



## ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ – ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Πρόγραμμα Μεταπτυχιακών Σπουδών  
«Κυβερνοασφάλεια και Επιστήμη Δεδομένων»

### Μεταπτυχιακή Διατριβή

Τίτλος Διατριβής	<b>Επιχειρηματική Αναλυτική στην Πράξη: Τμηματοποίηση Πελατειακής Βάσης και Ανάλυση Καλαθιού Αγορών για ένα ηλεκτρονικό κατάστημα μίας αλυσίδας στο χώρο του λιανεμπορίου</b> <b>Customer Analytics in Practice: Customer Segmentation and E-Shop Market Basket Analysis</b>
Όνοματεπώνυμο Φοιτητή	<b>Στεργιάννα Σιώλου</b>
Πατρώνυμο	<b>Γεώργιος</b>
Αριθμός Μητρώου	<b>ΜΠΚΕΔ21048</b>
Επιβλέπων	<b>Δημήτριος Αποστόλου, Καθηγητής</b>

Ημερομηνία Παράδοσης **Μάϊος 2024**

---

**Τριμελής Εξεταστική Επιτροπή**

Δημήτριος Αποστόλου  
Καθηγητής

Άγγελος Πικράκης  
Επίκουρος Καθηγητής

Διονύσιος Σωτηρόπουλος  
Επίκουρος Καθηγητής

## **Ευχαριστίες**

Θα ήθελα να εκφράσω την ειλικρινή μου εκτίμηση στον κ. **Ανδρέα Ζάρα**, Διδάσκοντα στο ΠΜΣ για την ιδιαίτερα σημαντική συμβολή του στην ολοκλήρωση της παρούσας μεταπτυχιακής διατριβής. Η επιστημονική και συμβουλευτική καθοδήγησή του σε όλα τα στάδια εκπόνησης της, από την έρευνα έως και τη συγγραφή, η αφοσίωσή του στην προώθηση της κριτικής σκέψης, της επεξηγηματικής προσέγγισης και της ακαδημαϊκής αριστείας έπαιξαν όλα καθοριστικό ρόλο στη διαμόρφωση της ποιότητας αυτής της εργασίας.

## ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ.....	6
ΕΙΣΑΓΩΓΗ.....	7
ΚΕΦΑΛΑΙΟ 1 – Τεχνητή νοημοσύνη και επιχειρήσεις .....	8
ΚΕΦΑΛΑΙΟ 2 – Customer Analytics.....	10
2.1 Τομείς εφαρμογής τεχνητής νοημοσύνης στο customer analytics.....	10
2.2 Οπτική ανάλυση.....	10
2.3 Το μοντέλο RFM.....	11
2.3.1 Γενική ανάλυση.....	11
2.3.2 Αλγόριθμοι Clustering στο RFM.....	13
2.3.3 Χρήση του μοντέλου RFM.....	17
2.4 Ανάλυση καλαθιού αγοράς.....	19
ΚΕΦΑΛΑΙΟ 3 – Εφαρμογή Αναλυτικής Πελατών σε πραγματικά επιχειρησιακά δεδομένα.....	20
3.1 Τα υπό ανάλυση δεδομένα.....	21
3.2 Αρχική εξερεύνηση δεδομένων .....	22
3.2.1 Όγκος συναλλαγών .....	22
3.2.2 Αξία συναλλαγών.....	26
3.2.3 Δημογραφικά χαρακτηριστικά και συναλλαγές.....	29
3.2.4 Συχνότητα εμφάνισης προϊόντων.....	33
3.2.5 Εξερεύνηση συσχετίσεων .....	37
3.3 Προετοιμασία δεδομένων .....	38
3.3.1 Εισαγωγή και μετατροπή βάσης δεδομένων .....	38
3.3.2 Μορφοποίηση και καθαρισμός δεδομένων .....	38
3.3.3 Ενοποίηση πινάκων.....	39
3.4 Εφαρμογή μοντέλου RFM.....	40
3.4.1 Υπολογισμός μετρικών RFM πελατών εταιρείας.....	40
3.4.2 Επεξεργασία αποτελεσμάτων RFM .....	40
3.4.3 Ταξινόμηση πελατών.....	44
3.4.4 Δημογραφική ανάλυση συστάδων πελατών .....	46
3.4.5 Ανάλυση αγορασμένων προϊόντων.....	49
3.5 Ανάλυση καλαθιού αγοράς.....	52
3.5.1 Προετοιμασία δεδομένων .....	53
3.5.2 Κανόνες συσχέτισης μεταξύ αγορασμένων προϊόντων .....	53
3.5.3 Κανόνες συσχέτισης μεταξύ αγορασμένων προϊόντων σε συστάδες πελατών .....	54
ΣΥΜΠΕΡΑΣΜΑΤΑ.....	57
ΠΑΡΑΡΤΗΜΑ.....	59
Εισαγωγή και μετατροπή βάσης δεδομένων (3.2.1) .....	59
Μορφοποίηση και καθαρισμός δεδομένων (3.2.2).....	59
Ενοποίηση πινάκων (3.2.3) .....	62
Υπολογισμός μετρικών RFM πελατών (3.3.1).....	62

Προετοιμασία δεδομένων (3.4.1).....	63
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	64

## ΠΕΡΙΛΗΨΗ

Η παρούσα εργασία αποσκοπεί στην εξερεύνηση τρόπων για την ενίσχυση των επιχειρηματικών λειτουργιών και των διαδικασιών λήψης αποφάσεων, μέσω της χρήσης Μεγάλων Δεδομένων και Τεχνητής Νοημοσύνης. Η εργασία επίσης εστιάζει στην εφαρμογή του μοντέλου RFM ως μέθοδο επιχειρηματικής αναλυτικής, που χρησιμοποιείται για την τμηματοποίηση των πελατών μίας επιχείρησης (customer segmentation) και την πρόβλεψη της μελλοντικής τους συμπεριφοράς. Τέλος, η εργασία ασχολείται με την μέθοδο Market Basket Analysis (MBA), η οποία δημιουργεί κανόνες συσχέτισης μεταξύ αγορασμένων προϊόντων από τους καταναλωτές, χρησιμοποιώντας ιστορικά δεδομένα αγορών. Για την εφαρμογή του μοντέλου RFM, της μεθόδου MBA και την εξαγωγή των αντίστοιχων αποτελεσμάτων και συμπερασμάτων, χρησιμοποιείται το λογισμικό SAS καθώς και ιστορικά δεδομένα της αλυσίδας ελληνικών καταστημάτων υπεραγοράς (super market) «Ατλάντικ», που δεν είναι πλέον ενεργή. Στο πρώτο κεφάλαιο, διερευνώνται οι πιθανές εφαρμογές των Business Intelligence και Business Analytics σε διάφορους τομείς της οικονομικής ζωής, με έμφαση στην Τεχνητή Νοημοσύνη. Στο δεύτερο κεφάλαιο, εξετάζονται οι εφαρμογές της Τεχνητής Νοημοσύνης στους τομείς του Customer Analytics, όπως την οπτική ανάλυση, την προγνωστική μοντελοποίηση και το customer segmentation. Περιγράφεται λεπτομερώς το μοντέλο RFM (Recency, Frequency and Monetary) και τα βασικά στοιχεία του, όπως το μοντέλο βαθμολόγησής του, ο σκοπός του και ο πιθανός συνδυασμός του με άλλες τεχνικές για τη βελτίωση των αποτελεσμάτων. Επίσης, περιγράφεται η μέθοδος MBA σε ότι αφορά το σκοπό της, τους αλγορίθμους που χρησιμοποιεί και τις έννοιες της υποστήριξης και εμπιστοσύνης, που παίζουν κεντρικό ρόλο σε αυτή. Στο τελευταίο κεφάλαιο, παρουσιάζεται μια μελέτη περίπτωσης για να καταδειχθεί πώς οι τεχνικές που περιγράφονται μπορούν να εφαρμοστούν πρακτικά στην ανάλυση αξίας πελατών (Customer Value) με τη χρήση του μοντέλου RFM, καθώς και στην ανάλυση καλαθιού αγοράς με χρήση της μεθόδου MBA. Πριν την εφαρμογή των τεχνικών, επιχειρείται μία βασική εξερεύνηση των δεδομένων της ανωτέρω επιχείρησης, με στόχο να καταδειχθεί η αξία της οπτικής ανάλυσης ως ενός απλού αλλά παράλληλα ισχυρού εργαλείου στον τομέα της επιχειρηματικής αναλυτικής. Η εργασία καταλήγει μέσα από την ανάλυση σε πολυάριθμα χρήσιμα αποτελέσματα και παρέχει δυνητικές προτάσεις για την αξιοποίησή τους. Στο παράρτημα της εργασίας, περιλαμβάνεται ο κώδικας που χρησιμοποιήθηκε εντός της εφαρμογής SAS, για την επεξεργασία των δεδομένων και τις διάφορες αναλύσεις.

## ABSTRACT

This paper aims to explore ways to enhance business operations and decision-making processes through the use of Big Data and Artificial Intelligence. The paper also focuses on the application of the RFM model as a business analytics method, used for customer segmentation and for predicting customers' future behavior. Finally, the paper discusses the Market Basket Analysis (MBA) method, which generates association rules between purchased products by consumers, using historical purchase data. For the application of the RFM model, the MBA method and the extraction of the corresponding results and conclusions, SAS software is used as well as historical data of the Greek supermarket chain "Atlantic", which is no longer active. In the first chapter, the potential applications of Business Intelligence and Business Analytics in various areas of economic life are explored, with an emphasis on Artificial Intelligence (AI). In the second chapter, the applications of AI in the areas of Customer Analytics, such as visual analytics, predictive modelling and customer segmentation, are examined. The RFM (Recency, Frequency and Monetary) model and its key elements, such as its scoring model, its purpose and its potential combination with other techniques to improve results, are described in detail. The MBA method in terms of its purpose, the algorithms used and the concepts of support and trust, which play a central role in it, are also described. In the last chapter, a case study is presented to demonstrate how the techniques described can be practically applied to Customer Value analysis using the RFM model, as well as to market basket analysis using the MBA method. Before applying the techniques, a basic exploration of the company's data is attempted, with the aim of demonstrating the value of visual analysis as a simple yet powerful tool in the field of business analytics. The paper concludes through the analysis to numerous useful results and provides potential recommendations for their use. In the appendix of the paper, the code used within the SAS application for data processing and various analyses is included.

## ΕΙΣΑΓΩΓΗ

Ζούμε σε έναν κόσμο πλημμυρισμένο από δεδομένα. Τα δεδομένα παράγονται σε όλους τους τομείς της ανθρώπινης δραστηριότητας. Κάθε φορά που κάποιος βλέπει ένα βίντεο, κάνει like σε ένα post, αγοράζει ένα προϊόν ή κάνει μία βιντεοκλήση μέσω διαδικτύου, παράγονται δεδομένα. Η πολυπλοκότητα και κυριότερα ο όγκος των δεδομένων αυτών, οδηγεί στην παραγωγή των μεγάλων δεδομένων (big data). Με τον σημερινό ρυθμό, δημιουργούνται καθημερινά 2,5 πεντάκις εκατομμύρια bytes δεδομένων, αλλά ο ρυθμός αυτός επιταχύνεται με την ανάπτυξη του Διαδικτύου των Πραγμάτων (IoT). Μόνο τα τελευταία δύο χρόνια δημιουργήθηκε το 90% των δεδομένων στον κόσμο.<sup>1</sup>

Ο όρος Μεγάλα δεδομένα ή Μεγα-δεδομένα (Big data) χρησιμοποιείται για να περιγράψει:

- α) σύνολα δεδομένων τόσο μεγάλα ή σύνθετα που ξεφεύγουν από τις δυνατότητες καταγραφής, αποθήκευσης και ανάλυσης των παραδοσιακών τεχνικών επεξεργασίας δεδομένων.<sup>2</sup>
- β) το πεδίο της επιστήμης της πληροφορικής που ασχολείται με τα μεγάλα δεδομένα.<sup>3</sup>

Τα μεγάλα δεδομένα συνδέθηκαν αρχικά με τρεις βασικές έννοιες: όγκος (volume), ποικιλία (variety) και ταχύτητα (velocity). Αργότερα, προστέθηκαν σε αυτές δύο επιπρόσθετες έννοιες: η ειλικρίνεια (veracity)<sup>4</sup> και η μεταβλητότητα (variability).<sup>5</sup> Οι ιδιότητες αυτές των μεγάλων δεδομένων (Volume, Variety, Velocity, Veracity, Value), κατέληξαν να αναφέρονται στη διεθνή κοινότητα ως "three Vs", "four Vs" και "five Vs".<sup>7</sup>

Οι οργανισμοί συλλέγουν και αναλύουν μεγάλα δεδομένα με στόχο να βελτιώσουν τις διαδικασίες τους, αλλά και τις αποφάσεις που λαμβάνουν.<sup>8</sup> Η αξιοποίηση των δεδομένων αυτών γίνεται μέσω του business intelligence και των business analytics.

Το Business intelligence (BI) συνδυάζει τη συλλογή δεδομένων, την αποθήκευση δεδομένων και τη διαχείριση γνώσεων με την ανάλυση για την παροχή input στη διαδικασία λήψης αποφάσεων. Το Business intelligence δίνει έμφαση στην ανάλυση μεγάλου όγκου δεδομένων, σχετικά με την επιχείρηση και τις δραστηριότητές της. Οι αναλύσεις κυμαίνονται από απλό reporting έως slice-and-dice, drill down, απάντηση σε ad hoc queries, real-time analysis και forecasting.<sup>9</sup>

Η επιχειρηματική αναλυτική (Business Analytics) είναι η διαδικασία μετατροπής δεδομένων σε πληροφορίες για τη βελτίωση των επιχειρηματικών αποφάσεων. Η διαχείριση δεδομένων, η οπτικοποίηση δεδομένων, η προγνωστική μοντελοποίηση, η εξόρυξη δεδομένων, η προσομοίωση προβλέψεων και η βελτιστοποίηση είναι μερικά από τα εργαλεία που χρησιμοποιούνται για τη δημιουργία πληροφοριών από δεδομένα. Ωστόσο, ενώ η επιχειρηματική αναλυτική στηρίζεται σε μεγάλο βαθμό στη στατιστική, ποσοτική και επιχειρησιακή ανάλυση, η ανάπτυξη οπτικοποιήσεων δεδομένων για την παρουσίαση των ευρημάτων και η διαμόρφωση επιχειρηματικών αποφάσεων είναι το τελικό αποτέλεσμα.

Στον πυρήνα της, η επιχειρηματική αναλυτική περιλαμβάνει έναν συνδυασμό των εξής:

- τον εντοπισμό νέων μοτίβων και σχέσεων με την εξόρυξη δεδομένων,
- τη χρήση ποσοτικής και στατιστικής ανάλυσης για τον σχεδιασμό επιχειρηματικών μοντέλων,
- διεξαγωγή δοκιμών A/B και δοκιμών πολλαπλών μεταβλητών με βάση τα ευρήματα,
- πρόβλεψη μελλοντικών επιχειρηματικών αναγκών, επιδόσεων και τάσεων του κλάδου με προγνωστικά μοντέλα- και
- κοινοποίηση των ευρημάτων σε ευκατανόητα reports σε συναδέλφους, διοίκηση και πελάτες.<sup>10</sup>

Προκειμένου να μπορέσουν όμως οι επιχειρήσεις, να εφαρμόσουν τις μεθοδολογίες που αφορούν το Business Intelligence και το Business Analytics, θα πρέπει να αξιοποιήσουν έναν άλλον, επίσης μεγάλο κλάδο της πληροφορικής: την Τεχνητή Νοημοσύνη. Ο κλάδος αυτός, ασχολείται με τη σχεδίαση και την υλοποίηση υπολογιστικών συστημάτων που μιμούνται στοιχεία της ανθρώπινης συμπεριφοράς τα οποία υπονοούν έστω και στοιχειώδη ευφυΐα: μάθηση, προσαρμοστικότητα, εξαγωγή συμπερασμάτων, κατανόηση από συμφραζόμενα, επίλυση προβλημάτων κλπ.<sup>11</sup>

Σκοπός της παρούσας μεταπτυχιακής εργασίας είναι να αναδείξει τομείς όπου η τεχνητή νοημοσύνη μπορεί να αξιοποιηθεί από τις επιχειρήσεις για τη βελτίωση των επιχειρηματικών τους δραστηριοτήτων και την αύξηση των εσόδων τους. Αυτό το επιτυγχάνει στο πρώτο κεφάλαιο, με την παρουσίαση μίας μη εξαντλητικής λίστας από τέτοιους τομείς, η οποία δημιουργήθηκε μέσω ανασκόπησης της υπάρχουσας βιβλιογραφίας. Ιδιαίτερη έμφαση δόθηκε στον τομέα του customer analytics και πιο συγκεκριμένα στη χρήση του μοντέλου RFM και του Market Basket Analysis ως μεθόδων επιχειρηματικής αναλυτικής, καθώς και της οπτικής ανάλυσης ως μέθοδο παρουσίασης των αποτελεσμάτων της. Προκειμένου να τεθεί το πλαίσιο εντός του οποίου αυτές οι μέθοδοι μπορούν να χρησιμοποιηθούν και να αποτυπωθούν δομικά

στοιχεία αυτών στο δεύτερο κεφάλαιο, έγινε χρησιμοποιήθηκε και πάλι ανασκόπηση της υπάρχουσας βιβλιογραφίας.

Παράλληλα, σκοπός της εργασίας είναι να αποδείξει ότι τα εργαλεία και οι μεθοδολογίες που αναφέρθηκαν στα προηγούμενα κεφάλαια, πράγματι λειτουργούν με ευεργετικά αποτελέσματα για τις επιχειρήσεις. Κάτι τέτοιο επιτυγχάνεται στο τρίτο κεφάλαιο, μέσω της εφαρμογής μεθόδων επιχειρηματικής αναλυτικής σε πραγματικά δεδομένα μία αληθινής επιχείρησης λιανεμπορίου. Πιο συγκεκριμένα, η εργασία χρησιμοποιεί την οπτική ανάλυση ως εργαλείο για την αποτύπωση των αποτελεσμάτων της αρχικής εξερεύνησης των δεδομένων της επιχείρησης. Έπειτα, υπολογίζονται οι μετρικές των μεταβλητών της μεθόδου RFM και οι πελάτες της επιχείρησης συσταδοποιούνται βάσει των αποτελεσμάτων. Για τον υπολογισμό των μετρικών RFM των πελατών, το recency εξήχθη από τη χρονοσφραγίδα της πιο πρόσφατης συναλλαγής του, το frequency από το άθροισμα των συναλλαγών του και το monetary από το συνολικό ποσό που έχει δαπανήσει. Χρησιμοποιώντας τα αποτελέσματα, οι πελάτες συσταδοποιήθηκαν με τη χρήση του αλγορίθμου k-means, ενώ για την εύρεση του αρχικού αριθμού συστάδων χρησιμοποιήθηκε ο αλγόριθμος Aligned Box Criterion (ABC) με κριτήριο εκτίμησης το Global Peak Value. Στη συνέχεια, έγινε ξεχωριστή δημογραφική ανάλυση και ανάλυση αγορασμένων προϊόντων για την εκάστοτε συστάδα πελατών. Προκειμένου να εξαχθούν κανόνες συσχέτισης μεταξύ προϊόντων, χρησιμοποιήθηκε η τεχνική του Market Basket Analysis. Ως ελάχιστο ποσοστό υποστήριξης (support), εφαρμόστηκε κατώφλι 0.05, ενώ ως ελάχιστο ποσοστό εμπιστοσύνης (confidence), εφαρμόστηκε κατώφλι 0.1. Η επιλογή των τιμών των δεικτών έγινε με βάση trial and error δοκιμές, προκειμένου να εξαχθεί αντιπροσωπευτικός αριθμός κανόνων συσχετίσεων, ικανών να παρέχουν χρήσιμη πληροφορία. Η εξαγωγή κανόνων συσχετίσεων επεκτάθηκε και ειδικότερα για κάθε μία από τις δημιουργηθείσες συστάδες πελατών. Η εφαρμογή όλων των μεθολογιών έγινε μέσω της πλατφόρμας επιχειρηματικής αναλυτικής SAS και τα αποτελέσματα των αναλύσεων αποτυπώθηκαν με τη χρήση γνωστών γραφημάτων οπτικής ανάλυσης.

Η εργασία τελικά επικυρώνει την αξία της οπτικής ανάλυσης, του μοντέλου RFM και του Market Basket Analysis ως μεθόδων επιχειρηματικής αναλυτικής που μπορούν να προσφέρουν ανταγωνιστικό πλεονέκτημα στην εκάστοτε επιχείρηση. Για να το επιτύχει αυτό, προτείνει δυνητικές χρήσεις των αποτελεσμάτων της ανάλυσης σε κάθε βήμα της.

## ΚΕΦΑΛΑΙΟ 1 – Τεχνητή νοημοσύνη και επιχειρήσεις

Η τεχνητή νοημοσύνη, μεγιστοποιεί τις δυνατότητες που έχουν οι επιχειρήσεις και οι οργανισμοί, για την υποστήριξη αποφάσεων και έχει εφαρμογές σε πολλούς τομείς. Ενδεικτικά, αναφέρονται οι παρακάτω:

- Finance:
  - Risk management: Ανάλυση μεγάλου όγκου δεδομένων και εντοπισμός μοτίβων και τάσεων, διευκολύνοντας την πρόβλεψη και τη διαχείριση των κινδύνων.<sup>12</sup>
  - Portfolio management<sup>13</sup>: Ανάλυση δεδομένων της αγοράς και εμφάνιση προτάσεων βέλτιστων χαρτοφυλακίων, που ευθυγραμμίζονται με τους στόχους και την ανοχή κινδύνου ενός επενδυτή.
  - Fraud detection: Εντοπισμός απάτης σε συναλλαγές και δραστηριότητες, όπως η νομιμοποίηση εσόδων από παράνομες δραστηριότητες και η κλοπή ταυτότητας.<sup>14</sup>
  - Credit risk/scoring: Αξιολόγηση της πιστοληπτικής ικανότητας των αιτούντων δανείου και λήψη αποφάσεων δανειοδότησης.<sup>15</sup>
  - Algorithmic trading<sup>16</sup>: Εκτέλεση συναλλαγών σε υψηλές ταχύτητες και λήψη αποφάσεων με βάση τα δεδομένα της αγοράς σε πραγματικό χρόνο.
  - Customer service: Δημιουργία εικονικών βοηθών που μπορούν να βοηθήσουν τους πελάτες στις οικονομικές τους ερωτήσεις και να τους παρέχουν εξατομικευμένες συμβουλές.<sup>17</sup>
  - Insurance: Διαχείριση των απαιτήσεων για αποζημιώσεις, καθορισμός των ασφαλιστικών τιμών και σύναψη συμβολαίων.<sup>18</sup>
- Operations:
  - Predictive pricing<sup>19</sup>: Πρόβλεψη της ζήτησης των πελατών και προσαρμογή των τιμών σε πραγματικό χρόνο, μεγιστοποιώντας τα έσοδα.



- Predictive maintenance: Ανάλυση δεδομένων αισθητήρων για την πρόβλεψη πότε ένα μηχάνημα είναι πιθανό να αποτύχει, επιτρέποντας την προληπτική συντήρηση και τη μείωση του χρόνου διακοπής λειτουργίας.<sup>20</sup>
  - Supply chain optimization: Βελτιστοποίηση των χρονοδιαγραμμάτων παραγωγής, διαχείριση των επιπέδων αποθεμάτων και πρόβλεψη της ζήτησης για τη βελτίωση της αποτελεσματικότητας της αλυσίδας εφοδιασμού.<sup>21</sup>
  - Quality control: Επιθεώρηση προϊόντων και εντοπισμός ελαττωμάτων, μειώνοντας τη σπατάλη και βελτιώνοντας την ποιότητα των προϊόντων.<sup>22</sup>
  - Workforce management: Προγραμματισμός της απασχόλησης των εργαζομένων και κατανομή των πόρων, βελτιώνοντας την παραγωγικότητα και μειώνοντας το κόστος εργασίας.
  - Process automation: Αυτοματοποίηση επαναλαμβανομένων εργασιών, όπως η καταχώρηση δεδομένων, απελευθερώνοντας εργαζόμενους για πιο πολύτιμες εργασίες.<sup>23</sup>
  - Inventory management: Βελτιστοποίηση των επιπέδων αποθεμάτων, μειώνοντας τη σπατάλη και βελτιώνοντας την ανταπόκριση στη ζήτηση των πελατών.<sup>24</sup>
  - Logistics optimization: Βελτιστοποίηση των διαδρομών και των χρονοδιαγραμμάτων παράδοσης, μειώνοντας το κόστος μεταφοράς και βελτιώνοντας τους χρόνους παράδοσης.<sup>25</sup>
- HR:
    - Recruitment: Αυτοματοποίηση και βελτιστοποίηση της διαδικασίας πρόσληψης, συμπεριλαμβανομένης της screening διαδικασίας βιογραφικών σημειωμάτων, του προγραμματισμού συνεντεύξεων και της αξιολόγησης υποψηφίων.<sup>26</sup>
    - Performance evaluation: Παρακολούθηση των επιδόσεων των εργαζομένων και παροχή πληροφόρησης βάσει δεδομένων για τη διαχείριση των επιδόσεων.
    - Employee engagement: Μέτρηση και βελτίωση της συμμετοχής των εργαζομένων με την ανάλυση δεδομένων ανατροφοδότησης και συναισθημάτων.<sup>27</sup>
    - Talent management: Εντοπισμός εργαζομένων με υψηλό δυναμικό και προβολή προτάσεων για ευκαιρίες ανάπτυξης.<sup>28</sup>
    - Learning and development: Εξατομίκευση των προγραμμάτων μάθησης και ανάπτυξης, με βάση τις ατομικές ανάγκες και τις επιδόσεις των εργαζομένων.
    - Diversity and inclusion: Εντοπισμός και αντιμετώπιση ασυνείδητων προκαταλήψεων σε διαδικασίες ανθρώπινου δυναμικού, όπως οι προσλήψεις και οι προαγωγές.
    - Compensation and benefits: Ανάλυση δεδομένων σχετικά με τις αποζημιώσεις και τις παροχές, ώστε να λαμβάνονται τεκμηριωμένες αποφάσεις και να διασφαλίζεται η ισότητα.
  - Δημόσιος τομέας:
    - Fraud detection: Εντοπισμός απάτης και διαφθοράς σε κυβερνητικά προγράμματα και πρωτοβουλίες.<sup>29</sup> Περίπου 10% του κρατικού προϋπολογισμού χάνεται σε fraud, waste και power abuse στις ΗΠΑ.<sup>30</sup>
    - Service delivery: Βελτίωση της παροχής υπηρεσιών με την αυτοματοποίηση καθηκόντων ρουτίνας, όπως η επεξεργασία αιτήσεων, και η παροχή πληροφοριών σε πραγματικό χρόνο στους πολίτες.
    - Decision making: Υποστήριξη της λήψης αποφάσεων με βάση τα αποδεικτικά στοιχεία, αναλύοντας μεγάλες ποσότητες δεδομένων και παρέχοντας πληροφορίες για τη δημόσια πολιτική.<sup>31</sup>
    - Emergency response: Βελτιστοποίηση των χρόνων απόκρισης σε καταστάσεις έκτακτης ανάγκης και κατανομή των πόρων, βελτιώνοντας τη δημόσια ασφάλεια.<sup>32</sup>
    - Predictive maintenance: Πρόβλεψη του πότε τα δημόσια περιουσιακά στοιχεία, όπως δρόμοι, γέφυρες και κτίρια, θα χρειαστούν συντήρηση, βελτιώνοντας τη διαχείριση των περιουσιακών στοιχείων και μειώνοντας το κόστος.
    - Crime prevention: Ανάλυση δεδομένων εγκληματικότητας και πρόβλεψη των σημείων όπου είναι πιθανό να συμβούν εγκλήματα, επιτρέποντας την προληπτική αστυνόμευση και τη μείωση των ποσοστών εγκληματικότητας.<sup>33</sup>
    - Traffic management: Βελτιστοποίηση της ροής της κυκλοφορίας και μείωση της συμφόρησης, βελτιώνοντας τις μεταφορές και μειώνοντας τις εκπομπές ρύπων.<sup>34</sup>

- Healthcare: Υποστήριξη πρωτοβουλιών δημόσιας υγείας με την ανάλυση δεδομένων υγείας, όπως μοτίβα ασθενειών, και την παροχή πληροφοριών σχετικά με την υγεία του πληθυσμού.<sup>35</sup>

## ΚΕΦΑΛΑΙΟ 2 – Customer Analytics

### 2.1 Τομείς εφαρμογής τεχνητής νοημοσύνης στο customer analytics

Σε ότι αφορά το business analytics, υπάρχουν πάρα πολλές εφαρμογές της τεχνητής νοημοσύνης και στα customer analytics (marketing και sales). Μερικοί από τους τομείς εφαρμογής είναι:

- Predictive modeling: Ανάλυση των δεδομένων των πελατών και πρόβλεψη της συμπεριφοράς τους, όπως τα αγοραστικά μοτίβα, το ρίσκο/πρόβλεψη churn και η αξία διάρκειας ζωής.<sup>36</sup>
- Customer segmentation: Τμηματοποίηση των πελατών με βάση τα δημογραφικά στοιχεία, τη συμπεριφορά και τις προτιμήσεις, επιτρέποντας εξατομικευμένο μάρκετινγκ και βελτιωμένη εμπειρία πελατών.<sup>37</sup>
- Customer service: Παροχή υποστήριξης πελατών σε πραγματικό χρόνο μέσω chatbots και εικονικών βοηθών, βελτιώνοντας την ικανοποίηση των πελατών και μειώνοντας το κόστος.<sup>38</sup>
- Marketing optimization: Βελτιστοποίηση εκστρατειών μάρκετινγκ με την πρόβλεψη της ανταπόκρισης των πελατών και την επιλογή των πιο αποτελεσματικών καναλιών, μηνυμάτων και προσφορών.<sup>39</sup> Για παράδειγμα, επιλογή βέλτιστου ύψους επέκτασης πιστωτικού ορίου (credit line) για κάθε πελάτη και προσφορά μέσω βέλτιστων καναλιών επικοινωνίας (personal bank adviser, local branch, email, call center, direct mail, website banner ad etc)<sup>40</sup>
- Customer feedback analysis: Ανάλυση της ανατροφοδότησης των πελατών, όπως οι αναρτήσεις και οι κριτικές στα μέσα κοινωνικής δικτύωσης, για να αποκτηθούν πληροφορίες σχετικά με το συναίσθημα και την εμπειρία των πελατών. Η ανάλυση των κριτικών δεν σταματάει στην απλή ανάλυση τυπικών λέξεων, αλλά επεκτείνεται ακόμα και στην ανάλυση μη-γραμματικών λέξεων (slang) και διαδικτυακή ορολογία (internet jargon).<sup>41</sup>
- Fraud detection: Εντοπισμός απάτης στις χρηματοπιστωτικές συναλλαγές και πρόληψη δόλιων δραστηριοτήτων, βελτιώνοντας την ασφάλεια και μειώνοντας τις απώλειες. Οι εταιρείες χρησιμοποιούν τεχνικές όπως anomaly detection, automated red flags, data visualization, predictive analytics και social network analysis, ως τα βασικά τους εργαλεία για το σκοπό αυτό.<sup>42</sup> Ωστόσο, οι τεχνολογίες που χρησιμοποιούνται πρέπει να μειώνουν όσο το δυνατόν πιθανά false negatives που μπορεί να οδηγήσουν σε απώλεια κερδών, αντί για το αντίστροφο.<sup>43</sup>
- Recommendation systems: Παροχή εξατομικευμένων συστάσεων στους πελάτες με βάση την προηγούμενη συμπεριφορά και τις προτιμήσεις τους, βελτιώνοντας τη δέσμευση και την αφοσίωση των πελατών.<sup>44</sup> Χρησιμοποιώντας μια αυτοματοποιημένη cross-selling προσέγγιση, δίνεται η δυνατότητα στους λιανοπωλητές να προσθέσουν αξία στα προϊόντα που έχουν ήδη επιλέξει οι πελάτες και τελικά να αυξήσουν τις πωλήσεις.<sup>45</sup>
- Pricing optimization: Πρόβλεψη της ζήτησης των πελατών και προσαρμογή των τιμών σε πραγματικό χρόνο, μεγιστοποιώντας τα έσοδα και την κερδοφορία.<sup>46</sup> Η βελτιστοποίηση της τιμής προσφοράς, μπορεί να επιφέρει και έμμεσα κέρδη κατά τη διάρκεια του κύκλου ζωής ενός προϊόντος, όπως μείωση μεταφερόμενου αποθέματος, μέσω προωθητικών ενεργειών ή εκκαθαρίσεων (clearances).<sup>47</sup>
- Market basket analysis: Αναγνώριση σχέσεων μεταξύ προϊόντων, όπως ποια προϊόντα αγοράζονται συνήθως μαζί, επιτρέποντας ευκαιρίες cross-selling και up-selling, καθώς και τη βελτίωση της τοποθέτησης των προϊόντων στα ράφια και της διαχείρισης των αποθεμάτων.<sup>48</sup>

### 2.2 Οπτική ανάλυση

Η οπτική ανάλυση (visual analytics), είναι ένας τομέας της επιστήμης των υπολογιστών και της πληροφορικής, που συνδυάζει την ανάλυση δεδομένων, με την οπτικοποίηση πληροφοριών, για την αντιμετώπιση πολύπλοκων προβλημάτων και την ανακάλυψη κρυμμένων μοτίβων ή σχέσεων. Βασίζεται στην ιδέα, ότι οι άνθρωποι επεξεργάζονται πληροφορίες καλύτερα και κατανοούν πιο εύκολα μεγάλα

σύνολα δεδομένων, όταν παρουσιάζονται με οπτικά εργαλεία όπως γραφικά, γραφήματα, διαγράμματα, χάρτες, εικόνες και άλλες απεικονίσεις ή οπτικά στοιχεία.

Ο στόχος της οπτικής ανάλυσης είναι να επιτρέψει στους χρήστες να εξερευνήσουν τα δεδομένα, να ανακαλύψουν συσχετίσεις, τάσεις και αποκλίσεις, να προβλέψουν μοτίβα και να εξάγουν συμπεράσματα. Οι αναλυτές χρησιμοποιούν εργαλεία οπτικής ανάλυσης, για να αλληλεπιδράσουν με τα δεδομένα, να δημιουργήσουν οπτικές αναπαραστάσεις των πληροφοριών και να εφαρμόσουν προηγμένες τεχνικές αναζήτησης και φιλτραρίσματος για να εξερευνήσουν τα δεδομένα πιο αποτελεσματικά και πιο γρήγορα.

Η οπτική ανάλυση, μπορεί να χρησιμοποιηθεί σε πολλούς τομείς και εφαρμογές. Στην επιχειρηματική ανάλυση, μπορεί να βοηθήσει στην ανακάλυψη κρυφών τάσεων αγοράς, στην ανίχνευση ανωμαλιών σε δεδομένα πωλήσεων ή στην παρακολούθηση της απόδοσης μιας επιχείρησης. Στα πλαίσια του customer analytics, η οπτική ανάλυση, χρησιμοποιείται για να κατανοήσει και να αναδείξει τα πρότυπα, τις τάσεις και τις συμπεριφορές των πελατών μιας επιχείρησης.

Οι πληροφορίες που συλλέγονται από διάφορες πηγές, όπως οι συναλλαγές πελατών, οι αξιολογήσεις, οι διαδικτυακές δραστηριότητες και άλλες πληροφορίες, αναλύονται και παρουσιάζονται με γραφικά και γραφήματα που είναι πιο προσιτά στην ανθρώπινη αντίληψη. Αυτό βοηθά τους αναλυτές να αντλήσουν ευκαιρίες, να εντοπίσουν προκλήσεις και να λάβουν αποφάσεις που βασίζονται σε δεδομένα.

Με τη χρήση της οπτικής ανάλυσης στον τομέα του customer analytics, μπορεί να επιτευχθούν τα εξής:

1. Αναγνώριση προτύπων: Η οπτική ανάλυση επιτρέπει στους αναλυτές να αναγνωρίσουν πρότυπα και τάσεις στις αγοραπωλησίες και τη συμπεριφορά των πελατών. Μπορούν να ανακαλύψουν ποια προϊόντα ή υπηρεσίες είναι πιο δημοφιλή, ποιοι πελάτες εμφανίζουν συγκεκριμένες συμπεριφορές αγοράς και πολλά άλλα. Για παράδειγμα, μπορούν να εξαχθούν πληροφορίες, όπως που γίνονται οι διάφορες συναλλαγές (π.χ. υποκατάστημα), πόσο συχνά και τι είδους συναλλαγές είναι.<sup>49</sup>
2. Ανάλυση συμπεριφοράς πελατών: Με την οπτική ανάλυση, μπορούν να δημιουργηθούν γραφήματα και αναφορές που απεικονίζουν τη συμπεριφορά των πελατών. Αυτό μπορεί να περιλαμβάνει τις αγορές που έχουν πραγματοποιήσει, τις προτιμήσεις τους, τη συχνότητα των αγορών και άλλα χαρακτηριστικά. Για παράδειγμα, σε ότι αφορά τον τομέα ασφαλίσεων ζωής, μπορεί να ανακαλύψει ανωμαλίες συμπεριφοράς, να παρέχει συστήματα υποστήριξης αποφάσεων και να οδηγήσει σε βελτιωμένες στρατηγικές διαχείρισης κινδύνου και αξιολόγησης αποζημιώσεων για την επιχείρηση ασφαλίσεων.<sup>50</sup>
3. Παρακολούθηση της απόδοσης: Με την οπτική ανάλυση, οι επιχειρήσεις μπορούν να παρακολουθούν και να αξιολογούν την απόδοση των καμπανιών μάρκετινγκ και των προϊόντων. Μπορούν να ανιχνεύσουν πώς οι πελάτες αντιδρούν σε συγκεκριμένες προσφορές ή προωθητικές ενέργειες και να προσαρμόσουν τις στρατηγικές τους ανάλογα.<sup>51 52</sup>
4. Ανάλυση κριτικών και συναισθήματος: Η οπτική ανάλυση επιτρέπει την γρήγορη απεικόνιση αποτελεσμάτων ανάλυσης των συναισθημάτων (sentiment analysis) και των κριτικών (reviews) των πελατών, σχετικά με την επιχείρηση.<sup>53 54</sup>

## 2.3 Το μοντέλο RFM

### 2.3.1 Γενική ανάλυση

Το μοντέλο RFM (Recency, Frequency and Monetary) έχει εφαρμοστεί ευρέως σε πολλούς επιχειρηματικούς τομείς, ιδίως στο άμεσο marketing. Με την υιοθέτηση του μοντέλου RFM, οι υπεύθυνοι λήψης αποφάσεων μπορούν να εντοπίσουν πολύτιμους πελάτες και στη συνέχεια να αναπτύξουν αποτελεσματική στρατηγική μάρκετινγκ.<sup>55</sup> Με λίγα λόγια, αποτελεί ένα εργαλείο customer analytics.

Η έννοια του RFM εισήχθη από τους Bult και Wansbeek το 1995.<sup>56</sup> Το RFM βασίζεται στη σκέψη ότι η προσέλκυση πελατών είναι ένα σημαντικό έργο, αλλά η διατήρηση πελατών είναι πιο σημαντική, δεδομένου ότι η απώλεια ενός πελάτη σημαίνει την απώλεια ολόκληρης της ροής των αγορών που ο πελάτης θα πραγματοποιούσε κατά τη διάρκεια της ζωής του.<sup>57</sup> Δεδομένου ότι δεν είναι όλοι οι πελάτες εξίσου ελκυστικοί οικονομικά για την εταιρεία, είναι εξαιρετικά σημαντικό να προσδιοριστεί πρώτα η κερδοφορία τους και στη συνέχεια να αναπτυχθούν οι πόροι στους πελάτες, σύμφωνα με τις αξίες τους.<sup>58</sup>

Καθώς τα αρχεία συναλλαγών μιας εταιρείας γίνονται πολύ μεγαλύτερα σε μέγεθος, είναι απαραίτητο να χωριστούν όλοι οι πελάτες σε κατάλληλο αριθμό ομάδων (clusters), με βάση κάποιες ομοιότητες τους για την τμηματοποίηση της αγοράς (customer segmentation).<sup>59</sup>

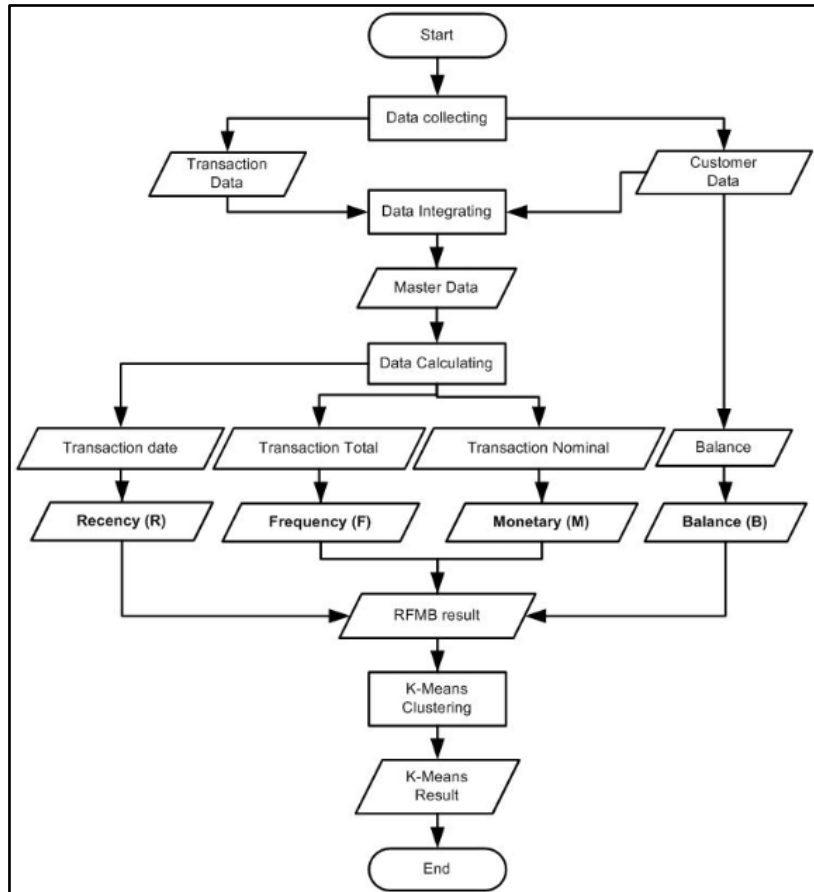
Οι ομοιότητες που χρησιμοποιούνται για την τμηματοποίηση αυτή, μπορεί να αφορούν διάφορα χαρακτηριστικά των πελατών. Το μοντέλο RFM χρησιμοποιεί τη συμπεριφορά των πελατών, ως κριτήριο τμηματοποίησής τους. Για να το επιτύχει αυτό, το RFM χρησιμοποιεί ένα scoring model που αποτελείται από τρία κριτήρια: πόσο πρόσφατα έχει αγοράσει ένας πελάτης (R - Recency), ποσό συχνά αγοράζει ένας πελάτης (F - Frequency) και πόσα ξοδεύει ένας πελάτης (M - Monetary).

Έπειτα, τα κριτήρια αυτά συνδυάζονται προκειμένου να τμηματοποιήσουν τους πελάτες, είτε με τον παραδοσιακό τρόπο (traditional), είτε με βάση τεχνικές εξόρυξης δεδομένων.

Κατά τον παραδοσιακό τρόπο, τα κριτήρια αυτά συνδυάζονται σε έναν τριψήφιο κωδικό κυψέλης RFM, που καλύπτει πέντε ίσα πεμπτημώρια (ομάδες 20%). Υπάρχουν πολλές προσεγγίσεις για τον ποσοτικό προσδιορισμό των τιμών RFM και η καλύτερη προσέγγιση εξαρτάται από το λεγόμενο “customer journey” και το επιχειρηματικό μοντέλο της εκάστοτε εταιρείας. Η βασική διαδικασία που ακολουθείται όμως, ακολουθεί τα εξής «μαθηματικά» βήματα:<sup>60</sup>

1. Οι πελάτες ταξινομούνται σε μια βάση δεδομένων με βάση καθεμία από τις τρεις διαστάσεις (συχνότητα, συχνότητα και χρηματικό ποσό) και χωρίζονται σε πέντε ίσες ομάδες (πεμπτημώρια). Με αυτόν τον τρόπο, γίνεται μέτρηση πόσο καλός είναι ο κάθε πελάτης σε σύγκριση με τους άλλους.
2. Κάθε πελάτης λαμβάνει μια βαθμολογία από το 1 έως το 5 με βάση τη θέση του σε καθεμία από τις τρεις διαστάσεις. Για το recency, οι πελάτες που αγόρασαν κάτι πιο πρόσφατα, λαμβάνουν υψηλότερη βαθμολογία. Για το frequency, οι πελάτες που αγοράζουν πράγματα πιο συχνά, λαμβάνουν υψηλότερη βαθμολογία. Για το monetary, οι πελάτες που ξοδεύουν περισσότερα χρήματα, λαμβάνουν υψηλότερη βαθμολογία.
3. Οι βαθμολογίες συνδυάζονται για να δημιουργήσουν έναν τριψήφιο αριθμό για κάθε πελάτη (π.χ. 555, 554, 553 κ.ο.κ.). Οι καλύτεροι πελάτες θα έχουν βαθμολογία 555 και οι χειρότεροι πελάτες θα έχουν βαθμολογία 111. Χρησιμοποιώντας αυτές τις βαθμολογίες, μπορεί να γίνει ομαδοποίηση (segmentation) των πελατών σε διαφορετικά τμήματα (segments) και να γίνει υπολογισμός ποιοι από αυτούς είναι πιο κερδοφόροι.

Αξίζει να σημειωθεί, ότι τα τρία αυτά κριτήρια δεν είναι πάντα επαρκή για να ικανοποιήσουν τις ανάγκες τμηματοποίησης πελατών. Για παράδειγμα, στον τραπεζικό τομέα όπου το υπόλοιπο (Balance) είναι ουσιαστικό στοιχείο της ανάλυσης πελατών, θα πρέπει θεωρητικά να ενταχθεί ως επιπρόσθετο κριτήριο στο μοντέλο, επεκτείνοντάς το σε RFM+B (recency-frequency-monetary-balance).<sup>61</sup> Όμως κάθε ανάλυση που χρησιμοποιεί το μοντέλο RFM ως βάση, θα πρέπει σε κάθε περίπτωση να λαμβάνει υπόψην της κατ’ ελάχιστο τα τρία αυτά κριτήρια.



**Εικόνα 1. RFM+B Flowchart. Τα κριτήρια Recency, Frequency και Monetary δεν είναι αρκετά για να ικανοποιήσουν τις ανάγκες τμηματοποίησης των τραπεζικών πελατών. Ως εκ τούτου, το υπόλοιπο (Balance) των πελατών θα πρέπει να εμπλακεί θεωρητικά στη διαδικασία ανάλυσης. Πηγή: U. Firdaus and D. Nugeraha Utama, Development of bank's customer segmentation model based on RFM+B approach, 2020, ICIC International 2021, ISSN 2185-2766, pp.17 - 26**

Η εξέλιξη της τεχνολογίας και συγκεκριμένα η ευρεία ανάπτυξη της τεχνητής νοημοσύνης, επιτρέπει πλέον την εφαρμογή τμηματοποίησης RFM με βάση την εξόρυξη δεδομένων (data-mining based segmentation). Σε αντίθεση με την προηγούμενη μαθηματική κατηγοριοποίηση των πελατών, χρησιμοποιούνται πλέον προχωρημένες στατιστικές μέθοδοι για την τμηματοποίηση RFM των πελατών. Αυτές επιτρέπουν την τμηματοποίηση των πελατών με πιο ουσιαστικό τρόπο σε σύγκριση με τις απλές μαθηματικές προσεγγίσεις. Οι πιο συχνά χρησιμοποιούμενοι αλγόριθμοι clustering, αναλύονται συνοπτικά παρακάτω:

### 2.3.2 Αλγόριθμοι Clustering στο RFM

#### Αλγόριθμοι Centroid Clustering

Ο αλγόριθμος K-means αποτελεί έναν από τους πιο διαδεδομένους αλγόριθμους clustering στο μοντέλο RFM.<sup>62</sup> Όπως μαρτυρά και το όνομά του, ο αλγόριθμος K-means χρησιμοποιείται για τον διαχωρισμό ενός συνόλου δεδομένων σε K συστάδες. Ανήκει στις μη ιεραρχικές μεθόδους ομαδοποίησης δεδομένων που χωρίζουν τα δεδομένα με τέτοιο τρόπο, έτσι ώστε η προκύπτουσα ομοιότητα εντός των συστάδων να είναι υψηλή (ελάχιστο άθροισμα τετραγώνων εντός των συστάδων), ενώ η ομοιότητα μεταξύ διαφορετικών συστάδων να είναι χαμηλή (μέγιστο άθροισμα τετραγώνων μεταξύ των συστάδων). Η μέθοδος K-Means είναι μία από τις πιο δημοφιλείς μεθόδους ομαδοποίησης, λόγω της απλότητας του αλγορίθμου και της ταχύτητας επιλογής του κέντρου της συστάδας.<sup>63</sup>

Αρχικά, ο αλγόριθμος ορίζει κάποια κέντρα με τυχαίο τρόπο. Σε μερικές παραλλαγές του, όπως μία που θα εξεταστεί παρακάτω, τα κέντρα αυτά μπορεί να επιλέγονται βάσει μίας ευριστικής μεθόδους (heuristics). Έπειτα, κάθε σημείο των δεδομένων ανατίθεται στην κοντινότερη συστάδα, βάσει ενός κριτηρίου απόστασης (π.χ. ευκλείδεια απόσταση). Στη συνέχεια, υπολογίζονται νέα κέντρα των συστάδων, λαμβάνοντας υπόψιν τον μέσο όρο των σημείων που ανήκουν σε κάθε συστάδα. Η διαδικασία επαναλαμβάνεται, είτε μέχρι να σταματήσουν να μετακινούνται τα κέντρα, είτε μέχρι να επιτευχθεί ένας μέγιστος αριθμός επαναλήψεων.

