

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΙΣΑΓΩΓΗ. Ο κόσμος των δεδομένων	4
Όγκος δεδομένων και η εκθετική τους αύξηση	4
Ορισμός των Big Data και τα χαρακτηριστικά τους.....	4
Τεχνολογίες αποθήκευσης και ανάλυσης.....	6
ΚΕΦΑΛΑΙΟ 1. Επιχειρησιακή αξιοποίηση δεδομένων - Business Intelligence (BI) και Business Analytics (BA)	7
1.1 Από το παρελθόν στο μέλλον: Insights μέσω BI και πρόβλεψη μέσω BA	7
1.2 Machine Learning και οι εφαρμογές του στις επιχειρήσεις	8
ΚΕΦΑΛΑΙΟ 2. Τεχνητή Νοημοσύνη - Ιστορική αναδρομή και είδη.....	10
2.1 Ορισμός τεχνητής νοημοσύνης.....	10
2.2 Σημαντικά παραδείγματα και επιπτώσεις στις επιχειρήσεις	12
ΚΕΦΑΛΑΙΟ 3. Επιχειρησιακές εφαρμογές των Big Data - Ανά τμήμα επιχείρησης.....	13
ΚΕΦΑΛΑΙΟ 4. Customer Segmentation & Market Basket Analysis	15
4.1 Τμηματοποίηση πελατών (Customer segmentation)	15
4.2 Ανάλυση Καλαθιού Αγοράς (Market Basket Analysis)	17
4.3 Literature review	18
ΚΕΦΑΛΑΙΟ 5. Ο κλάδος online delivery στην Ελλάδα	19
5.1 Ιστορική αναδρομή.....	19
5.2 Στρατηγικές και προκλήσεις	21
ΚΕΦΑΛΑΙΟ 6. Εφαρμογή ανάλυσης σε πραγματικά δεδομένα.....	22
6.1 Περιγραφή της βάσης δεδομένων	22
6.2 Προετοιμασία των δεδομένων	23
6.3 Διερευνητική ανάλυση δεδομένων (Exploratory data analysis)	24
6.3.1 Ανάλυση πωλήσεων.....	24
6.3.2 Ανάλυση πελατολογίου	34
6.3.3 Ανάλυση προϊόντων	40
ΚΕΦΑΛΑΙΟ 7. Εφαρμογή ανάλυσης RFM.....	50
7.1 Μεθοδολογία	50
7.2 Literature review	51
7.2.1 Κατανομές μεταβλητών R,F,M	52
7.2.2 Λογαριθμικός μετασχηματισμός	53
7.2.3 Box-Cox μετασχηματισμός.....	54
7.2.4 Σύγκριση αποτελεσμάτων Λογαριθμικού και Box-Cox μετασχηματισμού.....	57
7.2.5 Στατιστικοί έλεγχοι κανονικότητας	58
ΚΕΦΑΛΑΙΟ 8. Συσταδοποίηση.....	60
8.1 Προσδιορισμός βέλτιστου αριθμού ομάδων	62
8.1.1 Elbow method	62
8.1.2 Silhouette Analysis.....	63

8.1.3 Αλγόριθμος Aligned Box Criterion (ABC)	64
8.2 Εφαρμογή αλγορίθμου K-Means	66
8.3 Εφαρμογή αλγορίθμου DBSCAN	71
8.3.1 Σύγκριση συσταδοποιήσεων K-Means και DBSCAN	75
8.4 Αποτελέσματα συσταδοποίησης.....	76
8.4.1 Σύγκριση τμημάτων πελατών.....	85
ΚΕΦΑΛΑΙΟ 9. Υλοποίηση Market Basket Analysis.....	86
9.1 Μεθοδολογία	86
9.2 Εφαρμογή του Apriori αλγορίθμου.....	88
9.2.1 Σύγκριση αποτελεσμάτων MBA μεταξύ τμημάτων πελατών	95
9.3 Εφαρμογή Frequent Pattern (FP) Growth αλγορίθμου.....	99
9.3.1 Σύγκριση αποτελεσμάτων αλγορίθμων Apriori και FP-Growth	104
ΣΥΜΠΕΡΑΣΜΑΤΑ.....	105
ΠΑΡΑΡΤΗΜΑ	106
Υλοποίηση μετασχηματισμού και imputation με SQL	106
Υλοποίηση ανάλυσης με Python.....	118
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	148

ΠΕΡΙΛΗΨΗ

Στη σύγχρονη εποχή της ανάλυσης δεδομένων, οι επιχειρήσεις βασίζονται ολοένα και περισσότερο σε προηγμένες τεχνικές αναλυτικής για την κατανόηση της συμπεριφοράς των πελατών και τη βελτιστοποίηση των επιχειρησιακών τους στρατηγικών. Η παρούσα διατριβή εστιάζει σε δύο θεμελιώδεις μεθοδολογίες της επιχειρηματικής αναλυτικής: την Τμηματοποίηση Πελατών (Customer Segmentation) και την Ανάλυση Καλαθιού Αγορών (Market Basket Analysis - MBA), με εφαρμογή σε πραγματικά δεδομένα από διαδικτυακή πλατφόρμα παράδοσης ειδών σούπερ μάρκετ.

Αρχικά, διενεργείται Διερευνητική Ανάλυση Δεδομένων (EDA) με σκοπό την κατανόηση των προτύπων συναλλαγών, των αγοραστικών συμπεριφορών των πελατών και της κατανομής των πωλήσεων προϊόντων. Στη συνέχεια, η Τμηματοποίηση Πελατών πραγματοποιείται με χρήση του μοντέλου RFM (Recency, Frequency, Monetary) σε συνδυασμό με αλγορίθμους συσταδοποίησης, όπως ο K-Means και ο DBSCAN, για την κατηγοριοποίηση των πελατών σε διακριτά τμήματα. Τα αποτελέσματα αποκαλύπτουν τρεις κύριες ομάδες πελατών: Πιστοί Πελάτες (Loyalists), Δυνητικά Πιστοί Πελάτες (Potential Loyalists) και Πελάτες σε Κίνδυνο Αποχώρησης (Churners), παρέχοντας πολύτιμες πληροφορίες για στοχευμένες στρατηγικές μάρκετινγκ και διατήρησης πελατών.

Στη συνέχεια, η Ανάλυση Καλαθιού Αγορών υλοποιείται μέσω κανόνων συσχέτισης, με την εφαρμογή των αλγορίθμων Apriori και FP-Growth. Οι κανόνες αυτοί αναδεικνύουν σημαντικές σχέσεις μεταξύ των προϊόντων και προσφέρουν στρατηγικές για τη βελτιστοποίηση των πωλήσεων μέσω προτάσεων συνδυαστικών αγορών και προγραμμάτων προώθησης.

Η σύγκριση μεταξύ των αλγορίθμων συσταδοποίησης και των τεχνικών ανάλυσης συσχετίσεων επιτρέπει την αξιολόγηση των δυνατοτήτων και περιορισμών κάθε μεθόδου. Επιπλέον, η διασύνδεση των αποτελεσμάτων της τμηματοποίησης με την ανάλυση καλαθιού αγοράς προσφέρει εξαπομικευμένες προτάσεις σε επίπεδο πελατών, ενισχύοντας τη στρατηγική διατήρησης και ανάπτυξης εσόδων.

Η διατριβή ολοκληρώνεται με τη σύνθεση των κυριότερων ευρημάτων και την παρουσίαση προτάσεων για επιχειρησιακές στρατηγικές που στοχεύουν στην ενίσχυση της πιστότητας των πελατών, την εξαπομίκευση του μάρκετινγκ και τη βελτιστοποίηση των εσόδων. Μέσω της αξιοποίησης δεδομένων και αναλυτικών τεχνικών, η παρούσα μελέτη αναδεικνύει τη σημασία της επιχειρηματικής αναλυτικής ως εργαλείου στρατηγικού σχεδιασμού στον χώρο του διαδικτυακού λιανεμπορίου.

ABSTRACT

In an era of data-driven decision-making, businesses increasingly rely on advanced analytics to gain insights into customer behavior and optimize operational strategies. This thesis explores two fundamental methodologies in customer analytics: Customer Segmentation and Market Basket Analysis (MBA), applied to a real-world dataset from an online supermarket delivery platform.

The study begins with an Exploratory Data Analysis (EDA) to understand transaction patterns, customer behaviors, and product sales distribution. The Customer Segmentation approach utilizes the RFM (Recency, Frequency, Monetary) model combined with clustering algorithms such as K-Means and DBSCAN to categorize customers into meaningful segments. The results classify customers into Loyalists, Potential Loyalists, and Churners, enabling targeted marketing strategies.

Subsequently, the Market Basket Analysis employs association rule learning techniques, specifically Apriori and FP-Growth algorithms, to uncover purchase correlations and product affinities. The analysis identifies significant purchasing patterns and suggests optimal product bundling strategies to enhance cross-selling and upselling opportunities.

Comparative evaluations between clustering techniques and association rule methods provide insights into the strengths and limitations of each approach. Additionally, findings from the segmentation analysis are integrated into the MBA process, offering cluster-specific recommendations to optimize business outcomes.

The thesis concludes with a synthesis of key findings, outlining actionable business strategies for customer retention, personalized marketing, and revenue maximization. By leveraging data-driven methodologies, this study illustrates the potential of customer analytics to enhance strategic decision-making in the online retail sector.

ΕΙΣΑΓΩΓΗ. Ο κόσμος των δεδομένων

Όγκος δεδομένων και η εκθετική τους αύξηση

Ο όγκος των δεδομένων έχει γνωρίσει εκθετική αύξηση τις τελευταίες δεκαετίες, λόγω της προόδου της τεχνολογίας και της διάχυτης ενσωμάτωσης του διαδικτύου στην καθημερινή ζωή. Σύμφωνα με πρόσφατες εκθέσεις της Statista, ο συνολικός όγκος δεδομένων που παράγεται παγκοσμίως ήταν περίπου 64 zettabytes το 2020 και προβλέπεται να φθάσει τα 181 zettabytes μέχρι το 2025. Αυτή η εκπληκτική αύξηση τροφοδοτείται από διάφορους τομείς, όπως τα μέσα κοινωνικής δικτύωσης, το ηλεκτρονικό εμπόριο και η δημιουργία ψηφιακού περιεχομένου (Langer & Mukherjee, 2023).

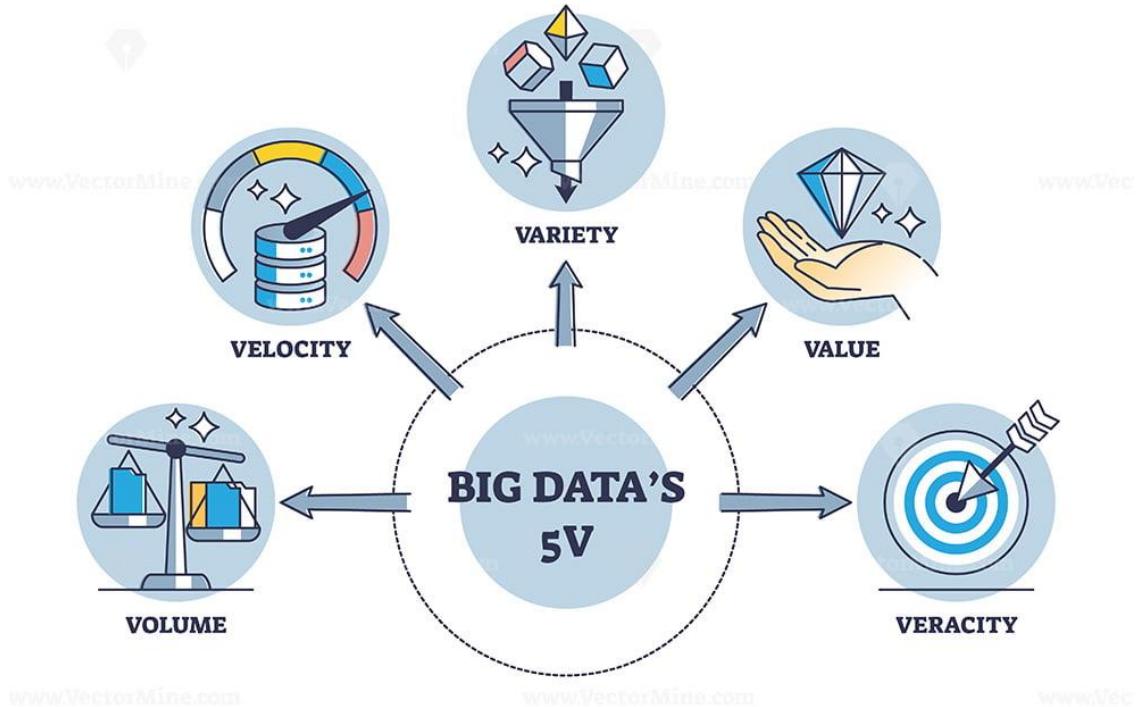
Οι πλατφόρμες των μέσων κοινωνικής δικτύωσης, ειδικότερα, έχουν καταστεί σημαντικοί συντελεστές αυτής της έκρηξης δεδομένων. Δισεκατομμύρια χρήστες συμμετέχουν ενεργά σε πλατφόρμες όπως το Facebook, το Instagram και το TikTok, δημιουργώντας καθημερινά τεράστιες ποσότητες δεδομένων μέσω αναρτήσεων, σχολίων, βίντεο και αλληλεπιδράσεων. Για παράδειγμα, μόνο το Facebook φέρεται να επεξεργάζεται πάνω από 4 petabytes δεδομένων κάθε μέρα. Ομοίως, οι πλατφόρμες ηλεκτρονικού εμπορίου συσσωρεύουν συνεχώς δεδομένα συναλλαγών και συμπεριφοράς, καθώς οι καταναλωτές πραγματοποιούν αγορές, περιηγούνται και αλληλεπιδρούν με προϊόντα στο διαδίκτυο. Αγορές όπως η Amazon και η Alibaba αξιοποιούν αυτά τα δεδομένα για να βελτιώσουν τους αλγορίθμους τους, να βελτιστοποιήσουν το απόθεμα και να εξατομικεύσουν τις συστάσεις (Ng et al., 2016).

Ο πολλαπλασιασμός των έξυπνων συσκευών και του «Διαδικτύου των Πραγμάτων» (Internet of Things-IoT) έχει εντείνει περαιτέρω την παραγωγή δεδομένων. Οι φορητές συσκευές, τα έξυπνα οικιακά συστήματα και τα συνδεδεμένα οχήματα συλλέγουν και μεταδίδουν συνεχώς δεδομένα, συμβάλλοντας σε ένα πλούσιο και πολύπλοκο ψηφιακό οικοσύστημα. Οι εκτιμήσεις της Cisco δείχνουν ότι μέχρι το 2025, οι συσκευές IoT θα παράγουν σχεδόν 80 zettabytes δεδομένων ετησίως (Deekshith, 2023).

Αυτή η εκθετική αύξηση των δεδομένων αναδεικνύει τόσο ευκαιρίες όσο και προκλήσεις. Οι οργανισμοί αξιοποιούν αυτά τα δεδομένα για να αποκτήσουν γνώσεις σχετικά με τη συμπεριφορά των καταναλωτών, να βελτιώσουν τις στρατηγικές μάρκετινγκ και να προωθήσουν την καινοτομία. Ωστόσο, η διαχείριση και η ανάλυση τέτοιων τεράστιων συνόλων δεδομένων απαιτεί προηγμένες τεχνολογικές υποδομές και αναλυτικές τεχνικές. Καθώς η ψηφιακή εποχή εξελίσσεται, η κατανόηση και η αποτελεσματική αξιοποίηση αυτών των δεδομένων θα παραμείνει καθοριστικός παράγοντας για τη διαμόρφωση ανταγωνιστικών και βιώσιμων επιχειρηματικών στρατηγικών (Deekshith, 2023).

