	-0,049	-1,91
{Κρεμμύδια κιλού.} → {ΦΡΕΣΚΑ ΦΡΟΥΤΑ}	2,196	2,154	-0,042	-1,91
{ΜΠΥΡΕΣ} → {ΑΝΑΨΥΚΤΙΚΑ}	2,026	1,986	-0,040	-2,00
{ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ} → {ΦΡΕΣΚΑ ΦΡΟΥΤΑ}	2,800	2,763	-0,037	-1,32
{ΕΞΩΤΙΚΑ ΦΡΟΥΤΑ} → {ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ}	2,034	2,013	-0,022	-1,07
{ΖΥΜΑΡΙΚΑ, ΛΑΧΑΝΙΚΑ ΠΡΑΣΙΝΑ ΠΟΙΚΙΛΙΑ} → {ΦΡΕΣΚΑ ΦΡΟΥΤΑ}	2,485	2,465	-0,020	-0,80
{ΞΗΡΟΙ ΚΑΡΠΟΙ ΣΥΣΚ/ΝΟΙ} → {SNACK}	2,211	2,193	-0,018	-0,84