Ωστόσο, ο K-means έχει κάποια προβλήματα και κάποιους περιορισμούς. Για παράδειγμα, η επιλογή του αριθμού των συστάδων K, πρέπει να γίνει εκ των προτέρων. Παράλληλα, ο αλγόριθμος είναι ευαίσθητος στην αρχικοποίηση των κέντρων. Τέλος, ο αλγόριθμος είναι ευαίσθητος στην έννοια της απόστασης και μπορεί να αντιμετωπίσει δυσκολίες στην εύρεση συστάδων με μη-σφαιρικά σχήματα.

Προκειμένου να επιλυθούν κάποια από αυτά τα προβλήματα, υπάρχουν μελέτες που προτείνουν τη χρήση του αλγόριθμου K-means++, μια βελτιωμένη παραλλαγή του K-means. Σε αντίθεση με τον K-means που επιλέγει τα αρχικά κέντρα των συστάδων κατά εντελώς τυχαίο τρόπο, ο K-means++ επιλέγει τα κέντρα αυτά, με έναν πιο «έξυπνο» τρόπο.<sup>64</sup>

- Αρχικά, επιλέγεται ένα από τα δείγματα των δεδομένων τυχαία ως το πρώτο κέντρο.
- Έπειτα, υπολογίζονται οι αποστάσεις όλων των άλλων δειγμάτων από το πρώτο κέντρο που επιλέχθηκε.
- Στη συνέχεια, επιλέγεται το επόμενο κέντρο με πιθανότητα που είναι ανάλογη του τετραγώνου της απόστασής του από το προηγούμενο κέντρο (ή τα προηγούμενα κέντρα). Δηλαδή, τα δείγματα που είναι πιο μακριά από τα υπάρχοντα κέντρα έχουν μεγαλύτερη πιθανότητα να επιλεγούν ως επόμενο κέντρο.
- Η διαδικασία επαναλαμβάνεται μέχρι να επιλεγούν όλα τα K κέντρα

Με αυτόν τον τρόπο, τα κέντρα επιλέγονται έτσι ώστε να είναι μακριά το ένα από το άλλο και διάσπαρτα μεταξύ τους σε όλο το χώρο των δεδομένων. Αυτό οδηγεί αφ' ενός σε ποιοτικότερα αποτελέσματα συσταδοποίησης, αφ' ετέρου σε πολύ μικρότερο χρόνο υλοποίησης της διαδικασίας συσταδοποίησης. Ενδεικτικά, σε μικρά dataset, ο K-means++ ολοκλήρωσε την συσταδοποίηση στον μισό χρόνο και παρήγαγε 20% καλύτερα αποτελέσματα, συγκριτικά από τον K-means. Σε μεγάλα dataset, ο K-means++ ολοκλήρωσε την συσταδοποίηση 70% ταχύτερα και παρήγαγε από 10 έως και 1000(!) φορές ποιοτικότερα αποτελέσματα, συγκριτικά από τον K-means.<sup>65</sup>

Μία άλλη παραλλαγή του K-means που προσπαθεί να μειώσει τους περιορισμούς του βασικού αλγορίθμου, είναι η K-medoids. Ο αλγόριθμος K-medoids, χρησιμοποιείται για να βρει τις πιο αντιπροσωπευτικές παρατηρήσεις (ή αλλιώς "medoids") από τα δεδομένα. Στην ουσία, ενώ ο αλγόριθμος K-means προσπαθεί να ελαχιστοποιήσει την ευκλείδεια απόσταση μεταξύ των σημείων μίας συστάδας και του κέντρου αυτής, ο αλγόριθμος K-medoids προσπαθεί να ελαχιστοποιήσει την απόσταση μεταξύ των σημείων της συστάδας και του medoid της συστάδας, το οποίο είναι ένα πραγματικό σημείο δεδομένων από το σύνολο δεδομένων.

Ο αλγόριθμος K-medoids είναι πιο ανθεκτικός στις ακραίες τιμές (outliers), καθώς τα κέντρα medoids είναι πραγματικά σημεία δεδομένων και επομένως μπορούν να αντιπροσωπεύουν καλύτερα τα δεδομένα. Επιπρόσθετα, μπορεί να χειριστεί διάφορες μετρικές απόστασης, σε αντίθεση με τον K-means που είναι περιορισμένος στην ευκλείδεια απόσταση. Από την άλλη πλευρά, ο αλγόριθμος K-medoids είναι πιο αργός και έχει μεγαλύτερες υπολογιστικές απαιτήσεις συγκριτικά με τον αλγόριθμο K-means, καθώς πρέπει να υπολογίσει τις αποστάσεις μεταξύ όλων των σημείων.

Η λειτουργία του αλγορίθμου είναι ως εξής:<sup>66</sup> Αφού επιλέξει τυχαία κάποια σημεία ως τα αρχικά κέντρα (medoids) των συστάδων, ο αλγόριθμος αναθέτει κάθε σημείο στην πλησιέστερη συστάδα, βάσει της ευκλείδεια απόστασης από τα τρέχοντα κέντρα. Για κάθε συστάδα, επιλέγει τυχαία ένα σημείο ως υποψήφιο για το νέο medoid της συστάδας και έπειτα υπολογίζει τις αποστάσεις όλων των σημείων της συστάδας από το νέο υποψήφιο medoid. Τέλος, ο αλγόριθμος υπολογίζει τη διαφορά μεταξύ των συνολικών αποστάσεων με το νέο και το παλιό medoid. Εάν η διαφορά είναι αρνητική, αντικαθιστά το παλιό medoid με το νέο.

Ο K-means αλγόριθμος είναι καλός για αριθμητικά δεδομένα, ενώ ο K-medoids αλγόριθμος είναι προσαρμοσμένος για κατηγορικά δεδομένα. Ο αλγόριθμος K-prototypes είναι ένας αλγόριθμος συσταδοποίησης (clustering) που συνδυάζει τον αλγόριθμο K-means και K-medoids για να διαχειριστεί ταυτόχρονα αριθμητικά και κατηγορικά δεδομένα. Το K-prototypes ενσωματώνει τα χαρακτηριστικά και των δύο αυτών αλγορίθμων για να δημιουργήσει μια ευέλικτη τεχνική συσταδοποίησης.

Ο αλγόριθμος επιλέγει  $k$  αρχικά "prototypes", καθένα από τα οποία αντιπροσωπεύει μια ομάδα. Κάθε "prototype" περιλαμβάνει μέσους όρους για τα αριθμητικά χαρακτηριστικά και τις πιο συχνές κατηγορίες για τα κατηγορικά χαρακτηριστικά. Έπειτα, ο αλγόριθμος αναθέτει κάθε δείγμα στο πλησιέστερο "prototype", με βάση μια απόσταση που συνδυάζει την Ευκλείδεια απόσταση για τα αριθμητικά χαρακτηριστικά και μια απόσταση κατηγορικής «ανομοιότητας» (π.χ. "hamming distance") για τα κατηγορικά χαρακτηριστικά. Στη συνέχεια, υπολογίζει καινούργια "prototypes" για κάθε ομάδα με βάση τα δεδομένα που έχουν ανατεθεί σε κάθε ομάδα. Η διαδικασία επαναλαμβάνεται μέχρι να συγκλίνουν τα "prototypes" ή μέχρι να επιτευχθεί ένας μέγιστος αριθμός επαναλήψεων.<sup>67</sup>

Εκτός από παραλλαγές του ίδιου του αλγόριθμου (K-means), έχουν ανακαλυφθεί και συνδυασμένες χρήσεις του με άλλες τεχνικές και μεθόδους, όπως η χρήση δεικτών εγκυρότητας (validity indexes) που βελτιστοποιούν τον αριθμό των αρχικά επιλεγμένων κέντρων.<sup>68</sup>

### Αλγόριθμοι Hierarchical Clustering

Η μέθοδος ιεραρχικής συσταδοποίησης (Hierarchical Clustering), αποσκοπεί στο να οργανώσει τα δεδομένα σε ένα δενδροειδές ιεραρχικό διάγραμμα, αντί για μια επίπεδη διαίρεση των δεδομένων σε ομάδες. Σε αντίθεση με την περίπτωση της συσταδοποίησης  $k$ -means, δεν απαιτείται αρχικά ο αριθμός των ομάδων. Είναι ιδιαίτερα χρήσιμο όταν τα δεδομένα εκφράζουν ιεραρχίες ή υπο-ομαδοποίηση, μιας και το δέντρο παρέχει μια δομημένη αναπαράσταση των δεδομένων, αποκαλύπτοντας την ιεραρχία και τις σχέσεις ανάμεσα στις ομάδες.

Η ιεραρχική συσταδοποίηση, χωρίζεται σε δύο βασικές υπο-κατηγορίες: την Συναθροιστική (Agglomerative) και την Διαχωριστική (Divisive).<sup>69</sup>

Στη συναθροιστική (agglomerative) συσταδοποίηση, ο αλγόριθμος δουλεύει «από μέσα προς τα έξω». Αρχικά, κάθε σημείο δεδομένων ξεκινάει ως μια ξεχωριστή ομάδα. Έπειτα, υπολογίζεται η απόσταση ανάμεσα σε όλες τις ομάδες χρησιμοποιώντας κάποια μετρική απόστασης (όπως η ευκλείδεια απόσταση, η απόσταση Manhattan, κλπ.) και εντοπίζεται η πιο κοντινή ζεύξη ομάδων. Στη συνέχεια, οι δύο πλησιέστερες ομάδες συγχωνεύονται σε μία. Η διαδικασία επαναλαμβάνεται μέχρι όλα τα σημεία δεδομένων να συγχωνευτούν σε μια μεγάλη ομάδα ή μέχρι να επιτευχθεί ο συγκεκριμένος αριθμός ομάδων.<sup>70</sup>

Οι agglomerative αλγόριθμοι, χρησιμοποιούν διάφορες τεχνικές σύνδεσης (linkage), για να υπολογίσουν την απόσταση μεταξύ διαφόρων συστάδων κατά τη διαδικασία της ιεραρχικής συσταδοποίησης. Μερικές τεχνικές σύνδεσης (linkage) είναι οι ακόλουθες:

- **Single Linkage (Nearest Neighbor Methods):** Η απόσταση μεταξύ δύο ομάδων υπολογίζεται βάσει της ελάχιστης απόστασης μεταξύ ενός σημείου από την πρώτη ομάδα και ενός σημείου από τη δεύτερη ομάδα. Δηλαδή, χρησιμοποιείται η απόσταση μεταξύ των πλησιέστερων γειτόνων από διαφορετικές ομάδες.
- **Complete Linkage (Furthest Neighbor Methods):** Η απόσταση μεταξύ δύο ομάδων υπολογίζεται βάσει της μέγιστης απόστασης μεταξύ ενός σημείου από την πρώτη ομάδα και ενός σημείου από τη δεύτερη ομάδα. Δηλαδή, χρησιμοποιείται η απόσταση μεταξύ των πιο απομακρυσμένων γειτόνων από διαφορετικές ομάδες.
- **Average Linkage (Between Groups Methods):** Η απόσταση μεταξύ δύο ομάδων υπολογίζεται ως ο μέσος όρος των αποστάσεων μεταξύ όλων των ζευγαριών σημείων, όπου το κάθε ζεύγος αποτελείται από ένα σημείο από την πρώτη ομάδα και ένα σημείο από τη δεύτερη ομάδα.
- **Ward Linkage:** Η απόσταση μεταξύ δύο ομάδων, υπολογίζεται βάσει της αύξησης της εσωτερικής διακύμανσης που θα προκύψει αν συγχωνευτούν οι δύο ομάδες. Συγκεκριμένα, γίνεται προσπάθεια να επιτευχθεί η συγχώνευση εκείνη, που θα οδηγήσει στην ελάχιστη αύξηση της συνολικής εσωτερικής διακύμανσης, δηλαδή του αθροίσματος των τετραγωνικών αποστάσεων των σημείων από τα κέντρα των αντίστοιχων clusters.

Στη διαχωριστική (divisive) συσταδοποίηση, ο αλγόριθμος δουλεύει «από έξω προς τα μέσα», δηλαδή ακολουθεί την ακριβώς αντίθετη προσέγγιση σε σχέση με τους αλγόριθμους agglomerative clustering. Όλα τα δεδομένα ξεκινούν ως μια μεγάλη ομάδα. Στη συνέχεια, αυτή η ομάδα πρόκειται να διαιρείται διαρκώς σε μικρότερες, σε κάθε βήμα του αλγόριθμου: Εντοπίζεται και διαιρείται η ομάδα που, αν διαιρεθεί, θα προκαλέσει τη μεγαλύτερη αύξηση στην συνολική διακύμανση (ή άλλο κριτήριο) μεταξύ των ομάδων. Αυτή η διαδικασία διαίρεσης επαναλαμβάνεται σε κάθε νέα ομάδα που προκύπτει, μέχρι όλες οι ομάδες

να περιέχουν μόνο ένα σημείο, ή μέχρι να επιτευχθεί ένας επιθυμητός αριθμός ομάδων.<sup>71</sup> Το Divisive Clustering μπορεί να είναι πιο χρονοβόρο από το Agglomerative Clustering, καθώς η διαίρεση μιας ομάδας σε υπο-ομάδες μπορεί να απαιτεί περισσότερους υπολογισμούς.

### Αλγόριθμοι Density-based Clustering

Ο DBSCAN (Density-Based Spatial Clustering of Applications with Noise) είναι ένας αλγόριθμος συσταδοποίησης που βασίζεται στην πυκνότητα των σημείων σε έναν πολυδιάστατο χώρο. Αντί να υποθέσει ότι οι συστάδες έχουν έναν συγκεκριμένο σχήμα (όπως κάνει ο αλγόριθμος k-means), ο DBSCAN βλέπει συστάδες ως περιοχές υψηλής πυκνότητας δεδομένων, που χωρίζονται μεταξύ τους από περιοχές χαμηλής πυκνότητας. Ο αλγόριθμος μπορεί να βρει οποιοδήποτε αριθμό συστάδων, με οποιοδήποτε σχήμα, και είναι ιδιαίτερα καλός στο να αντιμετωπίζει θόρυβο και outliers. Ο DBSCAN διαφέρει από τους περισσότερους αλγορίθμους συσταδοποίησης, καθώς δεν απαιτείται ο εκ των προτέρων καθορισμός του αριθμού των συστάδων<sup>72</sup> και μπορεί να δημιουργήσει ομάδες τυχαίων σχημάτων, με πολύ διαφορετικές εμφανίσεις και μεγέθη.

Τα βασικά βήματα του αλγόριθμου είναι τα εξής:<sup>73</sup>

1. Γίνεται αρχικοποίηση των παραμέτρων minPts (ελάχιστος αριθμός σημείων) και eps ( $\epsilon$ ) (η ακτίνα εντός της οποίας ψάχνουμε γείτονες).
2. Επιλέγεται τυχαία ένα σημείο p από το σύνολο δεδομένων ως το αρχικό σημείο.
3. Τα επόμενα βήματα επαναλαμβάνονται για κάθε σημείο στο σύνολο δεδομένων, μέχρι να έχουν επεξεργαστεί όλα τα σημεία.
4. Υπολογίζεται η απόσταση από το σημείο p σε όλα τα άλλα σημεία και αναγνωρίζονται τα σημεία που βρίσκονται εντός της ακτίνας eps.
5. Αν ο αριθμός των σημείων εντός της ακτίνας eps είναι μεγαλύτερος ή ίσος με το minPts, τότε το σημείο p καθορίζεται ως κεντρικό σημείο και αρχίζει μια νέα ομάδα. Το σημείο p συνδέεται με όλα τα άλλα σημεία εντός της ακτίνας eps.
6. Αν το σημείο p δεν είναι κεντρικό σημείο αλλά βρίσκεται εντός της ακτίνας eps ενός κεντρικού σημείου, τότε το σημείο p καθορίζεται ως συνοριακό σημείο και προστίθεται στην ομάδα του κεντρικού σημείου.
7. Αν το σημείο p δεν είναι ούτε κεντρικό ούτε συνοριακό σημείο, τότε επιλέγεται το επόμενο σημείο και επαναλαμβάνονται τα βήματα 4-6.

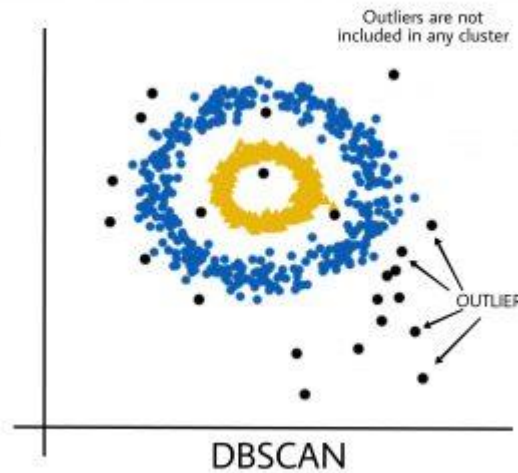
Με αυτόν τον τρόπο, ο αλγόριθμος DBSCAN «εξαπλώνεται» διασχίζοντας όλο το σύνολο δεδομένων, εντοπίζοντας ομάδες βάσει πυκνότητας και καθορίζοντας σημεία που είναι θόρυβος. Οι παράμετροι eps και minPts είναι κρίσιμες για την απόδοση του DBSCAN και απαιτείται πειραματισμός για να βρεθούν οι καλύτερες τιμές στο εκάστοτε σύνολο δεδομένων.

Με βάση τα ανωτέρω, αντιλαμβάνεται κανείς ότι σύμφωνα με τον DBSCAN, τα σημεία κατατάσσονται σε τρεις κατηγορίες:

1. Core Points (Κεντρικά Σημεία): Σημεία που έχουν τουλάχιστον minPts άλλα σημεία εντός μιας ακτίνας eps.
2. Border Points (Συνοριακά Σημεία): Σημεία που δεν έχουν minPts σημεία εντός της ακτίνας eps, αλλά βρίσκονται εντός της ακτίνας eps από ένα κεντρικό σημείο.
3. Noise Points (Σημεία Θορύβου ή Outliers): Σημεία που δεν είναι ούτε κεντρικά ούτε συνοριακά.

Τα σημεία θορύβου (outliers) ανιχνεύονται κατά την διάρκεια της διαδικασίας συσταδοποίησης, αλλά δεν ανατίθενται σε καμία ομάδα. Αντίθετα, παραμένουν απομονωμένα και αναγνωρίζονται απλά ως θόρυβος στα δεδομένα. Αυτό είναι ένα από τα πλεονεκτήματα του DBSCAN σε σύγκριση με άλλες μεθόδους ομαδοποίησης, όπως η ιεραρχική ομαδοποίηση ή ο K-means, που δεν διαχειρίζονται εύκολα τα outliers.<sup>74</sup>



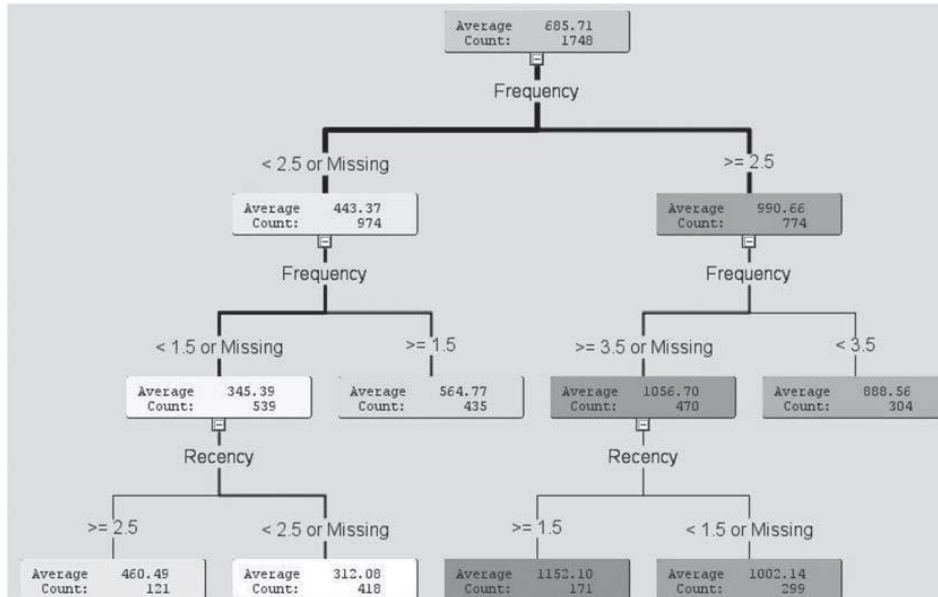


**Εικόνα 2. Επιρροή θορύβου (outliers) στον DBSCAN. Πηγή:**  
<https://www.geeksforgeeks.org/dbscan-clustering-in-ml-density-based-clustering/>

Αν και απαιτούνται περαιτέρω μετρήσεις σχετικά με την αποδοτικότητα του DBSCAN στην εφαρμογή του μοντέλου RFM, έρευνες έχουν δείξει ότι υπάρχουν περιπτώσεις που παράγει ποιοτικότερες συσταδοποιήσεις (π.χ. σε σύγκριση με τον K-means).<sup>75</sup>

### 2.3.3 Χρήση του μοντέλου RFM

Το μοντέλο RFM μπορεί να χρησιμοποιηθεί για τον υπολογισμό της αξίας των πελατών (customer value) και της αξίας διάρκειας ζωής (customer lifetime value), την εκτίμηση της πιθανότητας ανταπόκρισης για κάθε τύπο προσφοράς, την αξιολόγηση των διαδικτυακών κριτικών και τον προσδιορισμό της συχνότητας αποστολής email για τους ενεργούς πελάτες. Η τελική τμηματοποίηση των πελατών, μπορεί να γίνει σε διαφορετικές κατηγορίες (clusters), ανάλογα τον σκοπό εφαρμογής του μοντέλου και των αναγκών της επιχείρησης. Σε περιπτώσεις ποικιλόμορφης συστάδας (diverse cluster), μπορούμε να κάνουμε περαιτέρω τμηματοποίηση εντός αυτής με τη χρήση δέντρου αποφάσεων. Με αυτόν τον τρόπο μπορούν να δημιουργηθούν κάποια ένθετα τμήματα εσωτερικά της συστάδας και να καταστεί δυνατή η κατηγοριοποίηση των εν λόγω καταναλωτών σε κάποιες λογικές υποκατηγορίες.



**Εικόνα 3. Περαιτέρω τμηματοποίηση μίας συστάδας, χρησιμοποιώντας δέντρο αποφάσεων. Οι πελάτες μπορούν να χωριστούν σε κατηγορίες όπως “συχνότητα μεγαλύτερη από 2,5 με μέση χρηματική αξία 990,66” και “συχνότητα μεγαλύτερη από 2,5 και μικρότερη από 3,5 με μέση χρηματική αξία 1056,70” κ.ο.κ. Πηγή: Daqing Chen, Sai Laing Sain, Kun Guo, Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining, 2012, Database Marketing & Customer Strategy Management**

Το μοντέλο RFM έχει το θετικό ότι είναι μια μέθοδος που μπορεί να τμηματοποιήσει τους πελάτες με βάση τη συμπεριφορά τους, η οποία μπορεί εύκολα να εντοπιστεί μέσω του ιστορικού των συναλλαγών τους. Επιπρόσθετα, είναι οικονομικά αποδοτικό και εύκολα κατανοητό από τα στελέχη διοίκησης, επιτρέποντας στις εταιρείες να εντοπίζουν πολύτιμους πελάτες και να προβλέπουν την ανταπόκρισή τους στις στρατηγικές μάρκετινγκ. Ωστόσο, το μοντέλο εστιάζει μόνο στους καλύτερους πελάτες και αγνοεί τους υπόλοιπους, έχει περιορισμένες μεταβλητές επιλογής και προϋποθέτει ομοιογενή βάση δεδομένων πελατών. Για να αντιμετωπιστούν αυτές οι αδυναμίες, μπορεί να χρησιμοποιηθεί η υπο-τμηματοποίηση για τον εντοπισμό ανεκμετάλλευτων δυνητικών πελατών και να ληφθούν υπόψη οι σχεσιακές πληροφορίες. Αυτή είναι άλλη μία περίπτωση που η περαιτέρω τμηματοποίηση που περιεγράφηκε ανωτέρω, μπορεί να δώσει λύση.

Αφού η εργασία ομαδοποίησης χρησιμοποιηθεί για την εύρεση segments πελατών με παρόμοιες RFM τιμές, στη συνέχεια, με τη χρήση των segments αυτών και δημογραφικών μεταβλητών πελατών, ανακαλύπτονται κανόνες ταξινόμησης (classification rules) για την πρόβλεψη της μελλοντικής συμπεριφοράς και προφίλ των πελατών.

Ορισμένες μελέτες προσπάθησαν να βελτιώσουν την προβλεψιμότητα του μοντέλου RFM με την προσθήκη περισσότερων μεταβλητών ή την ανάπτυξη νέων μοντέλων. Κάποια από αυτά τα μοντέλα ενσωματώνουν μεταβλητές προηγούμενης αγοραστικής συμπεριφοράς, συμπεριλαμβανομένων των μεταβλητών RFM<sup>76</sup>, δημογραφικές μεταβλητές και πρόσθετες μεταβλητές πελατών για την πρόβλεψη της συμπεριφοράς τους<sup>77</sup>. Άλλα μοντέλα συνδέουν το RFM με το Customer Lifetime Value (CLV)<sup>78</sup> ή ενσωματώνουν την έννοια του RFM στη βιβλιογραφία του μάρκετινγκ για να ορίσουν τα διαδοχικά πρότυπα RFM<sup>79</sup>. Ορισμένες μελέτες προτείνουν ένα διευρυμένο μοντέλο RFM με την προσθήκη πρόσθετων παραμέτρων για την ταξινόμηση της αφοσίωσης των πελατών<sup>80</sup> και άλλες προτείνουν την αλλαγή του RFM σε RFC (recency, frequency and cost) για την τμηματοποίηση των πολιτών<sup>81</sup>. Κάποιοι ερευνητές χρησιμοποίησαν διάφορες τεχνικές ταξινόμησης, όπως δέντρα αποφάσεων, νευρωνικά δίκτυα και λογιστική παλινδρόμηση, για να αναλύσουν την ανταπόκριση των πελατών σε μια προώθηση προϊόντος<sup>82</sup>. Άλλες μελέτες έχουν συνδυάσει τα χαρακτηριστικά RFM με τη θεωρία των ακατέργαστων συνόλων για να εξορύξουν κανόνες ταξινόμησης και να ανακαλύψουν τα χαρακτηριστικά των πελατών που μπορούν να βοηθήσουν στη βελτίωση της διαχείρισης των πελατειακών σχέσεων (Customer Relationship Management)<sup>83</sup>. Επιπλέον, ορισμένες μελέτες έχουν χρησιμοποιήσει δημογραφικές

μεταβλητές όπως η ηλικία, το φύλο, το επάγγελμα και η οικογενειακή κατάσταση για να ανακαλύψουν κανόνες ταξινόμησης που προβλέπουν τις τιμές RFM των πελατών με την πάροδο του χρόνου<sup>84</sup>. Τα ευρήματα αυτών των μελετών, δείχνουν ότι ορισμένα από αυτά τα μοντέλα, παράγουν καλύτερα αποτελέσματα από το παραδοσιακό μοντέλο RFM.

Στο τελικό στάδιο, η παραπάνω διαδικασία μπορεί να χρησιμοποιηθεί για εξόρυξη κανόνων συσχέτισης (π.χ. για τη σύσταση προϊόντων) στις επιμέρους ομάδες πελατών. Οι κανόνες συσχέτισης στην εξόρυξη δεδομένων είναι μοτίβα που δείχνουν τη σχέση μεταξύ στοιχείων σε μια βάση δεδομένων. Η αριστερή πλευρά του κανόνα (X) είναι η συνθήκη και η δεξιά πλευρά (Y) είναι το αποτέλεσμα. Η εξόρυξη κανόνων συσχέτισης είναι το έργο της εύρεσης αυτών των μοτίβων και της μέτρησης της συχνότητάς τους.

## 2.4 Ανάλυση καλαθιού αγοράς

Η Ανάλυση Καλαθιού Αγοράς (Market Basket Analysis), είναι μια τεχνική εξόρυξης δεδομένων, που χρησιμοποιείται στο λιανικό εμπόριο και το μάρκετινγκ, για την αποκάλυψη συσχετίσεων και σχέσεων, μεταξύ προϊόντων που αγοράζονται συχνά μαζί, από τους πελάτες.<sup>85</sup> Είναι επίσης γνωστή, ως ανάλυση συγγένειας ή εξόρυξη κανόνων συσχέτισης (Association Rule Mining).

Η ανάλυση βασίζεται στην έννοια του "καλαθιού της αγοράς", το οποίο αναφέρεται σε μια συλλογή προϊόντων που αγοράζει ένας πελάτης κατά τη διάρκεια μίας συναλλαγής με την εταιρεία. Με την εξέταση μεγάλων συνόλων δεδομένων συναλλαγών, η ανάλυση καλαθιού αγοράς αποσκοπεί στον εντοπισμό μοτίβων, συσχετίσεων και εξαρτήσεων μεταξύ των ειδών που οι πελάτες τείνουν να αγοράζουν μαζί.<sup>86</sup>

Ο πρωταρχικός στόχος της ανάλυσης καλαθιού αγοράς είναι η αποκάλυψη συσχετίσεων μεταξύ προϊόντων για την απόκτηση γνώσεων σχετικά με τη συμπεριφορά των πελατών, τη βελτίωση των στρατηγικών μάρκετινγκ, τη βελτιστοποίηση της διάταξης (των προϊόντων) των καταστημάτων και την υποστήριξη των διαδικασιών λήψης αποφάσεων. Η ανάλυση μπορεί να αποκαλύψει πολύτιμες πληροφορίες, όπως ποια είδη αγοράζονται συνήθως μαζί, την ισχύ των συσχετίσεων και τον εντοπισμό συχνών συνόλων αντικειμένων και κανόνων συσχέτισης.

Τα συχνά σύνολα αντικειμένων είναι εκείνα που εμφανίζονται μαζί σε σημαντικό αριθμό καλαθιών αγοράς, ενώ οι κανόνες συσχέτισης είναι λογικές συνέπειες που περιγράφουν τις σχέσεις μεταξύ των αντικειμένων. Για παράδειγμα, ένας κανόνας συσχέτισης μπορεί να υποδεικνύει ότι οι πελάτες που αγόρασαν βούτυρο και γάλα είναι πολύ πιθανό να αγοράσουν και τυρί. Αυτοί οι κανόνες μπορούν να χρησιμοποιηθούν για διάφορους σκοπούς, όπως οι διασταυρούμενες πωλήσεις (cross-selling), οι τοποθετήσεις προϊόντων, οι στοχευμένες διαφημίσεις ή προωθήσεις και οι εξατομικευμένες συστάσεις.<sup>87</sup>

Η ανάλυση καλαθιού αγοράς εκτελείται συνήθως με τη χρήση αλγορίθμων, όπως ο αλγόριθμος Apriori ή ο αλγόριθμος FP-growth, που είναι οι πιο διαδεδομένοι<sup>88 89</sup>. Αυτοί οι αλγόριθμοι σαρώνουν τα δεδομένα συναλλαγών για την εύρεση συχνών συνόλων στοιχείων και δημιουργούν κανόνες συσχέτισης με βάση προκαθορισμένα κατώτατα όρια ελάχιστης υποστήριξης (support) και εμπιστοσύνης (confidence)<sup>90</sup>.

Γενικά, κατά την ανάλυση καλαθιού αγοράς, χρησιμοποιούνται κατά βάση οι εξής μετρικές, για τη μέτρηση της ισχύος και της σημασίας των κανόνων συσχέτισης μεταξύ στοιχείων: η υποστήριξη, η εμπιστοσύνη και η ανύψωση<sup>91</sup>.

Υποστήριξη (Support): Η υποστήριξη μετράει πόσο συχνά ένα σύνολο στοιχείων ή ένας κανόνας εμφανίζεται στο σύνολο δεδομένων. Υπολογίζει το ποσοστό των συναλλαγών που περιέχουν ένα συγκεκριμένο σύνολο στοιχείων ή κανόνα. Ο τύπος για την υποστήριξη είναι:

$$\text{Υποστήριξη}(X) = \frac{\text{Αριθμός συναλλαγών που περιέχουν } X}{\text{Συνολικός αριθμός συναλλαγών}}$$

Για παράδειγμα, αν έχουμε 1.000 συναλλαγές και 100 από αυτές περιέχουν το σύνολο στοιχείων {A, B}, τότε η υποστήριξη του {A, B} είναι  $100/1.000 = 0,1$  ή 10%.

Η υψηλή υποστήριξη υποδεικνύει ότι το σύνολο στοιχείων ή ο κανόνας είναι δημοφιλής ή εμφανίζεται συχνά στο σύνολο δεδομένων.

Εμπιστοσύνη (Confidence): Η εμπιστοσύνη μετρά την αξιοπιστία ή τη δύναμη ενός κανόνα συσχέτισης. Υπολογίζει την υπό συνθήκη πιθανότητα εύρεσης του επακόλουθου (Y) σε μια συναλλαγή δεδομένης της

παρουσίας του προηγούμενου (X). Με απλά λόγια, πόσο συχνά ο κανόνας είναι αληθής με βάση τα παρατηρούμενα δεδομένα.

Ο τύπος για την εμπιστοσύνη είναι ο εξής:

$$\text{Εμπιστοσύνη } (X \rightarrow Y) = \frac{\text{Υποστήριξη}(X \cap Y)}{\text{Υποστήριξη}(X)}$$

Για παράδειγμα, αν η υποστήριξη του {A} είναι 0,3 (30%) και η υποστήριξη του {A, B} είναι 0,1 (10%), η εμπιστοσύνη του κανόνα {A} → {B} θα είναι  $0,1/0,3 = 0,33$  ή 33%.

Οι τιμές εμπιστοσύνης κυμαίνονται από 0 έως 1, όπου το 1 υποδηλώνει τέλεια συσχέτιση μεταξύ του προηγούμενου και του επόμενου.

Ανύψωση (Lift): Η ανύψωση μετρά την ισχύ της συσχέτισης, μεταξύ του αριστερού και του δεξιού σκέλους ενός κανόνα, λαμβάνοντας υπόψη την ατομική υποστήριξη κάθε στοιχείου. Το Lift συγκρίνει την πιθανότητα εύρεσης του επακόλουθου στοιχείου (δεξί σκέλος) παρουσία του προηγούμενου στοιχείου (αριστερό σκέλος), με την πιθανότητα εύρεσης του επακόλουθου στοιχείου, ανεξάρτητα από το προηγούμενο στοιχείο. Ο τύπος για την ανύψωση είναι:

$$\text{Ανύψωση}(X \rightarrow Y) = \frac{\text{Υποστήριξη}(X \cap Y)}{\text{Υποστήριξη}(X) * \text{Υποστήριξη}(Y)}$$

Τιμές υποστήριξης μεγαλύτερες από 1, υποδηλώνουν θετική συσχέτιση μεταξύ των στοιχείων, δείχνοντας έτσι, ότι η εμφάνιση του προηγούμενου στοιχείου, αυξάνει την πιθανότητα εμφάνισης του επακόλουθου.

Για παράδειγμα, εάν η τιμή lift για τον κανόνα {A} → {B} είναι 2, σημαίνει ότι η εμφάνιση του {A} σε μια συναλλαγή καθιστά την εμφάνιση του {B} δύο φορές πιο πιθανή σε σύγκριση με την περίπτωση που το {A} απουσιάζει.

Εκτός από την υποστήριξη, την εμπιστοσύνη και την ανύψωση, έχουν εξεταστεί ανά καιρούς και άλλες μετρικές, όπως η ακρίβεια (accuracy), η συσχέτιση (correlation) και άλλες<sup>92</sup>, που όμως δεν θα εξετάσουμε στην παρούσα εργασία.

### ΚΕΦΑΛΑΙΟ 3 – Εφαρμογή Αναλυτικής Πελατών σε πραγματικά επιχειρησιακά δεδομένα

Η αξία των μεθόδων Customer Analytics, μπορεί να αποδειχθεί μόνο μέσω εφαρμογής τους σε πραγματικά δεδομένα. Για την εφαρμογή των μεθόδων της Οπτικής Ανάλυσης, του Μοντέλου RFM και της Ανάλυσης Καλαθιού Αγοράς χρησιμοποιήθηκαν ιστορικά δεδομένα της εταιρείας Atlantic.

Η Ατλάντικ ήταν αλυσίδα σούπερ μάρκετ αμιγώς ελληνικών συμφερόντων, που ιδρύθηκε το 1980 και δραστηριοποιούνταν σε όλη την ελληνική επικράτεια, τόσο στον τομέα της λιανικής, όσο και σε αυτόν της χονδρικής. Το 2009, κατείχε την 14η θέση ανάμεσα στις μεγαλύτερες εμπορικές εταιρείες, αλλά και την 6η θέση στον κλάδο της. Στο απόγειό της, η εταιρεία έφτασε να απαριθμεί πάνω από 150 καταστήματα και 4.000 υπαλλήλους, καθιστώντας την ελκυστική για χρήση των ιστορικών δεδομένων για μια πλούσια ανάλυση. Η εταιρεία δεν είναι πλέον ενεργή από το 2011.<sup>93</sup>

Η οπτική ανάλυση χρησιμοποιήθηκε για την αρχική εξερεύνηση των δεδομένων, υπό το πρίσμα εμφάνισης σημαντικών στοιχείων που αξίζουν περαιτέρω ανάλυσης. Χρησιμοποιήθηκε επίσης για να αναλύσει συμμετρίες και να εξάγει πληροφορία από τα αποτελέσματα του μοντέλου RFM και της συσταδοποίησης πελατών βάσει αυτού. Μεταξύ άλλων, χρησιμοποιήθηκαν γραμμικά γραφήματα (με και χωρίς δείκτες), ραβδογράμματα, χάρτες θερμότητας, ιστογράμματα, θηκογράμματα, γραφήματα πίτας και δακτυλίου, γραφήματα φυσαλίδας, γραφήματα διασποράς και πίνακες συσχετίσεων. Βοηθητικά στοιχεία όπως οριζόντιες γραμμές επιπέδων τιμών, γραμμές εμφάνισης τάσης και αλλαγές βάθους χρώματος ή μεγέθους σχήματος, χρησιμοποιήθηκαν προκειμένου να διευκολύνουν τους decision makers να λάβουν αποτελεσματικότερες αποφάσεις.

Το μοντέλο RFM χρησιμοποιήθηκε για την συσταδοποίηση των πελατών υπό τα τρία διαφορετικά δομικά του στοιχεία. Η εύρεση του recency, εξήχθη από τη χρονοσφραγίδα της πιο πρόσφατης συναλλαγής

του εκάστοτε πελάτη. Η εύρεση του frequency, εξήχθη από το άθροισμα των συναλλαγών του κάθε πελάτη. Το monetary, εξήχθη από το άθροισμα του ποσού που έχει δαπανήσει ο εκάστοτε πελάτης. Οι τιμές των RFM μεταβλητών που εμφανίζονταν λιγότερο από 1% των φορών, όταν ο συνολικός αριθμός διαφορετικών επιπέδων ήταν μεγαλύτερος από 25, αφαιρέθηκαν ως ακραίες. Η κύρια μέθοδος φιλτραρίσματος που χρησιμοποιήθηκε ήταν τα όρια μεταδεδομένων, με εναλλακτική την τυπική απόκλιση από το μέσο όρο, όταν δεν υπάρχουν αποθηκευμένα όρια στα μεταδεδομένα, όπου αφαιρούνται δεδομένα με απόσταση (διαφορά) τρεις τυπικές αποκλίσεις μακριά από το μέσο όρο. Επιπρόσθετα, οι παρουσιαζόμενες ασύμμετρες κατανομές στις τιμές των μεταβλητών, μετασχηματίστηκαν μέσω κανονικοποίησης. Για τη συσταδοποίηση των πελατών, βάσει των μεταβλητών RFM χρησιμοποιήθηκε ο αλγόριθμος k-means και η μέθοδος forgy για επιλογή τυχαίων αρχικών σημείων. Για την εκτίμηση του αριθμού των συστάδων, χρησιμοποιήθηκε ο αλγόριθμος Aligned Box Criterion (ABC), ενώ ως κριτήριο εκτίμησης, χρησιμοποιήθηκε το Global Peak Value επιλέγοντας τη μέγιστη τιμή μεταξύ όλων των τιμών κορυφής του gap statistics. Οι διάφορες συστάδες, δημιουργήθηκαν μέσω της σύγκρισης των τιμών τους με τις αντίστοιχες μέσες τιμές των μετρικών RFM από όλες τις συστάδες.

Η ανάλυση καλαθιού αγοράς χρησιμοποιήθηκε για την εξαγωγή κανόνων μεταξύ αγορασμένων προϊόντων. Η εύρεση βέλτιστων κατωφλίων ποσοστών υποστήριξης (support) και εμπιστοσύνης (confidence), έγινε μέσω πολλαπλών δοκιμών trial-and-error, έως ότου εξαχθούν κανόνες με ισχυρές συσχετίσεις μεταξύ προϊόντων, αρκετοί σε αριθμό ώστε να πραγματοποιηθούν αναλύσεις.

### 3.1 Τα υπό ανάλυση δεδομένα

Η βάση δεδομένων που θα χρησιμοποιηθεί από την προαναφερθείσα αλυσίδα καταστημάτων, συναποτελείται από πέντε διαφορετικούς πίνακες, σε μορφή αρχείων (.csv). Οι πίνακες και τα αναλυτικά πεδία αυτών φαίνονται παρακάτω:

- Customers (17.673 εγγραφές):
  - custCode: Κωδικός πελάτη
  - custAge: Ηλικία (το 2005)
  - custOccupation: Επάγγελμα
  - custGender: Φύλο (1 άντρες, 2 γυναίκες)
  - custNumChildren: Αριθμός παιδιών στην οικογένεια
  - custPrefecture: Νομός κατοικίας
  - custArea: Περιοχή κατοικίας
  - custPostalCode: Ταχ. κωδικός κατοικίας
  - registrationDate: Ημερομηνία εγγραφής (απόκτησης Loyalty κάρτας)
  - custStoreReg: Κατάστημα εγγραφής
- Items (102.143 εγγραφές):
  - itemID: Κωδικός προϊόντος (barcode)
  - itemName: Τίτλος προϊόντος
  - itemBasicID: Κωδικός βασικού προϊόντος (η μόνη διαφορά με το ItemID παρατηρείται σε περίπτωση που το προϊόν βγει σε κάποια προσφορά οπότε το ItemBasicID παραμένει το ίδιο αλλά δημιουργείται νέο ItemID)
  - itemBasicName: Τίτλος βασικού προϊόντος
  - itemSubgroup2ID: Κωδικός υποομάδας 2 που ανήκει το προϊόν
  - itemSubgroup2Name: Όνομα υποομάδας 2 που ανήκει το προϊόν
  - itemSubgroup1ID: Κωδικός υποομάδας 1 που ανήκει το προϊόν
  - itemSubgroup1Name: Όνομα υποομάδας 1 που ανήκει το προϊόν
  - itemGroupID: Κωδικός ομάδας που ανήκει το προϊόν
  - itemGroupName: Όνομα ομάδας που ανήκει το προϊόν
  - itemCategoryID: Κωδικός κατηγορίας που ανήκει το προϊόν
  - itemCategoryName: Όνομα κατηγορίας που ανήκει το προϊόν
  - isPrivateLabel: Ένδειξη αν είναι προϊόν ιδιωτικής ετικέτας
- Stores (6 εγγραφές):
  - storeCode: Κωδικός καταστήματος
  - storeArea: Περιοχή καταστήματος

- storePrefecture: Κωδικός νομου που ανήκει το κατάστημα (από πίνακα prefectures)
- storeZipCode: Ταχ. κωδικός καταστήματος
  
- TransactionItems (6.258.079 εγγραφές):
  - transID: Κωδικός συναλλαγής
  - transItem: Κωδικός προϊόντος
  - transQuantity: Ποσότητα
  - transValue: Αξία
  
- Transactions (1.057.077 εγγραφές):
  - transID: Κωδικός συναλλαγής-μοναδικός
  - transCode: Κωδικός συναλλαγής στο συγκεκριμένο κατάστημα
  - transCustomer: Κωδικός πελάτη
  - transStore: Κωδικός καταστήματος που πραγματοποιήθηκε η συναλλαγή
  - transTameio: Κωδικός οντότητας που πραγματοποιήθηκε η συναλλαγή
  - transDateTime: Ημερομηνία συναλλαγής

Η εφαρμογή του case study που αφορά customer analytics, θα υλοποιηθεί με τη χρήση του λογισμικού SAS Viya for learners. Το SAS Viya for Learners είναι μια προηγμένη πλατφόρμα ανάλυσης δεδομένων από την SAS, μια εταιρεία που εξειδικεύεται στην ανάλυση δεδομένων και τη στατιστική. Το λογισμικό παρέχει εργαλεία για την ανάλυση δεδομένων, τη δημιουργία αναφορών, την επεξεργασία και την οπτικοποίηση δεδομένων, ενώ υποστηρίζει επίσης προγραμματισμό σε διάφορες γλώσσες, όπως η SAS, η Python και η R.<sup>94</sup>

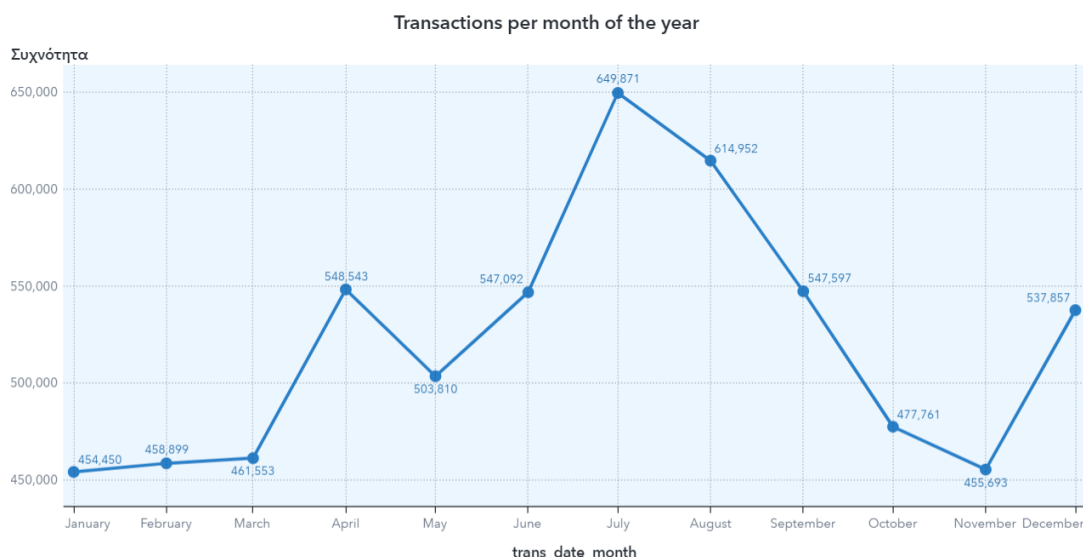
## 3.2 Αρχική εξερεύνηση δεδομένων

Πριν την εκτέλεση οποιασδήποτε ενέργειας επί των δεδομένων, είναι χρήσιμο να εκτελέσουμε μία αρχική εξερεύνηση στα δεδομένα μας. Η αρχική αυτή εξερεύνηση, θα μας δώσει χρήσιμα insights, που πιθανόν να αποκρυφθούν ή να μην είναι τόσο εμφανή μετά από οποιαδήποτε επεξεργασία επί των δεδομένων. Η επεξεργασία που θα πραγματοποιηθεί επί των δεδομένων, αναλύεται στις επόμενες ενότητες.