Ορισμός των Big Data και τα χαρακτηριστικά τους

Τα μεγάλα δεδομένα αναφέρονται στα τεράστια και πολύπλοκα σύνολα δεδομένων που οι παραδοσιακές μέθοδοι επεξεργασίας δεδομένων δυσκολεύονται να διαχειριστούν λόγω του μεγέθους, της ταχύτητας και της ποικιλομορφίας τους. Η έννοια έχει καταστεί κεντρική στη σύγχρονη ανάλυση, καθώς οι επιχειρήσεις και οι οργανισμοί επιδιώκουν να αντλήσουν σημαντικές πληροφορίες από τον πρωτοφανή όγκο δεδομένων που παράγονται καθημερινά. Τα καθοριστικά χαρακτηριστικά των μεγάλων δεδομένων συνήθως συμπυκνώνονται στο πλαίσιο 5Vs: Volume, Velocity, Variety, Veracity και Value (Oussous et al., 2018).



Εικόνα 1. Τα 5 V's των Big Data (πηγή: <https://vectormine.com/item/5vs-of-big-data-as-big-information-type-characteristics-outline-diagram/>)

Ο όγκος αφορά τον τεράστιο όγκο δεδομένων που παράγονται από διάφορες πηγές, συμπεριλαμβανομένων των πλατφορμών μέσων κοινωνικής δικτύωσης, των συσκευών IoT, των συναλλαγών ηλεκτρονικού εμπορίου και του ψηφιακού περιεχομένου. Αυτό το χαρακτηριστικό αναδεικνύει την ανάγκη για κλιμακούμενες λύσεις αποθήκευσης και προηγμένη υπολογιστική ισχύ για την αποτελεσματική διαχείριση και ανάλυση των αυξανόμενων συνόλων δεδομένων. Η ταχύτητα αντικατοπτρίζει την ταχύτητα με την οποία παράγονται και επεξεργάζονται τα δεδομένα, συχνά σε πραγματικό χρόνο. Η ροή δεδομένων από τις χρηματοπιστωτικές αγορές, τις διαδικτυακές αλληλεπιδράσεις και τα δίκτυα αισθητήρων αποτελεί παράδειγμα της ανάγκης ταχείας επεξεργασίας για τη διασφάλιση της έγκαιρης λήψης αποφάσεων και ανταπόκρισης (Emmanuel & Stanier, 2016).

Η ποικιλία δίνει έμφαση στους διάφορους τύπους διαθέσιμων δεδομένων, συμπεριλαμβανομένων δομημένων μορφών, όπως τα λογιστικά φύλλα, ημιδομημένων μορφών, όπως τα αρχεία XML, και αδόμητων μορφών, όπως βίντεο, εικόνες και κείμενο. Αυτή η ετερογένεια απαιτεί εξελιγμένα εργαλεία και αλγόριθμους για την ενσωμάτωση και την ανάλυση διαφορετικών τύπων δεδομένων. Η ειλικρίνεια αναφέρεται στην αξιοπιστία και την ακρίβεια των δεδομένων. Με την επικράτηση του θορύβου, των ασυνεπειών και της παραπληροφόρησης, η διασφάλιση της ποιότητας των δεδομένων καθίσταται κρίσιμο καθήκον για την αξιόπιστη ανάλυση (Kitchin & McArdle, 2016).

Τέλος, η Αξία υπογραμμίζει τον απώτερο σκοπό των μεγάλων δεδομένων: την εξαγωγή αξιοποιήσιμων πληροφοριών και τη δημιουργία οφελών. Ενώ τα δεδομένα από μόνα τους έχουν δυνατότητες, η πραγματική τους αξία έγκειται στην ικανότητά τους να ενημερώνουν για στρατηγικές αποφάσεις, να ενισχύουν την αποτελεσματικότητα και να αποκαλύπτουν ευκαιρίες. Μαζί, αυτά τα χαρακτηριστικά ορίζουν τον πολύπλοκο αλλά ισχυρό τομέα των μεγάλων δεδομένων, ο οποίος συνεχίζει να αναδιαμορφώνει τους κλάδους και να οδηγεί την καινοτομία σε έναν κόσμο με επίκεντρο τα δεδομένα (Günther et al., 2017).

Τεχνολογίες αποθήκευσης και ανάλυσης

Η εκθετική αύξηση των δεδομένων έχει καταστήσει αναγκαία την ύπαρξη προηγμένων τεχνολογιών αποθήκευσης και ανάλυσης ικανών να διαχειρίζονται τεράστια και πολύπλοκα σύνολα δεδομένων. Τα παραδοσιακά συστήματα βάσεων δεδομένων είναι συχνά ανεπαρκή για τη διαχείριση της κλίμακας και της ποικιλίας των σύγχρονων δεδομένων, οδηγώντας στην υιοθέτηση κατανεμημένων συστημάτων και πλαισίων όπως το Hadoop. Αυτές οι τεχνολογίες επιτρέπουν την αποτελεσματική αποθήκευση, επεξεργασία και ανάλυση δεδομένων σε πολλαπλούς διακομιστές, προσφέροντας επεκτασιμότητα και αξιοπιστία (Siddiqa et al., 2017).

Το Hadoop, ένα πλαίσιο ανοικτού κώδικα, έχει γίνει ακρογωνιαίος λίθος στη διαχείριση μεγάλων δεδομένων. Έχει σχεδιαστεί για την αποθήκευση μεγάλων συνόλων δεδομένων με κατανεμημένο τρόπο χρησιμοποιώντας το «κατανεμημένο σύστημα αρχείων Hadoop» (Hadoop Distributed File System-HDFS), το οποίο χωρίζει τα δεδομένα σε μικρότερα μπλοκ και τα διανέμει σε πολλούς διακομιστές. Η προσέγγιση αυτή όχι μόνο εξασφαλίζει πλεονασμό και ανοχή σε σφάλματα, αλλά διευκολύνει επίσης την παράλληλη επεξεργασία. Το μοντέλο προγραμματισμού MapReduce, βασικό συστατικό του Hadoop, επιτρέπει την αποτελεσματική ανάλυση με τη διάσπαση των εργασιών σε μικρότερες μονάδες που μπορούν να υποστούν ταυτόχρονη επεξεργασία (Gupta et al., 2022).

Τα κατανεμημένα συστήματα, γενικά, διαδραματίζουν κεντρικό ρόλο στη σύγχρονη αποθήκευση και ανάλυση δεδομένων. Αξιοποιώντας ένα δίκτυο διασυνδεδεμένων διακομιστών, τα συστήματα αυτά κατανέμουν τα φορτία εργασίας και εξασφαλίζουν υψηλή διαθεσιμότητα, ακόμη και σε περίπτωση βλαβών υλικού. Αυτή η αρχιτεκτονική είναι ιδιαίτερα επωφελής για οργανισμούς που ασχολούνται με δυναμικά και μεγάλης κλίμακας σύνολα δεδομένων, καθώς επιτρέπει την απρόσκοπτη κλιμάκωση και τη βελτίωση των επιδόσεων (Gupta et al., 2022).

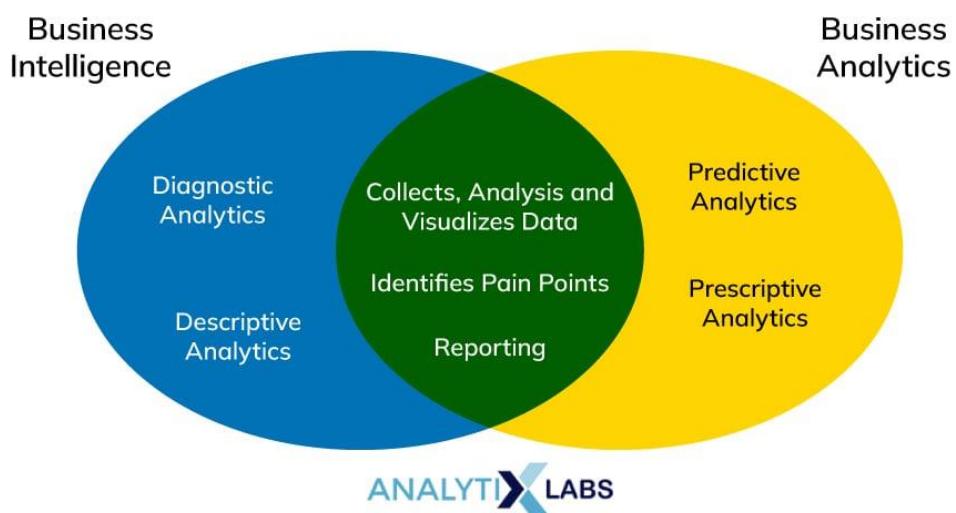
Η ενσωμάτωση αυτών των τεχνολογιών με προηγμένα αναλυτικά εργαλεία ενισχύει περαιτέρω τις δυνατότητές τους. Τα δεδομένα μπορούν να υποβάλλονται σε επεξεργασία και να αναλύονται σε πραγματικό χρόνο, επιτρέποντας στις επιχειρήσεις να αποκτούν γνώσεις και να ανταποκρίνονται γρήγορα στις μεταβαλλόμενες συνθήκες. Η υιοθέτηση πλαισίων αποθήκευσης και ανάλυσης, όπως το Hadoop και τα κατανεμημένα συστήματα, έχει φέρει επανάσταση στον τρόπο με τον οποίο οι οργανισμοί διαχειρίζονται και χρησιμοποιούν τα δεδομένα, παρέχοντας τη βάση για καινοτομίες στην τμηματοποίηση πελατών, την ανάλυση καλαθιού αγοράς και άλλες στρατηγικές βασισμένες στα δεδομένα που καθορίζουν το ανταγωνιστικό τοπίο της σημερινής ψηφιακής οικονομίας (Strohbach et al., 2016).

ΚΕΦΑΛΑΙΟ 1. Επιχειρησιακή αξιοποίηση δεδομένων - Business Intelligence (BI) και Business Analytics (BA)

1.1 Από το παρελθόν στο μέλλον: Insights μέσω BI και πρόβλεψη μέσω BA

Η «Επιχειρηματική Ευφυΐα» (Business Intelligence-BI) και η «Επιχειρηματική Ανάλυση» (Business Analytics-BA) έχουν καθοριστική σημασία για την εξέλιξη της χρήσης των δεδομένων στους οργανισμούς, αντικατοπτρίζοντας την εξέλιξη από τις περιγραφικές γνώσεις στις προγνωστικές και κανονιστικές δυνατότητες. Το BI επικεντρώνεται στην αξιοποίηση ιστορικών και τρεχόντων δεδομένων για την κατανόηση των προηγούμενων επιδόσεων και την ενημέρωση για τη λήψη αποφάσεων. Παρέχοντας δομημένες αναφορές, πίνακες οργάνων και οπτικοποιήσεις, το BI επιτρέπει στους οργανισμούς να εντοπίζουν τάσεις, να παρακολουθούν τις λειτουργίες και να διασφαλίζουν την ευθυγράμμιση με τους στρατηγικούς στόχους. Πρωταρχικός στόχος του είναι η ενίσχυση της επιχειρησιακής αποδοτικότητας και η υποστήριξη τακτικών και στρατηγικών αποφάσεων, προσφέροντας μια αναδρομική κατανόηση της οργανωτικής απόδοσης (Laursen & Thorlund, 2016).

Is Business Intelligence and Business Analytics Same?



Εικόνα 2. Ομοιότητες και διαφορές μεταξύ Business Intelligence και Business Analytics (πηγή: <https://www.analytixlabs.co.in/blog/business-intelligence-vs-business-analytics/>)

Η μετάβαση στην Επιχειρηματική Ανάλυση αντιπροσωπεύει μια μετατόπιση από την κατανόηση παρελθοντικών γεγονότων στην πρόβλεψη μελλοντικών αποτελεσμάτων. Η BA βασίζεται στα θεμέλια που έχει θέσει η BI, ενσωματώνοντας προηγμένα στατιστικά μοντέλα, αλγόριθμους μηχανικής μάθησης και τεχνικές εξόρυξης δεδομένων για την πρόβλεψη τάσεων, τον εντοπισμό προτύπων και τη σύσταση βέλτιστων δράσεων. Ενώ το BI ασχολείται με το «τι» και το «γιατί» της παρελθούσας και της παρούσας απόδοσης, το BA επιδιώκει να απαντήσει στα ερωτήματα «τι θα συμβεί» και «τι πρέπει να γίνει», ενισχύοντας έτσι την ικανότητα ενός

οργανισμού να προβλέπει τις μεταβολές της αγοράς, τη συμπεριφορά των πελατών και τις επιχειρησιακές προκλήσεις (Mohammed et al., 2024).

Η αλληλεπίδραση μεταξύ BI και BA είναι ιδιαίτερα σημαντική στην τμηματοποίηση πελατών και στην ανάλυση καλαθιού αγοράς. Το BI παρέχει πληροφορίες σχετικά με την ιστορική αγοραστική συμπεριφορά, τις προτιμήσεις των πελατών και τις τάσεις των πωλήσεων, διαμορφώνοντας το υπόβαθρο για τον εντοπισμό ζεχωριστών τμημάτων πελατών και αγοραστικών προτύπων. Η BA το προχωράει περαιτέρω με την πρόβλεψη των μελλοντικών αναγκών των πελατών, την πρόβλεψη της επιτυχίας στοχευμένων στρατηγικών μάρκετινγκ και τη βελτιστοποίηση της τοποθέτησης προϊόντων για τη μεγιστοποίηση της κερδοφορίας. Για παράδειγμα, στο λιανικό εμπόριο, το BI μπορεί να αναδείξει μια τάση αυξημένων πωλήσεων κατά τη διάρκεια συγκεκριμένων περιόδων, ενώ το BA θα προβλέψει μελλοντικές κορυφές πωλήσεων και θα προτείνει προσαρμογές αποθεμάτων ή στρατηγικές προώθησης (Phillips-Wren et al., 2021).

Η ενσωμάτωση των BI και BA οδηγεί επίσης στην καινοτομία σε δυναμικά μοντέλα τιμολόγησης, στην πρόβλεψη της ζήτησης και στο εξατομικευμένο μάρκετινγκ. Συνδυάζοντας τις περιγραφικές γνώσεις από το BI με τις προγνωστικές δυνατότητες του BA, οι οργανισμοί μπορούν να εφαρμόζουν στρατηγικές βασισμένες στα δεδομένα που προσαρμόζονται στις εξελισσόμενες συνθήκες της αγοράς και στις προσδοκίες των καταναλωτών. Οι προηγμένες τεχνικές BA, όπως τα νευρωνικά δίκτυα και τα δέντρα αποφάσεων, επιτρέπουν τη βαθύτερη διερεύνηση των πολύπλοκων σχέσεων δεδομένων, προωθώντας το ανταγωνιστικό πλεονέκτημα (Park et al., 2020).