Ο καλύτερος τρόπος για την αρχική εξερεύνηση επί των δεδομένων, είναι η οπτικοποίησή τους υπό διαφορετικά πρίσματα και οπτικές γωνίες που μπορούν να βοηθήσουν την εταιρεία να βγάλει συμπεράσματα. Η οπτική ανάλυση των δεδομένων είναι εύκολα κατανοητή από τον μέσο χρήστη, οδηγεί γρήγορα σε συμπεράσματα σχετικά με πιθανές διορθώσεις που μπορεί να απαιτηθούν στα δεδομένα και αποκαλύπτει τάσεις που οδηγούν τον αναλυτή να τις εξετάσει περαιτέρω με βαθύτερη ανάλυση.

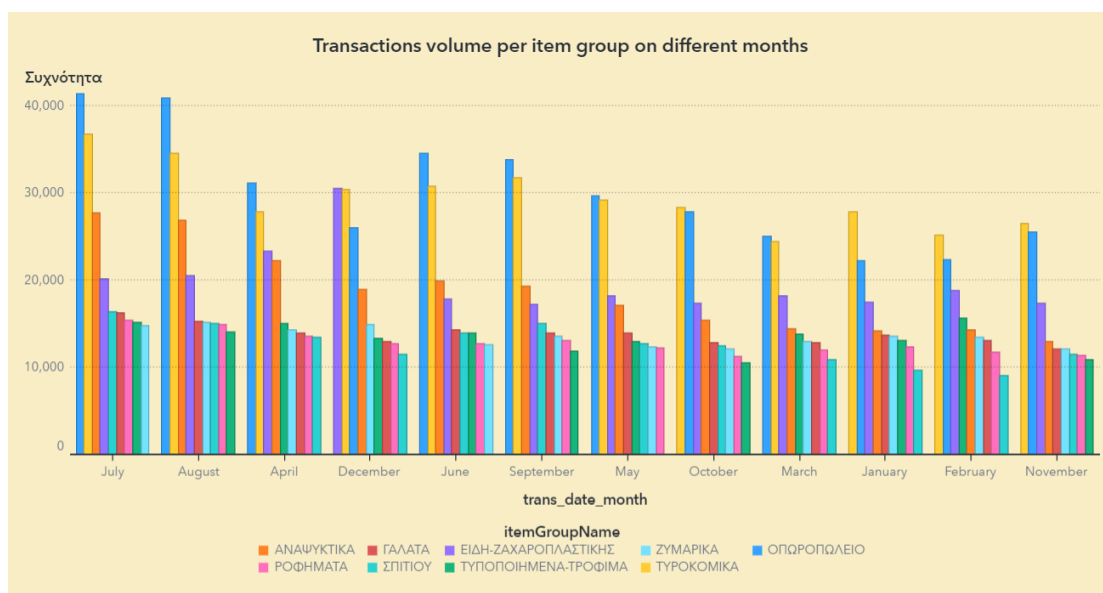
### 3.2.1 Όγκος συναλλαγών

Στα παρακάτω γραφήματα εξετάζεται ο όγκος συναλλαγών και πως αυτός μεταβάλλεται χρονικά.



**Εικόνα 4. Όγκος συναλλαγών ανά μήνα του έτους σε raw data**

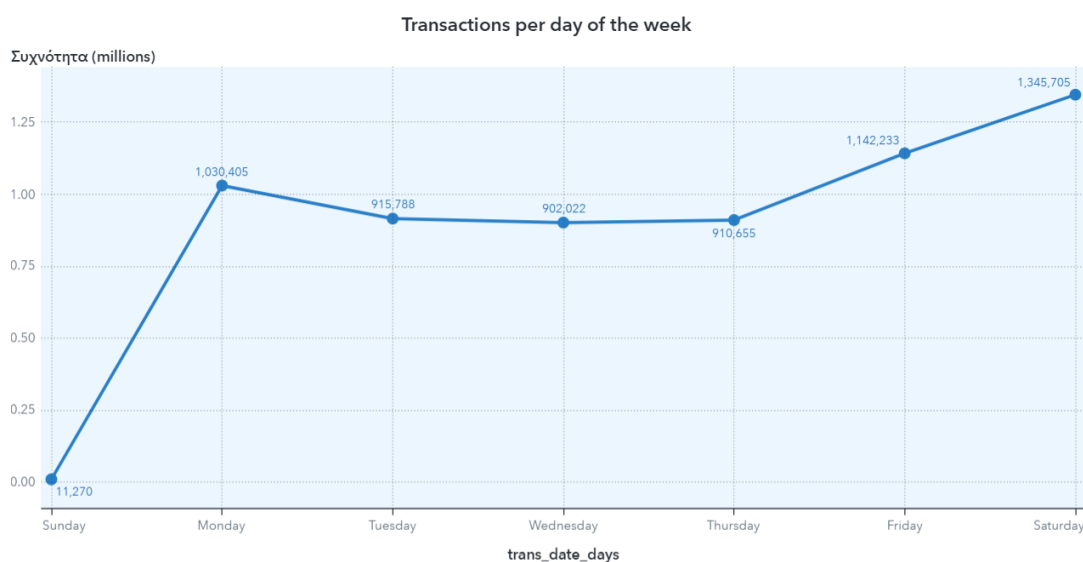
Η ανωτέρω εξερεύνηση των δεδομένων συναλλαγών, παρέχει χρήσιμα συμπεράσματα σχετικά με τον όγκο αυτών. Μακροσκοπικά, οι μήνες καλοκαιρινών διακοπών (Ιούνιος, Ιούλιος, Αύγουστος, Σεπτέμβριος) μαζί με τον Απρίλιο και τον Δεκέμβριο, εμφανίζουν τον μεγαλύτερο όγκο συναλλαγών. Ο τουρισμός και έθιμα που συνδέονται με γιορτινές περιόδους στους δύο τελευταίους μήνες, μπορεί να ευθύνονται για τον μεγάλο αυτό όγκο, όμως απαιτείται περαιτέρω αξιολόγηση.



**Εικόνα 5. Όγκος συναλλαγών ανά μήνα του έτους και ανά ομάδα προϊόντων**

Περαιτέρω εξερεύνηση του μηνιαίου όγκου συναλλαγών, αποτυπώνει τη διακύμανση της συμπεριφοράς των εννέα δημοφιλέστερων ομάδων προϊόντων. Με τον όρο δημοφιλέστερες, αναφερόμαστε στις ομάδες προϊόντων που έχουν τους μεγαλύτερους συνολικούς όγκους συναλλαγών μέσα στο έτος. Από την ανάλυση, αφαιρέθηκαν οι ομάδες προϊόντων «Γαλακτοκομικά» και «Snack-Μπισκότα-Ζαχαρώδη» οι οποίες αντιστοιχούν στις πρώτες δύο ομάδες αντίστοιχα, πλην όμως η κατάταξή τους παραμένει σταθερή εντός του έτους.

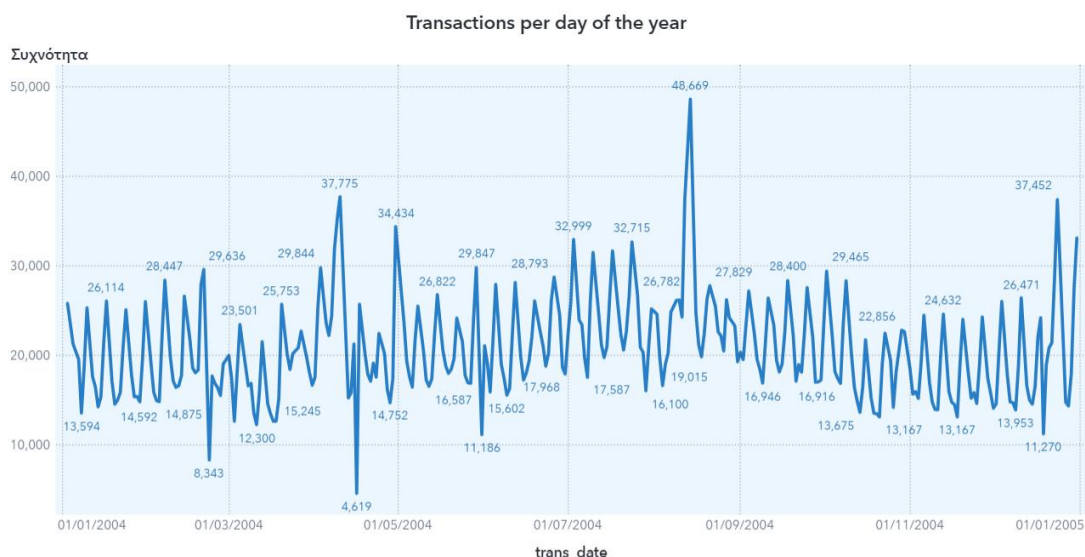
Στο ανωτέρω ομαδοποιημένο ραβδόγραμμα φαίνεται ότι υπάρχουν μεταβολές στην προτίμηση των καταναλωτών σε διάφορες ομάδες προϊόντων. Ενδεικτικά, παρόλο που η ομάδα 'ΟΠΩΡΟΠΩΛΕΙΟ' διατηρεί ως επί το πλείστον ένα από τα τρία υψηλότερα επίπεδα συναλλαγών, παρατηρείται μια σημαντική αλλαγή στην κυριαρχία της κατά τους μήνες Οκτώβριο, Ιανουάριο, Φεβρουάριο και Νοέμβριο, όπου υποχωρεί στη δεύτερη θέση υπολειπόμενη της ομάδας 'ΤΥΡΟΚΟΜΙΚΑ' και της ομάδας «ΕΙΔΗ-ΖΑΧΑΡΟΠΛΑΣΤΙΚΗΣ» κατά τον μήνα Δεκέμβριο. Μία αξιοσημείωτη αλλαγή είναι και εκείνη που αφορά την ομάδα προϊόντων «ΣΠΠΤΙΟΥ»: τα προϊόντα έχουν τις χαμηλότερες τιμές σε προτιμήσεις καταναλωτών τους μήνες Δεκέμβριο-Απρίλιο και τις υψηλότερες τιμές κατά τους μήνες Ιούλιο-Οκτώβριο.



**Εικόνα 6. Όγκος συναλλαγών ανά ημέρα της εβδομάδας σε raw data**

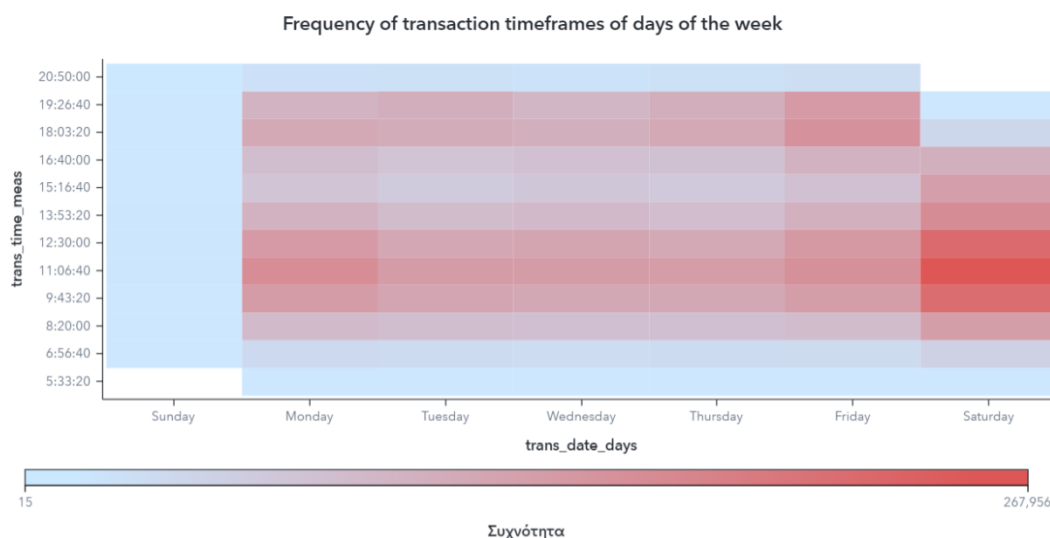
Ειδικεύοντας περαιτέρω το εύρος της ανάλυσης, οι ημέρες της εβδομάδας που εμφανίζουν τον μεγαλύτερο όγκο συναλλαγών είναι το Σάββατο, ακολουθούμενο από την Παρασκευή και τη Δευτέρα, σε αυτήν τη σειρά. Ένας εύκολος τρόπος να ερμηνεύσουμε ίσως τα αποτελέσματα αυτά είναι ότι το Σάββατο αποτελεί την πρώτη μη εργάσιμη μέρα της εβδομάδας, η Παρασκευή η πρώτη μέρα που οι καταναλωτές δεν έχουν να ξυπνήσουν νωρίς την επόμενη ημέρα για να πάνε στη δουλειά, ενώ η Δευτέρα είναι η πρώτη εργάσιμη μέρα της εβδομάδας, που απαιτεί τον ανεφοδιασμό για την υπόλοιπη εβδομάδα. Απαιτείται περαιτέρω ανάλυση επί των αρχικών αυτών υποθέσεων, για ασφαλή συμπεράσματα επ' αυτών.





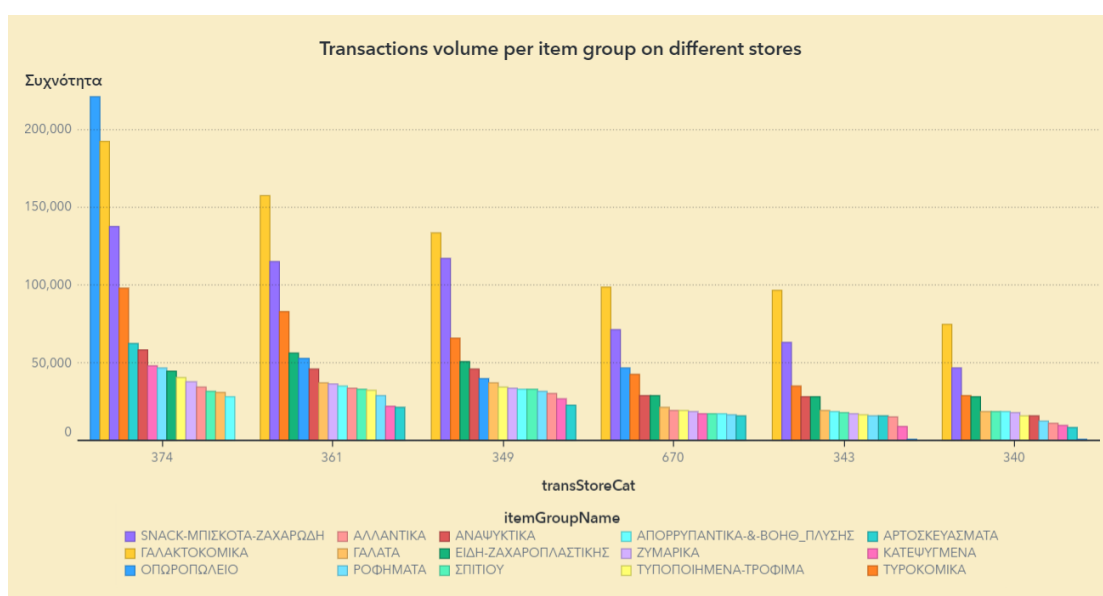
**Εικόνα 7. Όγκος συναλλαγών ανά ημερομηνία του έτους σε raw data**

Μικροσκοπική ανάλυση επί συγκεκριμένων ημερομηνιών του έτους, μας δείχνει ότι συγκεκριμένες ημερομηνίες κατέχουν ηγετική θέση σε όγκο συναλλαγών σε σύγκριση με άλλες. Ο όγκος συναλλαγών έφτασε στο ανώτατο επίπεδο των 48.669 συναλλαγών σε μία μόλις μέρα για το έτος 2004 την 14/08/2004, δηλαδή μία μέρα πριν την επίσημη αργία του δεκαπενταύγουστου. Ο όγκος έφτασε περισσότερο από 10 φορές πάνω(!) από το κατώτερο επίπεδο του ίδιου έτους, δηλαδή 4.619 ημερήσιες συναλλαγές την 16/04/2004. Σε αντιδιαστολή, μόλις μερικές μέρες προηγουμένως, δηλαδή την 10/04/2004, μία μέρα πριν το Πάσχα του 2004, είχαμε το δεύτερο μεγαλύτερο επίπεδο συναλλαγών του έτους, ήτοι 37.775 ημερήσιες συναλλαγές. Δεδομένου ότι και το τρίτο μεγαλύτερο επίπεδο όγκου ημερήσιων συναλλαγών (37.452) συμπίπτει με την 24/12/2004, δηλαδή μία μέρα πριν την αργία των Χριστουγέννων, μπορούμε να εξάγουμε ένα ασφαλές συμπέρασμα στο ότι οι μέρες που προηγούνται των μεγάλων επίσημων αργιών, κατέχουν τα μεγαλύτερα επίπεδα σε όγκο συναλλαγών.



**Εικόνα 8. Όγκος συναλλαγών ανά χρονικά διαστήματα (~1.5 ώρας) ανά ημέρα της εβδομάδας σε raw data**

Βαθύτερη μικροσκοπική εξερεύνηση, αυτήν την φορά επί χρονικών διαστημάτων εντός της ίδιας μέρας, αποκάλυπτουν τις ώρες που οι καταναλωτές τείνουν να συναλλάσσονται περισσότερο με τα καταστήματα της επιχείρησης για τις συναλλαγές τους. Στον ανωτέρω χάρτη θερμότητας (heatmap) το ωράριο εργασίας των καταστημάτων, διαιρέθηκε σε τμήματα διάρκειας περίπου μιάμιση ώρας και εξετάστηκε ο όγκος συναλλαγών σε καθένα από αυτά. Είναι εμφανές ότι κατά τα χρονικά διαστήματα μεταξύ 09:00 και 13:30 το πρωί και μεταξύ 17:20 και 20:00 το απόγευμα, οι καταναλωτές συναλλάσσονται περισσότερο με τα υποκαταστήματα της επιχείρησης σε όλες τις μέρες της εβδομάδας, πλην του Σαββάτου. Την Δευτέρα μεταξύ 10:25 και 11:50 αλλά και την Παρασκευή μεταξύ 17:20 με 18:50 παρατηρείται μία επιπρόσθετη αύξηση στις συναλλαγές της τάξεως του 25%, συγκριτικά με άλλες μέρες. Το Σάββατο οι ώρες που οι καταναλωτές συναλλάσσονται κυρίως με τα υποκαταστήματα είναι μεταξύ 09:00 και 16:00 με κορυφή (peak) μεταξύ 10:30 με 11:50.



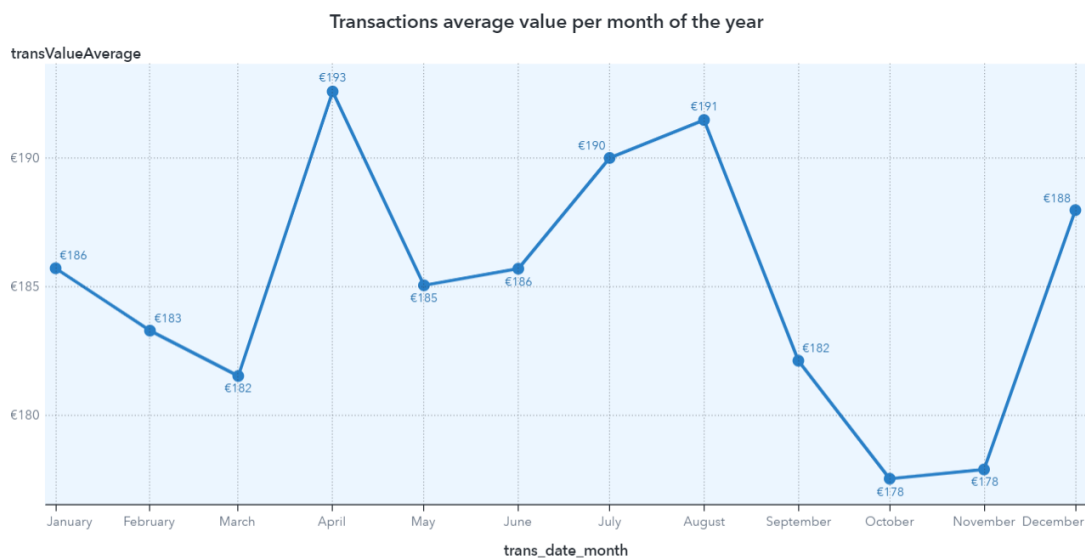
Εικόνα 9. Όγκος συναλλαγών ανά υποκατάστημα και ανά ομάδα προϊόντων

Εκτός από τη «χρονική» ανάλυση, η «τοπική» ανάλυση μπορεί να εξάγει επίσης σημαντική πληροφορία για την επιχείρηση. Στο παραπάνω γράφημα, αποτυπώνεται η προτίμηση των καταναλωτών, όπως αυτή εκφράζεται με τον όγκο συναλλαγών, για διαφορετικές ομάδες προϊόντων, σε διαφορετικά υποκαταστήματα της εταιρείας. Η εξερεύνηση δείχνει μεταξύ άλλων ότι η ομάδα προϊόντων «ΟΠΩΡΟΠΩΛΕΙΟ» έχει κυρίαρχη θέση στο υποκατάστημα 374, ενώ μειώνεται αισθητά η προτίμησή της από τους καταναλωτές στα υπόλοιπα υποκαταστήματα. Το ίδιο ισχύει αναλογικά και για την ομάδα «ΑΡΤΟΣΚΕΥΑΣΜΑΤΑ».

### 3.2.2 Αξία συναλλαγών

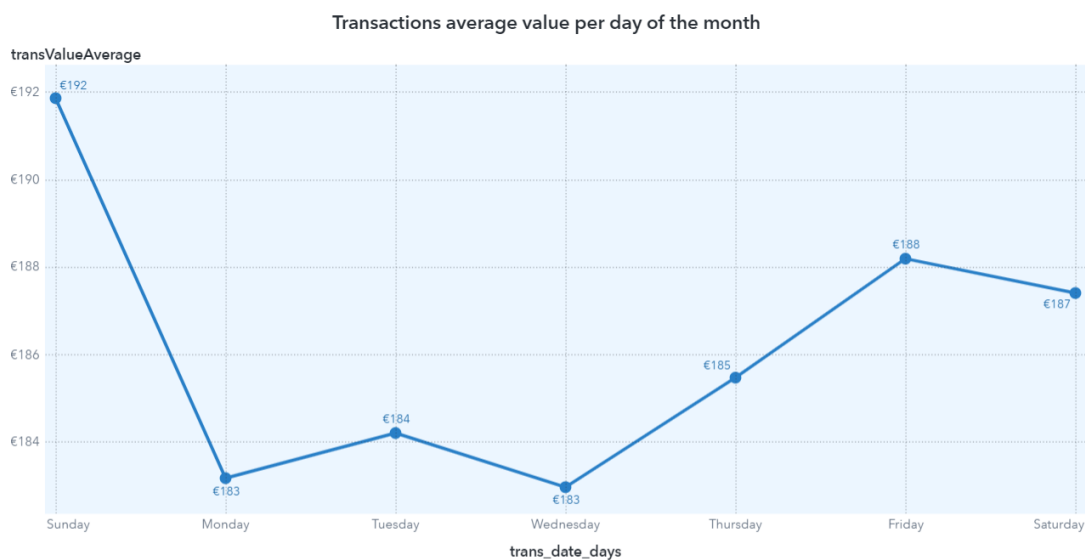
Η εξερεύνηση του όγκου των ημερήσιων συναλλαγών μπορεί να δώσει μία πιο ολοκληρωμένη εικόνα εάν συνδυαστεί με την εξερεύνηση της μέσης αξίας συναλλαγής στο εξεταζόμενο χρονικό διάστημα. Στο σημείο αυτό, είναι σημαντικό να τονιστεί ότι μετρήσεις όπως ο μέσος όρος ενδέχεται να επηρεάζονται από ακραίες τιμές, γεγονός που μπορεί να τους καθιστά λιγότερο αξιόπιστους ως δείκτες της συνολικής εικόνας. Η ευαισθησία του μέσου όρου σε αυτές τις τιμές απαιτεί προσοχή κατά την ερμηνεία των δεδομένων, καθώς μπορεί να μην αντανακλούν πλήρως την πραγματικότητα του συνόλου που μελετάται. Απαιτείται πάντα επιπρόσθετοι έλεγχοι των υποθέσεων για εξαγωγή ασφαλών συμπερασμάτων.

Στα παρακάτω γραφήματα εξετάζεται η (μέση) αξία συναλλαγών και πως αυτή μεταβάλλεται χρονικά.



**Εικόνα 10. Μέση αξία συναλλαγών ανά μήνα του έτους σε raw data**

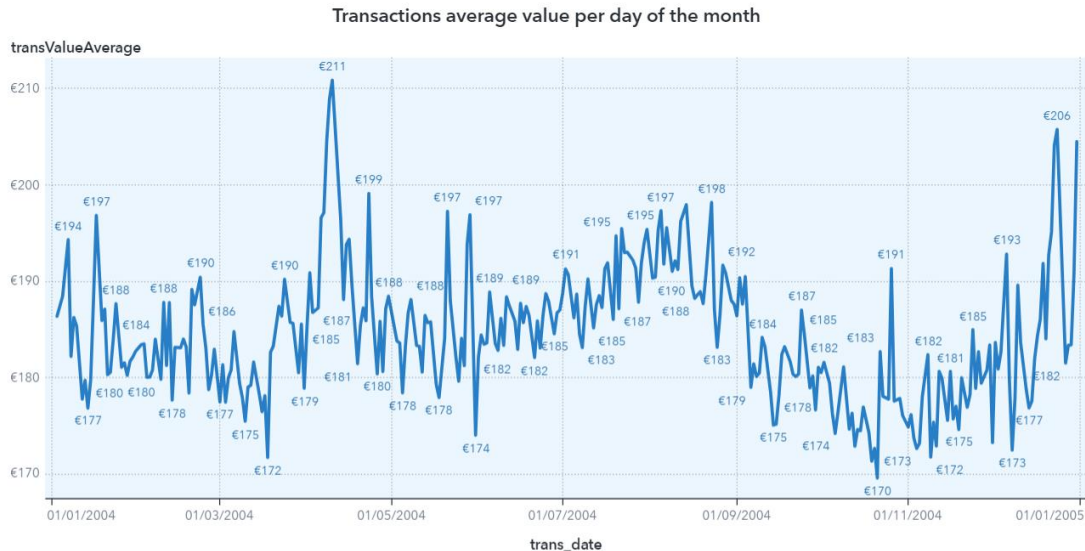
Μακροσκοπική εξερεύνηση της μέσης αξίας συναλλαγής στους μήνες του έτους, δείχνουν ότι η εικόνα της μέσης αξίας συναλλαγών δεν ακολουθεί παρόμοια εικόνα με εκείνη του όγκου συναλλαγών σε ότι αφορά τις κορυφές (peaks) και τα ανώτατα επίπεδα. Ο Απρίλιος έχει το υψηλότερο επίπεδο μέσης αξίας συναλλαγών που δείχνει ότι παρότι οι καταναλωτές συναλλάσσονται σπανιότερα από άλλους μήνες με τα υποκαταστήματα της επιχείρησης, ξοδεύουν περισσότερα ανά συναλλαγή. Κατά συνέπεια, η κερδοφορία της επιχείρησης είναι πιθανότατα πολύ ανώτερη εκείνον τον μήνα απ' ότι τους μήνες Ιούνιο, Σεπτέμβριο και Δεκέμβριο που έχουν ίδια επίπεδα όγκου συναλλαγών αλλά πολύ κατώτερη μέση αξία συναλλαγών.



**Εικόνα 11. Μέση αξία συναλλαγών ανά ημέρα της εβδομάδας σε raw data**

Εξειδικεύοντας περαιτέρω την εξερεύνηση, στις ημέρες της εβδομάδας, παρατηρείται ότι η Κυριακή κατέχει το μεγαλύτερο επίπεδο μέσης αξίας συναλλαγών. Αυτό δείχνει πιθανότατα ότι, παρά το γεγονός ότι κατά βάση οι συναλλαγές δεν είναι εφικτό να πραγματοποιηθούν τις Κυριακές, όταν είναι για

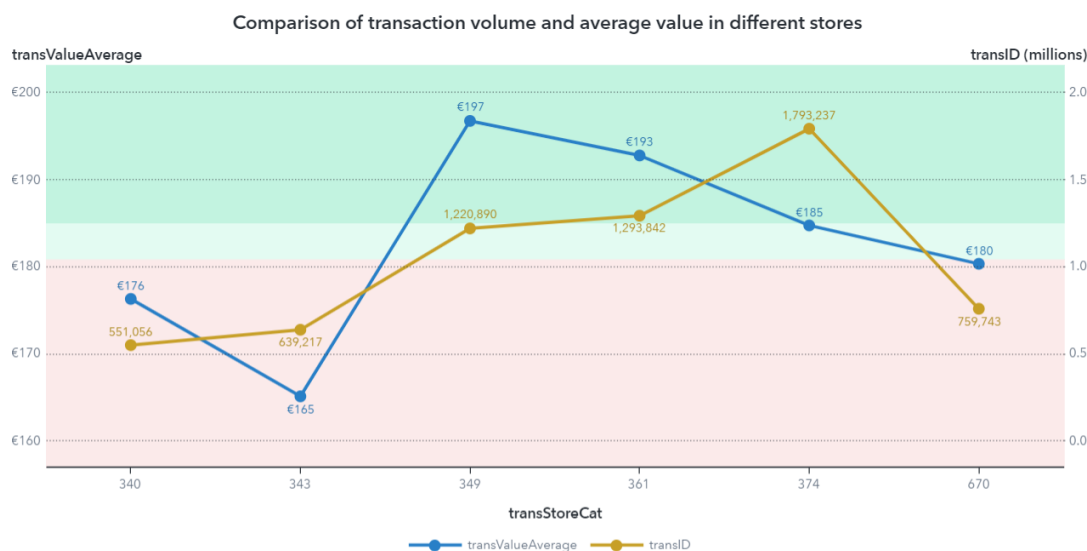
εξαιρετικούς λόγους διαθέσιμες (π.χ. εορτές), η μέση αξία καλαθίου αυξάνεται κατά έως και 10 ευρώ. Επιπρόσθετα, παρατηρούμε ότι παρόλο που το Σάββατο κατέχει μεγαλύτερο όγκο συναλλαγών από την Παρασκευή, οι συναλλαγές που πραγματοποιούνται την Παρασκευή είναι πιο κερδοφόρες για την εταιρεία.



**Εικόνα 12. Μέση αξία συναλλαγών ανά ημερομηνία του έτους σε raw data, με παράλληλη εμφάνιση τάσης**

Μικροσκοπική ανάλυση επί συγκεκριμένων ημερομηνιών του έτους, μας δείχνει ξανά ότι συγκεκριμένες ημερομηνίες κατέχουν ηγετική θέση σε μέση αξία συναλλαγών σε σύγκριση με άλλες. Οι ημερομηνίες 10/04/2004 και 24/12/2004 έχουν τις δύο υψηλότερες μέσες αξίες συναλλαγών, πολύ ανώτερες από την 14/08/2004, αντιστρέφοντας την σχέση που υπήρχε σε ότι αφορά τον όγκο ημερήσιων συναλλαγών. Δηλαδή, παρά το γεγονός ότι οι καταναλωτές συναλλάσσονται συχνότερα με την επιχείρηση πριν τον δεκαπενταύγουστο, συγκριτικά με τις άλλες δύο σημαντικές αργίες του έτους, ξοδεύουν τελικά πολύ λιγότερα ανά συναλλαγή. Μέτρα που βοηθάνε στην αύξηση των συναλλαγών των καταναλωτών αυτές τις ημέρες, θα οδηγήσουν σε μεγαλύτερα αναλογικά κέρδη απ' ότι σε άλλες ημέρες του έτους.

Είναι εξίσου σημαντικό να παρατηρηθεί η τάση που ακολουθείται στο μοτίβο μέσης αξίας συναλλαγών και η οποία απεικονίζεται στην *Εικόνα 12*. Μέση αξία συναλλαγών ανά ημερομηνία του έτους σε raw data, με παράλληλη εμφάνιση τάσης με τη μορφή ενός διακεκομμένου κόκκινου βέλους. Επαληθεύοντας τα ευρήματα της μακροσκοπικής ανάλυσης μηνών του έτους, οι καταναλωτές φαίνεται να ακολουθούν μία επαναλαμβανόμενη συμπεριφορά: είναι διατεθειμένοι να ξοδέψουν περισσότερο όσο πλησιάζουν σε αργίες αλλά ξοδεύουν λιγότερο όσο απομακρύνονται από αυτές. Το επαναλαμβανόμενο αυτό μοτίβο, μπορεί να χρησιμοποιηθεί για να προσαρμοστούν αντίστοιχα οι εταιρικές στρατηγικές σε ότι αφορά τις πωλήσεις.



**Εικόνα 13.** Συνδυαστική ανάλυση όγκου και μέσης αξίας συναλλαγών (καφέ και μπλε γραμμή αντίστοιχα) σε διαφορετικά υποκαταστήματα της εταιρείας. Το κόκκινο φόντο του διαγράμματος αντιπροσωπεύει τις τιμές κάτω από τους μέσους όρους των δύο μετρικών, το ανοιχτό πράσινο τις τιμές που βρίσκονται πάνω από το μέσο όρο μέσης αξίας συναλλαγών και το σκούρο πράσινο τις τιμές που είναι μεγαλύτερες από τους μέσους όρους και των δύο μετρικών συνδυαστικά.

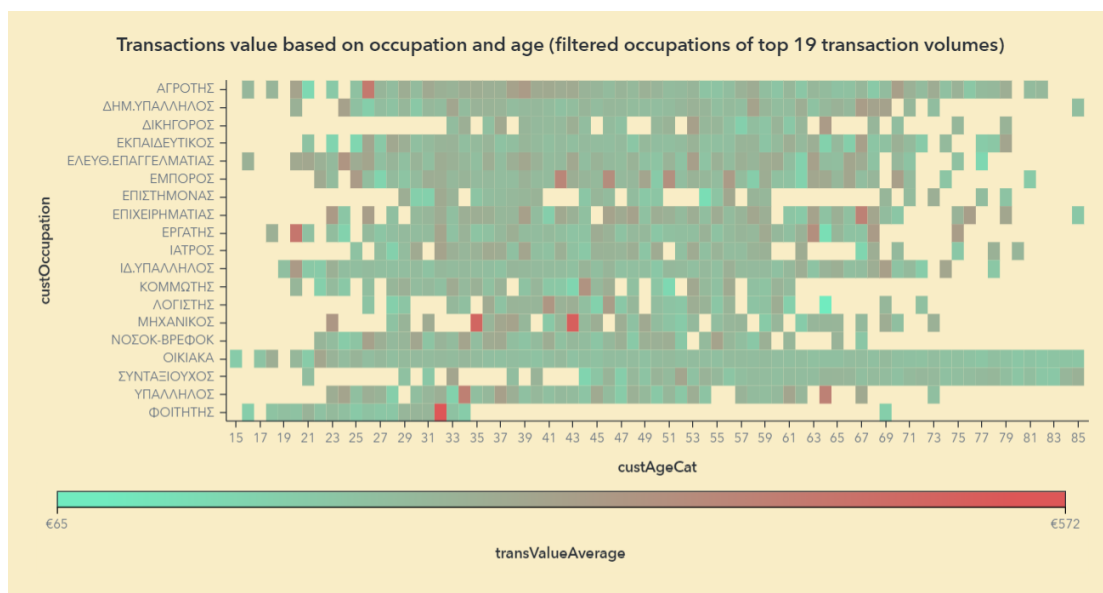
Το ανωτέρω γράφημα, απεικονίζει συνδυαστικά τον συνολικό όγκο συναλλαγών και τη μέση αξία συναλλαγών για τα διαφορετικά υποκαταστήματα της εταιρείας, για τα οποία υπάρχουν δεδομένα. Οι χρωματικές αποχρώσεις του φόντου του γραφήματος απεικονίζουν ποιες τιμές βρίσκονται κάτω και πάνω από τους δύο μέσους όρους αντίστοιχα, ως εξής:

- Σκούρο πράσινο: Όγκος και μέση αξία συναλλαγών μεγαλύτερα από τους αντίστοιχους μέσους όρους
- Ανοιχτό πράσινο: Μέση αξία συναλλαγών μεγαλύτερη του μέσου όρου, αλλά όγκος συναλλαγών μικρότερος του μέσου όρου
- Κόκκινο: Μέση αξία συναλλαγών και όγκος συναλλαγών μικρότερα από τους αντίστοιχους μέσους όρους

Με βάση τα παραπάνω, τα υποκαταστήματα 349, 361 και 374 είναι τα πιο κερδοφόρα υποκαταστήματα με μέση αξία συναλλαγών και όγκο συναλλαγών συνδυαστικά μεγαλύτερες από τους αντίστοιχους μέσους όρους. Αντίστροφα, τα καταστήματα 340, 343 και 670, έχουν χαμηλότερες τιμές από τους αντίστοιχους μέσους όρους, με το κατάστημα 343 να έχει την χειρότερη απόδοση μεταξύ των υποκαταστημάτων.

### 3.2.3 Δημογραφικά χαρακτηριστικά και συναλλαγές

Με βάση τα διαθέσιμα δεδομένα, φαίνεται εφικτό να πραγματοποιηθεί μία αρχική έστω εξερεύνηση των δημογραφικών χαρακτηριστικών των καταναλωτών της επιχείρησης. Η εξερεύνηση αυτή μπορεί να γίνει πιο εξειδικευμένα υπό το πρίσμα του πως τα χαρακτηριστικά αυτά μπορεί να επηρεάσουν τη συμπεριφορά των καταναλωτών σε ότι αφορά τις συνδιαλλαγές τους με την επιχείρηση. Η επιχείρηση μπορεί να αντλήσει τα αποτελέσματα της εξερεύνησης, προκειμένου να προσαρμόσει αντίστοιχα τις στρατηγικές μάρκετινγκ ή πωλήσεων και να αντιληφθεί ποιες υποθέσεις απαιτούν βαθύτερη διερεύνηση.

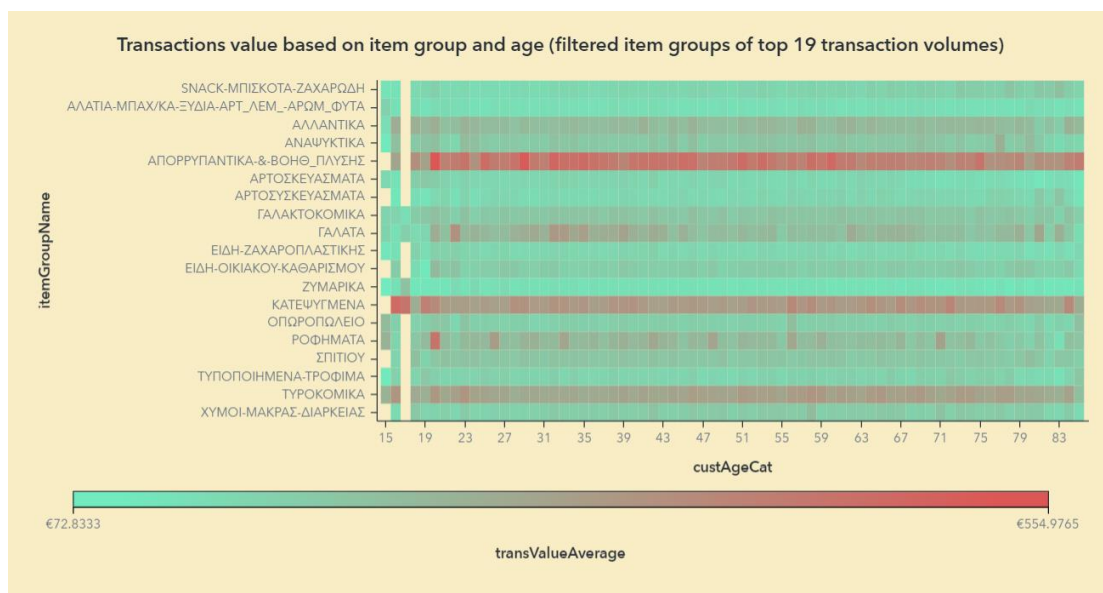


**Εικόνα 14. Θερμικός χάρτης μέσης αξίας συναλλαγών διαφορετικών επαγγελματιών και ηλικιών, βάση των δεκαεννέα επαγγελμάτων με τον μεγαλύτερο όγκο συναλλαγών**

Στον ανωτέρω θερμικό χάρτη (heatmap), αποτυπώνεται η μέση αξία συναλλαγών για διαφορετικά επαγγέλματα και ηλικίες. Για την απεικόνιση, επιλέχθηκαν τα δεκαεννέα επαγγέλματα με τον μεγαλύτερο όγκο συναλλαγών – και άρα τα δεκαεννέα εκείνα που έχουν μεγαλύτερη επιρροή στις συναλλαγές της επιχείρησης. Παρατηρούνται κάποιοι συνδυασμοί επαγγελμάτων-ηλικιών που έχουν πολύ μεγαλύτερη μέση αξία συναλλαγών (και ταυτόχρονα όγκο όπως ειπώθηκε προηγουμένως) από τους υπόλοιπους συνδυασμούς. Μερικοί τέτοιοι συνδυασμοί είναι:

- Φοιτητής - 32 ετών
- Υπάλληλος - 24, 34, 37, 61 και 64 ετών
- Μηχανικός - 35 και 43 ετών
- Συνταξιούχος - 52 ετών
- Οικιακά - 22 ετών
- Αγρότης 26 ετών
- Έμπορος – 42, 46, 49, 51, 56 και 70 ετών
- Λογιστής - 41 ετών κ.ο.κ.

Επένδυση στην προσέλκυση περισσότερων πελατών με το συνδυασμό αυτών των χαρακτηριστικών ή καλύτερης εκμετάλλευσης αυτών, μπορεί να έχει εξαιρετικά θετική επίπτωση στην εταιρική κερδοφορία.



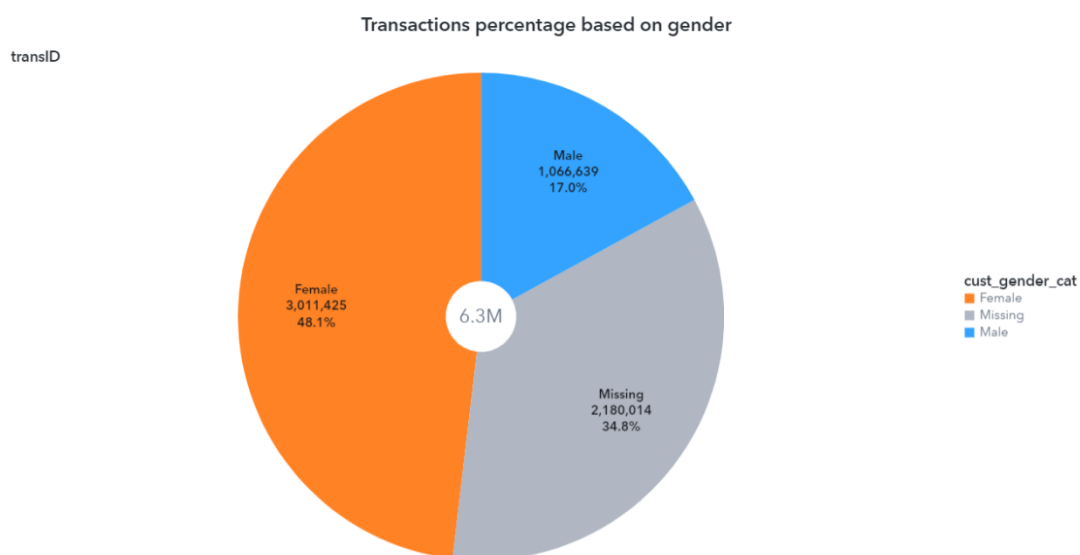
**Εικόνα 15. Θερμικός χάρτης μέσης αξίας συναλλαγών διαφορετικών ηλικιών για διαφορετικές ομάδες προϊόντων, βάση των δεκαεννέα ομάδων προϊόντων με τον μεγαλύτερο συνολικό όγκο συναλλαγών**

Στο ανωτέρω γράφημα απεικονίζεται η μέση αξία συναλλαγών για διαφορετικές ηλικίες και για διαφορετικές ομάδες προϊόντων. Με άλλα λόγια, εμφανίζονται οι προτιμήσεις καταναλωτών κάθε ηλικίας για επιμέρους ομάδες προϊόντων. Οι επιμέρους ομάδες προϊόντων που επιλέχθηκαν προς απεικόνιση, είναι εκείνες που κατέχουν τις δεκαεννέα υψηλότερες θέσεις σε συνολικό όγκο συναλλαγών. Είναι εμφανές ότι συγκεκριμένες ηλικίες έχουν μεγαλύτερη προτίμηση σε συγκεκριμένα προϊόντα από ότι άλλες. Μερικά παραδείγματα τέτοιων συνδυασμών είναι οι εξής:

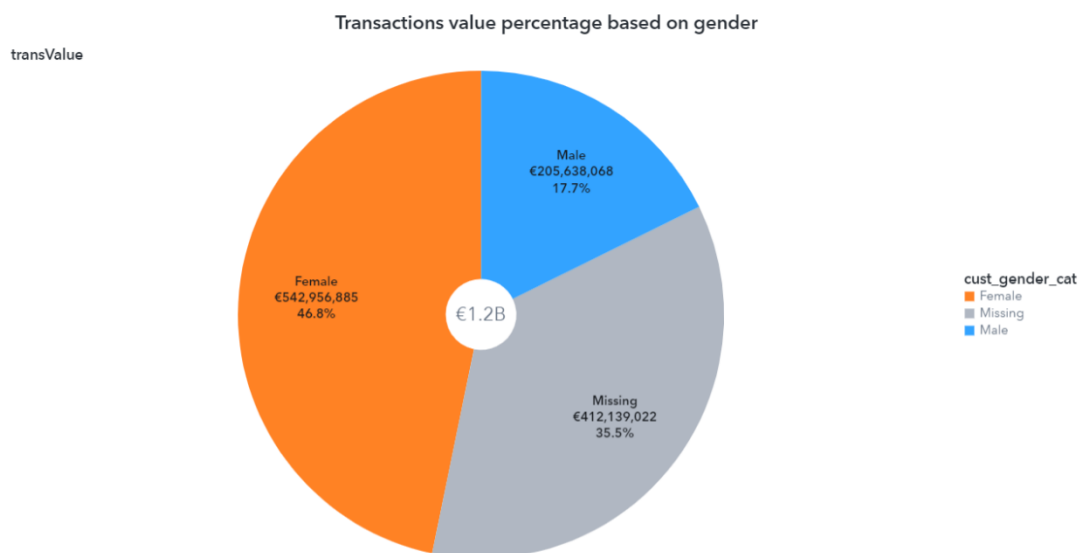
- Κατεψυγμένα – 16-20 ετών, 72,77 και 84 ετών
- Ροφήματα – 20, 26, 33, 48 56, 59 και 71 ετών
- Γάλατα – 20, 22, 29, 30, 32, 33, 35, 39, 62, 79, 81 και 83 ετών
- Είδη οικιακού καθαρισμού – 20, 22, 79 και 80 ετών κ.ο.κ.

Προωθητικές ενέργειες για αυτές τις ηλικίες στα αντίστοιχα προϊόντα μπορεί να έχουν αυξημένη αποτελεσματικότητα.