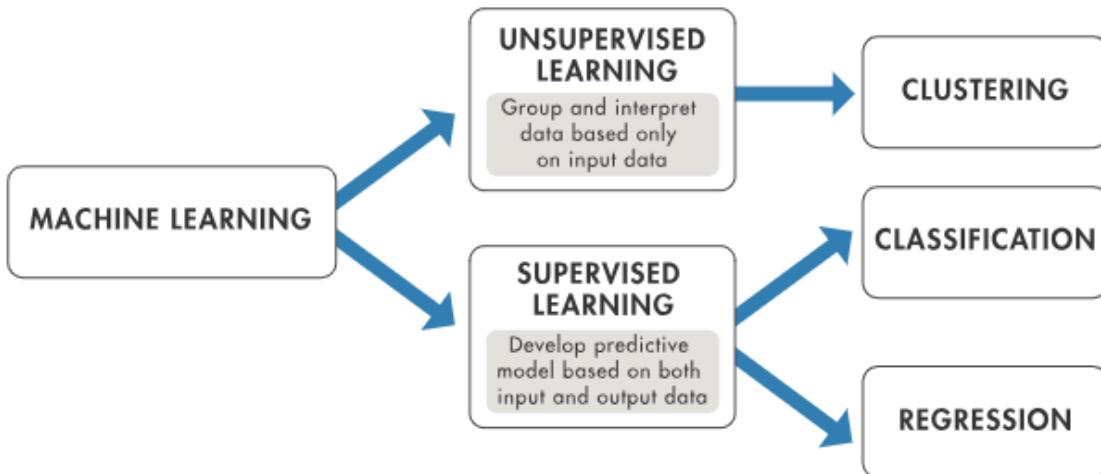
Οι τεχνολογικές εξελίξεις στην επεξεργασία και αποθήκευση δεδομένων έχουν ενισχύσει περαιτέρω τις δυνατότητες των BI και BA. Εργαλεία όπως η υπολογιστική νέφους, η τεχνητή νοημοσύνη και η ανάλυση σε πραγματικό χρόνο έχουν επεκτείνει το εύρος και την ταχύτητα της ανάλυσης, καθιστώντας δυνατή την άμεση άντληση αξιοποιήσιμων πληροφοριών από τεράστια σύνολα δεδομένων. Αυτή η εξέλιξη υποστηρίζει την απρόσκοπτη μετάβαση από την αντιδραστική λήψη αποφάσεων, που βασίζεται σε ιστορικές γνώσεις, σε προληπτικές στρατηγικές που ενημερώνονται από μελλοντική ανάλυση (Brijs, 2016).

Η σύγκλιση των BI και BA αντιπροσωπεύει μια μετασχηματιστική προσέγγιση για την αξιοποίηση των επιχειρησιακών δεδομένων, με επιπτώσεις που επεκτείνονται σε όλους τους κλάδους. Οι οργανισμοί που εκμεταλλεύονται αποτελεσματικά και τα δύο παραδείγματα είναι σε καλύτερη θέση να περιγγηθούν στην αβεβαιότητα, να αξιοποιήσουν τις αναδυόμενες ευκαιρίες και να διατηρήσουν την ανθεκτικότητά τους σε ανταγωνιστικά περιβάλλοντα. Με την ευθυγράμμιση των περιγραφικών δυνατοτήτων του BI με τις δυνάμεις πρόβλεψης και πρόβλεψης του BA, οι επιχειρήσεις μπορούν να επιτύχουν μια ολοκληρωμένη και προσανατολισμένη στο μέλλον προοπτική, εξασφαλίζοντας βιώσιμη ανάπτυξη και στρατηγική επιτυχία σε έναν όλο και περισσότερο επικεντρωμένο στα δεδομένα κόσμο (Tavera Romero et al., 2021).

1.2 Machine Learning και οι εφαρμογές του στις επιχειρήσεις

Η μηχανική μάθηση, ένα υποσύνολο της τεχνητής νοημοσύνης, έχει αναδειχθεί σε μετασχηματιστικό εργαλείο στον επιχειρηματικό τομέα, επιτρέποντας στους οργανισμούς να αντλούν αξιοποιήσιμες πληροφορίες από πολύπλοκα σύνολα δεδομένων και να αυτοματοποιούν τις διαδικασίες λήψης αποφάσεων. Αξιοποιώντας αλγορίθμους που βελτιώνουν την απόδοση μέσω της εμπειρίας, η μηχανική μάθηση προσφέρει ένα ισχυρό πλαίσιο για προγνωστική ανάλυση, αναγνώριση προτύπων και βελτιστοποίηση. Η εφαρμογή της εκτείνεται σε διάφορους

κλάδους, αντανακλώντας την προσαρμοστικότητά της και τη δυνατότητά της να φέρει επανάσταση στις παραδοσιακές επιχειρηματικές λειτουργίες (Lee & Shin, 2020).



Εικόνα 3. Machine learning techniques (πηγή: <https://www.mathworks.com/discovery/machine-learning.html>)

Στο πλαίσιο της τμηματοποίησης πελατών, οι αλγόριθμοι μηχανικής μάθησης αναλύουν τεράστιους όγκους δεδομένων πελατών για τον εντοπισμό διακριτών ομάδων βάσει συμπεριφορών, προτιμήσεων και δημογραφικών χαρακτηριστικών. Τεχνικές όπως η ομαδοποίηση k-means και η ιεραρχική ομαδοποίηση επιτρέπουν στις επιχειρήσεις να τμηματοποιούν την πελατειακή τους βάση με μεγαλύτερη ακρίβεια. Αυτή η τμηματοποίηση διευκολύνει την ανάπτυξη εξαπομικευμένων στρατηγικών μάρκετινγκ, βελτιώνοντας την ικανοποίηση και την αφοσίωση των πελατών. Επιπλέον, τα μοντέλα μηχανικής μάθησης είναι ικανά να εξελίσσονται με την πάροδο του χρόνου, προσαρμοζόμενα στις αλλαγές στη συμπεριφορά των πελατών και στη δυναμική της αγοράς, εξασφαλίζοντας έτσι τη συνάφεια των προσπαθειών τμηματοποίησης (Khan et al., 2020).

Μια άλλη εξέχουσα εφαρμογή της μηχανικής μάθησης στις επιχειρήσεις είναι η ανάλυση καλαθιού αγοράς, η οποία περιλαμβάνει την αποκάλυψη συσχετίσεων μεταξύ προϊόντων με βάση δεδομένα συναλλαγών. Αλγόριθμοι όπως η εκμάθηση κανόνων συσχέτισης, συμπεριλαμβανομένων των Apriori και FP-Growth, βοηθούν στον εντοπισμό συχνών συνόλων στοιχείων και στη δημιουργία κανόνων που αποκαλύπτουν τα αγοραστικά πρότυπα των καταναλωτών. Οι επιχειρήσεις χρησιμοποιούν αυτές τις γνώσεις για τη βελτιστοποίηση της τοποθέτησης προϊόντων, το σχεδιασμό στρατηγικών διασταύρουμενων πωλήσεων και τη βελτίωση της διαχείρισης αποθεμάτων. Για παράδειγμα, η κατανόηση του γεγονότος ότι οι πελάτες που αγοράζουν ένα προϊόν είναι πιθανό να αγοράσουν και ένα άλλο, επιτρέπει στις επιχειρήσεις να δημιουργούν στοχευμένες εκστρατείες προώθησης που αυξάνουν τα έσοδα (Whig, 2019).

Η μηχανική μάθηση διαδραματίζει επίσης κρίσιμο ρόλο στην προγνωστική ανάλυση, επιτρέποντας στις επιχειρήσεις να προβλέπουν μελλοντικές τάσεις και συμπεριφορές. Με την εφαρμογή μοντέλων παλινδρόμησης, δέντρων αποφάσεων και νευρωνικών δικτύων, οι οργανισμοί μπορούν να προβλέπουν τις πωλήσεις, να προβλέπουν την αποχώρηση των πελατών και να εκτιμούν τη ζήτηση. Αυτές οι προβλέψεις ενημερώνουν για τον στρατηγικό Customer Segmentation & Market Basket Analysis

σχεδιασμό και τις λειτουργικές προσαρμογές, διασφαλίζοντας ότι οι επιχειρήσεις παραμένουν ανταγωνιστικές και ανταποκρίνονται σε δυναμικά περιβάλλοντα (Canhoto & Clear, 2020).

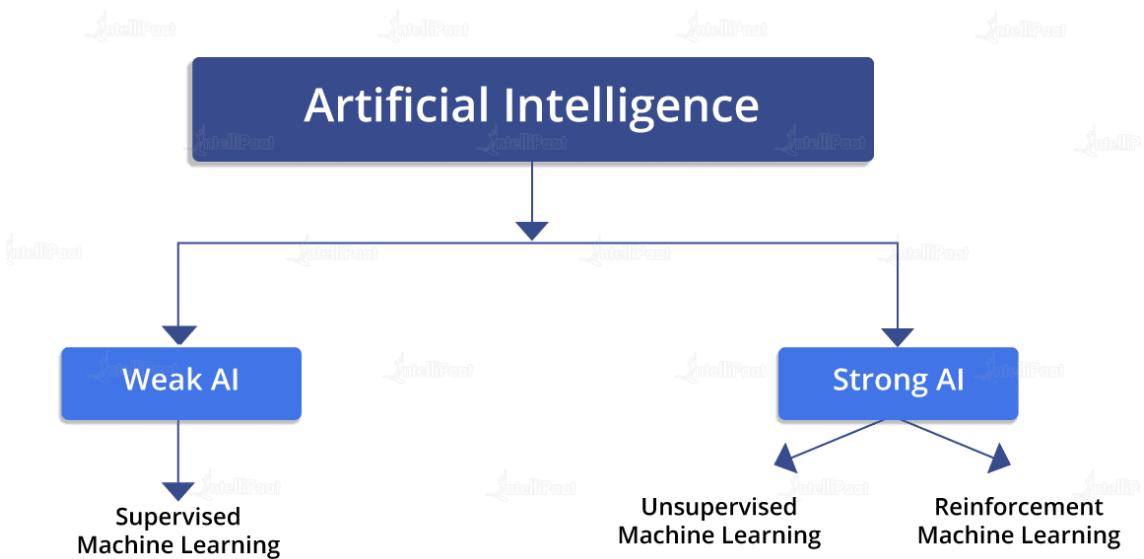
Επιπλέον, η μηχανική μάθηση έχει διευκολύνει τις εξελίξεις στις στρατηγικές δυναμικής τιμολόγησης, την ανίχνευση απάτης και τη βελτιστοποίηση της εφοδιαστικής αλυσίδας. Στη δυναμική τιμολόγηση, οι αλγόριθμοι αναλύουν σε πραγματικό χρόνο τις συνθήκες της αγοράς, την τιμολόγηση των ανταγωνιστών και την προθυμία των πελατών να πληρώσουν, επιτρέποντας στις επιχειρήσεις να προσαρμόζουν τις τιμές δυναμικά για τη μεγιστοποίηση της κερδοφορίας. Τα συστήματα ανίχνευσης απάτης αξιοποιούν τεχνικές επιβλεπόμενης και μη επιβλεπόμενης μάθησης για τον εντοπισμό ανωμαλιών σε δεδομένα συναλλαγών, προστατεύοντας τους οργανισμούς από οικονομικές απώλειες και απώλειες φήμης. Στη διαχείριση της εφοδιαστικής αλυσίδας, τα μοντέλα πρόβλεψης βελτιστοποιούν τα logistics, τα επίπεδα αποθεμάτων και τα χρονοδιαγράμματα παράδοσης, ενισχύοντας την αποδοτικότητα και μειώνοντας το κόστος (Hamzehi & Hosseini, 2022).

Η ενσωμάτωση της μηχανικής μάθησης στις επιχειρηματικές διαδικασίες υπογραμμίζει τη σημασία της ως μοχλού καινοτομίας και ανταγωνιστικότητας. Καθώς ο όγκος και η πολυπλοκότητα των δεδομένων συνεχίζουν να αυξάνονται, η μηχανική μάθηση θα παραμείνει καθοριστικής σημασίας για να μπορέσουν οι οργανισμοί να αξιοποιήσουν το δυναμικό των δεδομένων, να βελτιώσουν τη λήψη αποφάσεων και να επιτύχουν βιώσιμη ανάπτυξη σε ένα ταχέως εξελισσόμενο τοπίο (Bharadiya, 2023).

ΚΕΦΑΛΑΙΟ 2. Τεχνητή Νοημοσύνη - Ιστορική αναδρομή και είδη

2.1 Ορισμός τεχνητής νοημοσύνης

Η Τεχνητή Νοημοσύνη (Artificial Intelligence-AI) περιλαμβάνει ένα ευρύ φάσμα υπολογιστικών τεχνολογιών που έχουν σχεδιαστεί για να προσομοιώνουν την ανθρώπινη νοημοσύνη, επιτρέποντας στις μηχανές να εκτελούν εργασίες που παραδοσιακά απαιτούσαν ανθρώπινη γνώση. Η τεχνητή νοημοσύνη κατηγοριοποιείται συνήθως σε δύο κύρια παραδείγματα: «Ασθενής AI» (Weak AI) και «Ισχυρή AI» (Strong AI). Αυτές οι διακρίσεις είναι θεμελιώδεις για την κατανόηση του πεδίου εφαρμογής και των περιορισμών της τεχνητής νοημοσύνης όπως εφαρμόζεται σε διάφορους τομείς, συμπεριλαμβανομένων των επιχειρηματικών αναλύσεων (Flowers, 2019).



Εικόνα 4. Strong and Weak AI (πηγή: <https://intellipaat.com/blog/data-science-vs-deep-learning-vs-machine-learning-vs-artificial-intelligence/>)

Η ασθενής τεχνητή νοημοσύνη, η οποία αναφέρεται επίσης ως στενή τεχνητή νοημοσύνη, έχει σχεδιαστεί για την εκτέλεση συγκεκριμένων εργασιών ή την επίλυση συγκεκριμένων προβλημάτων εντός ενός καθορισμένου τομέα. Λειτουργεί υπό τους περιορισμούς των προπρογραμματισμένων αλγορίθμων και μοντέλων, εστιάζοντας στην επίτευξη σαφώς καθορισμένων στόχων χωρίς να διαθέτει πραγματική κατανόηση ή συνείδηση. Τα αδύναμα συστήματα AI είναι διαδεδομένα σε σύγχρονες εφαρμογές, από εικονικούς βοηθούς όπως η Siri και η Alexa έως συστήματα συστάσεων στο ηλεκτρονικό εμπόριο και σε πλατφόρμες ροής. Στις επιχειρήσεις, η Weak AI διευκολύνει διαδικασίες όπως η τμηματοποίηση πελατών και η ανάλυση καλαθιού αγοράς με την αυτοματοποίηση της ανάλυσης δεδομένων, τον εντοπισμό μοτίβων και την παροχή αξιοποίησιμων πληροφοριών. Παρά τη χρησιμότητά της, η ασθενής τεχνητή νοημοσύνη δεν έχει την ικανότητα να γενικεύει τη γνώση ή να λειτουργεί πέρα από τα προγραμματισμένα καθήκοντά της (Gil de Zúñiga et al., 2024).

Αντίθετα, η ισχυρή τεχνητή νοημοσύνη, γνωστή και ως γενική τεχνητή νοημοσύνη, αποτελεί μια φιλόδοξη έννοια στον τομέα της τεχνητής νοημοσύνης. Η ισχυρή τεχνητή νοημοσύνη στοχεύει στην επίτευξη νοημοσύνης ανθρώπινου επιπέδου, επιτρέποντας στις μηχανές να κατανοούν, να μαθαίνουν και να εκτελούν οποιαδήποτε διανοητική εργασία χωρίς προγραμματισμό σε συγκεκριμένο τομέα. Σε αντίθεση με την Weak AI, η Strong AI φιλοδοξεί να μιμηθεί το εύρος και το βάθος της ανθρώπινης νόησης, περιλαμβάνοντας τη συλλογιστική, την επίλυση προβλημάτων και τη συναισθηματική νοημοσύνη. Αν και θεωρητικής φύσης, η Strong AI παραμένει αντικείμενο σημαντικής ακαδημαϊκής και ηθικής συζήτησης, καθώς η υλοποίησή της θέτει βαθιές συνέπειες για την κοινωνία, την ηθική και την ανθρώπινη ταυτότητα. Επί του παρόντος, η ισχυρή τεχνητή νοημοσύνη υπάρχει μόνο ως θεωρητικό κατασκεύασμα, χωρίς να έχει επιτευχθεί πρακτική εφαρμογή (Liu, 2021).