**Εικόνα 16. Ποσόστωση όγκου συναλλαγών επί του συνόλου, βάσει φύλου, συμπεριλαμβανομένων missing values**



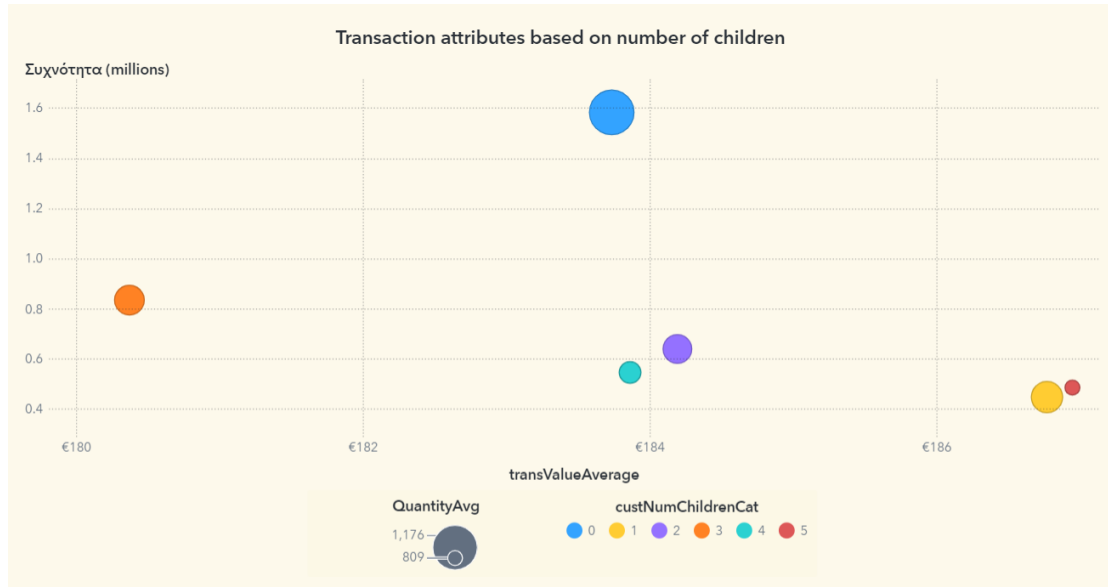
**Εικόνα 17. Ποσόστωση μέσης αξίας συναλλαγών επί του συνόλου, βάσει φύλου, συμπεριλαμβανομένων missing values**

Στα παραπάνω γραφήματα πίτας, αποτυπώνεται το ποσοστό κατά το οποίο συμβάλλει κάθε διαφορετικό φύλο, επί του συνόλου των συναλλαγών. Οι γυναίκες φαίνεται να συμβάλλουν σχεδόν κατά το ήμισυ τόσο σε όσον αφορά τον όγκο συναλλαγών (48,1%), όσο και σε όσον αφορά την συνολική αξία των συναλλαγών (46,8%). Αντίθετα, οι άνδρες φαίνεται να συμβάλλουν μόνο κατά 17% και 17,7% αντίστοιχα επί των συνολικών συναλλαγών, δηλαδή περίπου το ένα τρίτο σε βαρύτητα από εκείνη των γυναικών.

Επιπρόσθετα όμως παρατηρείται ότι οι απεικονίσεις αυτές δεν είναι καθόλου αξιόπιστες για τα δύο φύλα, διότι περίπου στο ένα τρίτο των παρατηρήσεων για τις συναλλαγές (34,8% και 35,5% αντίστοιχα), δεν υπάρχει καταχώριση σχετικά με το φύλο του καταναλωτή που πραγματοποίησε την αντίστοιχη συναλλαγή. Αυτό δημιουργεί πρόβλημα εξαγωγής ασφαλών συμπερασμάτων για την επιχείρηση και δείχνει ότι θα πρέπει να δώσει μεγαλύτερη βαρύτητα σε αυτήν τη καταχώριση, εάν θεωρεί αυτό το



δεδομένο ως σημαντικό για την ανάλυση των πελατών της. Οι τιμές που λείπουν μπορεί να έδιναν μία τελείως διαφορετική εικόνα σχετικά με την ισορροπία (ή μη) μεταξύ των δύο φύλων, να φανέρωναν νέα φύλα (binary ή άλλως προσδιορισμένο) και άλλα στοιχεία σημαντικά για αυτό το δημογραφικό χαρακτηριστικό και πως επηρεάζει ποσοστιαία τις συνολικές πωλήσεις.

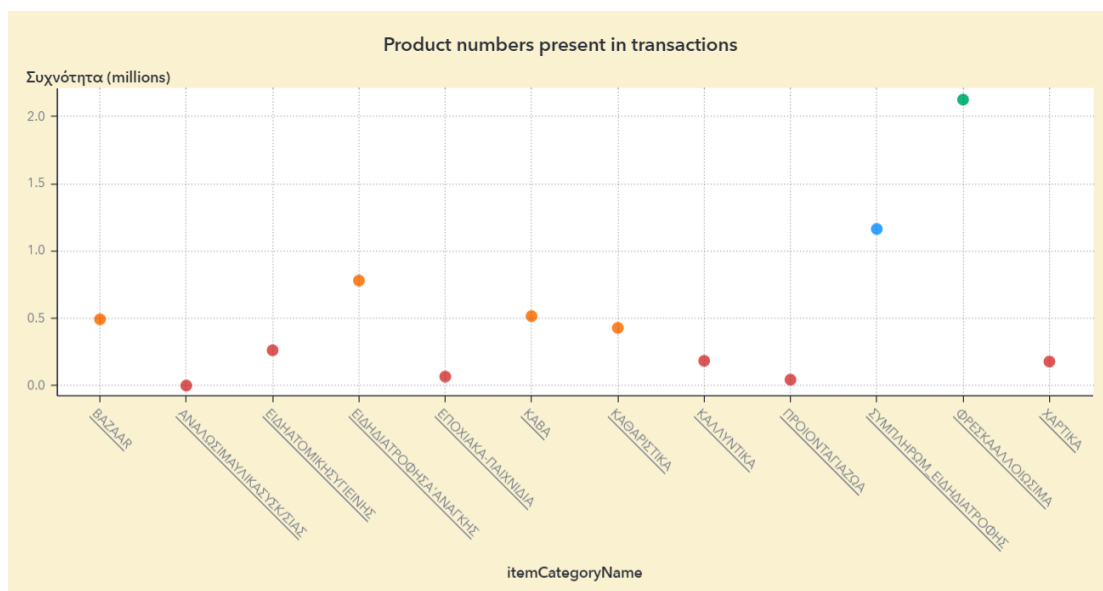


**Εικόνα 18. Διάγραμμα φυσαλίδας με όγκο συναλλαγών (άξονας Y), μέση αξία συναλλαγών (άξονας X), μέση συνολική ποσότητα προϊόντων καλαθιού (μέγεθος φυσαλίδας) και αριθμό παιδιών καταναλωτή (χρώμα φυσαλίδας)**

Στο παραπάνω γράφημα, αποτυπώνεται η σχέση μεταξύ του όγκου συναλλαγών, της μέσης αξίας συναλλαγών και της μέσης ποσότητας αγορασθέντων προϊόντων, από καταναλωτές με διαφορετικό αριθμό τέκνων. Παρατηρείται ότι οι πελάτες που δεν έχουν καθόλου τέκνα, έχουν τον μεγαλύτερο όγκο συναλλαγών αλλά φυσιολογική/μέτρια μέση αξία αγορών. Πελάτες της επιχείρησης σε ένα ή πέντε παιδιά έχουν υψηλή μέση αξία αγορών αλλά πολύ χαμηλή συχνότητα (όγκο) αγορών και μέτρια μέση ποσότητα αγορασμένων προϊόντων. Από την άλλη πλευρά, πελάτες με 2-4 τέκνα έχουν μέτρια με χαμηλή μέση αξία συναλλαγών, χαμηλό όγκο συναλλαγών και μέτρια μέση ποσότητα αγορασμένων προϊόντων. Απαιτείται περαιτέρω έρευνα για να μπορούμε να καταλήξουμε σε ασφαλή συμπεράσματα, σε ότι αφορά το ποιοι πελάτες έχουν τελικά τον θετικότερο αντίκτυπο στα εταιρικά έσοδα.

### 3.2.4 Συχνότητα εμφάνισης προϊόντων

Η σημαντικότητα των διάφορων προϊόντων, συγκριτικά με τα υπόλοιπα, μπορεί να επιτευχθεί και με την εξερεύνηση της συχνότητας εμφάνισής τους στις συναλλαγές. Η αντίστροφη αυτή προσέγγιση, σε σύγκριση με ότι είδαμε παραπάνω, γίνεται και με απόλυτους αριθμούς, αλλά και ποσοστιαία, όπως φαίνεται παρακάτω.



Εικόνα 19. Συχνότητα εμφάνισης κατηγοριών προϊόντων, εντός των συνολικών συναλλαγών

Στον κάθετο άξονα του παραπάνω γραφήματος κουκίδων (Dot Plot), εκφράζεται σε απόλυτο αριθμό η συχνότητα εμφάνισης κάθε κατηγορίας προϊόντων στις συναλλαγές. Η απόχρωση της κουκίδας της κάθε κατηγορίας, αντιστοιχεί στο ποσοστό συχνότητας εμφάνισης της κατηγορίας επί του συνόλου, ως εξής:

- Πράσινο χρώμα: Ποσοστό εμφάνισης  $\geq 30\%$
- Μπλε χρώμα:  $30\% > \text{Ποσοστό εμφάνισης} \geq 15\%$
- Πορτοκαλί χρώμα:  $15\% > \text{Ποσοστό εμφάνισης} \geq 5\%$
- Κόκκινο χρώμα:  $5\% > \text{Ποσοστό εμφάνισης}$

Παρατηρείται ότι η κατηγορία «Φρέσκα Αλλοιώσιμα» αποτελεί περισσότερο από το 30% των προϊόντων που μπαίνουν στο καλάθι των καταναλωτών. Αντίστοιχα, πάνω από 15% των συνολικά αγορασμένων προϊόντων, ανήκουν στην κατηγορία «Συμπληρωματικά είδη διατροφής».

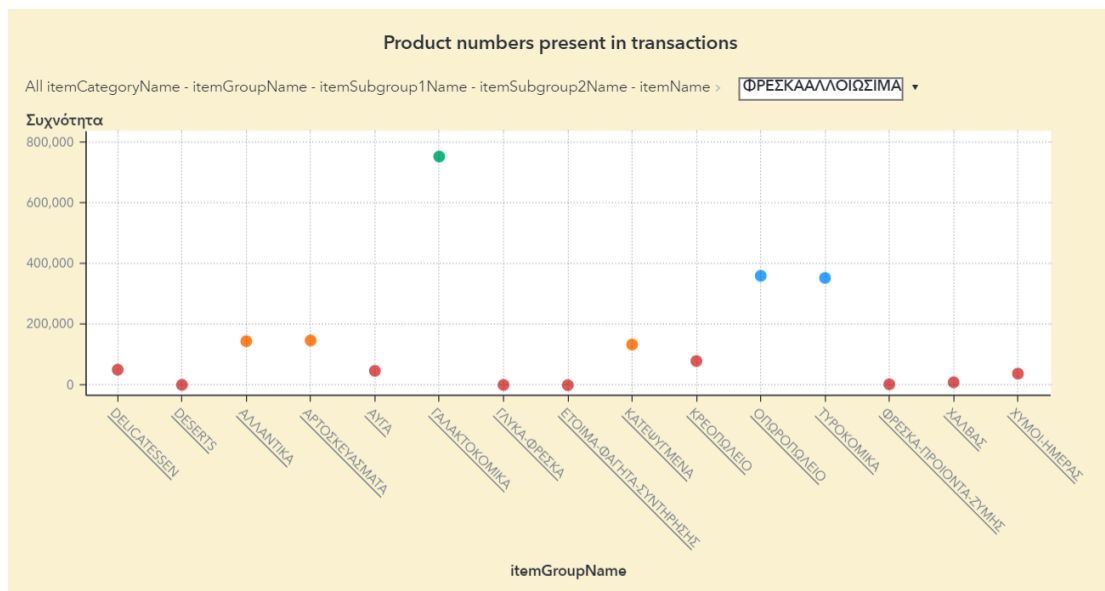
Η ιεραρχία που ακολουθείται στη δομή των προϊόντων στη βάση δεδομένων, μας επιτρέπει να εμβαθύνουμε την ανάλυση σε χαμηλότερες βαθμίδες αναπαράστασής τους. Πιο συγκεκριμένα, τα προϊόντα της βάσης δεδομένων ακολουθούν την εξής δομή:

Κατηγορία Προϊόντος > Ομάδα Προϊόντος > Υπο-ομάδα Προϊόντος > Όνομα Προϊόντος

Για παράδειγμα, υφίσταται η παρακάτω «διαδρομή»:

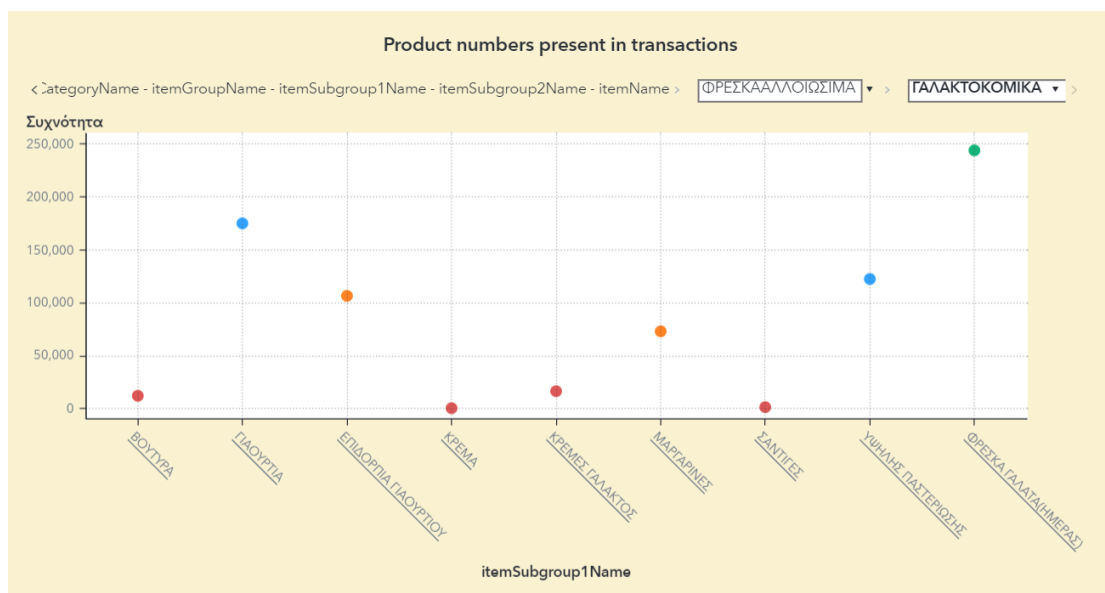
«Φρέσκα Αλλοιώσιμα» > «Γαλακτοκομικά» > «Γιαούρτια» > «Αγελαδινό»

Με βάση αυτά τα στοιχεία, δημιουργείται ένα διαδραστικό γράφημα, του οποίου στιγμιότυπα εφαρμογής του παραδείγματος αυτού φαίνονται παρακάτω.



Εικόνα 20. Συχνότητα εμφάνισης ομάδας προϊόντων, εντός των συνολικών συναλλαγών

Στο παραπάνω γράφημα, παρατηρούμε ότι η ομάδα «Γαλακτοκομικά» αποτελεί πάνω από 30% των προϊόντων εντός των συναλλαγών για προϊόντα που αφορούν την κατηγορία «Φρέσκα Αλλοιώσιμα», ενώ οι ομάδες «Τυροκομικά» και «Οπωροπωλείο» αποτελούν πάνω από 15% των αντίστοιχων προϊόντων.



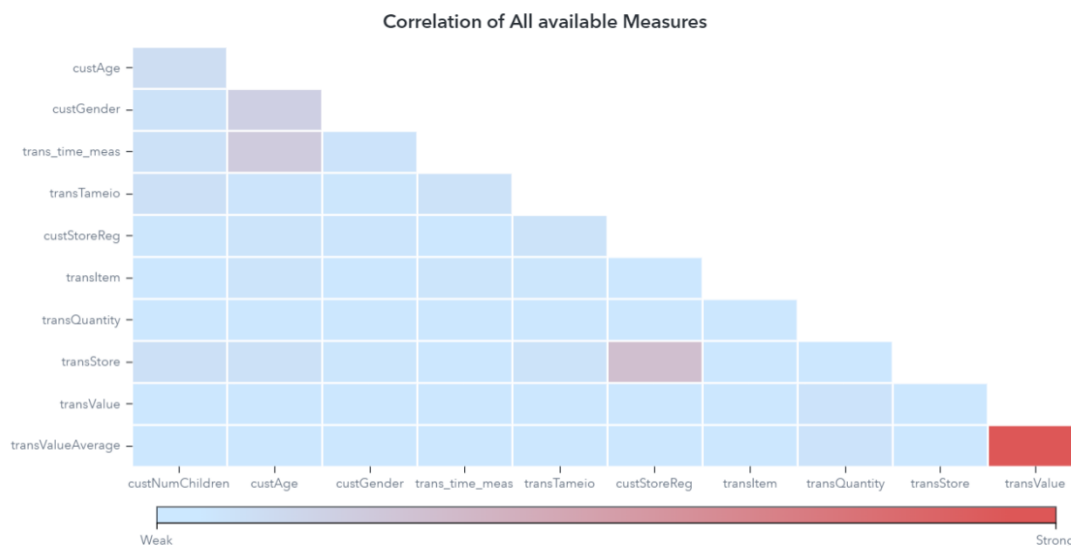
Εικόνα 21. Συχνότητα εμφάνισης υπο-ομάδας 1 προϊόντων, εντός των συνολικών συναλλαγών

Στο παραπάνω γράφημα, παρατηρούμε ότι η υπο-ομάδα 1 «Φρέσκα Γάλατα Ημέρας» αποτελεί πάνω από 30% των προϊόντων εντός των συναλλαγών για προϊόντα που αφορούν την ομάδα «Γαλακτοκομικά», ενώ οι υπο-ομάδες «Γιαούρτια» και «Υψηλής Παστερίωσης» αποτελούν πάνω από 15% των αντίστοιχων προϊόντων.



- Ποιες κατηγορίες/ομάδες και υποομάδες κυριαρχούν στις προτιμήσεις των καταναλωτών
- Ποια συγκεκριμένα προϊόντα καταλαμβάνουν τις ανώτερες θέσεις στις προτιμήσεις αυτές
- Πιθανή αναντιστοιχία μεταξύ σημαντικότητας μίας κατηγορίας/ομάδας/υπο-ομάδας και πραγματικότητας (π.χ. η διασημότητα ενός μόνο προϊόντος μπορεί να δώσει την ψευδαίσθηση ότι οι καταναλωτές προτιμούν όλη την αντίστοιχη ομάδα προϊόντων, αντί του συγκεκριμένου προϊόντος)

### 3.2.5 Εξερεύνηση συσχετίσεων



Εικόνα 24. Πίνακας συσχετίσεων των μεταβλητών της βάσης δεδομένων

Στον ανωτέρω πίνακα συσχετίσεων, παρατηρούμε τις συσχετίσεις μεταξύ διαφορετικών μεγεθών εντός της βάσης δεδομένων μας. Οι συσχετίσεις αυτές, που συνήθως μετρούνται μέσω του συντελεστή συσχέτισης του Pearson, μπορούν να αποκαλύψουν γραμμικές σχέσεις μεταξύ διαφόρων μετρικών και δεδομένων και πώς η μεταβολή μίας μεταβλητής ενδέχεται να συνδέεται με τη μεταβολή μιας άλλης. Είναι σημαντικό, ωστόσο, να γίνει αντιληπτό ότι η συσχέτιση δεν υποδηλώνει απαραίτητα αιτιώδη σχέση μεταξύ των αντίστοιχων μεγεθών. Αυτές οι πληροφορίες όμως, μπορούν να χρησιμοποιηθούν ως βάση για πιο πολύπλοκες αναλύσεις, όπως η πολυμεταβλητή ανάλυση (multivariate analysis) ή η δημιουργία προγνωστικών μοντέλων, προκειμένου να κατανοήσουμε καλύτερα πώς η αλλαγή ενός μεγέθους ενδέχεται να επηρεάσει άλλα μεγέθη.

Στον πίνακα, παρατηρείται ότι κατά βάση οι μεταβλητές δεν σχετίζονται γραμμικά μεταξύ τους. Παρατηρούμε μία μικρή συσχέτιση μεταξύ του καταστήματος εγγραφής του καταναλωτή και του καταστήματος όπου λαμβάνει χώρα μία συναλλαγή από αυτόν. Επιπρόσθετα, μία μικρή συσχέτιση παρατηρείται μεταξύ ηλικίας πελάτη και ώρας συναλλαγής αλλά και μεταξύ ηλικίας και φύλου πελάτη. Αυτές όμως οι συσχετίσεις, δεν μπορούν να δώσουν μία σημαντική πληροφορία που θα μας βοηθήσει σε μετέπειτα αναλύσεις.

Παρατηρείται ότι ακόμα και από μία γρήγορη εξέταση των δεδομένων, μπορούν συνήθως να εξαχθούν χρήσιμα συμπεράσματα για την επιχείρηση. Όμως θα πρέπει να εξεταστούν και πιο προηγμένες τεχνικές και πιο πολύπλοκες αναλύσεις, προκειμένου η επιχείρηση να αποκτήσει ανταγωνιστικό πλεονέκτημα. Τέτοιες αναλύσεις θα πραγματοποιηθούν παρακάτω, ακολουθώντας όσα αναλύθηκαν στο θεωρητικό τμήμα της παρούσας εργασίας.

### 3.3 Προετοιμασία δεδομένων

#### 3.3.1 Εισαγωγή και μετατροπή βάσης δεδομένων

Πριν από οποιαδήποτε ανάλυση των διαθέσιμων δεδομένων, απαιτείται η κατάλληλη προετοιμασία τους. Μιας και τα αρχικά δεδομένα έχουν μορφή .csv (comma separated values), το οποίο είναι σύνηθες σε πολλές βάσεις δεδομένων, θα πρέπει να μετατραπούν (μετά την εισαγωγή τους) σε μορφή που να μπορεί να χρησιμοποιηθεί από το λογισμικό διαχείρισης και ανάλυσης τους. Κατά συνέπεια, τα δεδομένα εισήχθησαν στο λογισμικό SAS, μέσω της ακολουθίας κώδικα που φαίνεται στο παράρτημα της εργασίας.

Ο κώδικας μετατρέπει τα αρχεία της βάσης δεδομένων σε αρχεία τύπου sas7bdat, που είναι αρχεία με δυαδική μορφή που μπορούν να διαβαστούν βέλτιστα από το λογισμικό σε μορφή πινάκων.

#### 3.3.2 Μορφοποίηση και καθαρισμός δεδομένων

Τα επιχειρησιακά δεδομένα που προέρχονται από αληθινές εμπορικές συναλλαγές, μπορούν να παρουσιάσουν πολλές δυσκολίες κατά τη διαχείρισή τους. Στην περίπτωση της αλυσίδας σουπερ μάρκετ, από την οποία αντλείται η υπάρχουσα βάση δεδομένων, τέτοια θέματα μπορεί να είναι:

##### α) Λάθος καταχώριση δεδομένων

Ο πελάτης μπορεί να συμπληρώσει λάθος δεδομένα σε λάθος πεδίο κατά την αίτηση εγγραφής του. Για παράδειγμα, μπορεί να σημειώσει το όνομά του στο πεδίο περιοχή ή μπορεί να αφήσει κενό ένα σημαντικό πεδίο που μπορεί να χρησιμοποιηθεί μετέπειτα για το σκοπό του customer analytics ή άλλης επιχειρηματικής ανάλυσης.

Πέραν του ίδιου του πελάτη της εταιρείας, η λάθος καταχώριση δεδομένων μπορεί να γίνει και από τον καταχωρητή των δεδομένων στο διαθέσιμο σύστημα, είτε αυτός είναι άνθρωπος (π.χ. υπάλληλος καταχώρισης), είτε αυτό είναι ένα αυτόματο σύστημα καταχώρισης δεδομένων (π.χ. σελίδα ή εφαρμογή σουπερ μάρκετ που επιτρέπει την εγγραφή νέων πελατών).

Η λάθος καταχώριση δεδομένων δεν αφορά φυσικά μόνο τους πελάτες της εταιρείας, αλλά και άλλα δεδομένα όπως συναλλαγές, περιγραφές προϊόντων κ.α. Σε αυτές τις περιπτώσεις έχουμε και δοκιμαστικές εγγραφές (π.χ. για δοκιμή λειτουργίας συστήματος καταχωρήσεων) που δεν ακολουθούν πρότυπα, εγγραφές από παλαιότερες βάσεις δεδομένων που δεν είχαν τον ίδιο αριθμό πεδίων, ίδια προϊόντα με μικρή διαφοροποίηση περιγραφής κ.ο.κ.

##### β) Αναντιστοιχία καταχωρίσεων δεδομένων

Ακόμα και σε περιπτώσεις σωστής καταγραφής των δεδομένων, η αντίληψη περί «σωστής» καταγραφής δεδομένων διαφέρει από άνθρωπο σε άνθρωπο. Εάν για παράδειγμα δεν υπάρχει μία κοινά αποδεκτή μορφή καταγραφής της περιοχής κατοικίας ενός πελάτη, μπορεί να έχουμε διαφορετικούς τύπους καταγραφής ίδιων δεδομένων. Ένας πελάτης μπορεί να καταχωρήσει ως κατοικία του την περιοχή «Άγιος Δημήτριος», ένας άλλος την ίδια περιοχή ως «Άγιος Δημήτριος Αττικής», ένας άλλος ως «Αγ. Δημήτριος Αθήνα» κ.ο.κ. Η αναντιστοιχία των μορφών απεικόνισης του ίδιου δεδομένου, δημιουργεί πρόβλημα ομαδοποίησης των δεδομένων κάτω από μία εγγραφή, αφού τα ίδια δεδομένα θα εμφανίζονται ως διαφορετικά.

##### γ) Ανομοιογένεια καταχωρίσεων δεδομένων

Στοιχεία όπως ημερομηνίες και χρονοσφραγίδες (timestamps) συναλλαγών, γλώσσες καταχώρισης στοιχείων (όνομα και προσωπικά στοιχεία), αριθμητικά δεδομένα (υποδιαστολές, εκθέτες κτλ), μετρικές (κιλά/λίβρες, γραμμάρια/ml κτλ), τύποι δεδομένων (κατηγορικές και αριθμητικές μεταβλητές), τρόπος κωδικοποίησης (UTF-8 κτλ), ειδικοί χαρακτήρες και σύμβολα κ.α. μπορεί να δημιουργήσουν πρόβλημα διότι καταχωρούνται ή κωδικοποιούνται διαφορετικά σε διαφορετικά σημεία της βάσης δεδομένων. Μαζική επεξεργασία και χρήση των δεδομένων αυτών θα αποτύχει εάν δεν ομογενοποιηθούν.

##### δ) Εσφαλμένες καταχωρήσεις δεδομένων

Σε περιπτώσεις σφαλμάτων του συστήματος, η καταχώριση δεδομένων μπορεί να γίνει με τρόπο που δεν ανταποκρίνεται στην πραγματικότητα. Για παράδειγμα, μπορεί κανείς να δει σε ένα πεδίο ονόματος πολλές συνεχόμενες τελείες ή σε πεδία που θα έπρεπε να υπάρχουν χαρακτήρες, να υπάρχουν αριθμοί. Σφάλματα συστήματος μπορεί να γίνουν για διάφορους λόγους, όπως απώλεια τάσης κατά την εγγραφή,

κακόβουλο λογισμικό, καταστροφή αρχείων βάσης δεδομένων από φυσικές καταστροφές, κακή αναβάθμιση συστήματος ή ενοποίηση ανομοιογενών βάσεων δεδομένων σε μία.

Η αρχική εξερεύνηση δεδομένων φανέρωσε σχεδόν όλα τα παραπάνω προβλήματα. Για το λόγο αυτό, έγιναν οι απαραίτητες μετατροπές χρησιμοποιώντας τις ακολουθίες κώδικα που βρίσκονται στο παράρτημα της εργασίας.

### 3.3.3 Ενοποίηση πινάκων

Όπως αναφέρθηκε ήδη, οι πίνακες που έχουμε στη διάθεσή μας, παρέχουν διαφορετικά δεδομένα τα οποία όμως διασυνδέονται μεταξύ τους με εγγραφές που δρουν ως εξωτερικά (ή δευτερεύοντα κλειδιά) εντός των πινάκων. Στο παρακάτω σχέδιο, απεικονίζεται συνοπτικά η δομή της βάσης δεδομένων μας, με βάση τα πρωτεύοντα και δευτερεύοντα κλειδιά κάθε πίνακα.



**Εικόνα 25. Οπτικοποίηση δομής βάσης δεδομένων (database schema). Οι πίνακες ενώνονται μέσω πρωτευόντων και δευτερευόντων κλειδιών.**

Προκειμένου να μπορεί να γίνει βαθύτερη ανάλυση στη συνέχεια, απαιτείται η διασύνδεση των πινάκων σε έναν ενιαίο πίνακα μέσω των πρωτευόντων και δευτερευόντων κλειδιών. Ο σχετικός κώδικας βρίσκεται στο παράρτημα της εργασίας.

Για να μειωθούν οι υπολογιστικές και αποθηκευτικές ανάγκες συστήματος, κάποιες διασυνδέσεις πινάκων που δεν φαίνονται εδώ, γίνονται σε μεταγενέστερο στάδιο παρακάτω.

### 3.4 Εφαρμογή μοντέλου RFM

#### 3.4.1 Υπολογισμός μετρικών RFM πελατών εταιρείας

Όπως αναφέρθηκε προηγουμένως (βλ. ενότητα «Μοντέλο RFM») το μοντέλο RFM, χρησιμοποιεί ως scoring model τρεις μετρικές: πόσο πρόσφατα έχει αγοράσει ένας πελάτης (recency), ποσό συχνά αγοράζει ένας πελάτης (frequency) και πόσα ξοδεύει ένας πελάτης (monetary).

Η εύρεση του recency, μπορεί να εξαχθεί από τη βάση δεδομένων μας, από τη χρονοσφραγίδα της πιο πρόσφατης συναλλαγής του εκάστοτε πελάτη. Υπάρχουν διαφορετικοί τρόποι μέτρησης της πιο «πρόσφατης» συναλλαγής, μιας και η μέτρηση μπορεί να γίνει βάσει ημερών, εβδομάδων ή ακόμα και μηνών (σε βάσεις δεδομένων που έχουν μεγάλο timeframe και για εταιρείες που στοχεύουν σε διαφορετικές στρατηγικές προσέγγισης). Στην παρούσα έρευνα, επιλέχτηκε η χρήση της εβδομάδας ως κλίμακα μέτρησης της προσφατότητας. Η εύρεση του frequency, μπορεί να υπολογιστεί από το άθροισμα των συναλλαγών που έχει διενεργήσει ο πελάτης με την επιχείρηση. Και τέλος, το monetary, μπορεί να εξαχθεί από το άθροισμα του ποσού που έχει δαπανήσει ο εκάστοτε πελάτης, για αγορές στην επιχείρηση.

Ο υπολογισμός μετρικών RFM των πελατών με βάση αυτήν την πληροφορία, έγινε με την σχετική ακολουθία κώδικα που βρίσκεται στο παράρτημα της εργασίας.

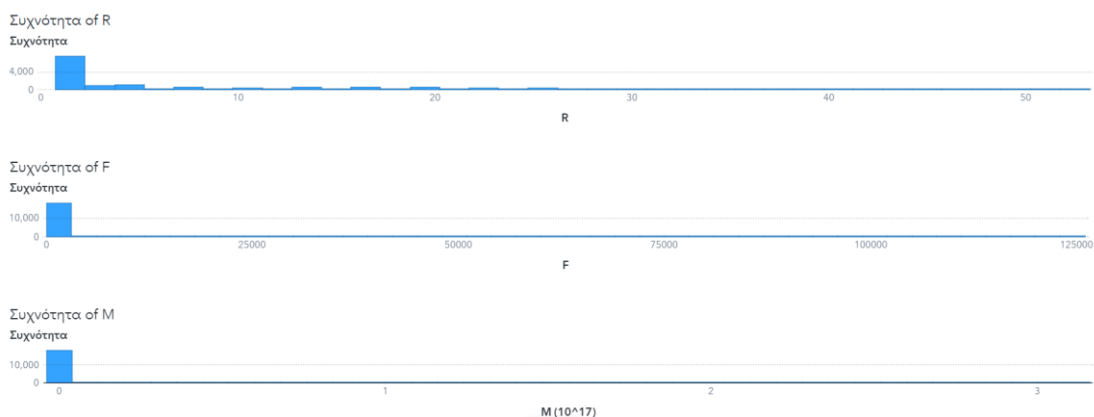
Η ημερομηνία 7 Ιανουαρίου 2005 επιλέχθηκε ως ημερομηνία έναρξης μέτρησης, διότι η τελευταία συναλλαγή που είναι αποθηκευμένη στο σύνολο των δεδομένων μας, διαθέτει την ημερομηνία 31 Δεκεμβρίου 2004. Συνεπώς η 7 Ιανουαρίου 2005 είναι η πρώτη μέρα κατά την οποία η πιο πρόσφατη συναλλαγή θα φέρει τιμή Recency ίση με 1 (εβδομάδα).

#### 3.4.2 Επεξεργασία αποτελεσμάτων RFM

Πριν γίνει οποιαδήποτε ταξινόμηση των πελατών με βάση το προφίλ RFM που υπολογίστηκε για αυτούς, απαιτείται αφαίρεση τυχόν ακραίων τιμών των τριών μετρικών. Η αφαίρεση των ακραίων τιμών (outliers), εξυπηρετεί τους ακόλουθους σκοπούς:

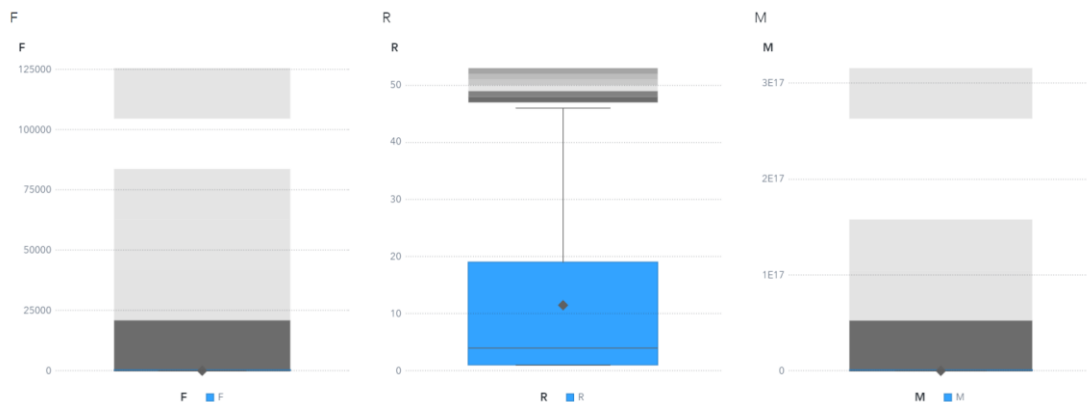
- Βελτιστοποίηση συστάδων: Οι ακραίες τιμές μπορεί να επηρεάσουν τη διαδικασία ταξινόμησης, οδηγώντας σε λιγότερο συνεκτικές ή λογικές συστάδες.<sup>95</sup>
- Αποφυγή παραμόρφωσης: Οι ακραίες τιμές μπορεί να επηρεάσουν με λάθος τρόπο τη μέση τιμή και άλλες στατιστικές ιδιότητες των δεδομένων, που μπορεί να χρησιμοποιηθούν για την ανάλυση των πελατών.

Η εξερεύνηση των αποτελεσμάτων για τις τρεις μεταβλητές του RFM, εμφανίζεται παρακάτω.



Εικόνα 26. Ιστόγραμμα κατανομής τιμών μεταβλητών RFM

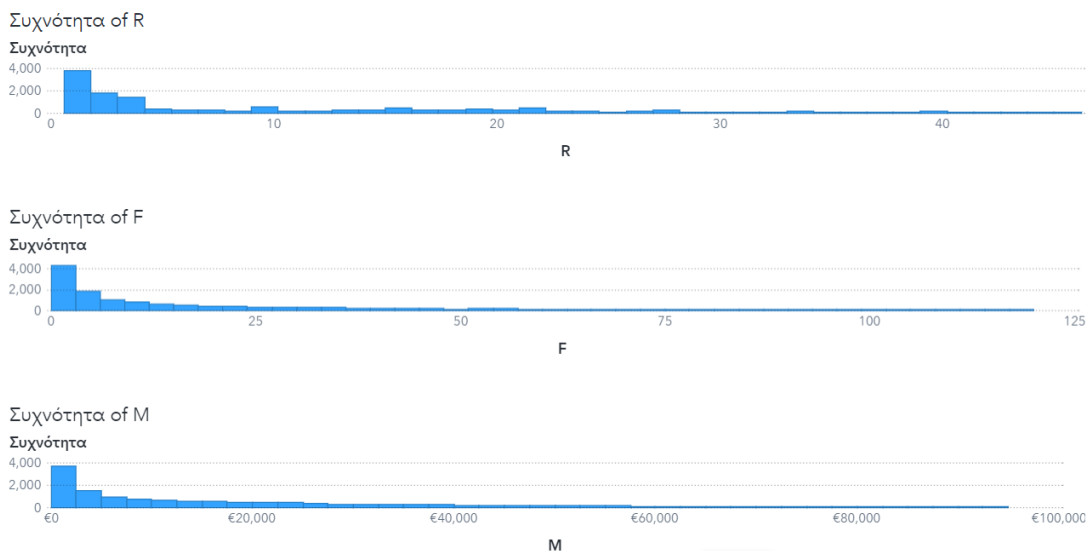




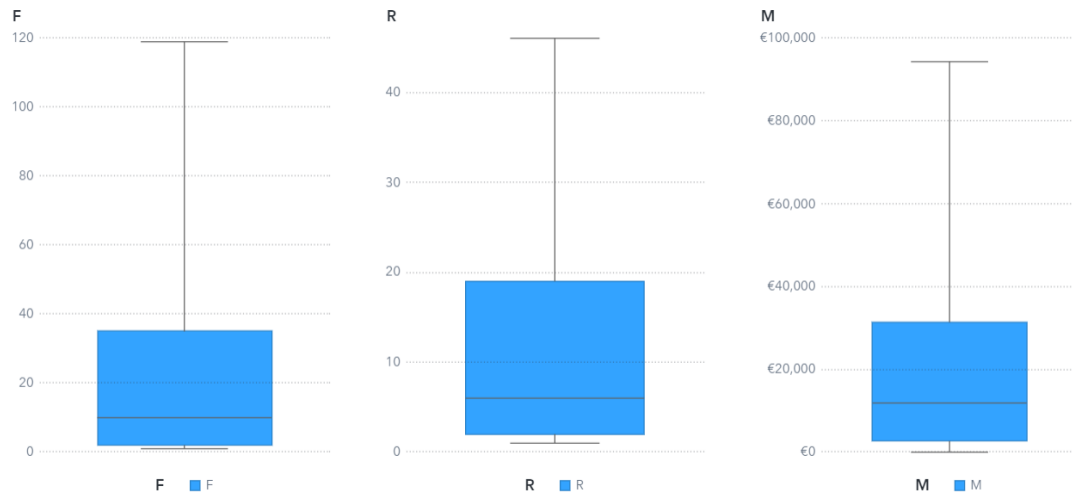
**Εικόνα 27. Ανάλυση θηκογραμμμάτων (box plots) με παρουσία ακραίων τιμών (γκρι χρώμα) των τριών μετρικών**

Η ανάλυση θηκογραμμμάτων που φαίνεται παραπάνω, μας επισημαίνει την ύπαρξη ακραίων τιμών εντός των μεταβλητών R, F και M. Προκειμένου να μην επηρεάσουν την ανάλυση με τρόπο που δεν μπορεί να εξάγει χρήσιμα συμπεράσματα, οι ακραίες αυτές τιμές, αφαιρέθηκαν. Μεταξύ άλλων, αφαιρέθηκαν και τα επίπεδα τιμών της εκάστοτε μεταβλητής που εμφανίζονται σπάνια, δηλαδή λιγότερο από 1% των φορών, όταν ο συνολικός αριθμός διαφορετικών επιπέδων είναι μεγαλύτερος από 25. Ως επίπεδο τιμών, νοείται εύρος διακριτών τιμών που μπορεί να λάβει η μεταβλητή. Η κύρια μέθοδος φιλτραρίσματος που χρησιμοποιήθηκε ήταν τα όρια μεταδεδομένων, με εναλλακτική την τυπική απόκλιση από το μέσο όρο, όταν δεν υπάρχουν αποθηκευμένα όρια στα μεταδεδομένα. Στην περίπτωση εφαρμογής της εναλλακτικής μεθόδου φιλτραρίσματος, αφαιρούνται δεδομένα με απόσταση (διαφορά) τρεις τυπικές αποκλίσεις μακριά από το μέσο όρο.

Μετά το φιλτράρισμα των ακραίων τιμών, η εικόνα που παρουσιάζουν οι τιμές των μεταβλητών, παρουσιάζεται παρακάτω.



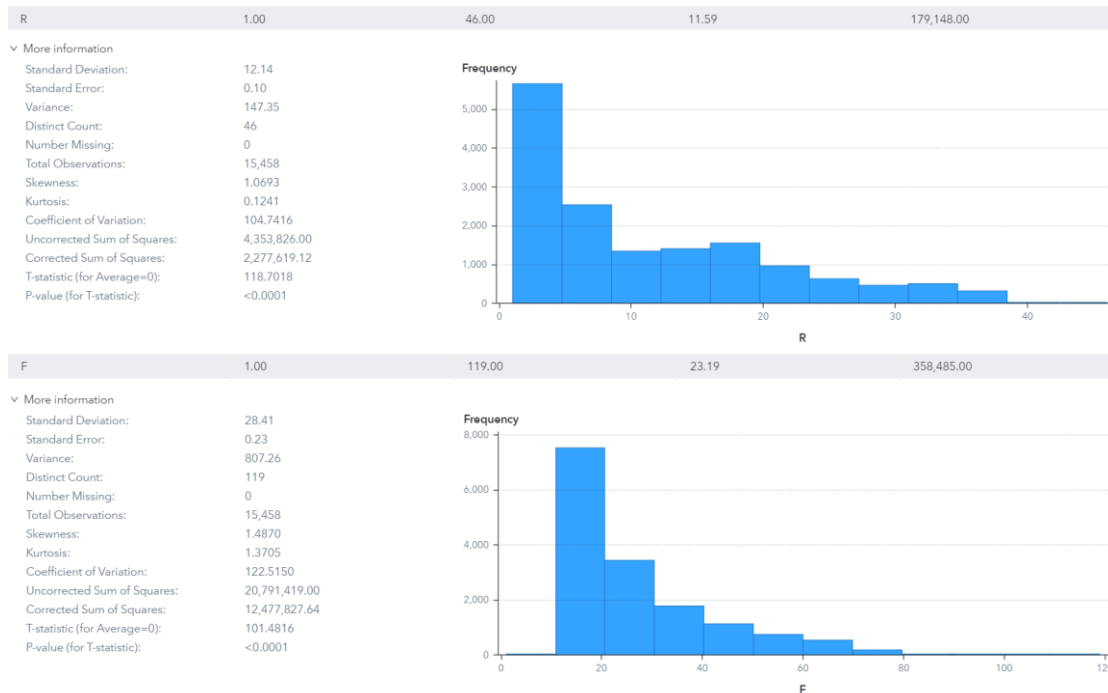
**Εικόνα 28. Ιστόγραμμα κατανομής τιμών μεταβλητών RFM μετά τη διαγραφή ακραίων τιμών**

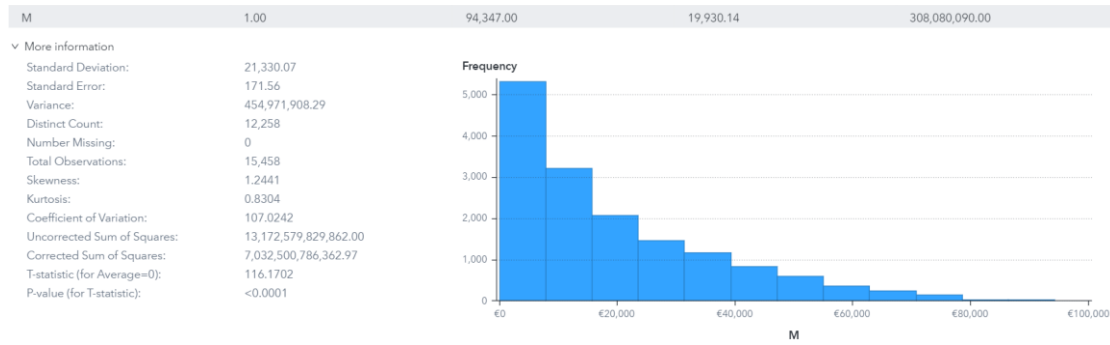


**Εικόνα 29. Ανάλυση θηκογραμμάτων τιμών μετρικών RFM μετά τη διαγραφή ακραίων τιμών**

Από τα ιστογράμματα κατανομής τιμών των μεταβλητών R,F και M, παρατηρείται ότι ακόμα και μετά το φιλτράρισμα των ακραίων τιμών τους, η κατανομή τους είναι ισχυρά ασύμμετρη προς τα αριστερά (skewed distribution). Πριν την ταξινόμηση των πελατών σε διακριτές συστάδες, πρέπει να γίνει μία προσπάθεια να κανονικοποιηθούν οι ασύμμετρες αυτές κατανομές, μέσω μετασχηματισμού τους. Αυτό συμβαίνει, διότι ασύμμετρες κατανομές μπορεί να παρεμβαίνουν στην ικανότητα της μεθόδου να αναγνωρίζει σωστά τις ομάδες.<sup>96</sup> Για παράδειγμα, οι αλγόριθμοι που βασίζονται σε αποστάσεις, όπως ο K-means<sup>97</sup>, μπορούν να επηρεαστούν σημαντικά από ασύμμετρες κατανομές δεδομένων. Η κανονικοποίηση των δεδομένων μέσω μετασχηματισμού, διασφαλίζει επίσης ότι κάθε μεταβλητή (R, F, M), θα έχει την ίδια (ίση) σημασία κατά τη διάρκεια της διαδικασίας συσταδοποίησης.<sup>98</sup> Χωρίς αυτή τη διαδικασία, μια μεταβλητή με πολύ μεγαλύτερες τιμές μπορεί να επηρεάσει υπερβολικά το αποτέλεσμα.

Επιβεβαίωση της ασυμμετρίας αυτής, μπορεί να γίνει με την διερεύνηση του συντελεστή ασυμμετρίας (skewness), για τις τρεις αυτές μετρικές.

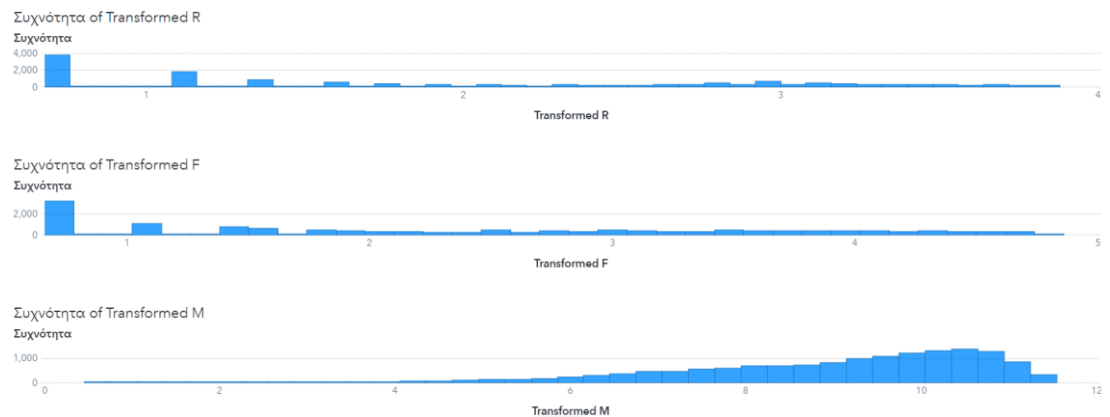




**Εικόνα 30. Ασύμμετρες κατανομές των μετρικών RFM**

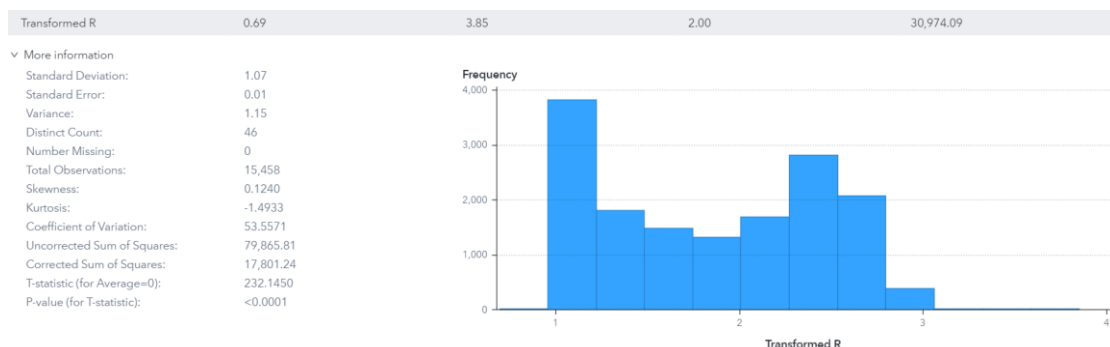
Ο συντελεστής ασυμμετρίας (skewness) των τριών μεταβλητών, έχει σε όλες τιμή μεγαλύτερη του 1, κάτι που υποδηλώνει ισχυρή (θετική) ασυμμετρία.