Η διαφοροποίηση μεταξύ της Weak AI και της Strong AI υπογραμμίζει τους διαφορετικούς βαθμούς πολυπλοκότητας και εφαρμοσιμότητας των συστημάτων τεχνητής νοημοσύνης. Η εξειδίκευση και ο προσανατολισμός της Weak AI σε εργασίες την καθιστούν ιδιαίτερα αποτελεσματική για άμεσες και πρακτικές περιπτώσεις χρήσης, ιδίως σε τομείς που βασίζονται

σε δεδομένα, όπως η επιχειρηματική ανάλυση. Η ικανότητά της να επεξεργάζεται ταχέως μεγάλους όγκους δεδομένων και να παράγει ιδέες έχει μεταμορφώσει τους κλάδους, παρέχοντας στους οργανισμούς εργαλεία για τη βελτίωση της λήψης αποφάσεων, τον εξορθολογισμό των λειτουργιών και την πρόβλεψη της συμπεριφοράς των καταναλωτών. Η ισχυρή τεχνητή νοημοσύνη, αν και δεν έχει ακόμη υλοποιηθεί, αντιπροσωπεύει την απόλυτη φιλοδοξία της δημιουργίας συστημάτων που μιμούνται την ανθρώπινη νοημοσύνη στο σύνολό της, προσφέροντας δυνητικές εξελίξεις και προκλήσεις πολύ πέρα από τις τρέχουσες δυνατότητες της αδύναμης τεχνητής νοημοσύνης (Bory et al., 2024).

Η εννοιολογική διάκριση μεταξύ αυτών των δύο μορφών τεχνητής νοημοσύνης είναι κρίσιμη για την κατανόηση των ρόλων και των περιορισμών τους τόσο στις τρέχουσες όσο και στις μελλοντικές εφαρμογές. Η Weak AI αποτελεί τη ραχοκοκαλία των πρακτικών εφαρμογών σήμερα, ενώ η Strong AI συνεχίζει να εμπνέει την πρόοδο της έρευνας, διαμορφώνοντας την πορεία της ανάπτυξης της τεχνητής νοημοσύνης. Η κατανόηση αυτών των παραδειγμάτων είναι απαραίτητη για την πλοιήγηση στο εξελισσόμενο τοπίο της τεχνητής νοημοσύνης και των επιπτώσεών της στις επιχειρήσεις και την κοινωνία (Russell & Norvig, 2021).

2.2 Σημαντικά παραδείγματα και επιπτώσεις στις επιχειρήσεις

Η AI έχει επηρεάσει σημαντικά τις επιχειρηματικές πρακτικές, με πολυάριθμες εφαρμογές που φέρνουν επανάσταση στη λήψη αποφάσεων, τη λειτουργική αποτελεσματικότητα και τη δέσμευση των πελατών. Η ενσωμάτωση της AI στις επιχειρηματικές αναλύσεις έχει επιφέρει μετασχηματιστικά αποτελέσματα, ιδίως στους τομείς της τμηματοποίησης πελατών και της ανάλυσης καλαθιού αγοράς. Αρκετά βασικά παραδείγματα καταδεικνύουν τις βαθιές επιπτώσεις της AI στις επιχειρήσεις, αναδεικνύοντας τον ρόλο της στην προώθηση του ανταγωνιστικού πλεονεκτήματος και της καινοτομίας (Akerkar, 2019).

Μια εξέχουσα εφαρμογή της AI στις επιχειρήσεις είναι η τμηματοποίηση πελατών. Οι προηγμένοι αλγόριθμοι AI αναλύουν τεράστια σύνολα δεδομένων για τον εντοπισμό σημαντικών τμημάτων πελατών με βάση συμπεριφορικά, δημογραφικά και συναλλακτικά χαρακτηριστικά. Για παράδειγμα, οι τεχνικές μηχανικής μάθησης, όπως η ομαδοποίηση και τα δέντρα αποφάσεων, επιτρέπουν στους οργανισμούς να τμηματοποιούν την πελατειακή τους βάση με ακρίβεια, επιτρέποντας στοχευμένες εκστρατείες μάρκετινγκ και εξατομικευμένες εμπειρίες. Η τμηματοποίηση πελατών με βάση την τεχνητή νοημοσύνη έχει υιοθετηθεί εκτενώς σε κλάδους όπως το λιανικό εμπόριο, οι τράπεζες και οι τηλεπικοινωνίες, όπου η κατανόηση των αναγκών και των προτιμήσεων των πελατών είναι απαραίτητη για τη βελτίωση της ικανοποίησης και της αφοσίωσης (Kalusivalingam et al., 2020).

Η ανάλυση καλαθιού αγοράς είναι ένας άλλος κρίσιμος τομέας όπου η AI έχει επιδείξει σημαντική αξία. Με τη χρήση αλγορίθμων μάθησης κανόνων συσχέτισης, όπως οι Apriori και FP-Growth, η AI εντοπίζει μοτίβα και συσχετίσεις μεταξύ προϊόντων που αγοράζονται συχνά μαζί. Αυτές οι γνώσεις καθοδηγούν τις επιχειρήσεις στη βελτιστοποίηση της τοποθέτησης προϊόντων, στο σχεδιασμό αποτελεσματικών στρατηγικών διασταύρωμάτων πωλήσεων και στη δημιουργία προωθητικών πακέτων που ενισχύουν τα έσοδα. Κολοσσοί του λιανεμπορίου, όπως η Amazon και η Walmart, αξιοποιούν την AI για την ανάλυση των δεδομένων συναλλαγών σε πραγματικό χρόνο, διασφαλίζοντας ότι οι συστάσεις και οι προσαρμογές των αποθεμάτων ευθυγραμμίζονται με τις απαιτήσεις των πελατών (Hoque et al., 2024).

Πέρα από αυτά τα παραδείγματα, οι εφαρμογές της AI στην προγνωστική ανάλυση έχουν εκτεταμένες επιπτώσεις στην επιχειρηματική στρατηγική. Τα μοντέλα πρόβλεψης, τα οποία τροφοδοτούνται από νευρωνικά δίκτυα και ανάλυση παλινδρόμησης, προβλέπουν τις τάσεις των

πωλήσεων, εντοπίζουν πιθανές αποχωρήσεις και προβλέπουν τη δυναμική της αγοράς. Αυτές οι δυνατότητες δίνουν τη δυνατότητα στις επιχειρήσεις να προσαρμόζονται προληπτικά στις αλλαγές, να βελτιστοποιούν την κατανομή των πόρων και να αναπτύσσουν στρατηγικές που μεγιστοποιούν την κερδοφορία. Για παράδειγμα, η πρόβλεψη της ζήτησης με βάση την AI βοηθά τους οργανισμούς να διατηρούν τα βέλτιστα επίπεδα αποθεμάτων, μειώνοντας το κόστος που σχετίζεται με την υπεραποθεματοποίηση ή τα αποθέματα (Kitsios & Kamariotou, 2021).

Οι επιπτώσεις της AI για τις επιχειρήσεις επεκτείνονται πέρα από τη λειτουργική αποδοτικότητα. Τα συστήματα AI επιτρέπουν δυναμικές στρατηγικές τιμολόγησης, ανίχνευση απάτης και ανάλυση συναισθήματος, μετασχηματίζοντας τον τρόπο με τον οποίο οι οργανισμοί αλληλεπιδρούν με τους πελάτες και ανταποκρίνονται στις συνθήκες της αγοράς. Τα δυναμικά μοντέλα τιμολόγησης, για παράδειγμα, αναλύουν μεταβλητές της αγοράς σε πραγματικό χρόνο για να καθορίζουν τιμές που μεγιστοποιούν την κερδοφορία, παραμένοντας παράλληλα ανταγωνιστικές. Τα συστήματα ανίχνευσης απάτης χρησιμοποιούν επιβλεπόμενη και μη επιβλεπόμενη μάθηση για τον εντοπισμό ανωμαλιών και την προστασία των επιχειρήσεων από οικονομικές απώλειες. Η ανάλυση συναισθήματος, που εφαρμόζεται στα μέσα κοινωνικής δικτύωσης και στα σχόλια των πελατών, παρέχει πολύτιμες πληροφορίες για τις αντιλήψεις του κοινού και τις αναδυόμενες τάσεις (Basal et al., 2024).

Αυτές οι εφαρμογές καταδεικνύουν την ευελιξία και τις μετασχηματιστικές δυνατότητες της AI στις επιχειρήσεις. Με την αυτοματοποίηση πολύπλοκων εργασιών και την ενίσχυση των αναλυτικών δυνατοτήτων, η AI όχι μόνο βελτιστοποιεί τις διαδικασίες λήψης αποφάσεων, αλλά και προάγει την καινοτομία και τη στρατηγική ανάπτυξη. Καθώς οι οργανισμοί συνεχίζουν να αξιοποιούν τις δυνατότητες της AI, οι επιπτώσεις της στο ανταγωνιστικό πλεονέκτημα και τη μακροπρόθεσμη βιωσιμότητα θα γίνονται όλο και πιο έντονες, διαμορφώνοντας το μέλλον των επιχειρήσεων στην ψηφιακή εποχή (Enholm et al., 2022).

ΚΕΦΑΛΑΙΟ 3. Επιχειρησιακές εφαρμογές των Big Data - Ανά τμήμα επιχείρησης

Τα μεγάλα δεδομένα έχουν μεταμορφώσει το τοπίο των σύγχρονων επιχειρήσεων, προσφέροντας τεράστιες δυνατότητες για την αύξηση της αποδοτικότητας, την ενίσχυση της λήψης αποφάσεων και τη δημιουργία αξίας σε διάφορα τμήματα. Μέσω προηγμένων αναλύσεων, οι οργανισμοί μπορούν πλέον να εξάγουν αξιοποιήσιμες πληροφορίες από τις τεράστιες ποσότητες δομημένων και μη δομημένων δεδομένων που συλλέγουν (Ajah & Nweke, 2019).

Στο μάρκετινγκ, η ανάλυση μεγάλων δεδομένων επιτρέπει την ανάπτυξη στοχευμένων εκστρατειών και την εξαγωγή βαθιών γνώσεων από τα δεδομένα των πελατών, αναδιαμορφώνοντας τον τρόπο με τον οποίο οι επιχειρήσεις συνεργάζονται με το κοινό τους. Τα τεράστια σύνολα δεδομένων που συλλέγονται από τις ψηφιακές αλληλεπιδράσεις, τις αγορές και τις προτιμήσεις επιτρέπουν τη δημιουργία εξαιρετικά εξατομικευμένων στρατηγικών μάρκετινγκ, αυξάνοντας την αφοσίωση των πελατών και οδηγώντας στην αύξηση των εσόδων. Η προγνωστική ανάλυση δίνει τη δυνατότητα στους οργανισμούς να προβλέπουν τη συμπεριφορά των πελατών, να εντοπίζουν πιθανές αποχωρήσεις και να σχεδιάζουν στρατηγικές διατήρησης που να ανταποκρίνονται αποτελεσματικά στις ατομικές ανάγκες. Επιπλέον, η ανάλυση συναισθήματος που προέρχεται από τα μέσα κοινωνικής δικτύωσης και τις πλατφόρμες κριτικών παρέχει ανατροφοδότηση σε πραγματικό χρόνο, βοηθώντας τους στις μεταβαλλόμενες προτιμήσεις των καταναλωτών. Αυτές οι δυνατότητες μειώνουν σημαντικά τις αναποτελεσματικότητες που σχετίζονται με τις ευρείες, μη

στοχευμένες προσπάθειες μάρκετινγκ, καθιστώντας κάθε αλληλεπίδραση με τον πελάτη πιο ουσιαστική και αποτελεσματική (Chaudhary & Alam, 2022).

Στο τμήμα του ανθρώπινου δυναμικού, τα μεγάλα δεδομένα διευκολύνουν τη βαθύτερη κατανόηση της δυναμικής του εργατικού δυναμικού, συμβάλλοντας στη βελτίωση της διατήρησης των εργαζομένων και στην αποτελεσματικότερη διαχείριση της μισθοδοσίας. Ο κύκλος εργασιών των εργαζομένων αποτελεί σημαντική πρόκληση για πολλούς οργανισμούς, με αποτέλεσμα συχνά υψηλό κόστος που συνδέεται με την πρόσληψη και την εκπαίδευση νέου προσωπικού. Αναλύοντας τα μοτίβα στη συμπεριφορά, τις επιδόσεις και τα σχόλια των εργαζομένων, οι οργανισμοί μπορούν να εντοπίσουν πρώιμα σημάδια δυσαρέσκειας ή αποδέσμευσης. Αυτό επιτρέπει τη λήψη προληπτικών μέτρων για την ενίσχυση της εργασιακής ικανοποίησης, όπως προσαρμογές στο φόρτο εργασίας, αναγνώριση των επιτευγμάτων ή ευκαιρίες για επαγγελματική ανάπτυξη (Zhang et al., 2021).

Επιπλέον, τα μοντέλα πρόβλεψης μπορούν να προβλέψουν τους κινδύνους κύκλου εργασιών, επιπρέποντας στα τμήματα ανθρώπινου δυναμικού να στοχεύουν αποτελεσματικά τις παρεμβάσεις. Όσον αφορά τη μισθοδοσία, η ανάλυση μεγάλων δεδομένων βελτιστοποιεί την κατανομή των πόρων και τη διαχείριση του κόστους, αναλύοντας τις τάσεις στις υπερωρίες, τη χρήση των παροχών και τα σημεία αναφοράς των αποζημιώσεων. Τέτοιες γνώσεις όχι μόνο εξορθολογίζουν τις διαδικασίες μισθοδοσίας, αλλά βοηθούν επίσης στην ευθυγράμμιση των στρατηγικών αποζημίωσης με τα πρότυπα του κλάδου, προωθώντας την ισότητα και την ικανοποίηση των εργαζομένων (Hamilton & Sodeman, 2020).

Η εφαρμογή των μεγάλων δεδομένων στην εφοδιαστική (logistics) έχει φέρει επανάσταση στη βελτιστοποίηση της δρομολόγησης, έναν κρίσιμο παράγοντα για την αποτελεσματικότητα της αλυσίδας εφοδιασμού. Η δυνατότητα ανάλυσης μεγάλου όγκου δεδομένων που σχετίζονται με τα πρότυπα κυκλοφορίας, τις καιρικές συνθήκες, την κατανάλωση καυσίμων και τα χρονοδιαγράμματα παράδοσης επιτρέπει στις εταιρείες να αναπτύσσουν εξαιρετικά αποδοτικές στρατηγικές δρομολόγησης. Οι στρατηγικές αυτές ελαχιστοποιούν τους χρόνους παράδοσης, μειώνουν το κόστος μεταφοράς και ενισχύουν την ικανοποίηση των πελατών. Οι προηγμένοι αλγόριθμοι μπορούν να προσαρμόζουν δυναμικά τις διαδρομές σε πραγματικό χρόνο, ανταποκρινόμενοι σε απρόβλεπτες διαταραχές, όπως ατυχήματα ή κλείσιμο δρόμων (Wang et al., 2016).

Επιπλέον, η ενσωμάτωση συσκευών IoT, όπως οι ανιχνευτές GPS και τα συστήματα τηλεματικής, δημιουργεί συνεχείς ροές δεδομένων που βελτιώνουν περαιτέρω τις λειτουργίες logistics. Οι τεχνολογίες αυτές διασφαλίζουν ότι τα εμπορεύματα μεταφέρονται με τον πιο αποδοτικό τρόπο, μειώνοντας όχι μόνο το κόστος αλλά και τις περιβαλλοντικές επιπτώσεις που σχετίζονται με την κατανάλωση καυσίμων και τις εκπομπές ρύπων. Η προληπτική συντήρηση, μια άλλη κρίσιμη πτυχή των εφαρμογών μεγάλων δεδομένων στην εφοδιαστική, διασφαλίζει ότι τα οχήματα συντηρούνται πριν από την εμφάνιση βλαβών, ελαχιστοποιώντας τον χρόνο διακοπής λειτουργίας και διατηρώντας την επιχειρησιακή συνέχεια (Jin & Kim, 2018).