Για την κανονικοποίηση των τιμών των μεταβλητών R,F και M, χρησιμοποιήθηκε ο λογαριθμικός μετασχηματισμός (log transformation)<sup>99</sup>. Οι κατανομές των τιμών των μεταβλητών, μετά την εφαρμογή του μετασχηματισμού, φαίνεται παρακάτω.



**Εικόνα 31. Αποτελέσματα κανονικοποίησης τιμών μετρικών R,F και M μέσω λογαριθμικού μετασχηματισμού**

Ο όρος «κανονικοποίηση» κατανομής, έχει σκοπό να δηλώσει τον μετασχηματισμό μίας κατανομής, ώστε να μοιάζει όσο το δυνατόν περισσότερο με μία κανονική κατανομή. Ο στόχος εδώ είναι να επιτευχθεί το χαρακτηριστικό σχήμα καμπάνας της γκαουσιανής κατανομής (Gaussian distribution). Στα παραπάνω γραφήματα, φαίνεται ότι όντως οι κατανομές των μετρικών του RFM, έχουν πλησιάσει πολύ περισσότερο την κανονική κατανομή. Οι τιμές των skewness των τριών κατανομών, όπως φαίνονται παρακάτω, μπορούν να ενισχύσουν το συμπέρασμά μας.





**Εικόνα 32. Συμμετρικές κατανομές των μετασχηματισμένων μετρικών RFM**

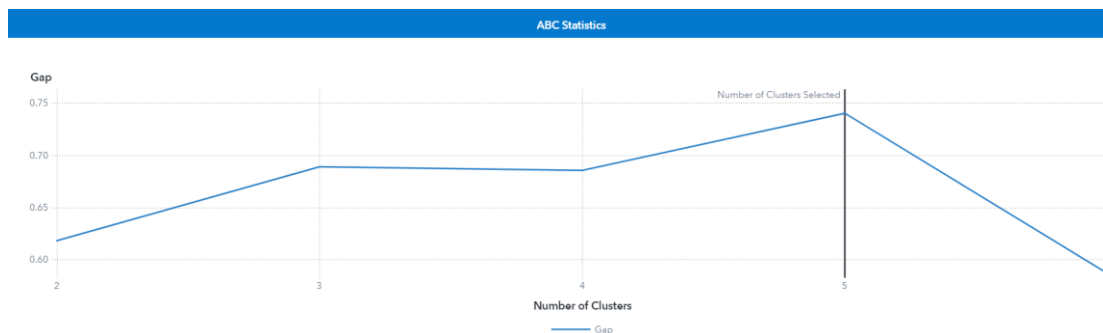
Ο συντελεστής ασυμμετρίας των μεταβλητών R και F, έχει πλέον σε όλες τιμή μεταξύ του -0.5 και 0.5, κάτι που υποδηλώνει συμμετρική κατανομή (not skewed).

Ο συντελεστής ασυμμετρίας της μεταβλητής M, έχει πλέον τιμή μεταξύ -0.5 και -1, επομένως η αντίστοιχη κατανομή της είναι μερικώς ασύμμετρη (moderately skewed). Αυτό αποτελεί σαφής βελτίωση από τη προηγούμενη κατάσταση της (highly skewed). Η παρούσα κατάσταση ασυμμετρίας της M είναι ικανοποιητική, προκειμένου να προχωρήσουμε στις απαραίτητες αναλύσεις.

### 3.4.3 Ταξινόμηση πελατών

Η αρχικοποίηση των συστάδων (clusters) πελατών έγινε με τη μέθοδο Forgy, η οποία επιλέγει τυχαία k σημεία από τα δεδομένα, ως τα αρχικά σημεία. Για την εκτίμηση του αριθμού των συστάδων, χρησιμοποιήθηκε ο αλγόριθμος Aligned Box Criterion (ABC) που βοηθάει τον υπολογιστή να καταλάβει πόσα διαφορετικά ομαδοποιημένα σύνολα υπάρχουν σε ένα σύνολο δεδομένων, σχεδιάζοντας εικονικά κουτιά γύρω από τις ομάδες αντικειμένων που ανήκουν μαζί. Ως μέγιστος αριθμός συστάδων, επιλέχθηκαν οι έξι συστάδες, ενώ ως μικρότερος αριθμός οι δύο συστάδες. Ως κριτήριο εκτίμησης, χρησιμοποιήθηκε το Global Peak Value που καθορίζει ότι ο αριθμός των συστάδων θα δίνεται από τη μέγιστη τιμή μεταξύ όλων των τιμών κορυφής του gap statistics.

Ο αλγόριθμος ABC, επέλεξε ως βέλτιστο αριθμό συστάδων τον αριθμό 5, όπως φαίνεται παρακάτω.



Εικόνα 33. Βέλτιστος αριθμός συστάδων, βάσει του αλγορίθμου ABC

Με βάση τη βαθμολογία του μοντέλου RFM στους πελάτες της εταιρείας, αυτοί ταξινομήθηκαν τελικά στα αντίστοιχα clusters, όπως φαίνεται στην παρακάτω αναφορά.

Cluster ID	Segment Profile	Συχνότητα	Ποσοστό συχνότητας	R	F	M
1	Potential Loyalists	2,866	18.54%	2.57	12.66	€14,736.86
2	Churners	2,533	16.39%	16.41	16.23	€18,706.13
3	High Value	4,604	29.78%	1.72	58.79	€44,688.43
4	Low Value	1,937	12.53%	23.39	1.19	€401.28
5	Low Value	3,518	22.76%	21.88	2.32	€3,393.63
Total		15,458	100.00%	11.59	23.19	€19,930.14

Monetary by Segment Profile  
Monetary



Εικόνα 34. Ανάλυση χαρακτηριστικών των πέντε διακριτών συστάδων πελατών βάση των τιμών μεταβλητών RFM

Οι μέσες τιμές των μεταβλητών RFM της εκάστοτε συστάδας, διαχωρίζουν τα χαρακτηριστικά στοιχεία της από εκείνα των υπολοίπων. Στην παραπάνω αναφορά, οι τιμές των μεταβλητών F και M που είναι πάνω από το μέσο όρο του συνολικού δείγματος, απεικονίζονται με πράσινο χρώμα, ενώ οι τιμές που βρίσκονται κάτω από το μέσο όρο, απεικονίζονται με κόκκινο. Αυτό ισχύει, διότι για αυτές τις μεταβλητές, είναι επιθυμητό οι τιμές να είναι όσο το δυνατόν μεγαλύτερες μιας και είναι σημαντικοί οι πελάτες που ξοδεύουν περισσότερα χρήματα στην επιχείρηση (M) και συναλλάσσονται πιο συχνά με την επιχείρηση (F). Ωστόσο, για την μεταβλητή R, ισχύει το αντίστροφο. Τιμές κάτω του μέσου όρου (με πράσινο χρώμα), είναι πιο επιθυμητές αφού αναζητούνται οι πελάτες που έχουν συνδιαλλαγή με την επιχείρηση πιο πρόσφατα από τους υπόλοιπους. Δηλαδή, όσο χαμηλότερες είναι οι τιμές στην μεταβλητή, τόσο πιο επιθυμητό.

Οι συστάδες που δημιουργήθηκαν, ονομάστηκαν με βάση τις τιμές των μετρικών RFM και αναλύονται συνοπτικά παρακάτω:

- “High-value”: Αυτοί οι πελάτες, αποτελούν το μεγαλύτερο μέγεθος πελατειακής βάσης (29,78% των πελατών) και ταυτόχρονα κατέχουν το μεγαλύτερο ποσοστό οικονομικής αξίας για την εταιρεία. Δηλαδή, οι πελάτες αυτοί είναι οι περισσότεροι σε σύγκριση με τους υπόλοιπους και η εταιρεία αποκομίζει τα περισσότερα κέρδη από αυτούς. Αυτοί οι πελάτες έχουν πραγματοποιήσει πρόσφατα αγορές, έχουν υψηλή συχνότητα αγορών και υψηλή χρηματική αξία. Οι πελάτες αυτοί είναι πιστοί στο Brand της εταιρείας, συμβάλλουν σημαντικά στα έσοδα της και πρέπει να δοθεί έμφαση στη διατήρησή του δεσμού που έχουν με την εταιρεία, προκειμένου η εταιρεία να έχει σταθερή πελατεία και μακρότερη διάρκεια ζωής. Η ανάλυση του προφίλ της ομάδας αυτής, μπορεί να δώσει χρήσιμα συμπεράσματα στην εταιρεία για το πώς να ενισχύσει τους δεσμούς της με τους πελάτες αυτούς και πώς να αυξήσει την οικονομική τους αξία στο χρόνο ζωής τους. Προτείνονται

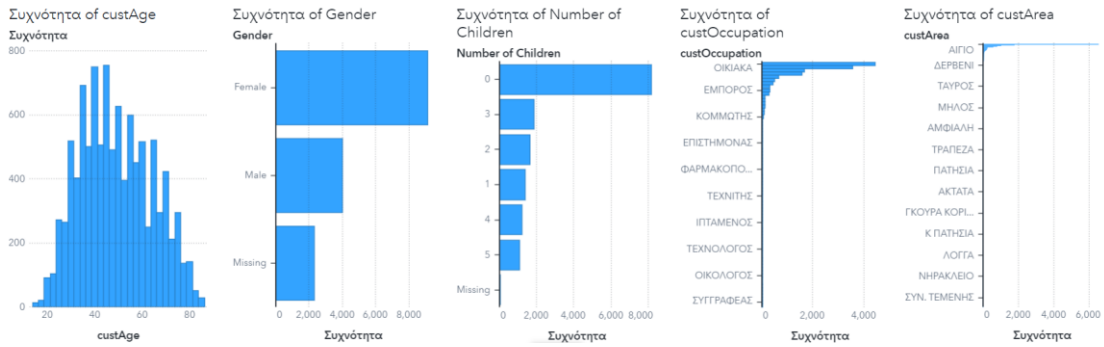
εξατομικευμένες/αποκλειστικές προσφορές, προσωποποιημένες εμπειρίες, ανώτερη υποστήριξη πελατών και άλλες υπηρεσίες προστιθέμενης αξίας, που θα διασφαλίζουν ότι οι πελάτες αυτοί παραμένουν ικανοποιημένοι και πιστοί στην εταιρεία.

- “Potential Loyalists”: Αυτοί οι πελάτες, αποτελούν ένα από τα μεγαλύτερα σε μέγεθος τμήματα πελατειακής βάσης της εταιρείας (18,54% των πελατών). Παράλληλα, τα έσοδα από τους πελάτες αυτούς είναι αρκετά υψηλά για την εταιρεία (18% των συνολικών εσόδων) και έχουν τη δυνατότητα να γίνουν πιστοί πελάτες, αυξάνοντας την πελατειακή βάση με την οποία η εταιρεία έχει δυνατούς και μακροχρόνιους δεσμούς. Αυτοί οι πελάτες έχουν πραγματοποιήσει πρόσφατες αγορές, παρουσιάζουν μέτρια συχνότητα αγορών και καλή χρηματική αξία, καθιστώντας τους νέους και εν δυνάμει πολύτιμους πελάτες της εταιρείας. Η συμπεριφορά των πελατών αυτών δείχνει ότι πράγματι είναι ελκυστικοί για μία προσπάθεια μετατροπής τους σε πιστούς High-Value πελάτες. Για το λόγο αυτό, η ανάλυση του προφίλ της ομάδας αυτής, μπορεί να δώσει χρήσιμα συμπεράσματα για τους τρόπους ενίσχυσης των εταιρικών δεσμών με τους πελάτες αυτούς και τους τρόπους αύξησης της οικονομικής τους αξία στο χρόνο ζωής τους. Για παράδειγμα, μπορεί να εφαρμοστούν προγράμματα πιστότητας, να τους προσφερθούν εξατομικευμένες προωθητικές ενέργειες και συστάσεις με βάση τις προηγούμενες αγορές τους ή κίνητρα για να επιστρέψουν και να κάνουν πρόσθετες αγορές, όπως αποκλειστικές εκπτώσεις ή γρηγορότερη πρόσβαση σε νέα προϊόντα.
- Churners (16,39% των πελατών): Αυτοί οι πελάτες αντιπροσωπεύουν το 35,29% της πελατειακής βάσης. Δεν έχουν πραγματοποιήσει πρόσφατες αγορές, έχουν χαμηλή συχνότητα αγορών, αλλά κατέχουν τη δεύτερη καλύτερη νομισματική αξία. Παρότι στο παρελθόν είχαν ίσως μια σταθερή παρουσία, τώρα είναι περισσότερο απομακρυσμένοι και έχουν παύσει να πραγματοποιούν αγορές. Υπάρχει κίνδυνος απώλειάς τους από τους ανταγωνιστές. Η ανάλυση του προφίλ της ομάδας αυτής μπορεί να βοηθήσει την εταιρεία να κατανοήσει τους λόγους της χαμηλής τους δραστηριότητας και να δημιουργήσει στρατηγικές για να τους επαναφέρει. Προτείνεται η εφαρμογή ερευνών ικανοποίησης πελατών, η εισαγωγή προσφορών επανασύνδεσης και επικοινωνιακών καμπανιών που στοχεύουν στην επανένταξη των πελατών αυτών.
- Low Value: Αυτοί οι πελάτες αντιπροσωπεύουν το 35,29% της πελατειακής βάσης. Δεν έχουν πραγματοποιήσει πρόσφατες αγορές, έχουν χαμηλή συχνότητα αγορών και χαμηλή με μέτρια χρηματική αξία. Αν και πολλοί σε αριθμό, δεν δεσμεύονται ενεργά και έχουν χαμηλή συνεισφορά στα έσοδα της εταιρείας. Συνήα πραγματοποιούν μόνο βασικές ή περιοδικές αγορές με χαμηλό όγκο. Μίας και η άμεση οικονομική αξία τους είναι χαμηλή, η εταιρεία θα πρέπει να σκεφτεί σοβαρά αν αξίζει να επενδύσει στην πιθανότητα αναβάθμισης της σχέσης τους με το brand, ή τελικά να τους παραβλέψει.

#### 3.4.4 Δημογραφική ανάλυση συστάδων πελατών

Είναι εφικτή μια πιο εξειδικευμένη ανάλυση επί των δημιουργηθέντων συστάδων πελατών, εκτός της γενικής που πραγματοποιήσαμε παραπάνω. Η δημογραφική ανάλυση πελατών είναι μία εξαιρετική ευκαιρία για την εταιρεία να γνωρίσει καλύτερα τους πελάτες της, αναπτύσσοντας τις σχέσεις της μαζί τους, αλλά και να επικεντρωθεί σε στρατηγικές επιλογές σχετικά με την προσέλκυση νέων. Για τη δημογραφική ανάλυση των πελατών κάθε συστάδας, επιλέχτηκε η οπτική ανάλυση, που παρέχει γρήγορα αλλά σημαντικά συμπεράσματα.

Τα δημογραφικά χαρακτηριστικά του συνολικού δείγματος, φαίνεται παρακάτω.



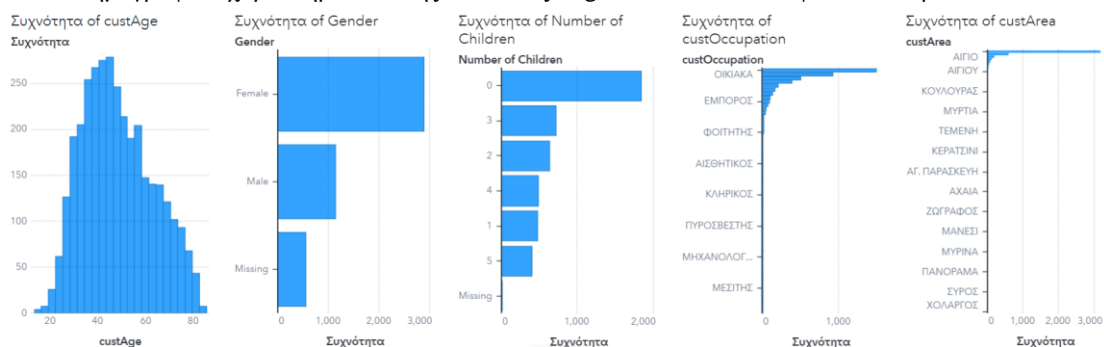
Εικόνα 35. Δημογραφική ανάλυση του συνολικού δείγματος πελατών

Από την οπτική δημογραφική ανάλυση της συνολικής πελατειακής βάσης, προκύπτουν τα παρακάτω ευρήματα:

- Συγκεκριμένα ηλικιακά διαστήματα φαίνεται να υπερéχουν κατά πολύ άλλων κοντινών σε τιμές με αυτά (π.χ. 29-31, 34-36, 39-41, 44-46). Είναι ένα ενδιαφέρον εύρημα το οποίο ίσως θα πρέπει να ερευνηθεί ως προς τις συνθήκες που το προκαλούν.
- Η γυναίκες πελάτες υπερτερούν έναντι των αντρών με διαφορά μεγαλύτερη από το διπλάσιο. Αυτό σημαίνει πως στα νοικοκυριά που επιλέγουν την εταιρεία, οι γυναίκες κάνουν τα ψώνια κατά κύριο λόγο. Προωθητικές ή διαφημιστικές ενέργειες που στοχεύουν τις γυναίκες πελάτες, μπορεί να έχουν μεγαλύτερη απόδοση κερδών.
- Οι πελάτες που δεν έχουν κανένα παιδί ελκύονται περισσότερο από τα προϊόντα και τις υπηρεσίες της αλυσίδας. Στοχευμένες διαφημιστικές και προωθητικές ενέργειες σε αυτούς τους πελάτες, μπορεί και εδώ να δημιουργήσουν καλύτερο Return of Investment (ROI).
- Οι πελάτες που δηλώνουν ένα από τα παρακάτω επαγγέλματα, υπερτερούν ξεκάθαρα σε ποσοστά έναντι των υπόλοιπων πελατών: Οικιακά (κυρίαρχη ομάδα), Ιδιωτικός Υπάλληλος, Συνταξιούχος, Δημόσιος Υπάλληλος, Εκπαιδευτικός, Αγρότης.
- Οι περιοχές προέλευσης των πελατών είναι κατά συντριπτική πλειοψηφία το Αίγιο και η Βάρη.

Με βάση αυτά τα δημογραφικά χαρακτηριστικά, μπορούμε να συγκρίνουμε και να αξιολογήσουμε τα αποτελέσματα που μπορεί εξειδικευμένα να παρέχει η δημογραφική ανάλυση κάθε συστάδας πελατών.

Τα δημογραφικά χαρακτηριστικά της συστάδας High-Value πελατών, φαίνεται παρακάτω.

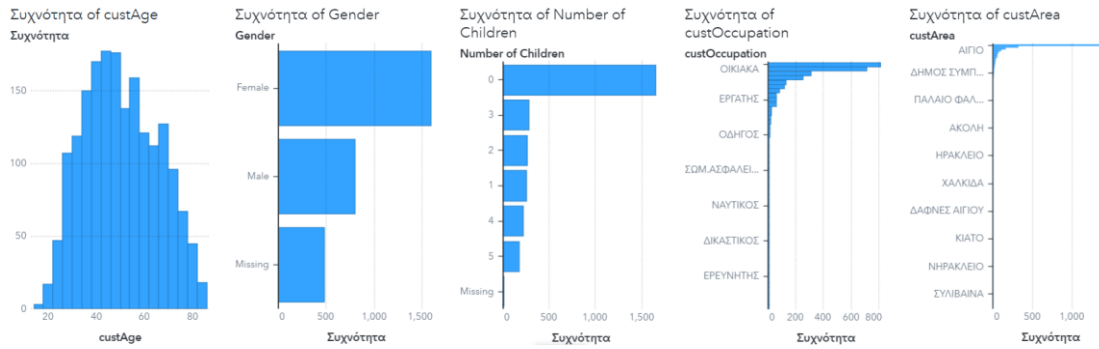


Εικόνα 36. Δημογραφική ανάλυση High-Value συστάδας πελατών

Παρατηρούμε ότι οι ηλικίες όπου οι High-Value πελάτες μεγιστοποιούνται, είναι μεταξύ 35 και 50 ετών. Κατά τα λοιπά, η συστάδα High-Value δεν φαίνεται να διαφοροποιείται ιδιαίτερα από τον συνολικό πληθυσμό.

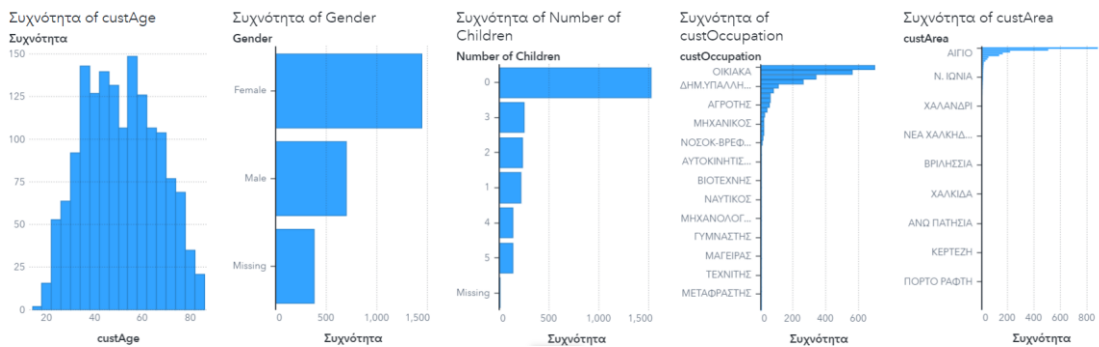
### Μεταπτυχιακή Διατριβή

### Στεργιάννα Σιώλου



**Εικόνα 37. Δημογραφική ανάλυση Potential Loyalists συστάδας πελατών**

Σε ότι αφορά τη συστάδα Potential Loyalists, παρατηρούμε μεγάλη δημογραφική ομοιότητα με τη συστάδα High Value. Η ηλικιακή τοποθέτηση των πελατών αυτής της ομάδας, κυμαίνεται στο ίδιο εύρος ηλικιών. Σε ότι αφορά τις περιοχές προέλευσης των πελατών, παρατηρούμε αυξημένες τιμές σε μερικές περιοχές που δεν παρατηρούνται έντονα στο σύνολο των συστάδων των πελατών: τη Βάρκιζα, το Κορωπί, τις Καμάρες και το Διακοπτό. Οι περιοχές αυτές προκαλούν ενδιαφέρον ως πιθανές αναδυόμενες αγορές που δημιουργούν το έδαφος για μετατροπή των πελατών αυτών σε High Value πελάτες. Κατά τα λοιπά, η συστάδα δεν φαίνεται να διαφέρει ιδιαίτερα από τον γενικό πληθυσμό. Παρατηρούμε μία αυξημένη τιμή στις missing τιμές σε αυτούς τους πελάτες στο επάγγελμα, που σημαίνει ότι αν θέλουμε να επενδύσουμε κάποια στρατηγική καμπάνια στοχευμένη σε επαγγέλματα, αυτή η ομάδα δημιουργεί εμπόδια με αυτές τις τιμές και θα πρέπει να βρεθεί μία λύση.



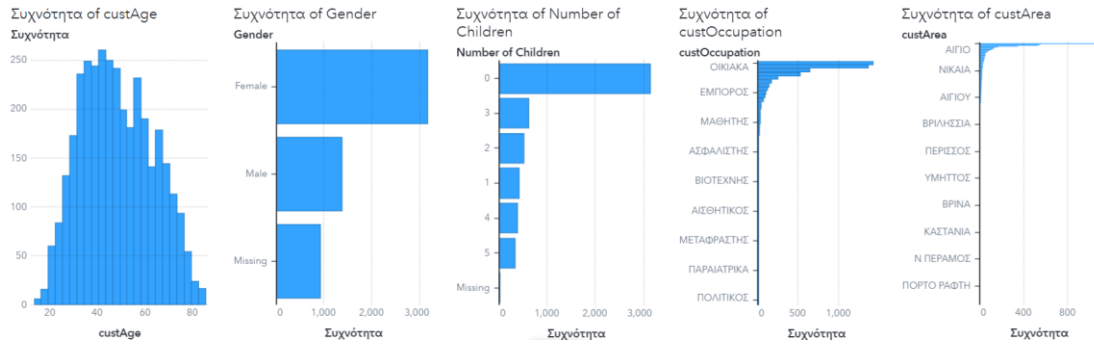
**Εικόνα 38. Δημογραφική ανάλυση Churners συστάδας πελατών**

Η συστάδα Churners δεν διαφέρει ιδιαίτερα από το συνολικό πληθυσμό. Η ηλικιακή κατανομή είναι πιο ομοιόμορφα κατανομημένη, με μοναδική μικρή πτώση στις ηλικίες μεταξύ 50-54, κάτι που μπορεί να μαρτυρά την έντονη θετική επιρροή της αλυσίδας σε αυτούς τους πελάτες. Απαιτείται περαιτέρω ανάλυση. Μία σημαντική παρατήρηση είναι ότι οι πελάτες με περιοχή προέλευσης τη Βάρη, φαίνεται πως κατέχουν πολύ ανώτερο ποσοστό απ' ό,τι σε άλλες συστάδες. Αυτή η πληροφορία είναι ένα ενδιαφέρον εύρημα που πρέπει να εξεταστεί περαιτέρω, διότι η ύπαρξη ανταγωνιστικής αλυσίδας στην περιοχή, μπορεί να οδηγήσει σε φυγή των πελατών προς ανταγωνιστικές εταιρείες.



### Μεταπτυχιακή Διατριβή

### Στεργιάννα Σιώλου



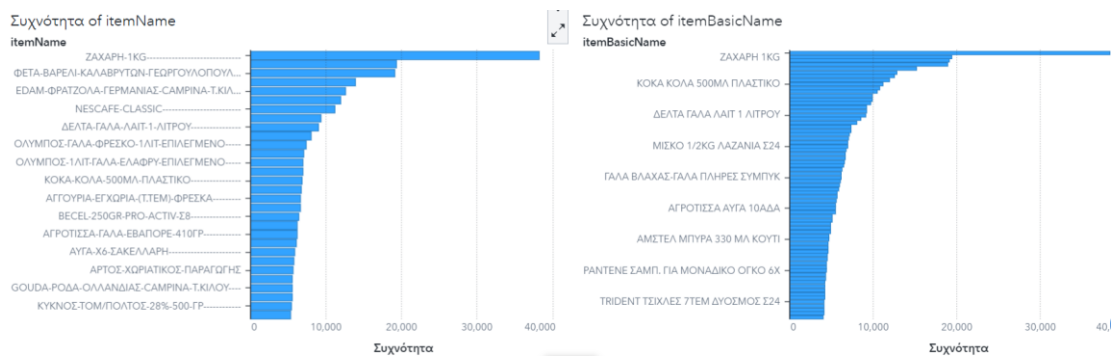
Εικόνα 39. Δημογραφική ανάλυση Low Value συστάδας πελατών

Η συστάδα Low Value δεν διαφέρει ιδιαίτερα από το συνολικό πληθυσμό. Η ηλικιακή κατανομή είναι επίσης πιο ομοιόμορφα κατανεμημένη. Σημαντική παρέκκλιση είναι η συγκριτικά αυξημένη τιμή της Αθήνας ως περιοχής προέλευσης των καταναλωτών. Καταδεικνύει πως η εταιρεία δεν έχει καθόλου ισχυρούς δεσμούς με πελάτες σε αυτές τις περιοχές και κατά συνέπεια οι πελάτες δεν την εμπιστεύονται.

Με βάση την ανωτέρω δημογραφική ανάλυση, είτε γενική είτε ειδική, μπορούν να υλοποιηθούν πιο στοχευμένες στρατηγικές ενέργειες προώθησης ή επιχειρησιακής φύσεως, ανάλογα με τη συστάδα πελατών που ενδιαφέρει περισσότερο ή τον επιδιωκόμενο εταιρικό σκοπό.

### 3.4.5 Ανάλυση αγορασμένων προϊόντων

Εκτός της δημογραφικής ανάλυσης πελατών, είναι χρήσιμο να υλοποιηθεί μία ανάλυση των πωληθέντων προϊόντων. Αρχικά, αναλύθηκαν οι πωλήσεις στο σύνολο των συστάδων πελατών, όπως φαίνεται παρακάτω.



Εικόνα 40. Ανάλυση όγκου πωλήσεων ανά προϊόν στο σύνολο των συστάδων πελατών

Η ανάλυση εμφανίζει τα πιο δημοφιλή προϊόντα, εκ των οποίων τις ανώτερες θέσεις καταλαμβάνουν:

- ΖΑΧΑΡΗ 1kg
- ΔΕΛΤΑ ΓΑΛΑ ΠΑΣΤ.ΧΑΡΤ.ΣΥΣΚ. 1 ΛΙΤΡΟ
- ΦΕΤΑ ΒΑΡΕΛΙ ΚΑΛΑΒΡΥΤΩΝ ΓΕΩΡΓΟΠΟΥΛΟΣ
- NESCAFE CLASSIC

Κ.Ο.Κ.

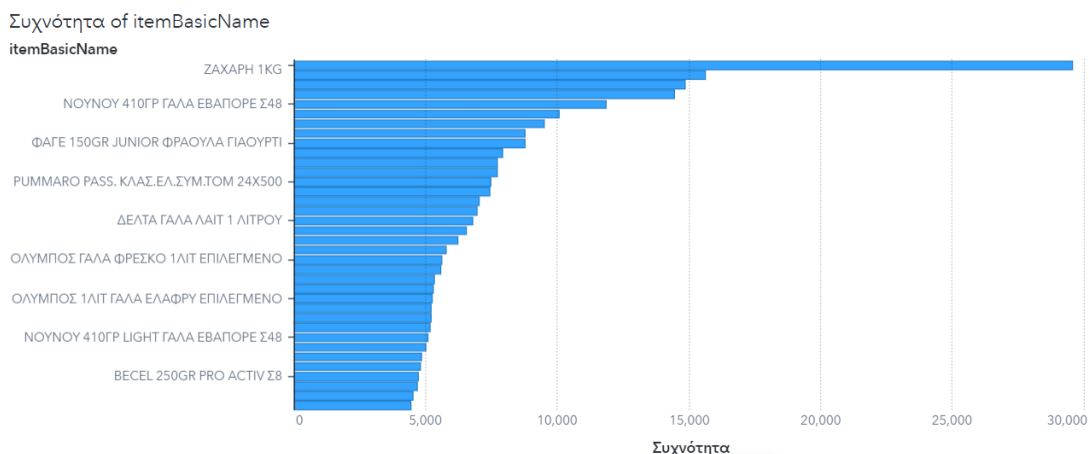
Η ανάλυση φυσικά έχει αξία και αντίστροφα, όπως θα δούμε σε λίγο. Δηλαδή να βρούμε ποια προϊόντα πωλούνται λιγότερο. Τις χαμηλότερες θέσεις καταλαμβάνουν:

- ΔΕΛΤΑ NATURAL ΜΟΥΣΛΙ 2Χ200 ΓΡ
- ΑΛΛΑΤΙΝΗ ΜΠΙΣΚΟΤΑ ΠΑΙΔ.200GR ΜΕ ΓΑΛΑ Σ16
- RILKEN STYLING EXTR.SHIN.MOUSS.ΔΥΝ.200ML
- NESTLE 500GR CH.MYLTILINGUAL Σ12

κ.ο.κ.

Παρατηρείται ότι η ίδια εταιρεία (ΔΕΛΤΑ) κατέχει προϊόντα που βρίσκονται στις πρώτες θέσεις των πωλήσεων, αλλά και ταυτόχρονα στις τελευταίες. Αυτό μπορεί να δείχνει πως δεν αρκεί το brand για να επιφέρει ισχυρές πωλήσεις σε ένα προϊόν, αλλά οι πελάτες πρέπει να βρίσκουν αξία σε αυτό. Μπορεί επίσης να δείχνει ότι κάποια προϊόντα απαιτούν περισσότερη διαφημιστική προβολή ή προωθήσεις για να ξεκινήσουν να δοκιμάζονται από τους πελάτες. Για παράδειγμα, στα προϊόντα με τις περισσότερες πωλήσεις, παρατηρούνται προϊόντα που είναι απλά και χρησιμοποιούνται εύκολα από ένα νοικοκυριό. Στον αντίποδα, τα προϊόντα που έχουν τις λιγότερες πωλήσεις, θα μπορούσαν να χαρακτηριστούν ως πιο ιδιαίτερα και άρα δύσκολα θα δοκιμαστούν από τους πελάτες για πρώτη φορά, ή εξυπηρετούν πολύ συγκεκριμένες προτιμήσεις. Φυσικά η ανάλυση αυτή συνδυάζεται με την ανάλυση καταναλωτών που έγινε σε προηγούμενο στάδιο. Εάν οι κυρίαρχες ομάδες πελατών έχουν κατά πλειοψηφία μικρές ηλικίες, πιο εύκολα θα πωληθούν ανθυγιεινά προϊόντα και πιο δύσκολα υγιεινά, όπως για παράδειγμα στην περίπτωση μας το μούσλι. Θα πρέπει να διερευνηθούν περαιτέρω οι αιτίες που κάποια προϊόντα υπερτερούν άλλων και μέσω αυτής της ανάλυσης, θα γίνει πιο εύκολα αντιληπτό που πρέπει να επικεντρωθεί η μελλοντική έρευνα.

Η ανάλυση πωληθέντων προϊόντων έχει αξία και όταν γίνεται σε εξειδικευμένες συστάδες πελατών. Αυτό ισχύει κυρίως όταν υπάρχει ενδιαφέρον επένδυσης σε στρατηγικές ενέργειες που στοχεύουν μία από τις συστάδες αυτές. Παρακάτω φαίνεται η ανάλυση πωλήσεων στην ομάδα πελατών High Value.



Εικόνα 41. Ανάλυση όγκου πωλήσεων ανά προϊόν για τη συστάδα πελατών High Value

Παρατηρείται ότι για την συστάδα High Value, δεν υπάρχει διαφοροποίηση από το γενικό σύνολο συστάδων πελατών. Αυτή η διαπίστωση αφορά τα προϊόντα με τις περισσότερες πωλήσεις και επιβεβαιώνει πρότερες μετρήσεις, που καταδείκνυαν την συστάδα αυτή ως κυρίαρχη συστάδα. Η συχνότητα αγορών και ο όγκος των αγορών φαίνεται ότι επιδρά όπως καμία άλλη στον όγκο πωλήσεων κάθε προϊόντος.

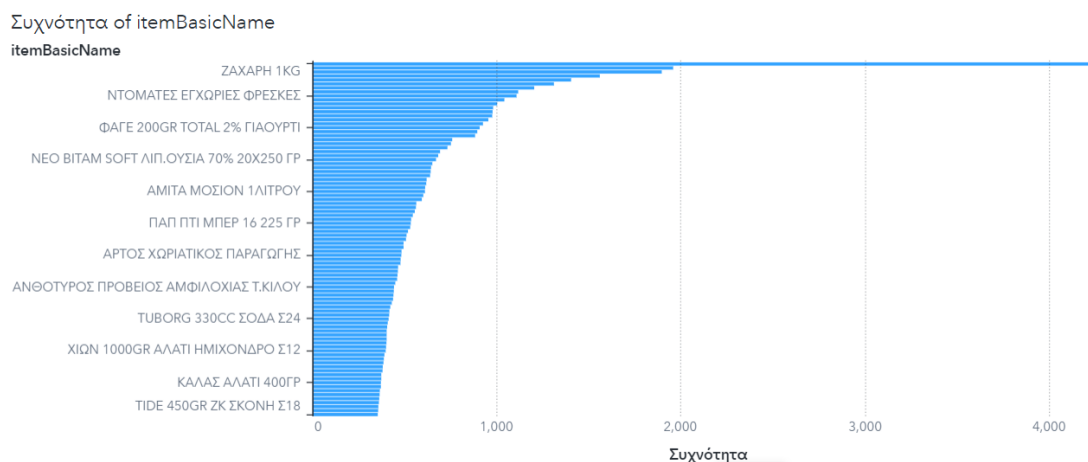
Αντίθετα, στα προϊόντα με τις λιγότερες πωλήσεις, συναντά κανείς προϊόντα όπως:

- HUBBA BUBBA ΤΣΙΧΛΑ SWEETY STRAWBERRY
- FERRERO KINDER ΜΕΓΑΛΗ ΕΚΠΛΗΞΗ
- ΕΛΙΤ ΚΕΙ ΒΑΝΙΛΙΑ ΚΑΚΑΟ 300-330ΓΡ

κ.ο.κ.

Οι χαμηλές πωλήσεις τέτοιων προϊόντων σχετίζονται ξεκάθαρα όπως ειπώθηκε προηγουμένως και με το εύρος ηλικίας πελατών. Η ανθυγιεινή διατροφή τείνει να μειώνεται σε μεγαλύτερες ηλικίες (εδώ η κυρίαρχη ομάδα είναι μεταξύ 35 και 50 ετών), πόσο μάλλον όταν η πλειοψηφία των πελατών δηλώνει πως δεν έχει παιδιά.

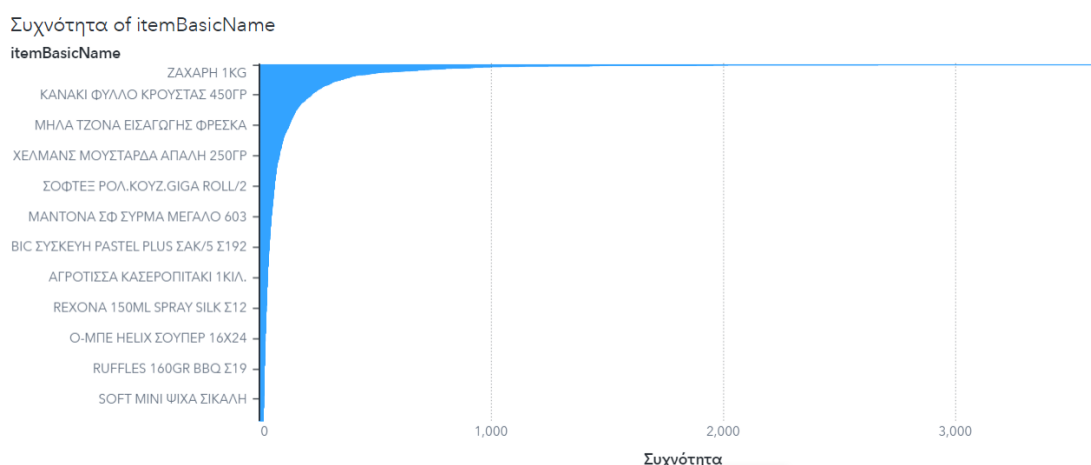
Ακολουθεί η ανάλυση πωλήσεων Potential Loyalists, όπως φαίνεται παρακάτω.



Εικόνα 42. Ανάλυση όγκου πωλήσεων ανά προϊόν για τη συστάδα πελατών Potential Loyalists

Η εικόνα που παρουσιάζεται στα προϊόντα με τις περισσότερες πωλήσεις παραμένουν ίδια με την High Value συστάδα πελατών, κάτι που δείχνει ξανά την ομοιότητα των δύο συστάδων και επιβεβαιώνει την ελκυστικότητα των πελατών αυτών ως μελλοντικούς εν δυνάμει High Value πελάτες. Τα προϊόντα που έχουν τις λιγότερες πωλήσεις δεν έχουν αξία ανάλυσης, μιας και υπάρχει μεγάλος αριθμός προϊόντων με μηδενικές πωλήσεις.

Παρακάτω, φαίνεται η ανάλυση πωλήσεων στην ομάδα πελατών Churners.

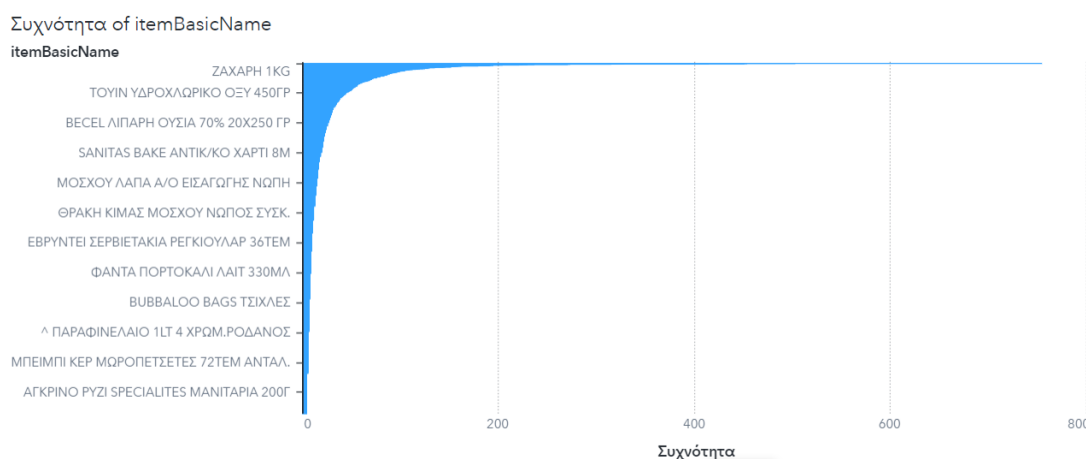


Εικόνα 43. Ανάλυση όγκου πωλήσεων ανά προϊόν για τη συστάδα πελατών Churners

Στην κορυφή της λίστας των προτιμήσεων αυτών των πελατών, φαίνεται πως βρίσκονται περισσότερα φρέσκα προϊόντα, απ' ό,τι στις υπόλοιπες ομάδες. Οι φρέσκιες ντομάτες βρίσκονται στην δεύτερη θέση, ενώ σε άλλες ομάδες δεν βρίσκονταν εντός της καλύτερης δεκάδας. Στα προτιμώμενα προϊόντα βρίσκονται επίσης υψηλά σε προτίμηση κρεμμύδια ξερά εγχώρια φρέσκα, πατάτες εγχώριες φρέσκες, ροδάκινα,

πεπόνια εγχώρια φρέσκα, καρότα εγχώρια φρέσκα, λεμόνια κ.ο.κ. Οι πελάτες αυτοί έχουν ξεκινήσει να απομακρύνονται από την επιχείρηση και κινδυνεύουν να απορροφηθούν από ανταγωνιστές. Οι προτιμήσεις τους σε φρέσκα προϊόντα μπορεί να δείχνει μία μείωση ποιότητας των φρέσκων τροφίμων της εταιρείας, που οδήγησε σε απώλεια πελατών. Στις χαμηλότερες θέσεις στις προτιμήσεις των πελατών αυτών βρίσκονται προϊόντα όπως μαχαίρι με πλαστική λαβή, champion γάντια, ζεβα ρολό κουζίνας κ.ο.κ. Μπορεί αυτό να δείχνει ότι οι καταναλωτές αυτοί έκαναν κυρίως τα καθημερινά ψώνια τους στην επιχείρηση και όχι τόσο αγορές με προϊόντα μεγαλύτερης διάρκειας ζωής. Παρατηρούμε λοιπόν ότι οι αναλύσεις αυτές παρέχουν στην εταιρεία πολύπλευρες εκτιμήσεις μίας κατάστασης, τις οποίες μπορεί να εκμεταλλευτεί για περαιτέρω έρευνα.

Ακολουθεί το γράφημα ανάλυσης πωλήσεων της συστάδας πελατών Low Value.



Εικόνα 44. Ανάλυση όγκου πωλήσεων ανά προϊόν για τη συστάδα πελατών Low Value

Τα προϊόντα αυτής της ομάδας δεν έχουν συνοχή μεταξύ τους και διαφέρουν πλήρως από των υπόλοιπων ομάδων. Τα νούμερα (όγκος πωλήσεων) είναι στατιστικά ασήμαντα και μη αντιπροσωπευτικά. Τα προϊόντα που αγοράζονται από αυτούς τους καταναλωτές είναι προϊόντα σπάνιας χρήσης – πιθανόν προϊόντα που δεν μπόρεσαν να βρουν σε άλλα καταστήματα. Η ομάδα αυτή δεν έχει φυσικά αξία περαιτέρω ανάλυσης, είναι όμως σημαντικό να επαληθεύουμε τα αποτελέσματα των προηγούμενων αναλύσεων. Η έντονη διαφοροποίηση στις προτιμήσεις των πελατών αυτών, συγκριτικά με τις υπόλοιπες, δείχνει και το λόγο που δεν κατέχουν έντονους δεσμούς με την εταιρεία. Επιπρόσθετα, ο μικρός όγκος συναλλαγών της συστάδας, επιβεβαιώνει το γεγονός ότι η επαναπροσέλκυση αυτών των πελατών δεν θα πρέπει να αποτελέσει προτεραιότητα της εταιρείας, έναντι των υπολοίπων συστάδων.

### 3.5 Ανάλυση καλαθιού αγοράς

Η πιο χρήσιμη όμως για την εταιρεία ανάλυση, σε ότι αφορά τις πωλήσεις προϊόντων, δεν είναι αυτή που πραγματοποιήθηκε παραπάνω. Διότι η εταιρεία, δεν επιθυμεί απλά να δει ποια προϊόντα της πωλούνται περισσότερο. Ενδιαφέρεται να δει και πως θα πουλήσει τελικά περισσότερα προϊόντα. Η επίτευξη αύξησης πωλήσεων και άρα εσόδων, προέρχεται από τεχνικές που ενθαρρύνουν τον καταναλωτή να προσθέσει επιπρόσθετα προϊόντα στο καλάθι του. Αυτό επιτυγχάνεται με τεχνικές όπως cross-selling, upselling, downselling, bundling κ.ο.κ. Οι περισσότερες από αυτές τις τεχνικές όμως απαιτούν μία πρότερη στατιστική ανάλυση των ιστορικών συναλλαγών των πελατών, προκειμένου να είναι επιτυχημένες. Και υπάρχει μία πολύ επιτυχημένη μέθοδος ανάλυσης για το σκοπό αυτό: η ανάλυση καλαθιού αγοράς.

### 3.5.1 Προετοιμασία δεδομένων

Όπως ειπώθηκε νωρίτερα, στην αρχική προετοιμασία δεδομένων, η σύνδεση μερικών πινάκων έγινε σε αυτό το στάδιο, για να μειωθούν οι υπολογιστικές και αποθηκευτικές ανάγκες του συστήματος.

Σε πρώτο στάδιο, έγινε η μετατροπή των αποτελεσμάτων του RFM analysis σε μορφή πίνακα, που μπορεί να επεξεργαστεί το λογισμικό SAS. Σε επόμενο στάδιο, έγινε ενοποίηση όλων των πινάκων που απέμειναν.

Και τα δύο στάδια υλοποιήθηκαν μέσω του κώδικα που βρίσκεται στο παράρτημα της εργασίας.

### 3.5.2 Κανόνες συσχέτισης μεταξύ αγορασμένων προϊόντων

Προκειμένου να εξαχθούν κανόνες συσχέτισης μεταξύ προϊόντων, χρησιμοποιήθηκε η τεχνική του Market Basket Analysis. Ως ελάχιστο ποσοστό υποστήριξης (support), εφαρμόστηκε κατώφλι 0.05, ενώ ως ελάχιστο ποσοστό εμπιστοσύνης (confidence), εφαρμόστηκε κατώφλι 0.1. Τα δύο κατώφλια επιλέχτηκαν να είναι λίγο πιο χαμηλά για δύο λόγους: Πρώτον, διότι ένα χαμηλότερο κατώφλι υποστήριξης μπορεί να αναδεικνύει συσχετίσεις που θα ήταν δυσδιάκριτες με υψηλότερο κατώφλι, σε πολλά δεδομένα. Μπορεί να ανακαλυφθούν συσχετίσεις που δεν είναι πολύ συνηθισμένες αλλά παρ' όλα αυτά είναι σημαντικές. Δεύτερον, ένα χαμηλότερο κατώφλι εμπιστοσύνης επιτρέπει την ανακάλυψη συσχετίσεων που μπορεί να μην είναι πολύ ισχυρές, αλλά είναι σημαντικές για την ανάλυση. Επιπρόσθετα, επιτρέπει την εξαγωγή περισσότερων κανόνων, δίνοντας μια ευρύτερη εικόνα των δυνατών συσχετίσεων μεταξύ διάφορων προϊόντων. Μερικές φορές οι εταιρείες μπορεί να επιλέγουν για διάφορους στρατηγικούς λόγους, να επικεντρώνουν το ενδιαφέρον τους σε συγκεκριμένα προϊόντα. Αν τα προϊόντα αυτά λείπουν από την ανάλυση, τότε η εταιρεία δεν κατέχει σημαντική γνώση, που θα μπορούσε να χρησιμεύσει για να αυξήσει τα κέρδη της.