Πέρα από το μάρκετινγκ, το ανθρώπινο δυναμικό και τα logistics, τα μεγάλα δεδομένα διαδραματίζουν καθοριστικό ρόλο στην ενίσχυση των διαδικασιών παραγωγής και πωλήσεων. Στην παραγωγή, η ενσωμάτωση της ανάλυσης δεδομένων επιτρέπει στους οργανισμούς να επιτύχουν μεγαλύτερη ακρίβεια και αποτελεσματικότητα. Αναλύοντας δεδομένα από αισθητήρες ενσωματωμένους σε μηχανήματα, οι κατασκευαστές μπορούν να παρακολουθούν την απόδοση του εξοπλισμού σε πραγματικό χρόνο, εντοπίζοντας ανωμαλίες που μπορεί να υποδεικνύουν πιθανές βλάβες. Η προληπτική συντήρηση διασφαλίζει ότι τα προβλήματα αντιμετωπίζονται πριν διαταράξουν την παραγωγή, μειώνοντας τον χρόνο διακοπής λειτουργίας και το σχετικό κόστος. Επιπλέον, τα μεγάλα δεδομένα διευκολύνουν τη βελτιστοποίηση των χρονοδιαγραμμάτων Customer Segmentation & Market Basket Analysis

παραγωγής αναλύοντας τα πρότυπα ζήτησης, τη διαθεσιμότητα πόρων και τους λειτουργικούς περιορισμούς. Αυτό διασφαλίζει ότι οι διαδικασίες παραγωγής ευθυγραμμίζονται με τις απαιτήσεις της αγοράς, ελαχιστοποιώντας τη σπατάλη και μεγιστοποιώντας την παραγωγή (Davenport, 2014).

Στις πωλήσεις, τα μεγάλα δεδομένα παρέχουν μια ολοκληρωμένη κατανόηση της συμπεριφοράς, των προτιμήσεων και των αγοραστικών προτύπων των πελατών. Οι ομάδες πωλήσεων μπορούν να αξιοποιήσουν αυτές τις γνώσεις για να αναπτύξουν αποτελεσματικότερες στρατηγικές, να εντοπίσουν ευκαιρίες διασταυρούμενων και ανοδικών πωλήσεων και να προβλέψουν τα έσοδα με μεγαλύτερη ακρίβεια. Οι αναλύσεις σε πραγματικό χρόνο δίνουν τη δυνατότητα στις ομάδες πωλήσεων να ανταποκρίνονται γρήγορα στις τάσεις της αγοράς, διασφαλίζοντας ότι οι προσπάθειές τους είναι τόσο έγκαιρες όσο και συναφείς (Hallikainen et al., 2020).

Οι μετασχηματιστικές δυνατότητες των μεγάλων δεδομένων επεκτείνονται και σε άλλες λειτουργίες εντός των οργανισμών, όπου υποστηρίζουν τη λήψη στρατηγικών αποφάσεων και τη λειτουργική αποδοτικότητα. Για παράδειγμα, στα χρηματοοικονομικά, η ανάλυση μεγάλων δεδομένων ενισχύει τη διαχείριση κινδύνων, εντοπίζοντας μοτίβα ενδεικτικά απάτης ή οικονομικών παρατυπών. Η παρακολούθηση των συναλλαγών και των δραστηριοτήτων λογαριασμών σε πραγματικό χρόνο επιτρέπει στους οργανισμούς να εντοπίζουν γρήγορα τις ανωμαλίες, μετριάζοντας τις πιθανές απώλειες. Ομοίως, στην έρευνα και την ανάπτυξη, τα μεγάλα δεδομένα επιταχύνουν την καινοτομία παρέχοντας πληροφορίες για τις τάσεις της αγοράς, τις ανάγκες των πελατών και τις αναδυόμενες τεχνολογίες. Αναλύοντας δεδομένα από διάφορες πηγές, οι οργανισμοί μπορούν να εντοπίζουν ευκαιρίες για ανάπτυξη και βελτίωση προϊόντων, διατηρώντας ανταγωνιστικό πλεονέκτημα στους κλάδους τους (Niu et al., 2021).

Επιπλέον, τα μεγάλα δεδομένα υποστηρίζουν πρωτοβουλίες περιβαλλοντικής βιωσιμότητας, επιτρέποντας στους οργανισμούς να παρακολουθούν και να βελτιστοποιούν την κατανάλωση ενέργειας, τη διαχείριση αποβλήτων και το αποτύπωμα άνθρακα. Αυτές οι εφαρμογές υπογραμμίζουν την ευελιξία των μεγάλων δεδομένων ως εργαλείο για την αύξηση της αξίας σε διάφορες οργανωτικές λειτουργίες (Bibri, 2018).

Η ενσωμάτωση των μεγάλων δεδομένων σε όλα τα επιχειρηματικά τμήματα υπογραμμίζει τις μετασχηματιστικές δυνατότητές τους, όχι μόνο για την ενίσχυση της αποδοτικότητας αλλά και για την προώθηση της καινοτομίας και της ανταγωνιστικότητας. Καθώς οι οργανισμοί συνεχίζουν να παράγουν και να αξιοποιούν τεράστιες ποσότητες δεδομένων, η ικανότητα εξαγωγής ουσιαστικών πληροφοριών θα παραμείνει κρίσιμος παράγοντας επιτυχίας. Αξιοποιώντας τις προηγμένες τεχνολογίες ανάλυσης και μηχανικής μάθησης, οι επιχειρήσεις μπορούν να ξεκλειδώσουν νέες ευκαιρίες, να ανταποκριθούν προληπτικά στις προκλήσεις και να προσφέρουν ανώτερη αξία στα ενδιαφερόμενα μέρη τους (Ajah & Nweke, 2019).

ΚΕΦΑΛΑΙΟ 4. Customer Segmentation & Market Basket Analysis

4.1 Τμηματοποίηση πελατών (Customer segmentation)

Η τμηματοποίηση πελατών είναι μια κρίσιμη αναλυτική διαδικασία στο μάρκετινγκ και την επιχειρηματική στρατηγική, η οποία επιτρέπει στους οργανισμούς να κατηγοριοποιούν την πελατειακή τους βάση σε διακριτές ομάδες με βάση κοινά χαρακτηριστικά ή συμπεριφορές. Η προσέγγιση αυτή επιτρέπει στις εταιρείες να προσαρμόζουν τις προσφορές, την επικοινωνία και Customer Segmentation & Market Basket Analysis

τη συνολική εμπειρία των πελατών τους ώστε να ανταποκρίνονται στις συγκεκριμένες ανάγκες κάθε τμήματος. Μια κοινή και αποτελεσματική μεθοδολογία για την τμηματοποίηση είναι η ανάλυση RFM, η οποία κατηγοριοποιεί τους πελάτες με βάση τρεις διαστάσεις: συχνότητα, συχνότητα και χρηματική αξία. Η συχνότητα μετρά το χρόνο που έχει παρέλθει από την τελευταία αγορά ενός πελάτη, η συχνότητα αξιολογεί πόσο συχνά ένας πελάτης ασχολείται με μια εταιρεία σε μια δεδομένη περίοδο και η χρηματική αξία αξιολογεί τις συνολικές δαπάνες ενός πελάτη. Αναλύοντας αυτές τις διαστάσεις, οι επιχειρήσεις μπορούν να εντοπίσουν πελάτες υψηλής αξίας, πελάτες που έχουν λήξει και πελάτες με προοπτικές ανάπτυξης. Για παράδειγμα, οι πελάτες που σημειώνουν υψηλή βαθμολογία και στις τρεις διαστάσεις μπορεί να είναι πρωταρχικοί στόχοι για προγράμματα αφοσίωσης, ενώ εκείνοι με υψηλή συχνότητα αλλά χαμηλή συχνότητα μπορεί να επωφεληθούν από εκστρατείες επαναπροσέγγισης (Christy, et al., 2021).

Εκτός από την ανάλυση RFM, οι αλγόριθμοι ομαδοποίησης διαδραματίζουν καθοριστικό ρόλο στην τμηματοποίηση πελατών, αξιοποιώντας τη δύναμη της μηχανικής μάθησης για την αποκάλυψη μοτίβων σε μεγάλα σύνολα δεδομένων. Μεταξύ των πιο ευρέως χρησιμοποιούμενων αλγορίθμων είναι ο αλγόριθμος K-means και η ιεραρχική ομαδοποίηση. Η ομαδοποίηση K-means χωρίζει ένα σύνολο δεδομένων σε έναν προκαθορισμένο αριθμό συστάδων ελαχιστοποιώντας τη διακύμανση εντός κάθε συστάδας. Αυτός ο αλγόριθμος είναι ιδιαίτερα αποδοτικός για μεγάλα σύνολα δεδομένων και χρησιμοποιείται συχνά για την τμηματοποίηση πελατών με βάση συμπεριφορικές και δημογραφικές μεταβλητές. Η ιεραρχική συσταδοποίηση, από την άλλη πλευρά, δημιουργεί μια δενδροειδή δομή συστάδων, επιτρέποντας στις επιχειρήσεις να διερευνήσουν διαφορετικά επίπεδα λεπτομέρειας στις προσπάθειες τμηματοποίησής τους. Σε αντίθεση με το K-means, η ιεραρχική συσταδοποίηση δεν απαιτεί προκαθορισμό του αριθμού των συστάδων, καθιστώντας την ευέλικτη επιλογή για διερευνητική ανάλυση. Αυτοί οι αλγόριθμοι επιπρέπουν στις επιχειρήσεις να τμηματοποιούν δυναμικά τους πελάτες και να προσαρμόζονται στις εξελισσόμενες συνθήκες της αγοράς, παρέχοντας μια βάση για εξατομικευμένες στρατηγικές μάρκετινγκ και κατανομή πόρων (Kansal et al., 2018).

Η πρακτική εφαρμογή της τμηματοποίησης πελατών είναι ιδιαίτερα εμφανής στο πεδίο των διαδικτυακών πλατφορμών, όπου παράγονται και συλλέγονται τεράστιες ποσότητες δεδομένων πελατών. Οι πλατφόρμες ηλεκτρονικού εμπορίου, για παράδειγμα, χρησιμοποιούν την τμηματοποίηση για να βελτιώσουν την εμπειρία του πελάτη και να αυξήσουν τις πωλήσεις. Αναλύοντας το ιστορικό περιήγησης, τα μοτίβα αγορών και τις δημογραφικές πληροφορίες, οι πλατφόρμες αυτές μπορούν να προτείνουν προϊόντα προσαρμοσμένα στις ατομικές προτιμήσεις, αυξάνοντας την πιθανότητα μετατροπής. Οι υπηρεσίες ροής χρησιμοποιούν παρόμοιες τεχνικές, χρησιμοποιώντας την τμηματοποίηση πελατών για να προτείνουν περιεχόμενο με βάση τις συνήθειες και τις προτιμήσεις προβολής. Αυτή η εξατομικευμένη προσέγγιση όχι μόνο βελτιώνει την ικανοποίηση των χρηστών, αλλά και προάγει τη μακροπρόθεσμη δέσμευση και αφοσίωση. Οι πλατφόρμες κοινωνικής δικτύωσης επωφελούνται επίσης από την τμηματοποίηση πελατών, προσαρμόζοντας τις διαφημίσεις και το περιεχόμενο σε συγκεκριμένες ομάδες χρηστών, μεγιστοποιώντας τη συνάφεια και την αποτελεσματικότητα των προσφορών τους (Ballestar et al., 2018).

Η τμηματοποίηση πελατών, μέσω τεχνικών όπως η ανάλυση RFM και οι αλγόριθμοι ομαδοποίησης, δίνει τη δυνατότητα στις επιχειρήσεις να κατανοήσουν και να αντιμετωπίσουν τις διαφορετικές ανάγκες της πελατειακής τους βάσης. Η εφαρμογή της σε διαδικτυακές πλατφόρμες καταδεικνύει τα απτά οφέλη αυτής της προσέγγισης, καθώς οι εξατομικευμένες εμπειρίες και οι στοχευμένες στρατηγικές έχουν καταστεί απαραίτητες στο σημερινό ανταγωνιστικό τοπίο. Καθώς τα δεδομένα συνεχίζουν να πολλαπλασιάζονται, η τελειοποίηση των τεχνικών τμηματοποίησης θα παραμείνει ακρογωνιαίος λίθος των πελατοκεντρικών επιχειρηματικών πρακτικών, οδηγώντας στην καινοτομία και την αξία σε όλους τους κλάδους (Doğan et al., 2018).

4.2 Ανάλυση Καλαθιού Αγοράς (Market Basket Analysis)

Η ανάλυση καλαθιού αγοράς είναι μια τεχνική εξόρυξης δεδομένων που χρησιμοποιείται για την αποκάλυψη συσχετίσεων μεταξύ στοιχείων σε σύνολα δεδομένων συναλλαγών, καθιστώντας την ένα θεμελιώδες εργαλείο για την κατανόηση της αγοραστικής συμπεριφοράς των καταναλωτών. Εντοπίζοντας μοτίβα συνύπαρξης μεταξύ προϊόντων, οι επιχειρήσεις μπορούν να αντλήσουν χρήσιμες πληροφορίες για τη βελτιστοποίηση της τοποθέτησης προϊόντων, των στρατηγικών τιμολόγησης και των δραστηριοτήτων προώθησης. Η τεχνική αυτή είναι ιδιαίτερα πολύτιμη στο λιανικό εμπόριο και το ηλεκτρονικό εμπόριο, όπου παράγονται καθημερινά τεράστιοι όγκοι δεδομένων συναλλαγών. Η σημασία της έγκειται στην ικανότητά της να αποκαλύπτει τις υποκείμενες σχέσεις μεταξύ ειδών, επιτρέποντας στις επιχειρήσεις να δημιουργούν συνδυασμένες προσφορές ή να προτείνουν συμπληρωματικά προϊόντα, αυξάνοντας έτσι τις πωλήσεις και βελτιώνοντας την εμπειρία του πελάτη (Ünvan, 2021).

Στην ανάλυση καλαθιού αγοράς χρησιμοποιούνται συνήθως διάφοροι αλγόριθμοι για τον εντοπισμό συχνών συνόλων στοιχείων και την εξαγωγή κανόνων συσχέτισης. Ο αλγόριθμος Apriori είναι μία από τις πιο ευρέως χρησιμοποιούμενες μεθόδους σε αυτό το πλαίσιο. Λειτουργεί με τον εντοπισμό συχνών μεμονωμένων στοιχείων στο σύνολο δεδομένων και στη συνέχεια την επέκτασή τους σε μεγαλύτερα σύνολα στοιχείων με βάση ένα κατώτατο όριο υποστήριξης. Ο αλγόριθμος χρησιμοποιεί μια συστηματική και επαναληπτική προσέγγιση για τη μείωση της υπολογιστικής πολυπλοκότητας, καθιστώντας τον αποτελεσματικό για μεγάλα σύνολα δεδομένων (Aldino et al., (Aldino, Pratiwi, Sintaro, & Putra, 2021)2021).

Ένας άλλος εξέχων αλγόριθμος είναι ο FP-Growth, που σημαίνει Frequent Pattern Growth. Σε αντίθεση με τον Apriori, ο FP-Growth εξαλείφει την ανάγκη δημιουργίας υποψηφίων συμπιέζοντας το σύνολο δεδομένων σε μια δενδρική δομή που ονομάζεται FP-tree. Αυτή η δομή επιτρέπει στον αλγόριθμο να εντοπίζει πιο αποτελεσματικά τα συχνά μοτίβα, ιδίως σε σενάρια όπου τα σύνολα δεδομένων είναι πυκνά. Ο Eclat, ή Equivalence Class Clustering and bottom-up Lattice Traversal, είναι μια άλλη εναλλακτική λύση που χρησιμοποιεί μια προσέγγιση αναζήτησης σε βάθος (depth-first search) για τη διερεύνηση του συνόλου δεδομένων. Εστιάζοντας στην κάθετη αναπαράσταση δεδομένων, το Eclat είναι αποτελεσματικό σε σενάρια όπου η οριζόντια αναπαράσταση καθίσταται υπολογιστικά δαπανηρή. Καθένας από αυτούς τους αλγορίθμους προσφέρει μοναδικά πλεονεκτήματα και η επιλογή τους εξαρτάται από τα συγκεκριμένα χαρακτηριστικά του συνόλου δεδομένων και τα επιθυμητά αποτελέσματα της ανάλυσης (Hossain et al., 2019; Kansal et al., 2023).