Ως γενικός κανόνας, αν ο όγκος των δεδομένων είναι χαμηλός ή αναζητούμε συχνές συσχετίσεις, επιλέγουμε υψηλότερα κατώφλια υποστήριξης. Αν όμως ο όγκος των δεδομένων είναι υψηλός ή αναζητούμε λιγότερο συνηθισμένες συσχετίσεις (όπως εδώ), τότε τα κατώφλια υποστήριξης μπορεί να είναι χαμηλότερα.

Αντίστοιχα, αν πρέπει να εντοπιστούν πολύ ισχυρές συσχετίσεις, επιλέγονται υψηλότερα κατώφλια εμπιστοσύνης. Αντίθετα, εάν πρέπει να εξερευνηθούν και λιγότερο ισχυρές συσχετίσεις, χρησιμοποιούνται χαμηλότερα κατώφλια εμπιστοσύνης (όπως εδώ).

Υλοποιώντας λοιπόν την μεθοδολογία δημιουργίας κανόνων συσχέτισης Market Basket Analysis, μέσω ανάλυσης των διαθέσιμων δεδομένων καλαθιών αγοράς, προκύπτει ο παρακάτω πίνακας:

	◎ COUNT	◎ SUPPORT	◎ CONF	◎ LIFT	△ ITEM1	△ ITEM2	△ ITEM3	△ RULE
1	219	0.061904222	66.565349544	426.88156228	ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΙΣΑΓΩΓΗ	ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ	ΠΙΠΕΡΙΕ.ΚΙΤΡΙΝΕΣ.ΕΞ	ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ & ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ &&& ΠΙΠΕΡΙΕ.ΚΙΤΡΙΝΕΣ.ΕΞ
2	219	0.061904222	39.17731968	426.88156228	ΠΙΠΕΡΙΕ.ΚΙΤΡΙΝΕΣ.ΕΙΣΑΓΩΓΗ	ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ	ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ	ΠΙΠΕΡΙΕ.ΚΙΤΡΙΝΕΣ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ &&& ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ &&& ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ
3	325	0.0906592023	59.85207035	383.8335475	ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ.ΕΙΣΑΓΩΓΗ	ΠΙΠΕΡΙΕ.ΚΙΤΡΙΝΕΣ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ	ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ	ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ &&& ΠΙΠΕΡΙΕ.ΚΙΤΡΙΝΕΣ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ &&& ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ
4	325	0.0906592023	58.139534884	383.8335475	ΠΙΠΕΡΙΕ.ΚΙΤΡΙΝΕΣ.ΕΙΣΑΓΩΓΗ	ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ	ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ	ΠΙΠΕΡΙΕ.ΚΙΤΡΙΝΕΣ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ &&& ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ &&& ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ
5	219	0.061904222	40.31491713	378.48782216	ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ.ΕΙΣΑΓΩΓΗ	ΠΙΠΕΡΙΕ.ΚΙΤΡΙΝΕΣ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ	ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΞ	ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ &&& ΠΙΠΕΡΙΕ.ΚΙΤΡΙΝΕΣ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ &&& ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΞ
6	219	0.061904222	57.32984292	378.48782216	ΠΙΠΕΡΙΕ.ΚΙΤΡΙΝΕΣ.ΕΙΣΑΓΩΓΗ	ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ	ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ	ΠΙΠΕΡΙΕ.ΚΙΤΡΙΝΕΣ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ &&& ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ &&& ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ
7	382	0.1065595492	48.35440308	310.0954915	ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΙΣΑΓΩΓΗ	ΠΙΠΕΡΙΕ.ΚΙΤΡΙΝΕΣ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ	ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΞ	ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ &&& ΠΙΠΕΡΙΕ.ΚΙΤΡΙΝΕΣ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ &&& ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΞ
8	382	0.1065595492	68.336314848	310.0954915	ΠΙΠΕΡΙΕ.ΚΙΤΡΙΝΕΣ.ΕΙΣΑΓΩΓΗ	ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ	ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΞ	ΠΙΠΕΡΙΕ.ΚΙΤΡΙΝΕΣ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ &&& ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ &&& ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΞ
9	219	0.061904222	67.384615385	305.77688413	ΠΙΠΕΡΙΕ.ΚΙΤΡΙΝΕΣ.ΕΙΣΑΓΩΓΗ	ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ	ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΞ	ΠΙΠΕΡΙΕ.ΚΙΤΡΙΝΕΣ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ &&& ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ &&& ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΞ
10	219	0.061904222	27.721518987	305.77688413	ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΙΣΑΓΩΓΗ	ΠΙΠΕΡΙΕ.ΚΙΤΡΙΝΕΣ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ	ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ	ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ &&& ΠΙΠΕΡΙΕ.ΚΙΤΡΙΝΕΣ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ &&& ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ
11	329	0.0917731091	41.64556962	274.9412896	ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΙΣΑΓΩΓΗ	ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ	ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΞ	ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ &&& ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ &&& ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΞ
12	329	0.0917731091	60.5893186	274.9412896	ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ.ΕΙΣΑΓΩΓΗ	ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ	ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΞ	ΠΙΠΕΡΙΕ.ΠΟΡΤΟΚΑΛΙ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ &&& ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΙΣΑΓΩΓΗ-ΦΡΕΣΚΕΣ &&& ΠΙΠΕΡΙΕ.ΚΟΚΚΙΝΕΣ.ΕΞ
13	203	0.058627195	48.91562651	227.14418815	ΑΜΠΤΑ.ΦΑΝ.200ΜΛ.ΦΡ.ΚΟΚ.ΣΤ.	ΑΜΠΤΑ.ΦΑΝ.ΜΥΛΕ.200ΜΛ	ΑΜΠΤΑ.ΦΑΝ.ΜΥΛΕ.20	ΑΜΠΤΑ.ΦΑΝ.200ΜΛ.ΦΡ.ΚΟΚ.ΣΤΑΦ.ΜΗΛ.ΑΧΛ. &&& ΑΜΠΤΑ.ΦΑΝ.ΜΥΛΕ.200ΜΛ &&& ΑΜΠΤΑ.ΦΑΝ.ΜΥΛΕ.20
14	203	0.058627195	26.29536788	227.14418815	ΑΜΠΤΑ.ΦΑΝ.200ΜΛ.ΑΠ.ΑΣΤ.ΣΤ.	ΑΜΠΤΑ.ΦΑΝ.200ΜΛ.ΦΡ.ΚΟΚ.ΣΤΑΦ.ΜΗΛ.ΑΧΛ.	ΑΜΠΤΑ.ΦΑΝ.ΜΥΛΕ.20	ΑΜΠΤΑ.ΦΑΝ.200ΜΛ.ΑΠ.ΑΣΤ.ΣΤΑΦ.ΜΗΛ.ΑΧΛ. &&& ΑΜΠΤΑ.ΦΑΝ.200ΜΛ.ΦΡ.ΚΟΚ.ΣΤΑΦ.ΜΗΛ.ΑΧΛ. &&& ΑΜΠΤΑ.ΦΑΝ.ΜΥΛΕ.20
15	213	0.059416712	40.22659748	216.2943276	ΔΕΛΤΑ.150GRK2.ADVANCEΣ.ΒΕ	ΔΕΛΤΑ.150GRK2.ADVANCEΣ.ΒΕΡ.ΑΧΛ.Δ20ΠΕΡ1	ΔΕΛΤΑ.150GRK2.ADVANCEΣ.ΒΕΡ.ΑΧΛ.Δ20ΠΕΡ1	ΔΕΛΤΑ.150GRK2.ADVANCEΣ.ΒΕΡ.ΑΧΛ.Δ20ΠΕΡ1 &&& ΔΕΛΤΑ.150GRK2.ADVANCEΣ.ΒΕΡ.ΑΧΛ.Δ20ΠΕΡ1 &&& ΔΕΛΤΑ.150GRK2.ADVANCEΣ.ΒΕΡ.ΑΧΛ.Δ20ΠΕΡ1
16	213	0.059416712	31.55555556	216.2943276	ΔΕΛΤΑ.150GRK2.ADVANCEΣ.ΜΗ	ΔΕΛΤΑ.150GRK2.ADVANCEΣ.ΒΕΡ.ΑΧΛ.Δ20ΠΕΡ1	ΔΕΛΤΑ.150GRK2.ADVANCEΣ.ΒΕΡ.ΑΧΛ.Δ20ΠΕΡ1	ΔΕΛΤΑ.150GRK2.ADVANCEΣ.ΜΗ.ΑΧΛ.Δ20ΠΕΡ1 &&& ΔΕΛΤΑ.150GRK2.ADVANCEΣ.ΒΕΡ.ΑΧΛ.Δ20ΠΕΡ1 &&& ΔΕΛΤΑ.150GRK2.ADVANCEΣ.ΒΕΡ.ΑΧΛ.Δ20ΠΕΡ1
17	203	0.058627195	51.133501239	196.25920783	ΑΜΠΤΑ.ΦΑΝ.200ΜΛ.ΑΠ.ΑΣΤ.ΣΤ.	ΑΜΠΤΑ.ΦΑΝ.ΜΥΛΕ.200ΜΛ	ΑΜΠΤΑ.ΦΑΝ.ΜΥΛΕ.20	ΑΜΠΤΑ.ΦΑΝ.200ΜΛ.ΑΠ.ΑΣΤ.ΣΤΑΦ.ΜΗΛ.ΑΧΛ. &&& ΑΜΠΤΑ.ΦΑΝ.ΜΥΛΕ.200ΜΛ &&& ΑΜΠΤΑ.ΦΑΝ.ΜΥΛΕ.20
18	203	0.058627195	21.734407375	196.25920783	ΑΜΠΤΑ.ΦΑΝ.200ΜΛ.ΦΡ.ΚΟΚ.ΣΤ.	ΑΜΠΤΑ.ΦΑΝ.200ΜΛ.ΑΠ.ΑΣΤ.ΣΤΑΦ.ΜΗΛ.ΑΧΛ.	ΑΜΠΤΑ.ΦΑΝ.ΜΥΛΕ.20	ΑΜΠΤΑ.ΦΑΝ.200ΜΛ.ΦΡ.ΚΟΚ.ΣΤΑΦ.ΜΗΛ.ΑΧΛ. &&& ΑΜΠΤΑ.ΦΑΝ.200ΜΛ.ΑΠ.ΑΣΤ.ΣΤΑΦ.ΜΗΛ.ΑΧΛ. &&& ΑΜΠΤΑ.ΦΑΝ.ΜΥΛΕ.20
19	203	0.058627195	62.26993865	190.14346234	ΑΜΠΤΑ.ΦΑΝ.200ΜΛ.ΑΠ.ΑΣΤ.ΣΤ.	ΑΜΠΤΑ.ΦΑΝ.200ΜΛ.ΦΡ.ΚΟΚ.ΣΤΑΦ.ΜΗΛ.ΑΧΛ.	ΑΜΠΤΑ.ΦΑΝ.ΜΥΛΕ.20	ΑΜΠΤΑ.ΦΑΝ.200ΜΛ.ΑΠ.ΑΣΤ.ΣΤΑΦ.ΜΗΛ.ΑΧΛ. &&& ΑΜΠΤΑ.ΦΑΝ.200ΜΛ.ΦΡ.ΚΟΚ.ΣΤΑΦ.ΜΗΛ.ΑΧΛ. &&& ΑΜΠΤΑ.ΦΑΝ.ΜΥΛΕ.20
20	203	0.058627195	17.291311955	190.14346234	ΑΜΠΤΑ.ΦΑΝ.ΜΥΛΕ.200ΜΛ	ΑΜΠΤΑ.ΦΑΝ.200ΜΛ.ΑΠ.ΑΣΤ.ΣΤΑΦ.ΜΗΛ.ΑΧΛ.	ΑΜΠΤΑ.ΦΑΝ.ΜΥΛΕ.20	ΑΜΠΤΑ.ΦΑΝ.ΜΥΛΕ.200ΜΛ &&& ΑΜΠΤΑ.ΦΑΝ.200ΜΛ.ΑΠ.ΑΣΤ.ΣΤΑΦ.ΜΗΛ.ΑΧΛ. &&& ΑΜΠΤΑ.ΦΑΝ.ΜΥΛΕ.20
21	212	0.0591377669	40.89978887	185.82300084	ΔΕΛΤΑ.ΑDVANCEΣ.ΒΕΡ.ΑΧΛ.Δ	ΔΕΛΤΑ.ΑDVANCEΣ.ΒΕΡ.ΑΧΛ.Δ.150GRK2.10ΠΕΡ1	ΔΕΛΤΑ.ΑDVANCEΣ.ΒΕΡ.ΑΧΛ.Δ.150GRK2.10ΠΕΡ1	ΔΕΛΤΑ.ΑDVANCEΣ.ΒΕΡ.ΑΧΛ.Δ.150GRK2.10ΠΕΡ1 &&& ΔΕΛΤΑ.ΑDVANCEΣ.ΒΕΡ.ΑΧΛ.Δ.150GRK2.10ΠΕΡ1 &&& ΔΕΛΤΑ.ΑDVANCEΣ.ΒΕΡ.ΑΧΛ.Δ.150GRK2.10ΠΕΡ1
22	212	0.0591377669	27.05639427	185.82300084	ΔΕΛΤΑ.ΑDVANCEΣ.ΜΗ.ΑΧΛ.Δ	ΔΕΛΤΑ.ΑDVANCEΣ.ΒΕΡ.ΑΧΛ.Δ.150GRK2.10ΠΕΡ1	ΔΕΛΤΑ.ΑDVANCEΣ.ΒΕΡ.ΑΧΛ.Δ.150GRK2.10ΠΕΡ1	ΔΕΛΤΑ.ΑDVANCEΣ.ΜΗ.ΑΧΛ.Δ.150GRK2.10ΠΕΡ1 &&& ΔΕΛΤΑ.ΑDVANCEΣ.ΒΕΡ.ΑΧΛ.Δ.150GRK2.10ΠΕΡ1 &&& ΔΕΛΤΑ.ΑDVANCEΣ.ΒΕΡ.ΑΧΛ.Δ.150GRK2.10ΠΕΡ1
23	326	0.09898264	42.22979275	142.07812795	ΑΜΠΤΑ.ΦΑΝ.200ΜΛ.ΑΠ.ΑΣΤ.ΣΤ.	ΑΜΠΤΑ.ΦΑΝ.200ΜΛ.ΦΡ.ΚΟΚ.ΣΤΑΦ.ΜΗΛ.ΑΧΛ.	ΑΜΠΤΑ.ΦΑΝ.ΜΥΛΕ.20	ΑΜΠΤΑ.ΦΑΝ.200ΜΛ.ΑΠ.ΑΣΤ.ΣΤΑΦ.ΜΗΛ.ΑΧΛ. &&& ΑΜΠΤΑ.ΦΑΝ.200ΜΛ.ΦΡ.ΚΟΚ.ΣΤΑΦ.ΜΗΛ.ΑΧΛ. &&& ΑΜΠΤΑ.ΦΑΝ.ΜΥΛΕ.20

Εικόνα 45. Κανόνες συσχέτισης μεταξύ αγορασμένων προϊόντων σε όλες τις συστάδες πελατών

Με βάση τα παραπάνω δεδομένα, γίνεται αντιληπτό ότι ένας πελάτης που έχει αγοράσει τα προϊόντα «Πιπεριές κόκκινες εισαγωγής φρέσκες» και «Πιπεριές πορτοκαλί εισαγωγής φρέσκες», είναι 426,88 φορές πιο πιθανό να αγοράσει το προϊόν «Πιπεριές κίτρινες εισαγωγής φρέσκες», από κάποιον άλλο πελάτη που δεν έχει αγοράσει τα δύο αυτά προϊόντα.



Οι τιμές ανύψωσης (Lift) μπορεί να είναι πολύ υψηλές, ειδικά όταν η πιθανότητα αγοράς του προϊόντος στη δεξιά πλευρά του κανόνα (εδώ οι κίτρινες πιπεριές) είναι πολύ χαμηλή. Εάν δηλαδή ένα στοιχείο αγοράζεται σπάνια μόνο του, αλλά συχνά αγοράζεται όταν δύο άλλα στοιχεία βρίσκονται στο καλάθι, η τιμή lift μπορεί να είναι αρκετά υψηλή. Ωστόσο, είναι ζωτικής σημασίας να σημειωθεί ότι ενώ μια τόσο υψηλή τιμή lift υποδηλώνει ισχυρή συσχέτιση, δεν συνεπάγεται απαραίτητα αιτιότητα.

Αντλώντας στοιχεία από χαμηλότερα lifts του πίνακα, παρουσιάζονται κι άλλα παραδείγματα, όπως αυτά που φαίνονται παρακάτω.

	⊙ COUNT	⊙ SUPPORT	⊙ CONF	⊙ LIFT	⊙ ITEM1	⊙ ITEM2	⊙ ITEM3	⊙ RULE
51	422	0.1177176172	20.220412075	46.258547687	ΦΑΓΕΚΙΝΤΕΡ-ΓΑΛΑΚΤΟΦΕΤΕΣ	ΦΑΓΕΚΙΝΤΕΡ-ΓΑΛΑΚΤΟΦΕΤΕΣ	ΦΑΓΕΚΙΝΤΕΡ-ΡΙΝΓΟΥ	ΦΑΓΕΚΙΝΤΕΡ-ΓΑΛΑΚΤΟΦΕΤΕΣ <=> ΦΑΓΕΚΙΝΤΕΡ-ΓΑΛΑΚΤΟΦΕΤΕΣ <=> ΦΑΓΕΚΙΝΤΕΡ-ΡΙΝΓΟΥ
52	422	0.1177176172	26.930440332	46.258547687	ΦΑΓΕΚΙΝΤΕΡ-ΓΑΛΑΚΤΟΦΕΤΕΣ	ΦΑΓΕΚΙΝΤΕΡ-ΓΑΛΑΚΤΟΦΕΤΕΣ	ΦΑΓΕΚΙΝΤΕΡ-ΡΙΝΓΟΥ	ΦΑΓΕΚΙΝΤΕΡ-ΓΑΛΑΚΤΟΦΕΤΕΣ <=> ΦΑΓΕΚΙΝΤΕΡ-ΓΑΛΑΚΤΟΦΕΤΕΣ <=> ΦΑΓΕΚΙΝΤΕΡ-ΡΙΝΓΟΥ
53	183	0.051048161	19.0625	46.204327931	ΗΛΙΟΣ 30GR-ΑΜΜΩΝΙΑ-ΦΑΚΕΛΟ Σ24	ΑΛΙΑΤΙΝΗ-1KG-ΑΛΕΥΡΙ-3ΤΕΜ-ΙΤΕΜ-Σ24	ΗΛΙΟΣ 30GR-ΑΜΜΩΝΙΑ-ΦΑΚΕΛΟ Σ24	ΗΛΙΟΣ 30GR-ΑΜΜΩΝΙΑ-ΦΑΚΕΛΟ Σ24 <=> ΑΛΙΑΤΙΝΗ-1KG-ΑΛΕΥΡΙ-3ΤΕΜ-ΙΤΕΜ-Σ24 <=> ΗΛΙΟΣ 30GR-ΑΜΜΩΝΙΑ-ΦΑΚΕΛΟ Σ24
54	183	0.051048161	6.9277108434	46.204327931	ΑΛΙΑΤΙΝΗ-1KG-ΑΛΕΥΡΙ-3ΤΕΜ-Ι	ΗΛΙΟΣ 30GR-ΑΜΜΩΝΙΑ-ΦΑΚΕΛΟ Σ24	ΑΛΙΑΤΙΝΗ-1KG-ΑΛΕΥΡΙ-3ΤΕΜ-Ι	ΑΛΙΑΤΙΝΗ-1KG-ΑΛΕΥΡΙ-3ΤΕΜ-Ι <=> ΗΛΙΟΣ 30GR-ΑΜΜΩΝΙΑ-ΦΑΚΕΛΟ Σ24 <=> ΑΛΙΑΤΙΝΗ-1KG-ΑΛΕΥΡΙ-3ΤΕΜ-Ι
55	207	0.0577430018	6.9277108434	43.800360171	ΠΙΠΕΡΙΕΣ-ΦΙΛΑΚΕΣ-ΕΓΧΩΡΙΕΣ	ΜΕΝΤΖΑΝΕΣ-ΦΙΛΑΚΕΣ-ΕΓΧΩΡΙΕΣ-ΦΡΕΣΚΕΣ	ΝΤΟΜΑΤΕΣ-ΕΓΧΩΡΙΕΣ	ΠΙΠΕΡΙΕΣ-ΦΙΛΑΚΕΣ-ΕΓΧΩΡΙΕΣ-ΦΡΕΣΚΕΣ <=> ΜΕΝΤΖΑΝΕΣ-ΦΙΛΑΚΕΣ-ΕΓΧΩΡΙΕΣ-ΦΡΕΣΚΕΣ <=> ΝΤΟΜΑΤΕΣ-ΕΓΧΩΡΙΕΣ
56	207	0.0577430018	36.507936508	43.800360171	ΜΕΝΤΖΑΝΕΣ-ΦΙΛΑΚΕΣ-ΕΓΧΩΡΙΕΣ	ΝΤΟΜΑΤΕΣ-ΕΓΧΩΡΙΕΣ-ΦΡΕΣΚΕΣ	ΠΙΠΕΡΙΕΣ-ΦΙΛΑΚΕΣ-ΕΓΧΩΡΙΕΣ	ΜΕΝΤΖΑΝΕΣ-ΦΙΛΑΚΕΣ-ΕΓΧΩΡΙΕΣ-ΦΡΕΣΚΕΣ <=> ΝΤΟΜΑΤΕΣ-ΕΓΧΩΡΙΕΣ-ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΙΕΣ-ΦΙΛΑΚΕΣ-ΕΓΧΩΡΙΕΣ
57	309	0.081960751	14.82013885	42.131635849	ΜΗΛΙΑΣΤΑΡΙΝΗ-ΕΓΧΩΡΙΑ-ΦΡΕΣΚΑ	ΑΧΛΑΔΙΑ-ΚΡΙΤΑΛΙΑ-ΕΓΧΩΡΙΑ-ΦΡΕΣΚΑ	ΜΗΛΙΑΣΤΑΡΙΝΗ-ΕΓΧΩΡΙΑ-ΦΡΕΣΚΑ	ΜΗΛΙΑΣΤΑΡΙΝΗ-ΕΓΧΩΡΙΑ-ΦΡΕΣΚΑ <=> ΑΧΛΑΔΙΑ-ΚΡΙΤΑΛΙΑ-ΕΓΧΩΡΙΑ-ΦΡΕΣΚΑ <=> ΜΗΛΙΑΣΤΑΡΙΝΗ-ΕΓΧΩΡΙΑ-ΦΡΕΣΚΑ
58	309	0.081960751	24.504361618	42.131635849	ΑΧΛΑΔΙΑ-ΚΡΙΤΑΛΙΑ-ΕΓΧΩΡΙΑ-ΦΡΕΣΚΑ	ΜΗΛΙΑΣΤΑΡΙΝΗ-ΕΓΧΩΡΙΑ-ΦΡΕΣΚΑ	ΜΗΛΙΑΣΤΑΡΙΝΗ-ΕΓΧΩΡΙΑ-ΦΡΕΣΚΑ	ΑΧΛΑΔΙΑ-ΚΡΙΤΑΛΙΑ-ΕΓΧΩΡΙΑ-ΦΡΕΣΚΑ <=> ΜΗΛΙΑΣΤΑΡΙΝΗ-ΕΓΧΩΡΙΑ-ΦΡΕΣΚΑ <=> ΜΗΛΙΑΣΤΑΡΙΝΗ-ΕΓΧΩΡΙΑ-ΦΡΕΣΚΑ
59	294	0.0820117997	35.041716329	41.971031334	ΦΑΓΕΤΖΟΝΙΟΡ-ΠΟΡΤΟΚΑΛΙ-1	ΦΑΓΕΤΖΟΝΙΟΡ-ΚΑΡΑΜΕΛΙΑ-ΦΡΟΝΤΑ-150GR	ΦΑΓΕΤΖΟΝΙΟΡ-ΠΟΡΤΟΚΑΛΙ-150GR	ΦΑΓΕΤΖΟΝΙΟΡ-ΠΟΡΤΟΚΑΛΙ-150GR <=> ΦΑΓΕΤΖΟΝΙΟΡ-ΚΑΡΑΜΕΛΙΑ-ΦΡΟΝΤΑ-150GR <=> ΦΑΓΕΤΖΟΝΙΟΡ-ΠΟΡΤΟΚΑΛΙ-150GR
60	294	0.0820117997	9.822920147	41.971031334	ΦΑΓΕΤΖΟΝΙΟΡ-ΚΑΡΑΜΕΛΙΑ-ΦΡΟΝΤΑ-150GR	ΦΑΓΕΤΖΟΝΙΟΡ-ΠΟΡΤΟΚΑΛΙ-150GR	ΦΑΓΕΤΖΟΝΙΟΡ-ΚΑΡΑΜΕΛΙΑ-ΦΡΟΝΤΑ-150GR	ΦΑΓΕΤΖΟΝΙΟΡ-ΚΑΡΑΜΕΛΙΑ-ΦΡΟΝΤΑ-150GR <=> ΦΑΓΕΤΖΟΝΙΟΡ-ΠΟΡΤΟΚΑΛΙ-150GR <=> ΦΑΓΕΤΖΟΝΙΟΡ-ΚΑΡΑΜΕΛΙΑ-ΦΡΟΝΤΑ-150GR
61	264	0.078432487	21.035890574	40.412856612	ΓΙΤΗΣ-ΕΓΧΩΡΙΕΣ-ΦΑΚΕΛΟ-ΒΑΝΙΛΙΑ	ΓΙΤΗΣ-ΓΚΑΡΙΝΗ-1+1-ΤΕΜ-ΠΡΟΣΦΟΡΑ	ΓΙΤΗΣ-ΕΓΧΩΡΙΕΣ-ΦΑΚΕΛΟ-ΒΑΝΙΛΙΑ	ΓΙΤΗΣ-ΕΓΧΩΡΙΕΣ-ΦΑΚΕΛΟ-ΒΑΝΙΛΙΑ <=> ΓΙΤΗΣ-ΓΚΑΡΙΝΗ-1+1-ΤΕΜ-ΠΡΟΣΦΟΡΑ <=> ΓΙΤΗΣ-ΕΓΧΩΡΙΕΣ-ΦΑΚΕΛΟ-ΒΑΝΙΛΙΑ
62	264	0.078432487	14.14709968	40.412856612	ΓΙΤΗΣ-ΕΓΧΩΡΙΕΣ-ΦΑΚΕΛΟ-ΒΑΝΙΛΙΑ	ΓΙΤΗΣ-ΕΓΧΩΡΙΕΣ-ΦΑΚΕΛΟ-ΒΑΝΙΛΙΑ	ΓΙΤΗΣ-ΕΓΧΩΡΙΕΣ-ΦΑΚΕΛΟ-ΒΑΝΙΛΙΑ	ΓΙΤΗΣ-ΕΓΧΩΡΙΕΣ-ΦΑΚΕΛΟ-ΒΑΝΙΛΙΑ <=> ΓΙΤΗΣ-ΕΓΧΩΡΙΕΣ-ΦΑΚΕΛΟ-ΒΑΝΙΛΙΑ <=> ΓΙΤΗΣ-ΕΓΧΩΡΙΕΣ-ΦΑΚΕΛΟ-ΒΑΝΙΛΙΑ
63	219	0.0610904222	14.98973306	40.161939541	ΡΟΥΛΙΑΚΑ-ΣΑΚΑΛΙ-1	ΡΟΥΛΙΑΚΑ-Νo7-ΣΑΚΤΡΟΦ-Χ25-ΓΥΔΑ-Ε40	ΡΟΥΛΙΑΚΑ-ΣΑΚΑΛΙ-1	ΡΟΥΛΙΑΚΑ-ΣΑΚΑΛΙ-1 <=> ΡΟΥΛΙΑΚΑ-Νo7-ΣΑΚΤΡΟΦ-Χ25-ΓΥΔΑ-Ε40 <=> ΡΟΥΛΙΑΚΑ-ΣΑΚΑΛΙ-1
64	219	0.0610904222	16.367713004	40.161939541	ΡΟΥΛΙΑΚΑ-Νo7-ΣΑΚΤΡΟΦ-Χ25-ΓΥΔΑ-Ε40	ΡΟΥΛΙΑΚΑ-ΣΑΚΑΛΙ-1	ΡΟΥΛΙΑΚΑ-Νo7-ΣΑΚΤΡΟΦ-Χ25-ΓΥΔΑ-Ε40	ΡΟΥΛΙΑΚΑ-Νo7-ΣΑΚΤΡΟΦ-Χ25-ΓΥΔΑ-Ε40 <=> ΡΟΥΛΙΑΚΑ-ΣΑΚΑΛΙ-1 <=> ΡΟΥΛΙΑΚΑ-Νo7-ΣΑΚΤΡΟΦ-Χ25-ΓΥΔΑ-Ε40
65	207	0.0577430018	12.699365053	38.811079933	ΜΕΝΤΖΑΝΕΣ-ΦΙΛΑΚΕΣ-ΕΓΧΩΡΙΕΣ	ΝΤΟΜΑΤΕΣ-ΕΓΧΩΡΙΕΣ-ΦΡΕΣΚΕΣ	ΠΙΠΕΡΙΕΣ-ΦΙΛΑΚΕΣ-ΕΓΧΩΡΙΕΣ	ΜΕΝΤΖΑΝΕΣ-ΦΙΛΑΚΕΣ-ΕΓΧΩΡΙΕΣ <=> ΝΤΟΜΑΤΕΣ-ΕΓΧΩΡΙΕΣ-ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΙΕΣ-ΦΙΛΑΚΕΣ-ΕΓΧΩΡΙΕΣ
66	207	0.0577430018	17.647058824	38.811079933	ΝΤΟΜΑΤΕΣ-ΕΓΧΩΡΙΕΣ-ΦΡΕΣΚΕΣ	ΠΙΠΕΡΙΕΣ-ΦΙΛΑΚΕΣ-ΕΓΧΩΡΙΕΣ-ΦΡΕΣΚΕΣ	ΜΕΝΤΖΑΝΕΣ-ΦΙΛΑΚΕΣ-ΕΓΧΩΡΙΕΣ	ΝΤΟΜΑΤΕΣ-ΕΓΧΩΡΙΕΣ-ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΙΕΣ-ΦΙΛΑΚΕΣ-ΕΓΧΩΡΙΕΣ-ΦΡΕΣΚΕΣ <=> ΜΕΝΤΖΑΝΕΣ-ΦΙΛΑΚΕΣ-ΕΓΧΩΡΙΕΣ
67	564	0.1572287585	20.04976893	38.863482597	ΑΝΘΟΣ-ΕΓΧΩΡΙΕΣ-ΙΤΕΜ-ΦΡ	ΚΡΕΜΜΥΔΙΑ-ΧΩΡΑ-ΕΓΧΩΡΙΑ-ΦΡΕΣΚΑ	ΑΝΘΟΣ-ΕΓΧΩΡΙΕΣ-ΙΤΕΜ-ΦΡΕΣΚΟΣ	ΑΝΘΟΣ-ΕΓΧΩΡΙΕΣ-ΙΤΕΜ-ΦΡΕΣΚΟΣ <=> ΚΡΕΜΜΥΔΙΑ-ΧΩΡΑ-ΕΓΧΩΡΙΑ-ΦΡΕΣΚΑ <=> ΑΝΘΟΣ-ΕΓΧΩΡΙΕΣ-ΙΤΕΜ-ΦΡΕΣΚΟΣ
68	564	0.1572287585	30.38891877	38.863482597	ΚΡΕΜΜΥΔΙΑ-ΧΩΡΑ-ΕΓΧΩΡΙΑ-ΦΡΕΣΚΑ	ΑΝΘΟΣ-ΕΓΧΩΡΙΕΣ-ΙΤΕΜ-ΦΡΕΣΚΟΣ	ΑΝΘΟΣ-ΕΓΧΩΡΙΕΣ-ΙΤΕΜ-ΦΡΕΣΚΟΣ	ΚΡΕΜΜΥΔΙΑ-ΧΩΡΑ-ΕΓΧΩΡΙΑ-ΦΡΕΣΚΑ <=> ΑΝΘΟΣ-ΕΓΧΩΡΙΕΣ-ΙΤΕΜ-ΦΡΕΣΚΟΣ <=> ΑΝΘΟΣ-ΕΓΧΩΡΙΕΣ-ΙΤΕΜ-ΦΡΕΣΚΟΣ
69	182	0.0507962093	21.08922636	37.87660689	ΜΟΛΤΟ-Β50GR-ΚΡΟΧΙΔΙΑ-ΚΕΡΑΣ-Σ20	ΜΟΛΤΟ-Β50GR-ΚΡΟΧΙΔΙΑ-ΑΤΟΜΙΚΟ-Σ20	ΜΟΛΤΟ-Β50GR-ΚΡΟΧΙΔΙΑ-ΚΕΡΑΣ-Σ20	ΜΟΛΤΟ-Β50GR-ΚΡΟΧΙΔΙΑ-ΚΕΡΑΣ-Σ20 <=> ΜΟΛΤΟ-Β50GR-ΚΡΟΧΙΔΙΑ-ΑΤΟΜΙΚΟ-Σ20 <=> ΜΟΛΤΟ-Β50GR-ΚΡΟΧΙΔΙΑ-ΚΕΡΑΣ-Σ20
70	182	0.0507962093	9.1182364729	37.87660689	ΜΟΛΤΟ-Β50GR-ΚΡΟΧΙΔΙΑ-ΑΤΟΜΙΚΟ-Σ20	ΜΟΛΤΟ-Β50GR-ΚΡΟΧΙΔΙΑ-ΚΕΡΑΣ-Σ20	ΜΟΛΤΟ-Β50GR-ΚΡΟΧΙΔΙΑ-ΑΤΟΜΙΚΟ-Σ20	ΜΟΛΤΟ-Β50GR-ΚΡΟΧΙΔΙΑ-ΑΤΟΜΙΚΟ-Σ20 <=> ΜΟΛΤΟ-Β50GR-ΚΡΟΧΙΔΙΑ-ΚΕΡΑΣ-Σ20 <=> ΜΟΛΤΟ-Β50GR-ΚΡΟΧΙΔΙΑ-ΑΤΟΜΙΚΟ-Σ20
71	187	0.0521699678	38.68091595	37.212094419	ΡΑΠΤΑ-1500CC-ΠΟΡΤΟΚΑΛΙ-Ε2	ΚΩΚΑ-ΚΟΛΑ-ΕΤΕ-ΟΝΙ-ΣΑΙΤ	ΡΑΠΤΑ-1500CC-Ε2	ΡΑΠΤΑ-1500CC-ΠΟΡΤΟΚΑΛΙ-Ε2 <=> ΚΩΚΑ-ΚΟΛΑ-ΕΤΕ-ΟΝΙ-ΣΑΙΤ <=> ΡΑΠΤΑ-1500CC-Ε2
...	...	...	...	...	...	...	...	...

Εικόνα 46. Κανόνες συσχέτισης μεταξύ αγορασμένων προϊόντων σε όλες τις συστάδες πελατών (συνέχεια)

Παρατηρείται ότι ένας πελάτης που έχει αγοράσει το προϊόν «Κρεμμύδια χλωρά εγχώρια φρέσκα», είναι 38,66 φορές πιο πιθανό να αγοράσει το προϊόν «Άνθος εγχώριος φρέσκος», από κάποιον άλλο πελάτη που δεν έχει αγοράσει το πρώτο προϊόν. Οι κανόνες συσχέτισης προϊόντων που εξήχθησαν δηλαδή, δεν αφορούν μόνο τριάδες προϊόντων αλλά και σχέσεις μεταξύ δύο προϊόντων.

Η τιμή count που εμφανίζεται κάθε φορά στην πρώτη στήλη αριστερά του πίνακα, εμφανίζει τον αριθμό που η συσχέτιση αυτή έχει λάβει ήδη χώρα στη βάση δεδομένων μας και φανερώνει πόσο σημαντικός είναι ο κανόνας που παρουσιάζεται.

Ακόμα περισσότερο σημαντική είναι η τιμή support. Όπως αναφέρθηκε παραπάνω, η μαθηματική φόρμουλα του support δίνεται από τον τύπο:

$$\text{Support} = \frac{\text{Αριθμός συναλλαγών όπου εμφανίζεται ένα σύνολο προϊόντων}}{\text{Συνολικός αριθμός συναλλαγών}}$$

Δηλαδή το support παρουσιάζει το ποσοστό των συναλλαγών στο σύνολο δεδομένων που περιέχουν το αντίστοιχο σύνολο προϊόντων. Κατά συνέπεια, το πρώτο παράδειγμα με τις πιπεριές, έχει πολύ χαμηλότερη συχνότητα εμφάνισης στο σύνολο δεδομένων (Support=0,061), απ' ότι το δεύτερο παράδειγμα με το κρεμμύδι και τον άνθηθο, όπου η συχνότητα εμφάνισης είναι 0,5.

Πρέπει να εξεταστεί κατά πόσο ένας κανόνας συσχέτισης είναι τελικά αρκετά χρήσιμος για να επηρεάσει ή όχι την εταιρική στρατηγική, αφού μια μικρή συχνότητα εμφάνισης του κανόνα στο γενικό πληθυσμό, μπορεί να μην έχει μεγάλη επιρροή στα έσοδα μιας εταιρείας.

### 3.5.3 Κανόνες συσχέτισης μεταξύ αγορασμένων προϊόντων σε συστάδες πελατών

Κανόνες συσχέτισης που μπορεί να ισχύουν σε όλες τις συστάδες των πελατών, μπορεί να μην ισχύουν σε μία συγκεκριμένη συστάδα πελατών, στην οποία η εταιρεία μπορεί να στοχεύει είτε με προωθήσεις ή άλλες στρατηγικές ενέργειες. Χρησιμοποιώντας τα αποτελέσματα της ανάλυσης RFM που ταξινόμησε τους πελάτες σε διαφορετικές συστάδες, μπορούμε να αντλήσουμε εξειδικευμένη γνώση σχετικά με τους κανόνες συσχέτισης προϊόντων σε κάθε μία από αυτές.

Επαναλήφθηκε η διαδικασία για την ανεύρεση συσχετίσεων μεταξύ προϊόντων, αυτήν τη φορά συγκεκριμένα για την ομάδα πελατών High Value. Τα αποτελέσματα φαίνονται παρακάτω.

	@ COUNT	@ SUPPORT	@ CONF	@ LIFT ↑	↳ ITEM1	↳ ITEM2	↳ ITEM3	↳ RULE
1	153	0.0565297391	67.105263158	462.14523905	ΠΙΠΕΡΙΕΣ-ΚΟΚΚΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕ...	ΠΙΠΕΡΙΕΣ-ΠΟΡΤΟΚΑΛΙ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕΣΚΕ...	ΠΙΠΕΡΙΕΣ-ΚΙΤΡΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-Φ...	ΠΙΠΕΡΙΕΣ-ΚΟΚΚΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕΣΚΕΣ... & ΠΙΠΕΡΙΕΣ-ΠΟΡΤΟΚΑΛΙ-ΕΙΣ...
2	153	0.0565297391	38.93129771	462.14523905	ΠΙΠΕΡΙΕΣ-ΚΙΤΡΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕ...	ΠΙΠΕΡΙΕΣ-ΚΟΚΚΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕΣΚΕ...	ΠΙΠΕΡΙΕΣ-ΠΟΡΤΟΚΑΛΙ-ΕΙΣΑΓΩΓΗΣ...	ΠΙΠΕΡΙΕΣ-ΚΟΚΚΙΝΕΣ-ΕΙΣ...
3	229	0.0864098709	40.582010582	417.22044509	ΠΙΠΕΡΙΕΣ-ΠΟΡΤΟΚΑΛΙ-ΕΙΣΑΓΩΓΗΣ-Φ...	ΠΙΠΕΡΙΕΣ-ΚΙΤΡΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕΣΚΕ...		ΠΙΠΕΡΙΕΣ-ΠΟΡΤΟΚΑΛΙ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕΣΚΕ... >>> ΠΙΠΕΡΙΕΣ-ΚΙΤΡΙΝΕΣ-ΕΙ...
4	229	0.0864098709	58.269720102	417.22044509	ΠΙΠΕΡΙΕΣ-ΚΙΤΡΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕ...	ΠΙΠΕΡΙΕΣ-ΠΟΡΤΟΚΑΛΙ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕΣΚ...		ΠΙΠΕΡΙΕΣ-ΚΙΤΡΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕΣΚΕ... >>> ΠΙΠΕΡΙΕΣ-ΠΟΡΤΟΚΑΛΙ-ΕΙ...
5	153	0.0565297391	40.476190476	407.23066383	ΠΙΠΕΡΙΕΣ-ΠΟΡΤΟΚΑΛΙ-ΕΙΣΑΓΩΓΗΣ-Φ...	ΠΙΠΕΡΙΕΣ-ΚΙΤΡΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕΣΚ...	ΠΙΠΕΡΙΕΣ-ΚΟΚΚΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ...	ΠΙΠΕΡΙΕΣ-ΠΟΡΤΟΚΑΛΙ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕΣΚΕ... >>> ΠΙΠΕΡΙΕΣ-ΚΙΤΡΙΝΕΣ-ΕΙ...
6	153	0.0565297391	56.87732342	407.23066383	ΠΙΠΕΡΙΕΣ-ΚΙΤΡΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕ...	ΠΙΠΕΡΙΕΣ-ΚΟΚΚΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕΣΚΕ...	ΠΙΠΕΡΙΕΣ-ΠΟΡΤΟΚΑΛΙ-ΕΙΣΑΓΩΓΗΣ...	ΠΙΠΕΡΙΕΣ-ΚΙΤΡΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕΣΚΕ... >>> ΠΙΠΕΡΙΕΣ-ΚΟΚΚΙΝΕΣ-ΕΙΣΑΓ...
7	269	0.0993888877	49.63099631	341.80223092	ΠΙΠΕΡΙΕΣ-ΚΟΚΚΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕ...	ΠΙΠΕΡΙΕΣ-ΚΙΤΡΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕΣΚ...		ΠΙΠΕΡΙΕΣ-ΚΟΚΚΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕ... >>> ΠΙΠΕΡΙΕΣ-ΚΙΤΡΙΝΕΣ-ΕΙΣ...
8	269	0.0993888877	68.44783715	341.80223092	ΠΙΠΕΡΙΕΣ-ΚΙΤΡΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕ...	ΠΙΠΕΡΙΕΣ-ΚΟΚΚΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕΣΚ...		ΠΙΠΕΡΙΕΣ-ΚΙΤΡΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕΣΚΕ... >>> ΠΙΠΕΡΙΕΣ-ΚΟΚΚΙΝΕΣ-ΕΙΣ...
9	153	0.0565297391	66.812227074	333.63462189	ΠΙΠΕΡΙΕΣ-ΠΟΡΤΟΚΑΛΙ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕ...	ΠΙΠΕΡΙΕΣ-ΚΙΤΡΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕΣΚ...	ΠΙΠΕΡΙΕΣ-ΚΟΚΚΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ...	ΠΙΠΕΡΙΕΣ-ΚΙΤΡΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕΣΚΕ... >>> ΠΙΠΕΡΙΕΣ-ΠΟΡΤΟΚΑΛΙ-ΕΙΣΑΓ...
10	153	0.0565297391	28.228782288	333.63462189	ΠΙΠΕΡΙΕΣ-ΚΟΚΚΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕ...	ΠΙΠΕΡΙΕΣ-ΚΙΤΡΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕΣΚ...	ΠΙΠΕΡΙΕΣ-ΠΟΡΤΟΚΑΛΙ-ΕΙΣΑΓΩΓΗΣ...	ΠΙΠΕΡΙΕΣ-ΚΟΚΚΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕΣΚΕ... >>> ΠΙΠΕΡΙΕΣ-ΚΙΤΡΙΝΕΣ-ΕΙΣ...
11	228	0.0842403955	42.066420664	301.20224917	ΠΙΠΕΡΙΕΣ-ΚΟΚΚΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕ...	ΠΙΠΕΡΙΕΣ-ΠΟΡΤΟΚΑΛΙ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕΣΚ...		ΠΙΠΕΡΙΕΣ-ΚΟΚΚΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕΣΚΕ... >>> ΠΙΠΕΡΙΕΣ-ΠΟΡΤΟΚΑΛΙ-ΕΙ...
12	228	0.0842403955	60.317460317	301.20224917	ΠΙΠΕΡΙΕΣ-ΠΟΡΤΟΚΑΛΙ-ΕΙΣΑΓΩΓΗΣ-Φ...	ΠΙΠΕΡΙΕΣ-ΚΟΚΚΙΝΕΣ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕΣΚΕ...		ΠΙΠΕΡΙΕΣ-ΠΟΡΤΟΚΑΛΙ-ΕΙΣΑΓΩΓΗΣ-ΦΡΕΣΚΕ... >>> ΠΙΠΕΡΙΕΣ-ΚΟΚΚΙΝΕΣ-ΕΙ...
13	138	0.0509876078	24	210.21649903	ΑΜΙΤΑ-ΦΑΝ-200ΜΛ-ΜΠ-ΑΠΙ-ΣΤΑΦ-Μ...	ΑΜΙΤΑ-ΦΑΝ-200ΜΛ-ΦΡ-ΚΟΚ-ΣΤΑΦ-ΜΗΛΟ...	ΑΜΙΤΑ-ΦΑΝ-ΜΠΛΕ-200ΜΛ...	ΑΜΙΤΑ-ΦΑΝ-200ΜΛ-ΜΠ-ΑΠΙ-ΣΤΑΦ-ΜΗΛΟ-ΑΧΛ... >>> ΑΜΙΤΑ-ΦΑΝ-200ΜΛ...
14	138	0.0509876078	44.660194175	210.21649903	ΑΜΙΤΑ-ΦΑΝ-200ΜΛ-ΦΡ-ΚΟΚ-ΣΤΑΦ-Μ...	ΑΜΙΤΑ-ΦΑΝ-ΜΠΛΕ-200ΜΛ...		ΑΜΙΤΑ-ΦΑΝ-200ΜΛ-ΜΠ-ΑΠΙ-ΣΤΑΦ-ΜΗΛΟ... >>> ΑΜΙΤΑ-ΦΑΝ-ΜΠΛΕ-200...
15	173	0.0639192475	31.977818854	198.50744459	ΔΕΛΤΑ-150GRΚ2-ADVANCE-ΜΗΛΟ/Μ...	ΔΕΛΤΑ-150GRΚ2-ADVANCE-ΒΕΡ/ΑΧΛΔ...		ΔΕΛΤΑ-150GRΚ2-ADVANCE-ΜΗΛΟ/ΜΠΑΝ/20ΠΕΡΤ... >>> ΔΕΛΤΑ-150GRΚ2...
16	173	0.0639192475	39.678890983	198.50744459	ΔΕΛΤΑ-150GRΚ2-ADVANCE-ΒΕΡ/ΑΧΛ...	ΔΕΛΤΑ-150GRΚ2-ADVANCE-ΜΗΛΟ/ΜΠΑΝ...		ΔΕΛΤΑ-150GRΚ2-ADVANCE-ΒΕΡ/ΑΧΛΔ/20ΠΕΡΤ... >>> ΔΕΛΤΑ-150GRΚ2...
17	138	0.0509876078	47.750865052	183.31861886	ΑΜΙΤΑ-ΦΑΝ-200ΜΛ-ΜΠ-ΑΠΙ-ΣΤΑΦ-Μ...	ΑΜΙΤΑ-ΦΑΝ-ΜΠΛΕ-200ΜΛ...	ΑΜΙΤΑ-ΦΑΝ-200ΜΛ-ΦΡ-ΚΟΚ-ΣΤΑ...	ΑΜΙΤΑ-ΦΑΝ-200ΜΛ-ΜΠ-ΑΠΙ-ΣΤΑΦ-ΜΗΛΟ-ΑΧΛ... & ΑΜΙΤΑ-ΦΑΝ-ΜΠΛΕ-20...
18	138	0.0509876078	19.374480805	183.31861886	ΑΜΙΤΑ-ΦΑΝ-200ΜΛ-ΦΡ-ΚΟΚ-ΣΤΑΦ-Μ...	ΑΜΙΤΑ-ΦΑΝ-200ΜΛ-ΜΠ-ΑΠΙ-ΣΤΑΦ-ΜΗΛ...	ΑΜΙΤΑ-ΦΑΝ-ΜΠΛΕ-200ΜΛ...	ΑΜΙΤΑ-ΦΑΝ-200ΜΛ-ΦΡ-ΚΟΚ-ΣΤΑΦ-ΜΗΛΟ... >>> ΑΜΙΤΑ-ΦΑΝ-200ΜΛ...
19	138	0.0509876078	14.634146341	177.61391228	ΑΜΙΤΑ-ΦΑΝ-ΜΠΛΕ-200ΜΛ...	ΑΜΙΤΑ-ΦΑΝ-200ΜΛ-ΜΠ-ΑΠΙ-ΣΤΑΦ-ΜΗΛ...	ΑΜΙΤΑ-ΦΑΝ-200ΜΛ-ΦΡ-ΚΟΚ-ΣΤΑ...	ΑΜΙΤΑ-ΦΑΝ-ΜΠΛΕ-200ΜΛ... >>> ΑΜΙΤΑ-ΦΑΝ-200ΜΛ-ΜΠ-ΑΠΙ...

Εικόνα 47. Κανόνες συσχέτισης μεταξύ αγορασμένων προϊόντων στη συστάδα πελατών High Value

Παρατηρούμε ότι τα αποτελέσματα κανόνων συσχέτισης μεταξύ προϊόντων της συστάδας High Value, είναι πολύ κοντινά των αποτελεσμάτων για όλες τις συστάδες. Και πάλι εδώ, αυτό έχει να κάνει φυσικά με την τεράστια επίδραση που έχει αυτή η ομάδα στον συνολικό όγκο πωλήσεων προϊόντων. Μάλιστα σε συγκεκριμένους κανόνες έχει μεγαλύτερο lift απ' ό,τι το γενικό σύνολο συστάδων. Για παράδειγμα, ένας πελάτης που έχει αγοράσει τα προϊόντα «Πιπεριές κόκκινες εισαγωγής φρέσκες» και «Πιπεριές πορτοκαλί εισαγωγής φρέσκες», είναι 462,14 φορές πιο πιθανό να αγοράσει το προϊόν «Πιπεριές κίτρινες εισαγωγής φρέσκες», από κάποιον άλλο πελάτη που δεν έχει αγοράσει τα δύο αυτά προϊόντα (σε αντίθεση με το 426,88 lift του γενικού συνόλου). Από τις συνολικές 219 συναλλαγές/καλάθια που περιείχαν αυτά τα τρία προϊόντα, τα 153 φαίνεται να προέρχονται από αυτήν τη συστάδα πελατών. Τα ευρήματα των προηγούμενων αναλύσεων μας, επιβεβαιώνονται ξανά.

Δεδομένου ότι η ομάδα αυτή έχει τόση επιρροή στα γενικά αποτελέσματα, είναι ενδιαφέρον να εξεταστεί και ο κανόνας με το μικρότερο lift, όπως φαίνεται παρακάτω.

	@ COUNT	@ SUPPORT	@ CONF	@ LIFT ↑	↳ ITEM1	↳ ITEM2	↳ ITEM3	↳ RULE
1	153	0.0565297391	5.765230992	1.000091326	TWIN-ΦΙΑΛΙΔΙΟ-190GR-E48	ΔΕΛΤΑ-ΓΑΛΑ-ΠΑΤ-ΚΑΡΤ-ΕΙΣΚ-1ΜΤ...		TWIN-ΦΙΑΛΙΔΙΟ-190GR-E48 >>> ΔΕΛΤΑ-ΓΑΛΑ-ΠΑΤ-ΚΑΡΤ-ΕΙΣΚ...
2	153	0.0565297391	0.9891409933	1.000091326	ΔΕΛΤΑ-ΓΑΛΑ-ΠΑΤ-ΚΑΡΤ-ΕΙΣΚ-1ΜΤ...	TWIN-ΦΙΑΛΙΔΙΟ-190GR-E48		ΔΕΛΤΑ-ΓΑΛΑ-ΠΑΤ-ΚΑΡΤ-ΕΙΣΚ-1ΜΤ... >>> TWIN-ΦΙΑΛΙΔΙΟ-190GR-E48
3	173	0.0639192475	5.8273178919	1.000091326	ΑΤΙΛΑΝΤΙΚΟΝ-ΥΠΕΡ-ΣΟΠΡΕΣ-ΣΤΡΗΝΚ...	ΔΕΛΤΑ-ΓΑΛΑ-ΠΑΤ-ΚΑΡΤ-ΕΙΣΚ-1ΜΤ...		ΑΤΙΛΑΝΤΙΚΟΝ-ΥΠΕΡ-ΣΟΠΡΕΣ-ΣΤΡΗΝΚ... >>> ΔΕΛΤΑ-ΓΑΛΑ-ΠΑΤ-ΚΑΡΤ...
4	173	0.0639192475	1.1002630384	1.000091326	ΔΕΛΤΑ-ΓΑΛΑ-ΠΑΤ-ΚΑΡΤ-ΕΙΣΚ-1ΜΤ...	ΑΤΙΛΑΝΤΙΚΟΝ-ΥΠΕΡ-ΣΟΠΡΕΣ-ΣΤΡΗΝΚ...		ΔΕΛΤΑ-ΓΑΛΑ-ΠΑΤ-ΚΑΡΤ-ΕΙΣΚ-1ΜΤ... >>> ΑΤΙΛΑΝΤΙΚΟΝ-ΥΠΕΡ-ΣΟΠΡ...
5	168	0.0620318704	5.847546119	1.01387084	ΜΑΡΟΝΙΑ-ΕΓΧΩΡΙΑ-ΤΤΕΜ-ΦΡΕΣΚΑ...	ΔΕΛΤΑ-ΓΑΛΑ-ΠΑΤ-ΚΑΡΤ-ΕΙΣΚ-1ΜΤ...		ΜΑΡΟΝΙΑ-ΕΓΧΩΡΙΑ-ΤΤΕΜ-ΦΡΕΣΚΑ... >>> ΔΕΛΤΑ-ΓΑΛΑ-ΠΑΤ-ΚΑΡΤ...
6	168	0.0620318704	0.7822318039	1.01387084	ΔΕΛΤΑ-ΓΑΛΑ-ΠΑΤ-ΚΑΡΤ-ΕΙΣΚ-1ΜΤ...	ΜΑΡΟΝΙΑ-ΕΓΧΩΡΙΑ-ΤΤΕΜ-ΦΡΕΣΚΑ...		ΔΕΛΤΑ-ΓΑΛΑ-ΠΑΤ-ΚΑΡΤ-ΕΙΣΚ-1ΜΤ... >>> ΜΑΡΟΝΙΑ-ΕΓΧΩΡΙΑ-ΤΤΕΜ...
7	152	0.0541602837	1.9491649905	1.0145877073	ΝΤΟΜΑΤΕΣ-ΕΓΧΩΡΙΕΣ-ΦΡΕΣΚΕΣ...	ΟΛΙΜΠΟΣ-1ΜΤ-ΓΑΛΑ-ΕΙΣΑΓΩΓΗΣ-ΕΠΙΣΤΗΜΕΝΟ...		ΝΤΟΜΑΤΕΣ-ΕΓΧΩΡΙΕΣ-ΦΡΕΣΚΕΣ... >>> ΟΛΙΜΠΟΣ-1ΜΤ-ΓΑΛΑ-ΕΙΣΑΓ...
8	152	0.0541602837	2.8935864183	1.0145877073	ΟΛΙΜΠΟΣ-1ΜΤ-ΓΑΛΑ-ΕΙΣΑΓΩΓΗΣ-ΕΠΙΣΤΗΜΕΝΟ...	ΝΤΟΜΑΤΕΣ-ΕΓΧΩΡΙΕΣ-ΦΡΕΣΚΕΣ...		ΟΛΙΜΠΟΣ-1ΜΤ-ΓΑΛΑ-ΕΙΣΑΓΩΓΗΣ-ΕΠΙΣΤΗΜΕΝΟ... >>> ΝΤΟΜΑΤΕΣ-ΕΓΧΩΡ...
9	191	0.070598005	1.225746216	1.0177177884	ΔΕΛΤΑ-ΓΑΛΑ-ΠΑΤ-ΚΑΡΤ-ΕΙΣΚ-1ΜΤ...	ΦΕΤΑ-ΔΟΚΙΟ-ΩΔΩΝΗ...		ΔΕΛΤΑ-ΓΑΛΑ-ΠΑΤ-ΚΑΡΤ-ΕΙΣΚ-1ΜΤ... >>> ΦΕΤΑ-ΔΟΚΙΟ-ΩΔΩΝΗ...
10	191	0.070598005	5.8094988322	1.0177177884	ΦΕΤΑ-ΔΟΚΙΟ-ΩΔΩΝΗ...	ΔΕΛΤΑ-ΓΑΛΑ-ΠΑΤ-ΚΑΡΤ-ΕΙΣΚ-1ΜΤ...		ΦΕΤΑ-ΔΟΚΙΟ-ΩΔΩΝΗ... >>> ΔΕΛΤΑ-ΓΑΛΑ-ΠΑΤ-ΚΑΡΤ-ΕΙΣΚ...
11	166	0.061329195	4.0937114673	1.017895621	ΑΠΡΟΣΣΑ-ΑΠΤΑ-10ΑΔΑ...	ΝΟΝΟΝ-410ΓΡ-ΓΑΛΑ-ΕΒΑΡΟΡΕ-Ε48...		ΑΠΡΟΣΣΑ-ΑΠΤΑ-10ΑΔΑ... >>> ΝΟΝΟΝ-410ΓΡ-ΓΑΛΑ-ΕΒΑΡΟΡ...
12	166	0.061329195	1.5230364511	1.017895621	ΝΟΝΟΝ-410ΓΡ-ΓΑΛΑ-ΕΒΑΡΟΡΕ-Ε48...	ΑΠΡΟΣΣΑ-ΑΠΤΑ-10ΑΔΑ...		ΝΟΝΟΝ-410ΓΡ-ΓΑΛΑ-ΕΒΑΡΟΡΕ-Ε48... >>> ΑΠΡΟΣΣΑ-ΑΠΤΑ-10ΑΔΑ...
13	139	0.0513579832	5.8724123363	1.0181882494	CAMPING-GAS-C230-ΦΙΑΛΙΔΙΟ-ΦΗΜΑ-ΣΑ...	ΔΕΛΤΑ-ΓΑΛΑ-ΠΑΤ-ΚΑΡΤ-ΕΙΣΚ-1ΜΤ...		CAMPING-GAS-C230-ΦΙΑΛΙΔΙΟ-ΦΗΜΑ-ΣΑ... >>> ΔΕΛΤΑ-ΓΑΛΑ-ΠΑΤ-ΚΑΡΤ...
14	139	0.0513579832	0.8904548366	1.0181882494	ΔΕΛΤΑ-ΓΑΛΑ-ΠΑΤ-ΚΑΡΤ-ΕΙΣΚ-1ΜΤ...	CAMPING-GAS-C230-ΦΙΑΛΙΔΙΟ-ΦΗΜΑ-ΣΑ...		ΔΕΛΤΑ-ΓΑΛΑ-ΠΑΤ-ΚΑΡΤ-ΕΙΣΚ-1ΜΤ... >>> CAMPING-GAS-C230-ΦΙΑΛ...
15	218	0.0805454413	0.7369995348	1.0186978853	ΖΑΧΑΡΗ-1ΚΩ...	ΑΝΘΟΣ-ΕΓΧΩΡΟΣ-ΤΤΕΜ-ΦΡΕΣΚΟΣ...		ΖΑΧΑΡΗ-1ΚΩ... >>> ΑΝΘΟΣ-ΕΓΧΩΡΟΣ-ΤΤΕΜ-ΦΡΕΣΚΟΣ...
16	218	0.0805454413	11.13381001	1.0186978853	ΑΝΘΟΣ-ΕΓΧΩΡΟΣ-ΤΤΕΜ-ΦΡΕΣΚΟΣ...	ΖΑΧΑΡΗ-1ΚΩ...		ΑΝΘΟΣ-ΕΓΧΩΡΟΣ-ΤΤΕΜ-ΦΡΕΣΚΟΣ... >>> ΖΑΧΑΡΗ-1ΚΩ...
17	203	0.07500351	0.68623131	1.0199706867	ΖΑΧΑΡΗ-1ΚΩ...	ΑΝΘΟΣ-ΕΓΧΩΡΟΣ-ΤΤΕΜ-ΦΡΕΣΚΟΣ...		ΖΑΧΑΡΗ-1ΚΩ... >>> ΑΝΘΟΣ-ΕΓΧΩΡΟΣ-ΤΤΕΜ-ΦΡΕΣΚΟΣ...
18	203	0.07500351	11.14721032	1.0199706867	ΑΝΘΟΣ-ΕΓΧΩΡΟΣ-ΤΤΕΜ-ΦΡΕΣΚΟΣ...	ΖΑΧΑΡΗ-1ΚΩ...		ΑΝΘΟΣ-ΕΓΧΩΡΟΣ-ΤΤΕΜ-ΦΡΕΣΚΟΣ... >>> ΖΑΧΑΡΗ-1ΚΩ...
19	159	0.0587465916	2.0598823125	1.0205148596	ΝΤΟΜΑΤΕΣ-ΕΓΧΩΡΙΕΣ-ΦΡΕΣΚΕΣ...	ΑΜΙΤΑ-ΜΟΔΟΝ-1ΛΙΤΡΩΝ...		ΝΤΟΜΑΤΕΣ-ΕΓΧΩΡΙΕΣ-ΦΡΕΣΚΕΣ... >>> ΑΜΙΤΑ-ΜΟΔΟΝ-1ΛΙΤΡΩΝ...
20	159	0.0587465916	2.9104887424	1.0205148596	ΑΜΙΤΑ-ΜΟΔΟΝ-1ΛΙΤΡΩΝ...	ΝΤΟΜΑΤΕΣ-ΕΓΧΩΡΙΕΣ-ΦΡΕΣΚΕΣ...		ΑΜΙΤΑ-ΜΟΔΟΝ-1ΛΙΤΡΩΝ... >>> ΝΤΟΜΑΤΕΣ-ΕΓΧΩΡΙΕΣ-ΦΡΕΣΚΕ...
21	321	0.1186016094	1.0851560123	1.0223826848	ΖΑΧΑΡΗ-1ΚΩ...	ΜΑΡΟΝΙΑ-ΕΓΧΩΡΙΑ-ΤΤΕΜ-ΦΡΕΣΚΑ...		ΖΑΧΑΡΗ-1ΚΩ... >>> ΜΑΡΟΝΙΑ-ΕΓΧΩΡΙΑ-ΤΤΕΜ-ΦΡΕΣΚΕ...
22	321	0.1186016094	11.172498906	1.0223826848	ΜΑΡΟΝΙΑ-ΕΓΧΩΡΙΑ-ΤΤΕΜ-ΦΡΕΣΚΑ...	ΖΑΧΑΡΗ-1ΚΩ...		ΜΑΡΟΝΙΑ-ΕΓΧΩΡΙΑ-ΤΤΕΜ-ΦΡΕΣΚΑ... >>> ΖΑΧΑΡΗ-1ΚΩ...
23	359	0.1326419753	11.229277448	1.0274327637	ΜΙΝΕΡ-ΟΛΙΑΝΘΙΑΣ-ΦΑΤ20ΔΑ-ΤΚΙΟΥ...	ΖΑΧΑΡΗ-1ΚΩ...		ΜΙΝΕΡ-ΟΛΙΑΝΘΙΑΣ-ΦΑΤ20ΔΑ-ΤΚΙΟΥ... >>> ΖΑΧΑΡΗ-1ΚΩ...

Εικόνα 48. Κανόνες συσχέτισης μεταξύ αγορασμένων προϊόντων πελατών High Value σε αύξουσα σειρά lift

Από τα αποτελέσματα παραπάνω, παρατηρείται ότι πράγματι δεν αρκεί η συχνότητα που εμφανίζεται σε ένα καλάθι ένα σύνολο προϊόντων. Για παράδειγμα, ο «κατώτατος» κανόνας συσχέτισης μεταξύ TWIN ΦΙΑΛΙΔΙΟ και ΔΕΛΤΑ ΓΑΛΑ ΠΑΣΤΕΡΙΩΜΕΝΟ, εμφανίζεται τον ίδιο αριθμό φορές στο ίδιο καλάθι σαν σύνολο, με τον «ανώτατο» κανόνα συσχέτισης που αφορά σύνολο με τις πιπεριές που αναλύθηκε προηγουμένως. Όμως τα δύο σύνολα έχουν τεράστια διαφορά lift.

Η επεξήγηση είναι απλή: Στην ανάλυση καθαθίου αγορών, το "lift" μετράει πόσο συχνά δύο προϊόντα αγοράζονται μαζί, σε σχέση με το πόσο συχνά θα αγοράζονταν ανεξάρτητα το ένα από το άλλο. Δύο

κανόνες με το ίδιο count (στην περίπτωση μας 153), μπορούν να έχουν διαφορετικό lift, ανάλογα με τη συχνότητα πώλησης των μεμονωμένων προϊόντων, που συμμετέχουν στον κάθε κανόνα συσχέτισης. Άρα, ένα υψηλό lift σημαίνει ότι τα προϊόντα αγοράζονται πολύ συχνότερα μαζί απ' ό,τι θα περιμέναμε με βάση την ανεξάρτητη συχνότητα πώλησης. Αντίθετα, ένα lift κοντά στο 1 σημαίνει ότι τα προϊόντα αγοράζονται μαζί όσο συχνά θα περιμέναμε να αγοραστούν, αν αγοράζονταν ανεξάρτητα.

Τα αντίστοιχα αποτελέσματα κανόνων συσχέτισεων, για τη συστάδα πελατών Potential Loylists, φαίνονται παρακάτω.

	⊙ COUNT	⊙ SUPPORT	⊙ CONF	⊙ LIFT ↓	⊙ ITEM1	⊙ ITEM2	⊙ ITEM3	⊙ RULE
1	22	0.060461572	40.74074741	422.26031746	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....
2	22	0.060461572	62.85714857	422.26031746	ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....
3	22	0.060461572	44.897959184	397.24838228	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....
4	22	0.060461572	53.98538585	397.24838228	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....
5	27	0.0744293748	50	370.14326531	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....
6	27	0.0744293748	55.10204816	370.14326531	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....
7	22	0.060461572	81.48148181	351.88359788	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....
8	22	0.060461572	26.19047619	351.88359788	ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....
9	41	0.113022889	75.9292926	327.89153439	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....
10	41	0.113022889	48.8052381	327.89153439	ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....
11	35	0.094825229	71.428571429	308.44938776	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....
12	35	0.094825229	41.66666667	308.44938776	ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....
13	33	0.090492359	55.93220339	277.9447112	ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΜΠ.ΑΔΠ.ΣΤΑΦ.ΜΗΛΟ.....	ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΦΡ.ΚΟΚ.ΣΤΑΦ.ΜΗΛΟ.....	ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΜΠ.ΑΔΠ.ΣΤΑΦ.ΜΗΛΟ.....	ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΜΠ.ΑΔΠ.ΣΤΑΦ.ΜΗΛΟ..... ↔ ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΦΡ.ΚΟΚ.ΣΤΑΦ.ΜΗΛΟ..... ↔ ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΜΠ.ΑΔΠ.ΣΤΑΦ.ΜΗΛΟ.....
14	33	0.090492359	45.20547452	277.9447112	ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΦΡ.ΚΟΚ.ΣΤΑΦ.ΜΗΛΟ.....	ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΜΠ.ΑΔΠ.ΣΤΑΦ.ΜΗΛΟ.....	ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΜΠ.ΑΔΠ.ΣΤΑΦ.ΜΗΛΟ.....	ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΦΡ.ΚΟΚ.ΣΤΑΦ.ΜΗΛΟ..... ↔ ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΜΠ.ΑΔΠ.ΣΤΑΦ.ΜΗΛΟ..... ↔ ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΜΠ.ΑΔΠ.ΣΤΑΦ.ΜΗΛΟ.....
15	27	0.0744293748	55.10204816	277.62244898	ΤΕΑΜΠΙΟΝ.ΚΡΟΝΙΔΕΑΝ.75ΓΡ.ΚΕΡΑΣ.....	ΤΕΑΜΠΙΟΝ.ΚΡΟΝΙΔΕΑΝ.75ΓΡ.ΚΕΡΑΣ.....	ΤΕΑΜΠΙΟΝ.ΚΡΟΝΙΔΕΑΝ.75ΓΡ.ΚΕΡΑΣ.....	ΤΕΑΜΠΙΟΝ.ΚΡΟΝΙΔΕΑΝ.75ΓΡ.ΚΕΡΑΣ..... ↔ ΤΕΑΜΠΙΟΝ.ΚΡΟΝΙΔΕΑΝ.75ΓΡ.ΚΕΡΑΣ..... ↔ ΤΕΑΜΠΙΟΝ.ΚΡΟΝΙΔΕΑΝ.75ΓΡ.ΚΕΡΑΣ.....
16	27	0.0744293748	37.5	277.62244898	ΤΕΑΜΠΙΟΝ.ΚΡΟΝΙΔΕΑΝ.75ΓΡ.ΚΕΡΑΣ.....	ΤΕΑΜΠΙΟΝ.ΚΡΟΝΙΔΕΑΝ.75ΓΡ.ΚΕΡΑΣ.....	ΤΕΑΜΠΙΟΝ.ΚΡΟΝΙΔΕΑΝ.75ΓΡ.ΚΕΡΑΣ.....	ΤΕΑΜΠΙΟΝ.ΚΡΟΝΙΔΕΑΝ.75ΓΡ.ΚΕΡΑΣ..... ↔ ΤΕΑΜΠΙΟΝ.ΚΡΟΝΙΔΕΑΝ.75ΓΡ.ΚΕΡΑΣ..... ↔ ΤΕΑΜΠΙΟΝ.ΚΡΟΝΙΔΕΑΝ.75ΓΡ.ΚΕΡΑΣ.....
17	22	0.060461572	41.509433962	268.89218229	ΑΡΤΕΝ.ΚΟΥΡΑΜΠΙΔΕΣ.ΧΥΜΑ.....	ΑΡΤΕΝ.ΜΕΛΟΜΑΚΑΡΟΝΑ.ΧΥΜΑ.....	ΑΡΤΕΝ.ΚΟΥΡΑΜΠΙΔΕΣ.ΧΥΜΑ.....	ΑΡΤΕΝ.ΚΟΥΡΑΜΠΙΔΕΣ.ΧΥΜΑ..... ↔ ΑΡΤΕΝ.ΜΕΛΟΜΑΚΑΡΟΝΑ.ΧΥΜΑ..... ↔ ΑΡΤΕΝ.ΚΟΥΡΑΜΠΙΔΕΣ.ΧΥΜΑ.....
18	22	0.060461572	39.285714286	268.89218229	ΑΡΤΕΝ.ΜΕΛΟΜΑΚΑΡΟΝΑ.ΧΥΜΑ.....	ΑΡΤΕΝ.ΚΟΥΡΑΜΠΙΔΕΣ.ΧΥΜΑ.....	ΑΡΤΕΝ.ΚΟΥΡΑΜΠΙΔΕΣ.ΧΥΜΑ.....	ΑΡΤΕΝ.ΜΕΛΟΜΑΚΑΡΟΝΑ.ΧΥΜΑ..... ↔ ΑΡΤΕΝ.ΚΟΥΡΑΜΠΙΔΕΣ.ΧΥΜΑ..... ↔ ΑΡΤΕΝ.ΜΕΛΟΜΑΚΑΡΟΝΑ.ΧΥΜΑ.....
19	28	0.0771860183	50	245.10810811	ΑΛΛ.ΝΑΚ.40.ΓΡ.ΩΔΑΦ.....	ΑΛΛ.2001.ΑΡΜΥΡΑ.40.ΓΡ.ΩΔΑΦ.....	ΑΛΛ.ΝΑΚ.40.ΓΡ.ΩΔΑΦ.....	ΑΛΛ.ΝΑΚ.40.ΓΡ.ΩΔΑΦ..... ↔ ΑΛΛ.2001.ΑΡΜΥΡΑ.40.ΓΡ.ΩΔΑΦ..... ↔ ΑΛΛ.ΝΑΚ.40.ΓΡ.ΩΔΑΦ.....
20	28	0.0771860183	83.87837838	245.10810811	ΑΛΛ.2001.ΑΡΜΥΡΑ.40.ΓΡ.ΩΔΑΦ.....	ΑΛΛ.ΝΑΚ.40.ΓΡ.ΩΔΑΦ.....	ΑΛΛ.ΝΑΚ.40.ΓΡ.ΩΔΑΦ.....	ΑΛΛ.2001.ΑΡΜΥΡΑ.40.ΓΡ.ΩΔΑΦ..... ↔ ΑΛΛ.ΝΑΚ.40.ΓΡ.ΩΔΑΦ..... ↔ ΑΛΛ.2001.ΑΡΜΥΡΑ.40.ΓΡ.ΩΔΑΦ.....
21	27	0.0744293748	38.571428571	237.15544794	ΑΜΙΤΑ.ΦΑΝ.ΜΠ.Ε.200ΜΛ.....	ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΜΠ.ΑΔΠ.ΣΤΑΦ.ΜΗΛΟ.....	ΑΜΙΤΑ.ΦΑΝ.ΜΠ.Ε.200ΜΛ.....	ΑΜΙΤΑ.ΦΑΝ.ΜΠ.Ε.200ΜΛ..... ↔ ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΜΠ.ΑΔΠ.ΣΤΑΦ.ΜΗΛΟ..... ↔ ΑΜΙΤΑ.ΦΑΝ.ΜΠ.Ε.200ΜΛ.....
22	27	0.0744293748	45.76271884	237.15544794	ΑΜΙΤΑ.ΦΑΝ.ΜΠ.Ε.200ΜΛ.....	ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΜΠ.ΑΔΠ.ΣΤΑΦ.ΜΗΛΟ.....	ΑΜΙΤΑ.ΦΑΝ.ΜΠ.Ε.200ΜΛ.....	ΑΜΙΤΑ.ΦΑΝ.ΜΠ.Ε.200ΜΛ..... ↔ ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΜΠ.ΑΔΠ.ΣΤΑΦ.ΜΗΛΟ..... ↔ ΑΜΙΤΑ.ΦΑΝ.ΜΠ.Ε.200ΜΛ.....
23	19	0.052376267	27.536231884	212.5283996	ΔΕΛΤΑ.ΑΔVANCE.ΜΗΛΟ.ΜΠΑΝ.150GRX210Π/ΠΕΡΙ.....	ΔΕΛΤΑ.ΑΔVANCE.ΒΕΡΥΚΟΚΟ.150GRX210Π/ΠΕΡΙ.....	ΔΕΛΤΑ.ΑΔVANCE.ΜΗΛΟ.ΜΠΑΝ.150GRX210Π/ΠΕΡΙ.....	ΔΕΛΤΑ.ΑΔVANCE.ΜΗΛΟ.ΜΠΑΝ.150GRX210Π/ΠΕΡΙ..... ↔ ΔΕΛΤΑ.ΑΔVANCE.ΒΕΡΥΚΟΚΟ.150GRX210Π/ΠΕΡΙ..... ↔ ΔΕΛΤΑ.ΑΔVANCE.ΜΗΛΟ.ΜΠΑΝ.150GRX210Π/ΠΕΡΙ.....

Εικόνα 49. Κανόνες συσχέτισης μεταξύ αγορασμένων προϊόντων πελατών Potential Loylists σε αύξουσα σειρά lift

Τα κορυφαία αποτελέσματα κανόνων συσχέτισης των Potential Loylists μοιάζουν με εκείνα των High Value, αλλά διαφέρουν ελάχιστα σε μερικά σημεία. Για παράδειγμα, για την ομάδα αυτή, ένας πελάτης που έχει αγοράσει το προϊόν «CHAMPION κρουασάν κεράσι», είναι 277,62 φορές πιο πιθανό να αγοράσει το προϊόν «CHAMPION κρουασάν βερύκοκο», από κάποιον άλλο πελάτη που δεν έχει αγοράσει το προϊόν αυτό.

Οι κανόνες συσχέτισης για τη συστάδα Churners, φαίνεται παρακάτω.

	⊙ COUNT	⊙ SUPPORT	⊙ CONF	⊙ LIFT ↓	⊙ ITEM1	⊙ ITEM2	⊙ ITEM3	⊙ ITEM4	⊙ RULE
1	23	0.05948819	39.455172414	362.26321839	ΔΕΛΤΑ.ΑΔVANCE.ΕΛΙΟ.ΜΠΑΝ.150GRX210Π/ΠΕΡΙ.....	ΔΕΛΤΑ.ΑΔVANCE.ΒΕΡΥΚΟΚΟ.150GRX210Π/ΠΕΡΙ.....	ΔΕΛΤΑ.ΑΔVANCE.ΕΛΙΟ.ΜΠΑΝ.150GRX210Π/ΠΕΡΙ.....	ΔΕΛΤΑ.ΑΔVANCE.ΕΛΙΟ.ΜΠΑΝ.150GRX210Π/ΠΕΡΙ.....	ΔΕΛΤΑ.ΑΔVANCE.ΕΛΙΟ.ΜΠΑΝ.150GRX210Π/ΠΕΡΙ..... ↔ ΔΕΛΤΑ.ΑΔVANCE.ΒΕΡΥΚΟΚΟ.150GRX210Π/ΠΕΡΙ..... ↔ ΔΕΛΤΑ.ΑΔVANCE.ΕΛΙΟ.ΜΠΑΝ.150GRX210Π/ΠΕΡΙ.....
2	23	0.05948819	51.11111111	362.26321839	ΔΕΛΤΑ.ΑΔVANCE.ΒΕΡΥΚΟΚΟ.150GRX210Π/ΠΕΡΙ.....	ΔΕΛΤΑ.ΑΔVANCE.ΕΛΙΟ.ΜΠΑΝ.150GRX210Π/ΠΕΡΙ.....	ΔΕΛΤΑ.ΑΔVANCE.ΕΛΙΟ.ΜΠΑΝ.150GRX210Π/ΠΕΡΙ.....	ΔΕΛΤΑ.ΑΔVANCE.ΒΕΡΥΚΟΚΟ.150GRX210Π/ΠΕΡΙ.....	ΔΕΛΤΑ.ΑΔVANCE.ΒΕΡΥΚΟΚΟ.150GRX210Π/ΠΕΡΙ..... ↔ ΔΕΛΤΑ.ΑΔVANCE.ΕΛΙΟ.ΜΠΑΝ.150GRX210Π/ΠΕΡΙ..... ↔ ΔΕΛΤΑ.ΑΔVANCE.ΒΕΡΥΚΟΚΟ.150GRX210Π/ΠΕΡΙ.....
3	35	0.0851395072	67.307692008	294.35658976	ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....
4	35	0.0851395072	37.34042553	294.35658976	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....
5	35	0.0851395072	36.842032623	256.70205174	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....
6	35	0.0851395072	59.32203898	256.70205174	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....
7	55	0.1337906541	57.894738842	253.19092945	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....
8	55	0.1337906541	58.510638299	253.19092945	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΠΟΡΤΟΚΑΛΙ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....
9	25	0.0608139337	29.761904762	235.28502747	ΔΕΛΤΑ.150GRX2.ΑΔVANCE.ΜΗΛΟ.ΜΠΑΝ.20Π/ΠΕΡΙ.....	ΔΕΛΤΑ.150GRX2.ΑΔVANCE.ΒΕΡΥΚΟΚΟ.150GRX210Π/ΠΕΡΙ.....	ΔΕΛΤΑ.150GRX2.ΑΔVANCE.ΜΗΛΟ.ΜΠΑΝ.20Π/ΠΕΡΙ.....	ΔΕΛΤΑ.150GRX2.ΑΔVANCE.ΜΗΛΟ.ΜΠΑΝ.20Π/ΠΕΡΙ.....	ΔΕΛΤΑ.150GRX2.ΑΔVANCE.ΜΗΛΟ.ΜΠΑΝ.20Π/ΠΕΡΙ..... ↔ ΔΕΛΤΑ.150GRX2.ΑΔVANCE.ΒΕΡΥΚΟΚΟ.150GRX210Π/ΠΕΡΙ..... ↔ ΔΕΛΤΑ.150GRX2.ΑΔVANCE.ΜΗΛΟ.ΜΠΑΝ.20Π/ΠΕΡΙ.....
10	25	0.0608139337	48.07692077	235.28502747	ΔΕΛΤΑ.150GRX2.ΑΔVANCE.ΒΕΡΥΚΟΚΟ.150GRX210Π/ΠΕΡΙ.....	ΔΕΛΤΑ.150GRX2.ΑΔVANCE.ΜΗΛΟ.ΜΠΑΝ.20Π/ΠΕΡΙ.....	ΔΕΛΤΑ.150GRX2.ΑΔVANCE.ΜΗΛΟ.ΜΠΑΝ.20Π/ΠΕΡΙ.....	ΔΕΛΤΑ.150GRX2.ΑΔVANCE.ΒΕΡΥΚΟΚΟ.150GRX210Π/ΠΕΡΙ.....	ΔΕΛΤΑ.150GRX2.ΑΔVANCE.ΒΕΡΥΚΟΚΟ.150GRX210Π/ΠΕΡΙ..... ↔ ΔΕΛΤΑ.150GRX2.ΑΔVANCE.ΜΗΛΟ.ΜΠΑΝ.20Π/ΠΕΡΙ..... ↔ ΔΕΛΤΑ.150GRX2.ΑΔVANCE.ΜΗΛΟ.ΜΠΑΝ.20Π/ΠΕΡΙ.....
11	31	0.0754092778	30.392158683	231.36873638	ΓΟΤΗΣ.ΣΤΥΛΗΣ.ΦΑΚΕΛΟ.ΒΑΝΙΛΙΟ.....	ΓΟΤΗΣ.ΑΝΘΟΣ.ΑΡΑΒ.ΣΤΥΛΗΣ.ΚΑΚΑΟ.ΦΑΚΕΛΟ.....	ΓΟΤΗΣ.ΣΤΥΛΗΣ.ΦΑΚΕΛΟ.ΒΑΝΙΛΙΟ.....	ΓΟΤΗΣ.ΣΤΥΛΗΣ.ΦΑΚΕΛΟ.ΒΑΝΙΛΙΟ.....	ΓΟΤΗΣ.ΣΤΥΛΗΣ.ΦΑΚΕΛΟ.ΒΑΝΙΛΙΟ..... ↔ ΓΟΤΗΣ.ΑΝΘΟΣ.ΑΡΑΒ.ΣΤΥΛΗΣ.ΚΑΚΑΟ.ΦΑΚΕΛΟ..... ↔ ΓΟΤΗΣ.ΣΤΥΛΗΣ.ΦΑΚΕΛΟ.ΒΑΝΙΛΙΟ.....
12	31	0.0754092778	57.40740707	231.36873638	ΓΟΤΗΣ.ΑΝΘΟΣ.ΑΡΑΒ.ΣΤΥΛΗΣ.ΚΑΚΑΟ.ΦΑΚΕΛΟ.....	ΓΟΤΗΣ.ΣΤΥΛΗΣ.ΦΑΚΕΛΟ.ΒΑΝΙΛΙΟ.....	ΓΟΤΗΣ.ΣΤΥΛΗΣ.ΦΑΚΕΛΟ.ΒΑΝΙΛΙΟ.....	ΓΟΤΗΣ.ΑΝΘΟΣ.ΑΡΑΒ.ΣΤΥΛΗΣ.ΚΑΚΑΟ.ΦΑΚΕΛΟ.....	ΓΟΤΗΣ.ΑΝΘΟΣ.ΑΡΑΒ.ΣΤΥΛΗΣ.ΚΑΚΑΟ.ΦΑΚΕΛΟ..... ↔ ΓΟΤΗΣ.ΣΤΥΛΗΣ.ΦΑΚΕΛΟ.ΒΑΝΙΛΙΟ..... ↔ ΓΟΤΗΣ.ΑΝΘΟΣ.ΑΡΑΒ.ΣΤΥΛΗΣ.ΚΑΚΑΟ.ΦΑΚΕΛΟ.....
13	45	0.109465806	36.585385854	208.88719512	ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΜΠ.ΑΔΠ.ΣΤΑΦ.ΜΗΛΟ.....	ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΦΡ.ΚΟΚ.ΣΤΑΦ.ΜΗΛΟ.....	ΑΜΙΤΑ.ΦΑΝ.ΜΠ.Ε.200ΜΛ.....	ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΜΠ.ΑΔΠ.ΣΤΑΦ.ΜΗΛΟ.....	ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΜΠ.ΑΔΠ.ΣΤΑΦ.ΜΗΛΟ..... ↔ ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΦΡ.ΚΟΚ.ΣΤΑΦ.ΜΗΛΟ..... ↔ ΑΜΙΤΑ.ΦΑΝ.ΜΠ.Ε.200ΜΛ.....
14	45	0.109465806	62.5	208.88719512	ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΦΡ.ΚΟΚ.ΣΤΑΦ.ΜΗΛΟ.....	ΑΜΙΤΑ.ΦΑΝ.ΜΠ.Ε.200ΜΛ.....	ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΜΠ.ΑΔΠ.ΣΤΑΦ.ΜΗΛΟ.....	ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΦΡ.ΚΟΚ.ΣΤΑΦ.ΜΗΛΟ.....	ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΦΡ.ΚΟΚ.ΣΤΑΦ.ΜΗΛΟ..... ↔ ΑΜΙΤΑ.ΦΑΝ.ΜΠ.Ε.200ΜΛ..... ↔ ΑΜΙΤΑ.ΦΑΝ.200ΜΛ.ΜΠ.ΑΔΠ.ΣΤΑΦ.ΜΗΛΟ.....
15	24	0.0583813763	45.283018848	198.03613007	ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΦΙΛΑΧΕΣ.ΕΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....	ΠΙΡΡΕΣ.ΚΟΚΚΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΦΙΛΑΧΕΣ.ΕΓΩΓΗΣ.ΦΡΕΣΚΕΣ..... ↔ ΠΙΡΡΕΣ.ΚΙΤΡΙΝΕΣ.ΕΣΑΓΩΓΗΣ.ΦΡΕΣΚΕΣ.....
16	24	0.0583813763	25.						



STRUDEL ΦΡ. ΔΑΣΟΥΣ», είναι 197,52 φορές πιο πιθανό να αγοράσει το προϊόν «ΣΕΒΕΝ ΝΤΕΙΣ STRUDEL ΜΗΛΟ-KΑΝΕΛΑ», από κάποιον άλλο πελάτη που δεν έχει αγοράσει το προϊόν αυτό.

Τέλος, οι κορυφαίοι κανόνες συσχέτισης της συστάδας πελατών «Low Value», φαίνονται στον πίνακα παρακάτω.

	◎ COUNT	◎ SUPPORT	◎ CONF	◎ LIFT 1	◎ ITEM1	◎ ITEM2	◎ ITEM3	◎ RULE
1	8	0.0765643265	88.88888889	773.7777778	ΚΟΥΒΑΣ ΕΣΟΥΓΓΑΡΙΣΜΑΤΟΣ ΕΤΡΟΥΤ	ΣΤΡΗΦΗ ΚΟΥΒΑΣ ΕΣΟΥΓΓΑΡΙΣΜΑΤΟΣ		ΚΟΥΒΑΣ ΕΣΟΥΓΓΑΡΙΣΜΑΤΟΣ ΕΤΡΟΥΤ <=> ΣΤΡΗΦΗ ΚΟΥΒΑΣ ΕΣΟΥΓΓΑΡΙΣΜΑΤΟΣ
2	8	0.0765643265	88.88888889	773.7777778	ΣΤΡΗΦΗ ΚΟΥΒΑΣ ΕΣΟΥΓΓΑΡΙΣΜΑΤΟΣ	ΚΟΥΒΑΣ ΕΣΟΥΓΓΑΡΙΣΜΑΤΟΣ ΕΤΡΟΥΤ		ΣΤΡΗΦΗ ΚΟΥΒΑΣ ΕΣΟΥΓΓΑΡΙΣΜΑΤΟΣ ΕΤΡΟΥΤ <=> ΚΟΥΒΑΣ ΕΣΟΥΓΓΑΡΙΣΜΑΤΟΣ ΕΤΡΟΥΤ
3	6	0.0574326539	66.15384615	468.74725275	ΗΜΗ-ΣΟ-ΥΕΛΛΩ-ΣΣΜΛ-ΣΣΤ	ΗΜΗ-ΣΟ-ΥΕΛΛΩ-ΣΣΜΛ-ΣΣΤ		ΗΜΗ-ΣΟ-ΥΕΛΛΩ-ΣΣΜΛ-ΣΣΤ <=> ΗΜΗ-ΣΟ-ΥΕΛΛΩ-ΣΣΜΛ-ΣΣΤ
4	6	0.0574326539	66.15384615	468.74725275	ΗΜΗ-ΣΟ-ΥΕΛΛΩ-ΣΣΜΛ-ΣΣΤ	ΗΜΗ-ΣΟ-ΥΕΛΛΩ-ΣΣΜΛ-ΣΣΤ		ΗΜΗ-ΣΟ-ΥΕΛΛΩ-ΣΣΜΛ-ΣΣΤ <=> ΗΜΗ-ΣΟ-ΥΕΛΛΩ-ΣΣΜΛ-ΣΣΤ
5	7	0.0670112942	38.88888889	406.23333333	ΠΙΠΕΡΕΣ ΑΤΡΗΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ ΠΟΡΤΟΚΑΛΙ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ	ΠΙΠΕΡΕΣ ΑΤΡΗΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ ΠΟΡΤΟΚΑΛΙ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ
6	7	0.0670112942	38.88888889	406.23333333	ΠΙΠΕΡΕΣ ΠΟΡΤΟΚΑΛΙ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ ΦΙΛΑΚΙΣ ΕΓΧΩΡΙΕΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ	ΠΙΠΕΡΕΣ ΠΟΡΤΟΚΑΛΙ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ ΦΙΛΑΚΙΣ ΕΓΧΩΡΙΕΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ
7	6	0.0574326539	66.15384615	401.74923077	ΤΩΜΥΣ ΕΣΟΡ-ΧΟΧΟ-ΚΡΟΝΙΣΙΑΝ ΒΑΝΗΛΙΑ-ΣΣΤ	ΤΩΜΥΣ ΕΣΟΡ-ΧΟΧΟ-ΚΡΟΝΙΣΙΑΝ ΚΑΚΙΑΟ-ΣΣΤ		ΤΩΜΥΣ ΕΣΟΡ-ΧΟΧΟ-ΚΡΟΝΙΣΙΑΝ ΒΑΝΗΛΙΑ-ΣΣΤ <=> ΤΩΜΥΣ ΕΣΟΡ-ΧΟΧΟ-ΚΡΟΝΙΣΙΑΝ ΚΑΚΙΑΟ-ΣΣΤ
8	6	0.0574326539	66.15384615	401.74923077	ΤΩΜΥΣ ΕΣΟΡ-ΧΟΧΟ-ΚΡΟΝΙΣΙΑΝ ΚΑΚΙΑΟ-ΣΣΤ	ΤΩΜΥΣ ΕΣΟΡ-ΧΟΧΟ-ΚΡΟΝΙΣΙΑΝ ΒΑΝΗΛΙΑ-ΣΣΤ		ΤΩΜΥΣ ΕΣΟΡ-ΧΟΧΟ-ΚΡΟΝΙΣΙΑΝ ΚΑΚΙΑΟ-ΣΣΤ <=> ΤΩΜΥΣ ΕΣΟΡ-ΧΟΧΟ-ΚΡΟΝΙΣΙΑΝ ΒΑΝΗΛΙΑ-ΣΣΤ
9	14	0.1340225924	77.77777778	386.88888889	ΠΙΠΕΡΕΣ ΑΤΡΗΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ ΠΟΡΤΟΚΑΛΙ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ		ΠΙΠΕΡΕΣ ΑΤΡΗΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ ΠΟΡΤΟΚΑΛΙ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ
10	7	0.0670112942	38.88888889	386.88888889	ΠΙΠΕΡΕΣ ΑΤΡΗΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ ΦΙΛΑΚΙΣ ΕΓΧΩΡΙΕΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ	ΠΙΠΕΡΕΣ ΑΤΡΗΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ ΦΙΛΑΚΙΣ ΕΓΧΩΡΙΕΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ
11	14	0.1340225924	66.66666667	386.88888889	ΠΙΠΕΡΕΣ ΠΟΡΤΟΚΑΛΙ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ ΑΤΡΗΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ		ΠΙΠΕΡΕΣ ΠΟΡΤΟΚΑΛΙ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ ΑΤΡΗΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ
12	7	0.0670112942	38.88888889	386.88888889	ΠΙΠΕΡΕΣ ΠΟΡΤΟΚΑΛΙ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ ΑΤΡΗΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ	ΠΙΠΕΡΕΣ ΠΟΡΤΟΚΑΛΙ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ ΑΤΡΗΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ
13	9	0.0861573808	64.285714286	373.07142857	ΠΙΠΕΡΕΣ ΚΟΚΚΙΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ ΠΟΡΤΟΚΑΛΙ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ	ΠΙΠΕΡΕΣ ΚΟΚΚΙΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ ΠΟΡΤΟΚΑΛΙ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ
14	9	0.0861573808	64.285714286	373.07142857	ΠΙΠΕΡΕΣ ΑΤΡΗΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ ΚΟΚΚΙΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ	ΠΙΠΕΡΕΣ ΑΤΡΗΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ ΚΟΚΚΙΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ
15	9	0.0861573808	42.85714286	344.37502637	ΠΙΠΕΡΕΣ ΠΟΡΤΟΚΑΛΙ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ ΑΤΡΗΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ	ΠΙΠΕΡΕΣ ΠΟΡΤΟΚΑΛΙ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ ΑΤΡΗΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ
16	9	0.0861573808	49.230749231	344.37502637	ΠΙΠΕΡΕΣ ΑΤΡΗΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ ΚΟΚΚΙΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ	ΠΙΠΕΡΕΣ ΑΤΡΗΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ ΚΟΚΚΙΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ
17	6	0.0574326539	54.54545455	316.54545455	ΠΙΠΕΡΕΣ ΚΟΚΚΙΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ ΦΙΛΑΚΙΣ ΕΓΧΩΡΙΕΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ	ΠΙΠΕΡΕΣ ΚΟΚΚΙΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ ΦΙΛΑΚΙΣ ΕΓΧΩΡΙΕΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ
18	6	0.0574326539	33.33333333	316.54545455	ΠΙΠΕΡΕΣ ΑΤΡΗΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ ΚΟΚΚΙΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ	ΠΙΠΕΡΕΣ ΑΤΡΗΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ ΚΟΚΚΙΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ
19	7	0.0670112942	63.63636364	316.54545455	ΠΙΠΕΡΕΣ ΚΟΚΚΙΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ ΦΙΛΑΚΙΣ ΕΓΧΩΡΙΕΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ	ΠΙΠΕΡΕΣ ΚΟΚΚΙΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ ΦΙΛΑΚΙΣ ΕΓΧΩΡΙΕΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ
20	7	0.0670112942	33.33333333	316.54545455	ΠΙΠΕΡΕΣ ΠΟΡΤΟΚΑΛΙ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ ΚΟΚΚΙΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ	ΠΙΠΕΡΕΣ ΠΟΡΤΟΚΑΛΙ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ ΚΟΚΚΙΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ
21	13	0.1344495501	72.22222222	279.41973309	ΠΙΠΕΡΕΣ ΑΤΡΗΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ ΚΟΚΚΙΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ		ΠΙΠΕΡΕΣ ΑΤΡΗΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ ΚΟΚΚΙΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ
22	13	0.1344495501	48.14814815	279.41973309	ΠΙΠΕΡΕΣ ΚΟΚΚΙΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ	ΠΙΠΕΡΕΣ ΑΤΡΗΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ		ΠΙΠΕΡΕΣ ΚΟΚΚΙΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ <=> ΠΙΠΕΡΕΣ ΑΤΡΗΝΕΣ ΕΣΑΓΩΓΗΣ ΦΡΕΣΚΕΣ
23	6	0.0574326539	40	278.56	ΑΜΥΤΑ-ΦΑΝ-ΣΟΜΜΑ-ΑΡΓΑΣΤΙ-ΣΤΑΦ-ΜΗΛΙ-ΑΛΥ	ΑΜΥΤΑ-ΦΑΝ-ΣΟΜΜΑ		ΑΜΥΤΑ-ΦΑΝ-ΣΟΜΜΑ-ΑΡΓΑΣΤΙ-ΣΤΑΦ-ΜΗΛΙ-ΑΛΥ <=> ΑΜΥΤΑ-ΦΑΝ-ΣΟΜΜΑ

Εικόνα 51. Κανόνες συσχέτισης μεταξύ αγορασμένων προϊόντων πελατών Low Value σε αύξουσα σειρά lift

Οι κανόνες συσχέτισης των Low Value, εμφανίζονται υπερβολικά λίγες φορές στα δεδομένα, όπως είναι αναμενόμενο. Ωστόσο, κάποιον από αυτούς έχουν υπερβολικά μεγάλα lift. Ενδεικτικά, ο κορυφαίος κανόνας συσχέτισης, μεταξύ των προϊόντων «κουβάς σφουγγαρίσματος» και «στριφτήρι κουβά σφουγγαρίσματος», έχει lift 773,77. Αυτό σημαίνει ότι ένας πελάτης της συστάδας που έχει αγοράσει ένα από τα δύο προϊόντα, είναι 773,77 φορές πιο πιθανό να αγοράσει το άλλο προϊόν.