Οι πρακτικές εφαρμογές της ανάλυσης καλαθιού αγοράς παραδειγματίζονται στον τομέα των ηλεκτρονικών παραδόσεων, όπου η κατανόηση της αγοραστικής συμπεριφοράς των πελατών είναι ζωτικής σημασίας για τη βελτιστοποίηση των λειτουργιών και την αύξηση των εσόδων. Οι διαδικτυακές πλατφόρμες παράδοσης ειδών παντοπωλείου, για παράδειγμα, χρησιμοποιούν την ανάλυση καλαθιού αγοράς για να προτείνουν είδη που αγοράζονται συχνά μαζί, όπως ψωμί και βιούτυρο ή καφέ και κρέμα γάλακτος. Αυτές οι συστάσεις όχι μόνο βελτιώνουν την εμπειρία του χρήστη απλοποιώντας τη διαδικασία αγορών, αλλά και ενθαρρύνουν υψηλότερες τιμές καλαθιού μέσω στρατηγικών διασταύρουμενων πωλήσεων (Kurniawan et al., 2018).

Οι πλατφόρμες παράδοσης φαγητού χρησιμοποιούν παρόμοιες τεχνικές για να προτείνουν συμπληρωματικά είδη, όπως ποτά ή επιδόρπια, παράλληλα με τις παραγγελίες κύριων πιάτων. Αξιοποιώντας τις γνώσεις που προκύπτουν από τα δεδομένα συναλλαγών, οι πλατφόρμες αυτές μπορούν επίσης να βελτιστοποίησουν τη διαχείριση των αποθεμάτων, διασφαλίζοντας ότι τα συχνά συνδυασμένα είδη είναι επαρκώς εφοδιασμένα. Επιπλέον, η ανάλυση καλαθιού αγοράς βοηθά στο σχεδιασμό αποτελεσματικών εκστρατειών προώθησης, όπως η προσφορά εκπτώσεων σε συνδυασμένα είδη ή στοχευμένων προσφορών με βάση το ιστορικό αγορών.

Αυτές οι εφαρμογές υπογραμμίζουν την αξία της ανάλυσης καλαθιού αγοράς για την προώθηση της λειτουργικής αποδοτικότητας και την ικανοποίηση των πελατών στο ανταγωνιστικό τοπίο των ηλεκτρονικών παραδόσεων (Kumar et al., 2024).

4.3 Literature review

Οι Akshara & Jain (2024) διερεύνησαν τον ρόλο της περιγραφικής, διαγνωστικής, προβλεπτικής και προτρεπτικής αναλυτικής στη βελτιστοποίηση των πωλήσεων στο ηλεκτρονικό εμπόριο. Η μελέτη ανέδειξε ότι η περιγραφική αναλυτική βοηθά τις επιχειρήσεις να κατανοήσουν την προηγούμενη απόδοση μέσω της ανάλυσης τάσεων και προτύπων, ενώ η διαγνωστική αναλυτική εξετάζει τα αίτια πίσω από αυτές τις τάσεις, αποκαλύπτοντας δυνατά και αδύναμα σημεία. Η προβλεπτική αναλυτική, χρησιμοποιώντας στατιστικά μοντέλα και μηχανική μάθηση (ML), προβλέπει μελλοντικές προτιμήσεις πελατών, διακυμάνσεις στη ζήτηση και πιθανές ευκαιρίες πωλήσεων, επιτρέποντας στις επιχειρήσεις να λαμβάνουν προδραστικές αποφάσεις. Η προτρεπτική αναλυτική βασίζεται στις προβλεπτικές πληροφορίες και προτείνει εφαρμόσιμες στρατηγικές, όπως εξατομικευμένες προτάσεις προϊόντων, στοχευμένες καμπάνιες μάρκετινγκ και δυναμικές προσαρμογές τιμών. Η μελέτη κατέληξε στο συμπέρασμα ότι η ενσωμάτωση αυτών των τεσσάρων τύπων αναλυτικής επιτρέπει στις πλατφόρμες ηλεκτρονικού εμπορίου να βελτιώσουν την εμπειρία των πελατών, να βελτιστοποίησουν τις στρατηγικές μάρκετινγκ και τη διαχείριση αποθεμάτων, οδηγώντας τελικά σε αύξηση της ανάπτυξης και της κερδοφορίας. Οι συγγραφείς πρότειναν ότι μελλοντικές έρευνες θα πρέπει να αντιμετωπίσουν προκλήσεις όπως η προστασία δεδομένων, η ενοποίηση διαφορετικών πηγών δεδομένων και η διεύρυνση των αναλυτικών δυνατοτήτων. Επιπλέον, η αξιοποίηση τεχνητής νοημοσύνης (AI) και μηχανικής μάθησης (ML) μπορεί να βελτιώσει περαιτέρω την αποτελεσματικότητα της ανάλυσης δεδομένων στο ηλεκτρονικό εμπόριο.

Οι Omol et al. (2024) εξέτασαν την εφαρμογή του αλγορίθμου Apriori και της Ανάλυσης Market Basket (MBA) για την αποκάλυψη αγοραστικών προτύπων καταναλωτών σε ένα σουπερμάρκετ στην Κένυα. Η μελέτη χρησιμοποίησε βιβλιοθήκες του αλγορίθμου Apriori σε προγραμματιστική γλώσσα Python για την ανάλυση των συναλλακτικών δεδομένων, εντοπίζοντας 42 κανόνες συσχέτισης που ανέδειξαν σημαντικές σχέσεις προϊόντων και τάσεις ταυτόχρονων αγορών. Τα βασικά ευρήματα υπέδειξαν ότι η στρατηγική τοποθέτηση προϊόντων που αγοράζονται συχνά, μπορεί να ενισχύσει τη δημιουργία εσόδων. Για παράδειγμα, η συχνή συνύπαρξη κέικ ή γλυκισμάτων με καφέ ή τσάι υποδηλώνει ευκαιρίες στοχευμένων προωθήσεων και πωλήσεων βασισμένων στην ευκολία του καταναλωτή. Η έρευνα υπογράμμισε ότι η ανάλυση MBA προσφέρει εφαρμόσιμες πληροφορίες, επιτρέποντας την ανάπτυξη στρατηγικών διασταυρούμενων πωλήσεων (cross-selling) μέσω συνδυαστικών προσφορών και τοποθέτησης συμπληρωματικών προϊόντων στον ίδιο χώρο. Ανέδειξε επίσης τη σημασία της εποχικότητας στην αγοραστική συμπεριφορά, δίνοντας τη δυνατότητα στους λιανοπωλητές να προσαρμόζουν τις προσφορές τους ανάλογα με τις διακυμάνσεις της ζήτησης κατά τη διάρκεια της ημέρας. Τα ευρήματα κατέληξαν στο συμπέρασμα ότι η MBA με τη χρήση του αλγορίθμου Apriori μπορεί να βελτιστοποιήσει τις λειτουργίες των σουπερμάρκετ, να αυξήσει την ικανοποίηση των πελατών και να οδηγήσει σε αύξηση των εσόδων, εξασφαλίζοντας έτσι ένα ανταγωνιστικό πλεονέκτημα στον κλάδο του λιανεμπορίου.

Οι Hery & Widjaja (2024) πραγματοποίησαν συγκριτική ανάλυση των αλγορίθμων Apriori και FP-Growth για την Ανάλυση Market Basket (MBA), χρησιμοποιώντας 20.507 συναλλαγές από το αρτοποιείο The Bread Basket. Η μελέτη αξιολόγησε βασικούς δείκτες απόδοσης, αποκαλύπτοντας ότι ο αλγόριθμος Apriori είχε χρόνο εκτέλεσης 4,08 δευτερόλεπτα και κατανάλωση μνήμης 45,36 MiB, ενώ ο FP-Growth παρουσίασε ελαφρώς μεγαλύτερο χρόνο

εκτέλεσης (4,15 δευτερόλεπτα) αλλά πολύ χαμηλότερη κατανάλωση μνήμης (0,08 MiB). Η ανάλυση των κανόνων συσχέτισης έδειξε ότι ο FP-Growth είναι πιο αποδοτικός σε μνήμη χάρη στη δομή δεδομένων FP-tree, η οποία μειώνει την ανάγκη για επαναλαμβανόμενες σαρώσεις της βάσης δεδομένων, ενώ εντόπισε σημαντικούς κανόνες συσχέτισης. Τα ευρήματα μπορούν να συμβάλουν στη βελτιστοποίηση της τοποθέτησης προϊόντων, των στοχευμένων προωθήσεων και της διαχείρισης αποθεμάτων. Η μελέτη καταλήγει στο συμπέρασμα ότι ο FP-Growth είναι προτιμότερος για μεγάλα σύνολα δεδομένων, ενώ ο Apriori παραμένει χρήσιμος για αναλύσεις μικρότερης κλίμακας. Παράλληλα, η έρευνα πρότεινε περαιτέρω μελέτες που να εξετάζουν υβριδικές προσεγγίσεις, ανάλυση δεδομένων σε πραγματικό χρόνο και επέκταση της μελέτης σε διαφορετικούς τομείς λιανεμπορίου για τη βελτίωση των τεχνικών MBA. Παρά την αποτελεσματικότητά του, οι περιορισμοί της μελέτης έγκεινται σε ζητήματα που αφορούν την ποιότητα των δεδομένων και το γεγονός ότι η ανάλυση περιορίστηκε σε ένα μόνο αριθμό αριθμού.

Οι Hoque et al. (2024) παρουσίασαν την Πελατοκεντρική Ανάλυση Καλαθιού Αγοράς (Customer-Centric Market Basket Analysis, CC-MBA), μια προηγμένη προσέγγιση που ενσωματώνει την ανάλυση RFM με τον αλγόριθμο K-Means clustering για τη βελτίωση της εξόρυξης κανόνων συσχέτισης στην ανάλυση αγοραστικής συμπεριφοράς. Σε αντίθεση με τους παραδοσιακούς αλγορίθμους Apriori και FP-Growth, οι οποίοι παρουσιάζουν προβλήματα υπολογιστικής απόδοσης και παράγουν υπερβολικά αδύναμους κανόνες συσχέτισης σε χαμηλά όρια υποστήριξης, η CC-MBA δίνει προτεραιότητα σε πελάτες υψηλής αξίας, με αποτέλεσμα την εξαγωγή πιο σχετικών και χρήσιμων κανόνων. Η μελέτη έδειξε ότι η CC-MBA βελτίωνει την ποιότητα των κανόνων συσχέτισης, όπως καταδεικνύεται από υψηλότερες τιμές υποστήριξης (support), εμπιστοσύνης (confidence), ανύψωσης (lift) και πεποίθησης (conviction). Τα ευρήματα υποδηλώνουν ότι η CC-MBA μπορεί να βοηθήσει τις επιχειρήσεις να βελτιώσουν τη στρατηγική τοποθέτησης προϊόντων, να σχεδιάσουν στοχευμένες προωθητικές ενέργειες και να αναπτύξουν πελατοκεντρικές στρατηγικές μάρκετινγκ, οδηγώντας σε αυξημένη ικανοποίηση των καταναλωτών και αύξηση των εσόδων. Ταυτόχρονα, επιτρέπει στους λιανοπωλητές να εντοπίζουν πελάτες υψηλής αξίας, να αναγνωρίζουν κρυφές προτιμήσεις προϊόντων και να βελτιστοποιούν τις στρατηγικές προώθησης και συνδυασμού προϊόντων. Η μελέτη αναγνώρισε επίσης προκλήσεις στην εφαρμογή της CC-MBA, όπως η διαθεσιμότητα δεδομένων, το κόστος υποδομής και η τεχνική εξειδίκευση που απαιτείται για την ενσωμάτωσή της στα συστήματα CRM. Ως μελλοντική έρευνα, προτείνεται η διερεύνηση των πρακτικών επιπτώσεων της CC-MBA σε διάφορους κλάδους καθώς και η ανάπτυξη στρατηγικών για την αντιμετώπιση των προκλήσεων εφαρμογής της.

ΚΕΦΑΛΑΙΟ 5. Ο κλάδος online delivery στην Ελλάδα

5.1 Ιστορική αναδρομή

Ο κλάδος των ηλεκτρονικών παραδόσεων στην Ελλάδα έχει γνωρίσει σημαντική ανάπτυξη, λόγω της εμφάνισης πολλών βασικών παικτών και της αυξανόμενης υιοθέτησης των ψηφιακών πλατφορμών από τους καταναλωτές. Μεταξύ των σημαντικότερων εταιρειών του κλάδου είναι οι Efood, Wolt και Box, καθεμία από τις οποίες έχει διαδραματίσει σημαντικό ρόλο στη διαμόρφωση του τοπίου της αγοράς. Η Efood, που ιδρύθηκε το 2012, είναι η μεγαλύτερη και πιο αναγνωρίσιμη πλατφόρμα στην Ελλάδα, προσφέροντας ένα ευρύ φάσμα υπηρεσιών παράδοσης που περιλαμβάνουν γεύματα, είδη παντοπωλείου και άλλα είδη πρώτης ανάγκης για το νοικοκυρίο. Αξιοποιώντας το εκτεταμένο δίκτυο συνεργατών της και μια ισχυρή τεχνολογική υποδομή, η Efood έχει διατηρήσει ηγετική θέση, εξυπηρετώντας εκατομμύρια πελάτες σε εθνικό επίπεδο. Η Wolt, μια διεθνής υπηρεσία παράδοσης που εισήλθε στην ελληνική αγορά το 2019, έχει διακριθεί με

την εστίασή της στην κορυφαία ποιότητα υπηρεσιών και το φιλικό προς το χρήστη περιβάλλον εργασίας. Με έμφαση στους γρήγορους χρόνους παράδοσης και τις συνεργασίες της με εστιατόρια υψηλής ποιότητας, η Wolt απέκτησε γρήγορα μια πιστή πελατειακή βάση. Ομοίως, η Box, μια υπηρεσία που ξεκίνησε από την Cosmote, έχει αξιοποιήσει την ενσωμάτωσή της στο ευρύτερο οικοσύστημα της Cosmote, προσφέροντας απρόσκοπτες ψηφιακές εμπειρίες και αξιοποιώντας τις υφιστάμενες πελατειακές σχέσεις για να επεκτείνει την εμβέλειά της (Βασιλάκη, 2017).



**Εικόνα 5. Online delivery πλατφόρμες στην Ελλάδα (πηγή:
<https://www.thetotalbusiness.com/2022/06/22/poso-apechoun-oi-platformes-food-delivery-apo-ton-kalupsoun-oles-tis-paraggelies-stin-ellada/>)**

Η ανάπτυξη της αγοράς online delivery στην Ελλάδα υπογραμμίζεται από εντυπωσιακά στατιστικά στοιχεία ανάπτυξης, που αντανακλούν την αυξανόμενη εξάρτηση των καταναλωτών από αυτές τις πλατφόρμες. Τα τελευταία χρόνια, ο τομέας γνώρισε εκθετική ανάπτυξη, με σημαντική επιτάχυνση κατά τη διάρκεια της πανδημίας COVID-19. Καθώς τα μέτρα εγκλεισμού και τα πρότυπα κοινωνικής αποστασιοποίησης περιόρισαν τις παραδοσιακές επιλογές για φαγητό και ψώνια, οι υπηρεσίες ηλεκτρονικής παράδοσης έγιναν κρίσιμη σανίδα σωτηρίας τόσο για τους καταναλωτές όσο και για τις επιχειρήσεις. Σύμφωνα με εκθέσεις της αγοράς, ο τομέας της online παράδοσης τροφίμων στην Ελλάδα αυξήθηκε κατά περισσότερο από 50% το 2020, με σταθερά ανοδική πορεία τα επόμενα χρόνια. Η συνολική αξία της αγοράς εκτιμάται ότι θα ξεπεράσει τα 500 εκατ. ευρώ ετησίως, με συνεχή επέκταση λόγω των εξελισσόμενων προτιμήσεων των καταναλωτών και των εξελίξεων στην τεχνολογία (Skalkos et al., 2023).

Η εξάπλωση της χρήσης των smartphone και τα υψηλά ποσοστά διείσδυσης του διαδικτύου έχουν τροφοδοτήσει περαιτέρω την υιοθέτηση των υπηρεσιών ηλεκτρονικής παράδοσης στην Ελλάδα. Περίπου το 85% του πληθυσμού κατέχει smartphone και η πλειονότητα αυτών των χρηστών εμπλέκεται με ψηφιακές πλατφόρμες για τις καθημερινές τους ανάγκες. Αυτός ο ψηφιακός μετασχηματισμός επέτρεψε σε εταιρείες όπως η Efood, η Wolt και η Box να ενσωματώσουν εξελιγμένες τεχνολογίες, όπως παρακολούθηση σε πραγματικό χρόνο, συστάσεις με τεχνητή νοημοσύνη και ασφαλή συστήματα πληρωμών, βελτιώνοντας τη συνολική εμπειρία Customer Segmentation & Market Basket Analysis

του χρήστη. Επιπλέον, η άνοδος των συνδρομητικών μοντέλων και των προγραμμάτων επιβράβευσης έχει δώσει κίνητρα για επαναλαμβανόμενη χρήση, προωθώντας τη διαρκή αύξηση της δέσμευσης και της διατήρησης των πελατών (Βασιλάκη, 2017; Avagianou et al., 2024).

Η ιστορική εξέλιξη του κλάδου των ηλεκτρονικών παραδόσεων στην Ελλάδα αναδεικνύει τον σημαντικό αντίκτυπό του στην ευρύτερη οικονομία και τη συμπεριφορά των καταναλωτών. Προσαρμοζόμενες στις μεταβαλλόμενες δυναμικές της αγοράς και αξιοποιώντας τις τεχνολογικές καινοτομίες, οι εταιρίες αυτού του κλάδου όχι μόνο μεταμόρφωσαν τον τρόπο παράδοσης αγαθών και υπηρεσιών, αλλά και δημιούργησαν νέες ευκαιρίες τόσο για τις επιχειρήσεις όσο και για τους ιδιώτες. Καθώς ο κλάδος συνεχίζει να εξελίσσεται, παραμένει ένα κρίσιμο συστατικό της ψηφιακής οικονομίας της Ελλάδας, αντανακλώντας μια ευρύτερη παγκόσμια τάση προς την ευκολία και τη συνδεσιμότητα (Ganou et al., 2022).

5.2 Στρατηγικές και προκλήσεις

Ο κλάδος της ηλεκτρονικής παράδοσης στην Ελλάδα χαρακτηρίζεται από την ταχεία επέκτασή του και την υιοθέτηση καινοτόμων στρατηγικών που αποσκοπούν στην ενίσχυση της εξατομικευμένης εμπειρίας του πελάτη, αντιμετωπίζοντας παράλληλα τις προκλήσεις κόστους και λειτουργίας. Η εξατομίκευση έχει αναδειχθεί σε ακρογωνιαίο λίθο του ανταγωνιστικού πλεονεκτήματος, καθοδηγούμενη από την ενσωμάτωση προηγμένης ανάλυσης δεδομένων και τεχνητής νοημοσύνης. Αξιοποιώντας τα δεδομένα πελατών, οι εταιρίες μπορούν να προσαρμόζουν τις προσφορές τους στις ατομικές προτιμήσεις, εξασφαλίζοντας μια πιο ελκυστική και ικανοποιητική εμπειρία χρήστη. Πλατφόρμες όπως η Efood, η Wolt και η Box χρησιμοποιούν αλγόριθμους που αναλύουν το ιστορικό αγορών, τη συμπεριφορά περιήγησης και δημογραφικές πληροφορίες για να παρέχουν στοχευμένες συστάσεις και προωθητικές ενέργειες. Αυτή η προσέγγιση όχι μόνο ενισχύει την αφοσίωση των πελατών, αλλά οδηγεί επίσης σε υψηλότερα ποσοστά μετατροπής και αυξημένες τιμές παραγγελιών (Katsikas & Gritzalis, 2017).

Επιπλέον, η χρήση δυναμικών στρατηγικών τιμολόγησης και ενημερώσεων σε πραγματικό χρόνο διασφαλίζει ότι οι πελάτες λαμβάνουν έγκαιρες και σχετικές πληροφορίες, βελτιώνοντας τη συνολική αλληλεπίδρασή τους με την πλατφόρμα. Η εξατομίκευση επεκτείνεται πέρα από τις συστάσεις προϊόντων, περιλαμβάνοντας χαρακτηριστικά όπως οι προσαρμόσιμες επιλογές παράδοσης και η ανταποκρινόμενη υποστήριξη πελατών, τα οποία συμβάλλουν συλλογικά σε μια απρόσκοπτη και επικεντρωμένη στον χρήστη εμπειρία (Polydoropoulou et al., 2022).

Η διαχείριση του κόστους και των λειτουργιών αποτελεί έναν κρίσιμο τομέα εστίασης για τις πλατφόρμες ηλεκτρονικής παράδοσης που δραστηριοποιούνται στην Ελλάδα, όπου η διατήρηση της κερδοφορίας εν μέσω έντονου ανταγωνισμού αποτελεί μόνιμη πρόκληση. Οι εταιρίες πρέπει να περιηγηθούν στην πολυπλοκότητα της διαχείρισης των δικτύων υλικοτεχνικής υποδομής, της βελτιστοποίησης των διαδρομών παράδοσης και της εξισορρόπησης του εργατικού κόστους. Οι προηγμένες τεχνολογίες, συμπεριλαμβανομένης της μηχανικής μάθησης και της προγνωστικής ανάλυσης, διαδραματίζουν καθοριστικό ρόλο στον εξορθολογισμό των λειτουργιών. Για παράδειγμα, οι αλγόριθμοι βελτιστοποίησης διαδρομών ελαχιστοποιούν τους χρόνους παράδοσης και την κατανάλωση καυσίμων αναλύοντας σε πραγματικό χρόνο τα δεδομένα κυκλοφορίας και τις καιρικές συνθήκες. Τέτοια μέτρα όχι μόνο μειώνουν το λειτουργικό κόστος, αλλά ευθυγραμμίζουν επίσης με τους στόχους βιωσιμότητας, καλύπτοντας την αυξανόμενη ζήτηση των καταναλωτών για περιβαλλοντικά συνειδητές πρακτικές. Επιπλέον, η υιοθέτηση ευέλικτων μοντέλων εργατικού δυναμικού, συμπεριλαμβανομένων των συνεργασιών με ανεξάρτητους ταχυμεταφορείς, επιτρέπει στις εταιρίες να κλιμακώνουν τις λειτουργίες τους ανάλογα με τη διακύμανση της ζήτησης. Ωστόσο, η προσέγγιση αυτή εγείρει προκλήσεις που σχετίζονται με την ικανοποίηση και τη διατήρηση του εργατικού δυναμικού, καθιστώντας αναγκαία Customer Segmentation & Market Basket Analysis

την εφαρμογή δίκαιων δομών αποζημίωσης και μηχανισμών υποστήριξης (Samanta-Rounti et al., 2016; Fragkos et al., 2024).

Παρά τις εξελίξεις αυτές, ο κλάδος της διαδικτυακής παράδοσης στην Ελλάδα αντιμετωπίζει συνεχείς προκλήσεις όσον αφορά την εξισορρόπηση του κόστους των τεχνολογικών επενδύσεων με την ανάγκη προσφοράς ανταγωνιστικών τιμών. Η ενσωμάτωση λύσεων αιχμής, ενώ είναι απαραίτητη για την επιχειρησιακή αποτελεσματικότητα, απαιτεί σημαντικές οικονομικές δαπάνες που μπορεί να επιβαρύνουν τους μικρότερους παίκτες της αγοράς. Επιπλέον, το αυξανόμενο κόστος των πρώτων υλών και οι πληθωριστικές πιέσεις προσθέτουν περαιτέρω πολυπλοκότητα στις στρατηγικές διαχείρισης του κόστους. Η αντιμετώπιση αυτών των προκλήσεων απαιτεί μια λεπτή ισορροπία μεταξύ κανονομίας, αποδοτικότητας και ικανοποίησης των πελατών. Καθώς ο κλάδος συνεχίζει να εξελίσσεται, η ικανότητά του να προσαρμόζεται σε αυτές τις δυναμικές θα καθορίσει τη μακροπρόθεσμη βιωσιμότητά του και την επιτυχία του να ανταποκρίνεται στις απαιτήσεις της ελληνικής αγοράς (Fragkos et al., 2024).

ΚΕΦΑΛΑΙΟ 6. Εφαρμογή ανάλυσης σε πραγματικά δεδομένα

6.1 Περιγραφή της βάσης δεδομένων

Το dataset που προκύπτει από τα στοιχεία πωλήσεων της online πλατφόρμας delivery, αποτελείται από έναν πίνακα που περιλαμβάνει 574.855 γραμμές και 10 στήλες, οι οποίες παρουσιάζουν τα κάτωθι χαρακτηριστικά:

Όνομα στήλης	Αριθμός μη κενών τιμών	Μοναδικές τιμές	Τύπος μεταβλητής	Περιγραφή
order_id	574.855	95.480	object	Κωδικός παραγγελίας
customer_id	574.855	10.000	object	Κωδικός πελάτη
order_timestamp	574.855	95.140	object	Ημερομηνία και ώρα της παραγγελίας
product_name	574.720	4.043	object	Ονομασία του προϊόντος
product_category	574.850	375	object	Κατηγορία του προϊόντος
product_price	574.855	954	float64	Τιμή του προϊόντος
items_sold	574.855	33	int64	Αριθμός πωληθέντων τεμαχίων
geographical_region_of_user	574.855	14	object	Γεωγραφική περιοχή (πόλη) της παραγγελίας
user_last_order_date	574.855	152	object	Ημερομηνία της πιο πρόσφατης παραγγελίας του πελάτη

f0_	574.855	-	object	Στήλη που δημιουργείται αυτόματα κατά την εξαγωγή των δεδομένων (δεν χρησιμοποιείται στην ανάλυση)
-----	---------	---	--------	--

Πίνακας 1. Βάση δεδομένων πλατφόρμας

6.2 Προετοιμασία των δεδομένων

Κατά τη διερεύνηση των παρεχόμενων δεδομένων δεν παρατηρήθηκαν στρεβλές ή λανθασμένες καταχωρίσεις σε καμία εκ των στηλών των πίνακα. Το μοναδικό ζήτημα που έπρεπε να αντιμετωπιστεί ήταν οι κενές (null) τιμές που εμπεριέχονταν στις στήλες «product_name» και «product_category», όπως είδαμε στον Πίνακα 1 παραπάνω. Η μεθοδολογία που αποφασίστηκε να ακολουθηθεί για την απόδοση ονόματος (imputation), ήταν καταρχάς η προσπάθεια να εντοπιστεί το όνομα του ανώνυμου προϊόντος με βάση την καταχωρημένη τιμή και κατηγορία του, σε προϊόντα αντίστοιχης τιμής και κατηγορίας. Για παράδειγμα, αρχικά προέκυπτε μια κενή τιμή στην κατηγορία «Healthy Αλμυρά Snacks» με τιμή 1.6. Αναζητώντας άλλα προϊόντα που ανήκουν σε αυτήν την κατηγορία με τιμή 1.6, βρέθηκαν τα εξής δύο:

product_name	Αριθμός εγγραφών
Nestle Fitness Multigrain Chips Δημητριακών Θαλασσινό Αλάτι 75g	3
Nestle Fitness Multigrain Chips Δημητριακών Sweet Chili 75g	1

Πίνακας 2. Προϊόντα κατηγορίας «Healthy Αλμυρά Snacks» με τιμή 1,6

Εφόσον προέκυψαν περισσότερα του ενός προϊόντα, αποφασίστηκε η ανάθεση του ονόματος σε εκείνο με τις περισσότερες εμφανίσεις. Στην προκειμένη περίπτωση, του πρώτου.

Με την αντίστοιχη λογική αντικαταστάθηκαν όλες οι κενές τιμές στα ονόματα των προϊόντων, πλην εκείνων για τα οποία δεν βρέθηκαν κατά προσέγγιση καταχωρήσεις. Συνολικά απέμειναν 10 προϊόντα στα οποία δεν κατέστη δυνατό να ανατεθεί υφιστάμενο προϊόν και ως εκ τούτου μετονομάστηκαν σε 'NoName'. Αντίστοιχα, πέντε κατηγορίες προϊόντων που δεν ήταν δυνατόν να ταυτιστούν, μετονομάστηκαν σε 'NoCategory'. Η τελική λίστα των προϊόντων και κατηγοριών χωρίς όνομα, εμφανίζεται στον παρακάτω πίνακα:

product_name	product_category	Αριθμός εγγραφών
NoName	Make A Wish	2
NoName	Αρνί	2
NoName	Αυγά	18
NoName	Έτοιμες Σαλάτες	3
NoName	Κρέατα & Πουλερικά	1
NoName	Κρουασάν & Τσουρέκια	22
NoName	Προϊόντα Κρεάτων & Πουλερικών	9

NoName	Σοκολάτες	28
NoName	Χοιρινό	1
NoName	NoCategory	5

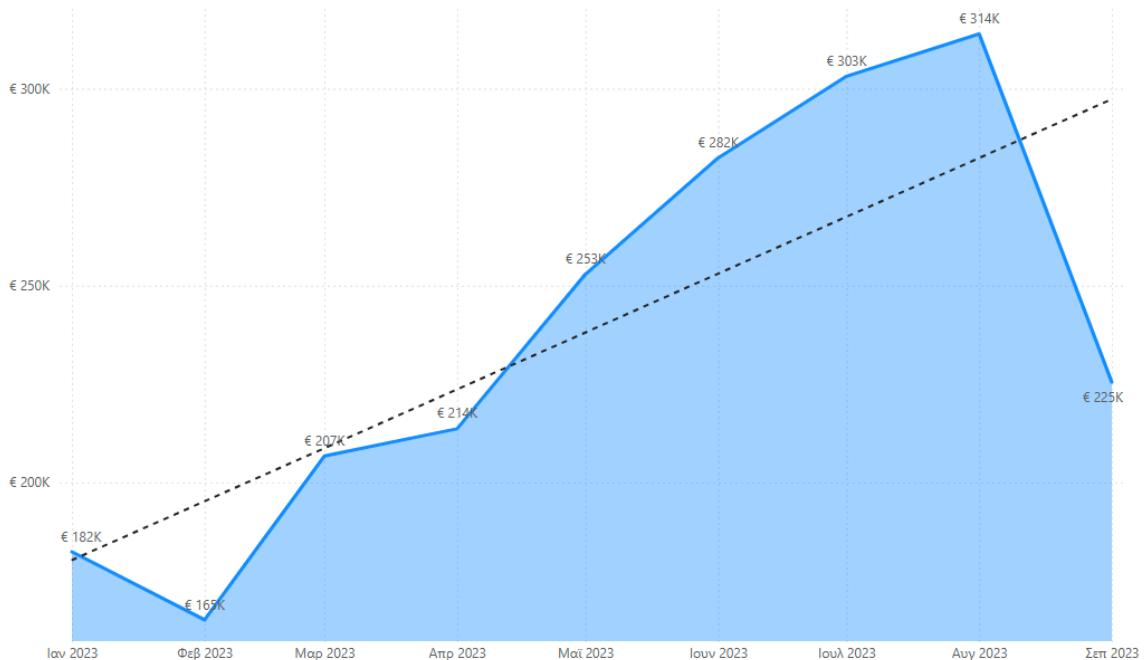
Πίνακας 3. Τελική λίστα προϊόντων και κατηγοριών χωρίς εξειδικευμένη ονομασία

Για την εκτέλεση της διαδικασίας του imputation χρησιμοποιήθηκε η υπολογιστή γλώσσα SQL. Κατά την ολοκλήρωσή της δημιουργήθηκε το νέο, ενημερωμένο dataset, το οποίο υπέστη την περαιτέρω επεξεργασία για τις ανάγκες της παρούσας ανάλυσης στη γλώσσα Python, με χρήση Jupyter Notebooks και μια σειρά από βιβλιοθήκες που αναφέρονται αναλυτικά στο Παράρτημα του κώδικα. Στην Python υλοποιήθηκαν κι όλες οι οπτικοποιήσεις που περιλαμβάνονται στην άσκηση, εκτός εκείνων που αφορούν τη διερευνητική ανάλυση δεδομένων κι έχουν δημιουργηθεί με τη βοήθεια εργαλείο οπτικοποιήσεων δεδομένων, Power BI.