Με βάση τα παραπάνω ευρήματα, προτείνεται να χρησιμοποιηθούν οι κανόνες συσχέτισης που δημιουργήθηκαν, προκειμένου να υλοποιηθούν προωθητικές ενέργειες σχεδιασμένες ειδικά για κάθε ένα από τα εκάστοτε εστιαζόμενα τμήματα πελατών. Για όσους πελάτες του τμήματος, έχουν ήδη αγοράσει σωρευτικά τα προϊόντα που εμφανίζονται στο αριστερό μέρος (LHS), η προώθηση του προϊόντος που εμφανίζεται στο δεξί μέρος (RHS), είναι πολύ περισσότερο πιθανό να αποδώσει και να πείσει τον καταναλωτή να αυξήσει τη χρηματική αξία του καλαθιού του, αγοράζοντάς το. Ως αποτέλεσμα θα αυξηθούν τα έσοδα, ενώ παράλληλα θα ενδυναμωθούν οι σχέσεις των πελατών με την εταιρεία. Τέλος, με αυτόν τον τρόπο, τα προϊόντα που για οποιοδήποτε λόγο δεν αγοράστηκαν, θα πουληθούν και δεν θα παραμένουν σε μεγάλο απόθεμα, ελευθερώνοντας χώρο στην αποθήκη και μειώνοντας τα κόστη συντήρησης.

## ΣΥΜΠΕΡΑΣΜΑΤΑ

Μέσα στον ωκεανό δεδομένων που οι επιχειρήσεις καλούνται να πλοηγηθούν, η χρήση των μεγάλων δεδομένων είναι πιο επιτακτική από ποτέ. Το πεδίο των μεγάλων δεδομένων έχει εξελιχθεί σε έναν κρίσιμο πυλώνα στον τομέα του Business Intelligence, διαδραματίζοντας πλέον ένα καθοριστικό ρόλο στη σύγχρονη επιχειρηματική σκέψη και στρατηγική. Επιχειρήσεις που αξιοποιούν έξυπνα την ανάλυση των μεγάλων δεδομένων, όχι μόνο αποκτούν μια βαθύτερη κατανόηση των πελατών τους, αλλά καταφέρνουν επίσης να αποκτήσουν μία πολύτιμη εικόνα για τις συναλλαγές και τη συμπεριφορά τους στην αγορά. Η επιχειρηματική αναλυτική οδηγεί αναπόφευκτα στην υιοθέτηση της τεχνητής νοημοσύνης για την διαρκή εξέλιξη των μεθοδολογιών της. Η παρούσα εργασία, μέσω της πρακτικής εφαρμογής καινοτόμων εργαλείων και μεθοδολογιών όπως το Visual Analytics, το μοντέλο RFM (Recency, Frequency, Monetary value) και το Market Basket Analysis, απέδειξε την αξία των τους στην επιβίωση και την ενίσχυση της ανταγωνιστικότητας για κάθε φιλόδοξη επιχείρηση.

Η ανάλυση που παρουσιάστηκε στην εργασία αυτή αποτελεί μόνο μια εισαγωγή σε μερικές από τις ατελείωτες δυνατότητες που προσφέρουν τα μεγάλα δεδομένα. Αξιοποιώντας τα ευρήματα της

επιχειρηματικής αναλυτικής, οι επιχειρήσεις retail (και όχι μόνο) μπορούν να επιλέξουν αποτελεσματικές στρατηγικές σχετικά με τα προϊόντα τους, όπως το ποια προϊόντα θα πρέπει να πακετάρουν μαζί, ποια προϊόντα θα πρέπει να τοποθετούνται σε διπλανά ράφια, ποια προϊόντα θα πρέπει να τοποθετούνται στο πίσω μέρος του καταστήματος ώστε να αναγκάζουν τους πελάτες να διασχίσουν συγκεκριμένες διαδρομές για να φτάσουν σε αυτά, ποια θα πρέπει να τοποθετούνται σε εύκολη πρόσβαση και ποια όχι. Οι επιχειρήσεις επίσης όμως μπορούν να επιλέξουν και στρατηγικές που αφορούν και άλλα στοιχεία της επιχειρηματικής τους δραστηριότητας, όπως το να επιλέγουν τον αριθμό που θα κατέχουν από το κάθε προϊόν στις αποθήκες τους και στα κέντρα διανομής τους, πότε πρέπει να ενισχύουν ή να μειώνουν το απασχολούμενο προσωπικό τους, ποιες περιοχές και ομάδες πελατών είναι ελκυστικοί για επέκταση επιρροής, ποιες στρατηγικές προώθησης είναι πιο αποτελεσματικές και για ποιους πελάτες τους και ούτω καθ' εξής. Η αξία των αποτελεσμάτων μίας ανάλυσης σταματάει μόνο εκεί που σταματά η φαντασία της κάθε επιχείρησης στο πως μπορεί να τα διαχειριστεί.

Η ικανότητα όμως ενός οργανισμού να εξάγει ποιοτικά συμπεράσματα από τα δεδομένα αυτά είναι άμεσα συνδεδεμένη με την ποιότητα των δεδομένων που συλλέγει. Συνεπώς, η σημασία της επιλογής και της χρήσης ποιοτικών δεδομένων, καθώς και της αξιοποίησης των προαναφερθέντων εργαλείων και μεθοδολογιών, είναι περισσότερο επίκαιρη από ποτέ. Για αυτό το λόγο, απαιτείται πολύ καλή προετοιμασία και εξερεύνηση των δεδομένων, πριν τη διενέργεια οποιασδήποτε ανάλυσης. Στο παράδειγμα που εξετάστηκε, φάνηκε ότι απαιτήθηκαν εκτενείς τέτοιες ενέργειες παρά το γεγονός ότι η βάση δεδομένων ήταν περιορισμένη χρονικά και ποσοτικά.

Τα στελέχη επιχειρήσεων που λαμβάνουν αποφάσεις δεν έχουν συνήθως ούτε την τεχνική γνώση ούτε το χρόνο για περίπλοκες αναλύσεις. Η οπτική ανάλυση αποδεικνύεται ως το ιδανικότερο εργαλείο για την αντιμετώπιση αυτού του προβλήματος, παρέχοντας μια ολοκληρωμένη εικόνα με γρήγορο και αξιόπιστο τρόπο. Ακόμα και στα χέρια των τεχνικά έμπειρων αναλυτών, η οπτική ανάλυση μπορεί να προσφέρει τις κατάλληλες υποθέσεις ή ερωτήματα που θέτουν τη βάση για περαιτέρω σε βάθος ανάλυση των δεδομένων. Στην περίπτωση που εξετάζεται, η οπτική ανάλυση συνέβαλε ακόμη και στην προετοιμασία των δεδομένων πριν από την ανάλυση. Το πιο σημαντικό στοιχείο της τεχνικής ανάλυσης είναι ότι διαφορετικές αναπαραστάσεις των ίδιων δεδομένων μπορούν να οδηγήσουν σε πολύ διαφορετικά, αλλά εξίσου χρήσιμα αποτελέσματα και συμπεράσματα.

Το μοντέλο RFM αξιοποιεί τρεις μεταβλητές, οι οποίες είναι εύκολα προσβάσιμες καθώς τα δεδομένα που χρησιμοποιούνται είναι διαθέσιμα σε κάθε εταιρεία. Επιπλέον, η λειτουργία του μοντέλου είναι απλή και κατανοητή ακόμα και για τα στελέχη που λαμβάνουν αποφάσεις μέσα στην εταιρεία. Παρά την απλότητά του, το μοντέλο RFM, όπως φάνηκε από την εφαρμογή του, παράγει εξαιρετικά αποτελέσματα για την ταξινόμηση των πελατών. Μέσω αυτής της ταξινόμησης, η εταιρεία μπορεί να αναγνωρίζει τις ομάδες στις οποίες ανήκουν οι πελάτες της, να εστιάζει στις κατάλληλες ομάδες πελατών και να καθορίζει τις στρατηγικές της βάσει αυτών.

Τέλος, η ανάλυση του καλαθιού αγορών κατάφερε να αποκαλύψει με εξαιρετικά κατανοητό τρόπο τις σχέσεις μεταξύ διαφορετικών προϊόντων στο καλάθι του ίδιου καταναλωτή. Η στατιστική επεξεργασία των μεμονωμένων συναλλαγών, με τις κατάλληλες μεταβλητές που εξυπηρετούν τους στόχους κάθε επιχείρησης, μπορεί να μεγιστοποιήσει τα κέρδη και ταυτόχρονα να ελαχιστοποιήσει τα κόστη. Οι εταιρείες μπορούν να προσαρμόσουν τη μεθοδολογία και να επιτύχουν ποικίλα και χρήσιμα αποτελέσματα, ανάλογα με το αν θέλουν να δώσουν έμφαση σε ισχυρούς κανόνες ή σε μεγαλύτερο αριθμό κανόνων.

Αναμφίβολα, οι επιχειρήσεις που εστιάζουν στη συλλογή και ανάλυση ποιοτικών μεγάλων δεδομένων, έχουν τη δυνατότητα να αποκτήσουν σημαντικό ανταγωνιστικό πλεονέκτημα. Η εξελιγμένη ανάλυση και ερμηνεία των δεδομένων αυτών, μπορεί να αποτελέσει τη βάση για επιχειρησιακές αποφάσεις, δημιουργώντας ένα πλαίσιο για μελλοντική εξέλιξη και καινοτομία. Είναι σαφές ότι οι μεθοδολογίες και τα εργαλεία που αναπτύχθηκαν στην παρούσα εργασία, δεν αποτελούν απλά τεχνικές εργασίες ανάλυσης δεδομένων, αλλά συνιστούν βασικές στρατηγικές, που μπορούν να υιοθετήσουν οι επιχειρήσεις, για να οδηγηθούν μελλοντικά προς μια πιο δυναμική και καινοτόμο πορεία.

Καθώς προχωράμε στο μέλλον, η εξέλιξη της τεχνολογίας μεγάλων δεδομένων αναμένεται να καταλάβει μια ακόμα πιο κεντρική θέση στον κόσμο του Business Intelligence. Η συνεχής ενσωμάτωση προηγμένων τεχνολογιών όπως η τεχνητή νοημοσύνη και η μηχανική μάθηση, ανοίγει νέους ορίζοντες για την ανάλυση και ερμηνεία των δεδομένων. Παράλληλα, αυτή η τεχνολογική πρόοδος φέρει μαζί της σημαντικές ηθικές και νομικές προκλήσεις, κυρίως όσον αφορά την ιδιωτικότητα και την ασφάλεια των δεδομένων. Οι επιχειρήσεις και οι ερευνητές πρέπει να προχωρούν με σύνεση, συνδυάζοντας την πρωτοποριακή τεχνολογική ανάπτυξη με μια ακλόνητη δέσμευση στην ηθική χρήση των δεδομένων.

## ΠΑΡΑΡΤΗΜΑ

Στο παρόν παράρτημα, βρίσκονται οι ακολουθίες κώδικα που χρησιμοποιήθηκαν και αναφέρονται εντός της εργασίας.

### Εισαγωγή και μετατροπή βάσης δεδομένων (3.2.1)

Εισαγωγή δεδομένων στο λογισμικό SAS.

```
Code
1 /* transformation files from csv to sas7bdat */
2 data thesis.customers;
3   set casuser.customers;
4 run;
5
6 data thesis.items;
7   set casuser.items;
8 run;
9
10 data thesis.prefectures;
11   set casuser.prefectures;
12 run;
13
14 data thesis.properties;
15   set casuser.properties;
16 run;
17
18 data thesis.stores;
19   set casuser.stores;
20 run;
21
22 data thesis.sysdiagrams;
23   set casuser.sysdiagrams;
24 run;
25
26 data thesis.transactionItems1;
27   set casuser.transactionItems1;
28 run;
29
30 data thesis.transactionItems2;
31   set casuser.transactionItems2;
32 run;
33
34 data thesis.transactionItems3;
35   set casuser.transactionItems3;
36 run;
37
38 data thesis.transactions;
39   set casuser.transactions;
40 run;
41
42 data thesis.transactionItems_total;
43   set thesis.transactionItems1 thesis.transactionItems2 thesis.transactionItems3;
44 run;
```

Εικόνα 52. Ακολουθία κώδικα για την μετατροπή αρχείων τύπου .csv της βάσης δεδομένων σε native αρχεία τύπου .sas7bdat

### Μορφοποίηση και καθαρισμός δεδομένων (3.2.2)

Οι απαραίτητες μετατροπές δεδομένων, έγιναν μέσω των παρακάτω ακολουθιών κώδικα.

Code

```

1  data thesis.transactions;
2      set thesis.transactions;
3      trans_date=input(transDateTime, ANYDTDTE19.);
4      trans_time=input(transDateTime, ANYDTTME19.);
5      format trans_date ddmmyy10. trans_time TIME8.;
6  run;
7
8  data thesis.customers;
9      set thesis.customers;
10     registr_date=input(registrationDate, ANYDTDTE19.);
11     format registr_date ddmmyy10.;
12 run;

```

Εικόνα 53. Μετατροπή χαρακτήρων στο πεδίο ημερομηνίας σε μορφή ημερομηνία-ώρα

Code

```

1  proc freq data=thesis.customers nlevels;
2      tables custAge custOccupation custGender custNumChildren custPrefecture custArea custPostalCode;
3  run;
4
5  proc freq data=thesis.customers nlevels;
6      tables custCode;
7  run;
8
9  proc freq data=thesis.customers nlevels;
10     tables custArea;
11 run;
12
13 data thesis.customers_clean1;
14     set thesis.customers;
15     if substr(custArea,1,1)='.' then do;
16         substr(custArea,1,1)='';
17         custArea=catt(custArea);
18     end;
19     if substr(custArea,1,1)='' then do;
20         substr(custArea,1,1)='';
21         custArea=catt(custArea);
22     end;
23 run;
24
25 proc freq data=thesis.customers_clean1 nlevels;
26     tables custArea;
27 run;
28
29 data thesis.customers_clean2;
30     set thesis.customers_clean1;
31     custArea=compress(custArea, '(),1234567890');
32 run;
33
34 proc freq data=thesis.customers_clean2 nlevels;
35     tables custArea;
36 run;
37
38 data thesis.customers_clean3;
39     set thesis.customers_clean2;
40     if substr(custArea||FLENGTH(custArea),1)='.' or substr(custArea||FLENGTH(custArea),1)='- ' then do;

```

Εικόνα 54. Καθαρισμός δεδομένων

```

Code
41     substr(custArea,LENGTH(custArea),1)='';
42     end;
43     run;
44
45     proc freq data=thesis.customers_clean3 nlevels;
46     tables custArea;
47     run;
48
49     data thesis.customers_clean4;
50     set thesis.customers_clean3;
51     if LENGTH(custArea)<=4 then
52     custArea='';
53     run;
54
55     proc freq data=thesis.customers_clean4 nlevels;
56     tables custArea;
57     run;
58
59     data thesis.customers_clean5;
60     set thesis.customers_clean4;
61     if substr(custArea,1,1)='/' then
62     custArea='';
63     else if substr(custArea,1,1)='.' then
64     custArea='';
65     drop '.';
66     run;
67
68     proc freq data=thesis.customers_clean5 nlevels;
69     tables custArea;
70     run;
71
72     data thesis.customers_clean6;
73     set thesis.customers_clean5;
74     if custArea in ('FFFFFF', 'ΑΓΝΟΣΤΗ', 'ΚΕ'Ε', 'ΗΗΗΗΗΗ', 'ΚΚΚΚΚΚΚΚ', 'ΚΑΚ'Α', 'ΜΜΜΜΜΜΜΜ', 'ΕΚΑ', 'ΕΜΕ', 'ΕΕΕΕΕΕΕΕΕΕ', 'ΕΕ
75     run;
76
77     proc freq data=thesis.customers_clean6 nlevels;
78     tables custArea;
79     run;
80

```

Εικόνα 55. Καθαρισμός δεδομένων (συνέχεια)

```

81     data thesis.customers_clean7;
82     set thesis.customers_clean6;
83     if find(custArea, 'ΓΛΥΦΑΔΑ') ^= 0 then custArea='ΓΛΥΦΑΔΑ';
84     if find(custArea, 'ΔΗΜΗΤΡΙΟ') ^= 0 then custArea='ΑΓ. ΔΗΜΗΤΡΙΟΣ';
85     if find(custArea, 'ΜΑΡ') ^= 0 and find(custArea, 'ΑΓ') ^= 0 then custArea='ΑΓ. ΜΑΡΙΝΑ ΚΟΡΩΠΙΟΥ';
86     if find(custArea, 'ΡΕΝΤΗΣ') ^= 0 then custArea='ΑΓ. Ι. ΡΕΝΤΗΣ';
87     if find(custArea, 'ΠΑΡΑΣΚΕΥΗ') ^= 0 then custArea='ΑΓ. ΠΑΡΑΣΚΕΥΗ';
88     if find(custArea, 'ΑΙΓΑΛ') ^= 0 then custArea='ΑΙΓΑΛΕΩ';
89     if find(custArea, 'ΑΚΤΗ') ^= 0 then custArea='ΓΑΛΑΖΙΑ ΑΚΤΗ';
90     if find(custArea, 'ΓΑΛΑΤΣΙ') ^= 0 then custArea='ΓΑΛΑΤΣΙ';
91     if find(custArea, 'ΜΠΟΛ') ^= 0 then custArea='ΔΗΜΟΣ ΣΥΜΠΟΛΙΤΕΙΑΣ';
92     if find(custArea, 'ΧΕΡΩΜΑ') ^= 0 then custArea='ΧΕΡΩΜΑ ΒΑΡΗΣ';
93     if find(custArea, 'ΣΚΑΡΠΙΖΑ') ^= 0 then custArea='ΣΚΑΡΠΙΖΑ';
94     if find(custArea, 'ΡΟΔΙΑ') ^= 0 then custArea='ΡΟΔΙΑ ΑΙΓΙΟΥ';
95     if find(custArea, 'ΣΜΥΡΝΗ') ^= 0 then custArea='ΝΕΑ ΣΜΥΡΝΗ';
96     if find(custArea, 'ΧΑΛΚΗ') ^= 0 then custArea='ΝΕΑ ΧΑΛΚΗΔΟΝΑ';
97     if find(custArea, 'ΜΙΛΑ') ^= 0 then custArea='ΜΗΛΑΔΕΖΑ ΒΑΡΗΣ';
98     if find(custArea, 'ΜΑΡΟΥΣΙ') ^= 0 then custArea='ΜΑΡΟΥΣΙ';
99     if find(custArea, 'ΚΟΡΜ') ^= 0 then custArea='ΚΟΡΜΠΙ ΒΑΡΗΣ';
100    if find(custArea, 'ΚΙΤΣΙ') ^= 0 then custArea='ΚΙΤΣΙ';
101    if find(custArea, 'ΝΙΚΗ') ^= 0 then custArea='ΘΕΣΣΑΛΟΝΙΚΗ';
102    if find(custArea, 'ΑΙΓΙΟ') ^= 0 then custArea='ΑΙΓΙΟ';
103    if find(custArea, 'ΒΡ') ^= 0 and find(custArea, 'ΣΙΑ') ^= 0 then custArea='ΒΡΙΑΛΗΣΣΙΑ';
104    if find(custArea, 'ΔΙΑΚΟ') ^= 0 then custArea='ΔΙΑΚΟΠΤΟ';
105    if find(custArea, 'ΚΑΛΥΒΙ') ^= 0 then custArea='ΚΑΛΥΒΙΑ';
106    if find(custArea, 'ΛΑΓ') ^= 0 then custArea='ΛΑΓΟΝΗΣΙ';
107    if find(custArea, 'ΦΑΛ') ^= 0 and find(custArea, 'Π') ^= 0 then custArea='ΠΑΛΑΙΟ ΦΑΛΗΡΟ';
108    if find(custArea, 'ΣΑΡΩΝΙΑΔΑ') ^= 0 then custArea='ΣΑΡΩΝΙΑΔΑ ΑΝΑΒΥΣΣΟΥ';
109    run;
110
111    proc freq data=thesis.customers_clean7 nlevels;
112    tables custArea;
113    run;

```

Εικόνα 56. Καθαρισμός δεδομένων (συνέχεια)

### Ενοποίηση πινάκων (3.2.3)

Η διασύνδεση πινάκων πραγματοποιήθηκε μέσω του παρακάτω κώδικα.

```
Code
1  data thesis.transactions_1;
2      set thesis.transactions
3          (rename=(transCustomer=custCode));
4  run;
5
6  proc sort data=thesis.transactions_1 out=thesis.transactions_1s;
7      by custCode;
8  run;
9
10 proc sort data=thesis.customers_clean7 out=thesis.customers_clean7s;
11     by custCode;
12 run;
13
14 data thesis.cust_trans;
15     merge thesis.transactions_1s (in=a) thesis.customers_clean7s (in=b);
16     by custCode;
17     if a=1 and b=1;
18 run;
19
20 proc sort data=thesis.cust_trans out=thesis.cust_trans_s;
21     by transID;
22 run;
23
24 proc sort data=thesis.transactionitems_total out=thesis.transactionitems_tot_s;
25     by transID;
26 run;
27
28 data thesis.cust_trans_tranit;
29     merge thesis.cust_trans_s (in=a) thesis.transactionitems_tot_s (in=b);
30     by transID;
31     if a=1 and b=1;
32     drop Var1 transTameio transDateTime trans_time custPostalCode registrationDate custStoreReg registr_date registr_time;
33 run;
```

Εικόνα 57. Ταξινόμηση και διασύνδεση πινάκων μέσω των πρωτευόντων και δευτερευόντων κοινών κλειδιών τους

### Υπολογισμός μετρικών RFM πελατών (3.3.1)

Ο υπολογισμός μετρικών RFM των πελατών με βάση αυτήν την πληροφορία, έγινε με την ακολουθία κώδικα που φαίνεται παρακάτω.

```
Code
1  /* Create RFM dataset using PROC SQL */
2  proc sql;
3      /* Select customer ID (custCode), age (custAge), gender (custGender),
4         Region (custPrefecture, custArea), Number of Children (custNumChildren) and Occupation (custOccupation)
5         from cust_trans_tranit dataset */
6      /* Calculate monetary value (M) as the sum of distinct invoice total values */
7      /* Calculate frequency (F) as the count of distinct invoice numbers */
8      /* Calculate recency (R) as the number of weeks between the most recent invoice date and a
9         fixed date (7Jan2005) */
10     /* Set a constant value of 1 for T */
11     create table thesis.RFM_model as
12     select
13         custCode, custAge, custGender, custNumChildren, custPrefecture, custArea, custOccupation,
14         sum(distinct transValue) FORMAT=BEST. as 'M'n,
15         count(distinct 'transID'n) as F,
16         intck('week', max(trans_date), '7Jan2005'd) as R,
17         1 as T
18     from
19         thesis.cust_trans_tranit
20     group by
21         custCode, custAge, custGender, custNumChildren, custPrefecture, custArea, custOccupation;
22     quit;
```

Εικόνα 58. Υπολογισμός μετρικών πελατών με βάση το μοντέλο RFM (Recency, Frequency, Monetary)

### Προετοιμασία δεδομένων (3.4.1)

Η μετατροπή των αποτελεσμάτων του RFM analysis σε μορφή πίνακα, που μπορεί να επεξεργαστεί το λογισμικό SAS, έγινε μέσω του παρακάτω κώδικα.

```
Code
1 data thesis.RFM_Results_Atlantic;
2   set casuser.RFM_Results_Atlantic;
3   keep custCode _cluster_ID_;
4 run;
```

Εικόνα 59. Μετατροπή αποτελεσμάτων ανάλυσης RFM από μορφή sashdat σε μορφή sas7bdat

Η ενοποίηση όλων των πινάκων που απέμειναν, έγινε μέσω του παρακάτω κώδικα.

```
5 proc sort data=thesis.cust_trans_tranit out=thesis.cust_trans_tranit_s;
6   by custCode;
7 run;
8
9 proc sort data=thesis.RFM_Results_Atlantic out=thesis.RFM_Results_Atlantic_s;
10  by custCode;
11 run;
12
13 data thesis.cust_trans_tranit_rfm;
14  merge thesis.cust_trans_tranit_s (in=a) thesis.RFM_Results_Atlantic_s (in=b);
15  by custCode;
16  if a=1 and b=1;
17  keep transID transItem custCode _CLUSTER_ID_ ;
18 run;
19
20 data thesis.items_1;
21  set thesis.items
22    (rename=(itemID=transItem));
23 run;
24
25 proc sort data=thesis.cust_trans_tranit_rfm out=thesis.cust_trans_tranit_rfm_s;
26  by transItem;
27 run;
28
29 proc sort data=thesis.items_1 out=thesis.items_1_s;
30  by transItem;
31 run;
32
33 data thesis.mba;
34  merge thesis.cust_trans_tranit_rfm_s (in=a) thesis.items_1_s (in=b);
35  by transItem;
36  if a=1 and b=1;
37  keep transID transItem itemName itemBasicID itemBasicName custCode _CLUSTER_ID_ ;
38 run;
```

Εικόνα 60. Ενοποίηση υπόλοιπων πινάκων σε έναν ενιαίο, που περιλαμβάνει τα αποτελέσματα της ανάλυσης RFM

**BIBΛΙΟΓΡΑΦΙΑ**

- [1] <https://bernardmarr.com/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/>
- [2] Laudon, K. και Laudon, J., Πληροφοριακά Συστήματα Διοίκησης: 11η αμερικανική έκδοση (μετάφρ. Αρκουδέας Π, επιμ. Μάνθου Β.), (Κλειδάριθμος 2015) σελ 283-4, 350, ISBN 978-960-461-623-7
- [3] EKN Report: Big Data in Retail, SAS whitepaper, 2013 (<https://www.sas.com/en/whitepapers/ekn-retail-leveraging-big-data.html>)
- [4] Cappa, Francesco; Oriani, Raffaele; Peruffo, Enzo; McCarthy, Ian (2021). "Big Data for Creating and Capturing Value in the Digitalized Environment: Unpacking the Effects of Volume, Variety, and Veracity on Firm Performance". *Journal of Product Innovation Management*.
- [5] Dedić, N.; Stanier, C. (2017). "Towards Differentiating Business Intelligence, Big Data, Data Analytics and Knowledge Discovery". *Innovations in Enterprise Information Systems Management and Engineering. Lecture Notes in Business Information Processing*. Vol. 285. Berlin; Heidelberg: Springer International Publishing. pp. 114–122.
- [6] [https://en.wikipedia.org/wiki/Big\\_data](https://en.wikipedia.org/wiki/Big_data)
- [7] "The 5 V's of big data". *Watson Health Perspectives*. 17 September 2016.
- [8] [https://el.wikipedia.org/wiki/Μεγάλα\\_δεδομένα](https://el.wikipedia.org/wiki/Μεγάλα_δεδομένα)
- [9] Negash, S., Gray, P. (2008). *Business Intelligence*. In: *Handbook on Decision Support Systems 2. International Handbooks Information System*. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-540-48716-6\\_9](https://doi.org/10.1007/978-3-540-48716-6_9)
- [10] Wake Forest University (<https://business.wfu.edu/masters-in-business-analytics/articles/what-is-analytics/>)
- [11] [https://el.wikipedia.org/wiki/Τεχνητή\\_νοημοσύνη](https://el.wikipedia.org/wiki/Τεχνητή_νοημοσύνη)
- [12] Arshi Naim. (2022). *ROLE OF ARTIFICIAL INTELLIGENCE IN BUSINESS RISK MANAGEMENT*. *American Journal of Business Management, Economics and Banking, 1*, 55–66. Retrieved from <https://www.americanjournal.org/index.php/ajbmeb/article/view/37>
- [13] Chan, MC., Wong, CC., Tse, W.F., Cheung, B.K.S., Tang, G.YN. (2002). *Artificial Intelligence in Portfolio Management*. In: Yin, H., Allinson, N., Freeman, R., Keane, J., Hubbard, S. (eds) *Intelligent Data Engineering and Automated Learning — IDEAL 2002*. IDEAL 2002. *Lecture Notes in Computer Science*, vol 2412. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/3-540-45675-9\\_60](https://doi.org/10.1007/3-540-45675-9_60)
- [14] Nick F. Ryman-Tubb, Paul Krause, Wolfgang Garn, *How Artificial Intelligence and machine learning research impacts payment card fraud detection: A survey and industry benchmark*, *Engineering Applications of Artificial Intelligence*, Volume 76, 2018, Pages 130-157, ISSN 0952-1976, <https://doi.org/10.1016/j.engappai.2018.07.008>
- [15] Ahmad Ghodselahe and Ashkan Amirmadhi, *Application of Artificial Intelligence Techniques for Credit Risk Evaluation*, *International Journal of Modeling and Optimization*, Vol. 1, No. 3, August 2011
- [16] Armin Beverungen, *Algorithmic Trading, Artificial Intelligence and the Politics of Cognition*, 2009
- [17] Ana Fernandez, *Artificial Intelligence in Financial Services*, Bank of Spain, 2019
- [18] Park SH, Choi J, Byeon JS. *Key Principles of Clinical Validation, Device Approval, and Insurance Coverage Decisions of Artificial Intelligence*. *Korean J Radiol*. 2021 Mar;22(3):442-453. doi: 10.3348/kjr.2021.0048. PMID: 33629545; PMCID: PMC7909857
- [19] SUSSY BAYONA-ORÉ, RINO CERNA, EDUARDO TIRADO HINOJOZA, *Machine Learning for Price Prediction for Agricultural Products*, *WSEAS TRANSACTIONS on BUSINESS and ECONOMICS*, 2021
- [20] Wo Jae Lee, Haiyue Wu, Huitaek Yun, Hanjun Kim, Martin B.G. Jun, John W. Sutherland, *Predictive Maintenance of Machine Tool Systems Using Artificial Intelligence Techniques Applied to Machine Condition Data*, *Procedia CIRP*, Volume 80, 2019, Pages 506-511, ISSN 2212-8271
- [21] Rupa Dash, Mark McMurtrey, Carl Rebman, Upendra K. Kar, *Application of Artificial Intelligence in Automation of Supply Chain Management*, *Journal of Strategic Innovation and Sustainability* Vol. 14(3) 2019
- [22] A. Chouchene, A. Carvalho, T. M. Lima, F. Charrua-Santos, G. J. Osório and W. Barhoumi, "Artificial Intelligence for Product Quality Inspection toward Smart Industries: Quality Control of Vehicle Non-Conformities," *2020 9th International Conference on Industrial Technology and Management (ICITM)*, Oxford, UK, 2020, pp. 127-131, doi: 10.1109/ICITM48982.2020.9080396.



- [23] Somayya Madakam, Rajesh M. Holmukhe, Durgesh Kumar Jaiswal, *The Future Digital Work Force: Robotic Process Automation (RPA)*, 2018
- [24] Praveen K B, Pradyumna Kumar, Prateek J, Pragathi G, *Inventory Management using Machine Learning*, *International Journal of Engineering Research & Technology (IJERT)*, Vol. 9 Issue 06, June-2020, ISSN: 2278-0181
- [25] Yun Zhang, *The application of artificial intelligence in logistics and express delivery*, 2019 *J. Phys.: Conf. Ser.* 1325 012085, doi 10.1088/1742-6596/1325/1/012085
- [26] Bilal Hmoud, Varallyai LASZLO, *WILL ARTIFICIAL INTELLIGENCE TAKE OVER HUMANRESOURCES RECRUITMENT AND SELECTION?*, *Network Intelligence Studies*, 2019
- [27] Jennifer R. Burnett, *The Future of Employee Engagement: Real-Time Monitoring and Digital Tools for Engaging a Workforce*, *International Studies of Management & Organization*, Volume 94, 2019
- [28] Gonzalez, Manuel F.; Capman, John F.; Oswald, Frederick L.; Theys, Evan R.; and Tomczak, David L. (2019) "“Where’s the I-O?” Artificial Intelligence and Machine Learning in Talent Management Systems," *Personnel Assessment and Decisions: Number 5 : Iss. 3 , Article 5*. DOI: <https://doi.org/10.25035/pad.2019.03.005>
- [29] Torres Berru, Y., López Batista, V.F., Torres-Carrión, P., Jimenez, M.G. (2020). *Artificial Intelligence Techniques to Detect and Prevent Corruption in Procurement: A Systematic Literature Review*. In: Botto-Tobar, M., Zambrano Vizueté, M., Torres-Carrión, P., Montes León, S., Pizarro Vásquez, G., Durakovic, B. (eds) *Applied Technologies. ICAT 2019. Communications in Computer and Information Science*, vol 1194. Springer, Cham. [https://doi.org/10.1007/978-3-030-42520-3\\_21](https://doi.org/10.1007/978-3-030-42520-3_21)
- [30] *Reducing Government Budget Deficits by Attacking Fraud, Waste and Abuse*, 2019, SAS whitepaper
- [31] Assunta Di Vaio, Rohail Hassan, Claude Alavoine, *Data intelligence and analytics: A bibliometric analysis of human–Artificial intelligence in public sector decision-making effectiveness*, *Technological Forecasting and Social Change*, Volume 174, 2022, 121201, ISSN 0040-1625, <https://doi.org/10.1016/j.techfore.2021.121201>
- [32] Ben Ortiz, Laura Kahn, Marc Bosch, Philip Bogden, Viveca Pavon-Harr, Onur Savas, Ian McCulloh, *Improving Community Resiliency and Emergency Response With Artificial Intelligence*, 2020
- [33] Y. -L. Lin, T. -Y. Chen and L. -C. Yu, "Using Machine Learning to Assist Crime Prevention," 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), Hamamatsu, Japan, 2017, pp. 1029-1030, doi: 10.1109/IIAI-AAI.2017.46.
- [34] Chowdhury, M., Sadek, A., Ma, Y., Kanhere, N., & Bhavsar, P. (2006). *Applications of Artificial Intelligence Paradigms to Decision Support in Real-Time Traffic Management*. *Transportation Research Record*, 1968(1), 92–98. <https://doi.org/10.1177/0361198106196800111>
- [35] M Matheny, ST Israni, M Ahmed, *Artificial intelligence in health care: The hope, the hype, the promise, the peril*, 2019
- [36] Bingquan Huang, Mohand Tahar Kechadi, Brian Buckley, *Customer churn prediction in telecommunications*, *Expert Systems with Applications*, Volume 39, Issue 1, 2012, Pages 1414-1425, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2011.08.024>
- [37] Chenguang Wang, *Efficient customer segmentation in digital marketing using deep learning with swarm intelligence approach*, *Information Processing & Management*, Volume 59, Issue 6, 2022, 103085, ISSN 0306-4573, <https://doi.org/10.1016/j.ipm.2022.103085>
- [38] U. Arsenijevic and M. Jovic, "Artificial Intelligence Marketing: Chatbots," 2019 International Conference on Artificial Intelligence: Applications and Innovations (IC-AIAI), Belgrade, Serbia, 2019, pp. 19-193, doi: 10.1109/IC-AIAI48757.2019.00010.
- [39] Nanayakkara, N. W. O. K. D. S. P., *Application of Artificial Intelligence in Marketing Mix: A Conceptual Review (November 19, 2020)*. *Proceedings of the International Conference on Business & Information (ICBI) 2020*, Available at SSRN: <https://ssrn.com/abstract=3862936> or <http://dx.doi.org/10.2139/ssrn.3862936>
- [40] *Using SAS® Marketing Optimization to improve credit-line optimization*, 2017, SAS whitepaper
- [41] Sunghong Park, Junhee Cho, Kanghee Park, Hyunjung Shin, *Customer sentiment analysis with more sensibility*, *Engineering Applications of Artificial Intelligence*, Volume 104, 2021, 104356, ISSN 0952-1976, <https://doi.org/10.1016/j.engappai.2021.104356>
- [42] *ANTI-FRAUD TECHNOLOGY Benchmarking Report*, 2019, SAS whitepaper
- [43] *Balancing Fraud Detection and the Customer Experience Why it can be more important to identify the good customers than the fraudulent ones*, 2019, SAS whitepapers
- [44] Parvinder Singh, Mandeep Singh, *Fraud Detection by Monitoring Customer Behavior and Activities*, *International Journal of Computer Applications (0975 – 8887)*, Volume 111 – No 11, February 2015

- [45] Wayne Thompson, Recommendation Systems. An Overview of System Types and Benefits – and Why You Should Implement With SAS®, 2015, SAS whitepaper
- [46] Kris Johnson Ferreira, Bin Hong Alex Lee, David Simchi-Levi (2015) Analytics for an Online Retailer: Demand Forecasting and Price Optimization. *Manufacturing & Service Operations Management* 18(1):69-88.
- [47] Life Cycle Price Optimization: Pricing strategies for maximum profit, 2020, SAS whitepaper
- [48] Iqra Rehman, Dr. Hamid Ghous, Structured Critical Review on Market Basket Analysis using Deep Learning & Association Rules, *International Journal of Scientific & Engineering Research*, Volume 12, Issue 1, January-2021, 1 ISSN 2229-5518
- [49] Daniela Garza Gutiérrez, Juan Ignacio González Espinosa, Luz María Valdez de la Rosa, Applicable Models of Customer Analytics for a Retail Company in Mexico, *Proceedings of the International Conference on Industrial Engineering and Operations Management Pilsen, Czech Republic, July 23-26, 2019*
- [50] Md Rafqul Islam (2021), Customer Behavior Analytics and Visualization, University of Technology Sydney
- [51] D. Zaratsian, M. Osborne, J. Plumley (2013), Uncovering Patterns in Textual Data with SAS® Visual Analytics and SAS Text Analytics, SAS Institute Inc
- [52] V. Kumar, Transformative Marketing: The Next 20 Years
- [53] Kostiantyn Kucher, Carita Paradis, Andreas Kerren (2018), The State of the Art in Sentiment Visualization
- [54] Gaur et al. (2021), Capitalizing on big data and revolutionary 5G technology: Extracting and visualizing ratings and reviews of global chain hotels, *Computers and Electrical Engineering* 95
- [55] Jo-Ting Wei, Shih-Yen Lin and Hsin-Hung Wu, A review of the application of RFM model,
- [56] Bult, Jan & Wansbeek, Tom. (1995). Optimal Selection for Direct Mail. *Marketing Science*. 14. 378-394. 10.1287/mksc.14.4.378.
- [57] Kotler P, Armstrong G (2006), *Principles of Marketing*, Eleventh Edition, Pearson Prentice Hall, New Jersey.
- [58] Sohrabi B, Khanlari A (2007). Customer lifetime value (CLV) measurement based on RFM model. *Iranian Acc. Aud. Rev.*, 14(47): 7-20.
- [59] Hung C, Tsai C (2008). Market segmentation based on hierarchical self-organizing map for markets of multimedia on demand. *Expert Syst. Appl.*, 34(1): 780-787
- [60] J. Wei, S. Lin, H. Wu, A review of the application of RFM model, 2010, *African Journal of Business Management* Vol. 4(19), pp. 4199-4206, <http://www.academicjournals.org/AJBM>, ISSN 1993-8233 ©2010 Academic Journals
- [61] U. Firdaus and D. Nugeraha Utama, Development of bank's customer segmentation model based on RFM+B approach, 2020, *ICIC International 2021*, ISSN 2185-2766, pp.17 - 26
- [62] D. Chen, S. Sain, K. Guo, "Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining", 2012, *Journal of Database Marketing & Customer Strategy Management* (2012) 19, 197 – 208., doi: 10.1057/dbm.2012.17
- [63] D. Arthur, S. Vassilvitskii, "K-means++: the advantages of careful seeding", *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms (SODA '07)*, Society for Industrial and Applied Mathematics, USA, 2007, pp. 1027–1035
- [64] Z. Rahman, M. S. Hossain, M. Hasan, A. Imteaj, "An enhanced method of initial cluster center selection for K-means algorithm", 2021 *Innovations in Intelligent Systems and Applications Conference (ASYU)*, Elazig, Turkey, 2021, pp. 1-6, doi: 10.1109/ASYU52992.2021.9599017
- [65] Arthur, David & Vassilvitskii, Sergei. (2007). K-Means++: The Advantages of Careful Seeding. *Proc. of the Annu. ACM-SIAM Symp. on Discrete Algorithms*. 8. 1027-1035. 10.1145/1283383.1283494.
- [66] T. Wang, Q. Li, D. Bucci, Y. Liang, B. Chen, P. K. Varshney, "K-Medoids Clustering of Data Sequences With Composite Distributions", *IEEE Transactions on Signal Processing*, vol. 67, pp. 2093-2106, 2019, doi: 10.1109/TSP.2019.2901370
- [67] Ji, Jinchao & Pang, Wei & Zheng, Y. & Wang, Z. & Ma, Z. & Zhang, L.. (2015). A novel cluster center initialization method for the k-prototypes algorithms using centrality and distance. *Applied Mathematics and Information Sciences*. 9. 2933-2942. 10.12785/amis/090621.
- [68] R. Gustriansyah, N. Suhandi, F. Antony, Clustering optimization in RFM analysis based on k-means, *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 18, No. 1, April 2020, pp. 470-477, ISSN: 2502-4752, DOI: 10.11591/ijeecs.v18.i1.pp470-477

- [69] M. Z. Rodriguez, C. H. Comin, D. Casanova, O. M. Bruno, D. R. Amancio, L. d. F. Costa, et al., "Clustering algorithms: A comparative approach", PLoS ONE, vol. 14, no. 1, e0210236, 2019, doi: 10.1371/journal.pone.0210236
- [70] Muhidin, A. (2017). Analysis of Hierarchical Clustering and K-Mean Methods with LRFMP Model on Customer Segmentation. SIGMA, Journal of Technology Pelita Bangsa, 7(1), 82–83.
- [71] Dhandayudam, Prabha, and Ilango Krishnamurthi. "An improved clustering algorithm for customer segmentation." *International Journal of Engineering Science and Technology (IJEST)* 4.2 (2012): 695-702
- [72] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise", Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD'96), AAAI Press, 1996, pp. 226–231.
- [73] R. Berahmana, F. Mohammed, K. Chairuang, Customer Segmentation Based on RFM Model Using K-Means, K-Medoids, and DBSCAN Methods, LONTAR KOMPUTER VOL. 11, NO. 1 APRIL 2020, DOI : 10.24843/LKJITI.2020.v11.i01.p04
- [74] S. Monalisa, F. Kurnia, Analysis of DBSCAN and K-means algorithm for evaluating outlier on RFM model of customer behaviour, TELKOMNIKA, Vol.17, No.1, February 2019, pp.110~117, ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v17i1.9394
- [75] Idowu, Surefunmi & Kattukottai, Srivatsav. (2019). Customer Segmentation Based on RFM Model Using K-Means, Hierarchical and Fuzzy C- Means Clustering Algorithms. 10.13140/RG.2.2.15379.71201.
- [76] Buckinx W, Poel DV (2005). Customer base analysis: Partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. Eur. J. Oper. Res., 164: 252-268
- [77] Coussement K, Poel DV (2008). Improving customer attrition prediction by integrating emotions from client / company interaction emails and evaluating multiple classifiers. Expert Syst. Appl. 36: 6127-6134
- [78] Fader PS, Hardie BGS, Lee KL (2005). RFM and CLV: Using Iso-Value curves for customer base analysis. J. Mark. Res. 415-430
- [79] Chen YL, Kuo MH, Wu SY, Tang K (2009). Discovering recency, frequency, and monetary (RFM) sequential patterns from customers' purchasing data. Electron Commerce Res. Appl., 8: 241-251
- [80] Hosseini SMS, Maleki A, Gholarmian MR (2010). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. Expert Syst. Appl., 37: 5259-5264
- [81] King SF (2007). Citizens as customers: Exploring the future of CRM in UK local government. Gov. Inform Q., 24: 47-63
- [82] Olson, D.L.; Cao, Q.; Gu, C. & Lee, D. (2009). Comparison of customer response models, *Service Business*, Vol. 3, No. 2, (June 2009) 117-130, ISSN: 1862-8516
- [83] Cheng, C-H. & Chen, Y-S. (2009). Classifying the segmentation of customer value via RFM model and RS theory, *Expert Systems with Applications*, Vol. 36, No. 3, (April 2009) 4176-4184, ISSN: 0957-4174
- [84] Derya Birant (2011). *Data Mining Using RFM Analysis, Knowledge-Oriented Applications in Data Mining*, Prof. Kimito Funatsu (Ed.), ISBN: 978-953-307-154-1
- [85] Rao, A.B., Kiran, J.S. & G, P. Application of market–basket analysis on healthcare. *Int J Syst Assur Eng Manag* (2021)
- [86] L. Cavique, "A scalable algorithm for the market basket analysis", *Journal of Retailing and Consumer Services*, vol. 14, no. 6, pp. 400-407, 2007, doi: 10.1016/j.jretconser.2007.02.003
- [87] Manpreet Kaur and Shivani Kang (2016), *Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining*, *Procedia Computer Science*
- [88] A. A. Aldino, E. D. Pratiwi, Setiawansyah, S. Sintaro and A. Dwi Putra, "Comparison Of Market Basket Analysis To Determine Consumer Purchasing Patterns Using Fp-Growth And Apriori Algorithm," 2021 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE), Banyuwangi, Indonesia, 2021, pp. 29-34
- [89] M. Hossain, A. H. M. S. Sattar and M. K. Paul, "Market Basket Analysis Using Apriori and FP Growth Algorithm," 2019 22nd International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 2019, pp. 1-6
- [90] Qisman et al. (2020), *Market basket analysis using apriori algorithm to find consumer patterns in buying goods through transaction data (case study of Mizan computer retail stores)*, Department of Computer Science
- [91] K. Tatiana and M. Mikhail (2018), *Market basket analysis of heterogeneous data sources for recommendation system improvement*
- [92] Ling Guo, Songtao Guo and Xintao Wu (2007), *Privacy Preserving Market Basket Data Analysis*, University of North Carolina at Charlotte
- [93] <https://el.wikipedia.org/wiki/Ατλάντικ>

<sup>[94]</sup> [https://www.sas.com/en\\_us/software/viya-for-learners.html](https://www.sas.com/en_us/software/viya-for-learners.html)

<sup>[95]</sup> L. Ott, L. Pang, F. Ramos, S. Chawla, “On Integrated Clustering and Outlier Detection”, NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 1, December 2014, Pages 1359–1367, ISBN: 9781510800410

<sup>[96]</sup> Hindawi, “Retracted: An Empirical Study on Customer Segmentation by Purchase Behaviors Using a RFM Model and K-Means Algorithm”, Mathematical Problems in Engineering, Vol. 2023, Article ID 9873736, 1 page <https://doi.org/10.1155/2023/9873736>

<sup>[97]</sup> J. Liang, L. Bai, C. Dang, F. Cao, “The K-Means-Type Algorithms Versus Imbalanced Data Distributions”, IEEE Transactions on Fuzzy Systems, vol. 20, no. 4, pp. 728-745, Aug. 2012, doi: 10.1109/TFUZZ.2011.2182354.

<sup>[98]</sup> S. Monalisa, P. Nadya, R. Novita, “Analysis for Customer Lifetime Value Categorization with RFM Model”, Procedia Computer Science, Vol. 161, 2019, 834-840, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2019.11.190>.

<sup>[99]</sup>

[https://documentation.sas.com/doc/en/vdmlcdc/v\\_017/vdmlref/p0z9zh68v63r8gn1lh5qcu6xhvk4.htm](https://documentation.sas.com/doc/en/vdmlcdc/v_017/vdmlref/p0z9zh68v63r8gn1lh5qcu6xhvk4.htm)