6.3 Διερευνητική ανάλυση δεδομένων (Exploratory data analysis)

6.3.1 Ανάλυση πωλήσεων

Εφόσον ολοκληρώθηκε η προεπεξεργασία των δεδομένων, ακολουθεί η ανάλυση τους μέσω οπτικοποιήσεων προκειμένου να γίνει κατανοητή η δομή τους και να εξαχθούν χρήσιμες πληροφορίες για τις συναλλαγές. Η αρχή πραγματοποιείται με τα γενικά στοιχεία των πωλήσεων.

**Εικόνα 6. Συνολικός τζίρος ανά μήνα του έτους**

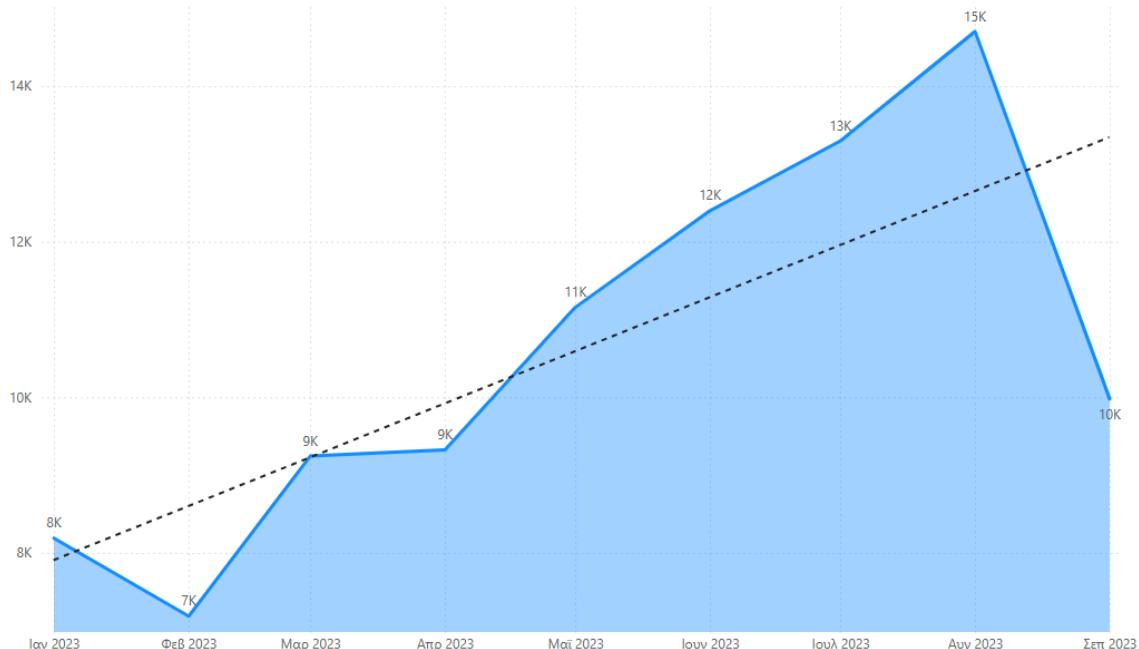
Η Εικόνα 1 παρουσιάζει τις μηνιαίες τάσεις τζίρου για την περίοδο από τον Ιανουάριο έως τον Σεπτέμβριο του 2023, υπογραμμίζοντας σημαντικές εποχικές διακυμάνσεις στα έσοδα που παράγει η εταιρεία. Παρατηρούμε ότι τα χαμηλότερα έσοδα παρατηρήθηκαν τον Φεβρουάριο, φτάνοντας τις €165 χιλ., γεγονός που πιθανόν αντικατοπτρίζει μια συνηθισμένη μείωση κατανάλωσης μετά την περίοδο των εορτών. Από τον Μάρτιο κι έπειτα, παρατηρείται μια σαφής ανοδική τάση, με τα έσοδα να αυξάνονται σταθερά κάθε μήνα. Η αύξηση αυτή κορυφώνεται τον Customer Segmentation & Market Basket Analysis

Αύγουστο, όταν η εταιρεία κατέγραψε τα υψηλότερα έσοδα ύψους €314 χιλ. Ακολουθεί μια σημαντική πτώση τον Σεπτέμβριο, με τα έσοδα να υποχωρούν στα €225 χιλ., καταγράφοντας μείωση της τάξης του 28% σε σύγκριση με τον προηγούμενο μήνα.

Η απόδοση των εσόδων κατά την εξεταζόμενη περίοδο υποδεικνύει ξεχωριστά εποχικά πρότυπα, με το πρώτο τρίμηνο (Ιανουάριος έως Μάρτιος) να χαρακτηρίζεται από χαμηλότερα επίπεδα πωλήσεων. Αντίθετα, το δεύτερο τρίμηνο και η αρχή του τρίτου τριμήνου (Απρίλιος έως Αύγουστος) παρουσιάζουν ισχυρή ανάπτυξη, με την αύξηση των εσόδων να αποδίδεται πιθανόν σε ενισχυμένη καταναλωτική δραστηριότητα κατά τους ανοιξιάτικους και καλοκαιρινούς μήνες. Αυτές οι περίοδοι ενδέχεται να συμπίπτουν με πρωθητικές ενέργειες, αυξημένη συχνότητα παραγγελιών ή υψηλότερη ζήτηση για εποχικά προϊόντα. Η εξαιρετική απόδοση του Αυγούστου πιθανόν να αντικατοπτρίζει τη θετική επίδραση της καταναλωτικής συμπεριφοράς που συνδέεται με το καλοκαίρι, όπως ο αυξημένος τουρισμός ή οι δαπάνες διακοπών.

Οστόσο, η σημαντική πτώση των εσόδων κατά τον Σεπτέμβριο απαιτεί περαιτέρω διερεύνηση. Η μείωση αυτή θα μπορούσε να αποδοθεί σε μειωμένη ζήτηση μετά την καλοκαιρινή κορύφωση ή στην απουσία στοχευμένων καμπανιών για τη διατήρηση της αφοσίωσης των πελατών κατά τη μετάβαση στους φθινοπωρινούς μήνες.

Η ανάλυση αυτών των τάσεων υπογραμμίζει αρκετές ευκαιρίες για βελτίωση. Για να αντιμετωπιστεί η στασιμότητα εσόδων κατά το πρώτο τρίμηνο, η εταιρεία θα μπορούσε να εξετάσει την εφαρμογή στοχευμένων πρωθητικών ενέργειών ή προγραμμάτων επιβράβευσης για να ενθαρρύνει τις αγορές στις αρχές του έτους. Παρομοίως, η απότομη μείωση τον Σεπτέμβριο υποδεικνύει την ανάγκη στρατηγικών παρεμβάσεων, όπως καμπάνιες για την έναρξη της σχολικής περιόδου ή πρωθήσεις προϊόντων προσαρμοσμένων στις ανάγκες του φθινοπώρου. Η κατανόηση των παραγόντων που οδήγησαν στην κορύφωση του Αυγούστου θα είναι επίσης απαραίτητη για την αναπαραγωγή της επιτυχίας της συγκεκριμένης περιόδου τα επόμενα χρόνια. Αυτό θα μπορούσε να περιλαμβάνει την ανάλυση της απόδοσης συγκεκριμένων κατηγοριών προϊόντων, την αξιολόγηση των πρωθητικών ενέργειών ή τη διερεύνηση των γεωγραφικών διαφορών στη συμπεριφορά των πελατών. Αξιοποιώντας αυτά τα ευρήματα, η εταιρεία μπορεί να βελτιώσει την απόδοση των εσόδων της καθ' όλη τη διάρκεια του έτους, εξασφαλίζοντας βιώσιμη ανάπτυξη και αυξημένη αφοσίωση πελατών.

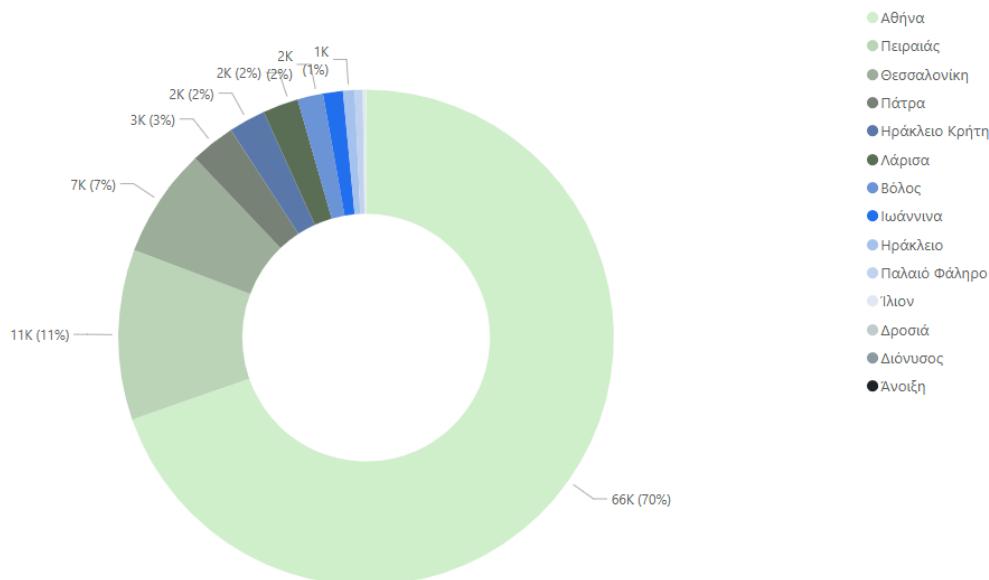


Εικόνα 7. Συνολικός αριθμός παραγγελιών ανά μήνα του έτους

Στην Εικόνα 2 παρουσιάζεται η μηνιαία τάση στον συνολικό αριθμό παραγγελιών για την ίδια περίοδο, καταγράφοντας τις εποχικές μεταβολές στη συχνότητα αγορών των πελατών. Ο Φεβρουάριος παρουσιάζει το χαμηλότερο επίπεδο παραγγελιών, με περίπου 7 χιλ. Στην αντίπερα όχθη, ο Αύγουστος αποτελεί την κορύφωση της περιόδου, με περίπου 15 χιλ. παραγγελίες, γεγονός που αντικατοπτρίζει την αυξημένη ζήτηση που συχνά συνοδεύει τις καλοκαιρινές διακοπές και τις σχετικές ανάγκες.

Η ανάλυση δείχνει μια σταθερή αύξηση στον αριθμό παραγγελιών από τον Μάρτιο και μετά, με ιδιαίτερη ένταση τους μήνες Μάιο έως Αύγουστο, προτού σημειωθεί μια απότομη πτώση περίπου 33% τον Σεπτέμβριο. Εξετάζοντας τις συσχετίσεις με τον Τζίρο, είναι φανερό ότι η αύξηση κατά τους καλοκαιρινούς μήνες οφείλεται κυρίως στη συχνότερη τοποθέτηση παραγγελιών. Η μικρή διαφοροποίηση στη μέση αξία παραγγελίας υποδηλώνει ότι οι αλλαγές στον όγκο των πωλήσεων διαδραματίζουν καθοριστικό ρόλο στη συνολική απόδοση της εταιρείας.

Για τη βελτίωση της απόδοσης κατά τις περιόδους χαμηλής δραστηριότητας, όπως το πρώτο τρίμηνο και ο Σεπτέμβριος, η εταιρεία θα μπορούσε να επικεντρωθεί σε στρατηγικές που ενισχύουν τη συχνότητα παραγγελιών. Ενδεικτικά, πρωθητικές ενέργειες με θεματικές πωλήσεις (π.χ. προσφορές επιστροφής στα σχολεία) θα μπορούσαν να τονώσουν τη ζήτηση και να ενισχύσουν την αφοσίωση των πελατών. Επιπλέον, η κατανόηση των παραγόντων που ενίσχυσαν τις πωλήσεις τον Αύγουστο θα προσφέρει χρήσιμες κατευθύνσεις για την αντιγραφή αυτής της επιτυχίας σε άλλες περιόδους.



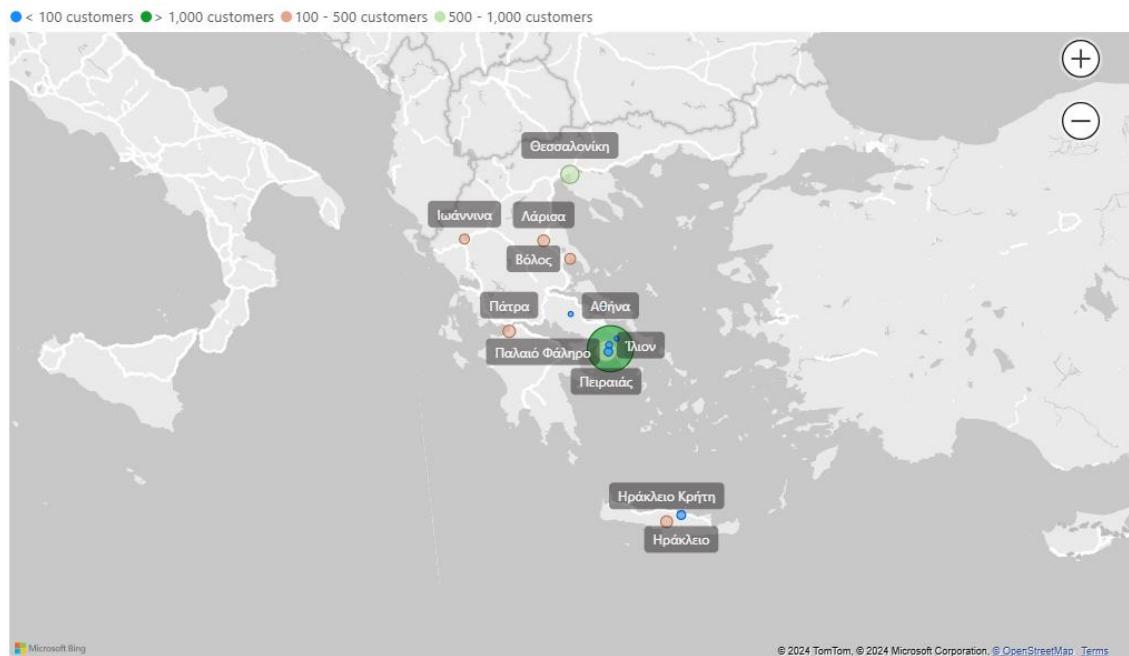
Εικόνα 8. Συνολικός αριθμός και ποσοστό παραγγελιών ανά περιοχή

Το ανωτέρω pie chart (Εικόνα 3) απεικονίζει την κατανομή του συνολικού αριθμού παραγγελιών ανά γεωγραφική περιοχή, υπογραμμίζοντας την έντονη συγκέντρωση δραστηριότητας σε αστικά κέντρα. Η Αθήνα κυριαρχεί, συνεισφέροντας το 70% του συνολικού όγκου παραγγελιών (66 χιλ.), ενώ ακολουθούν ο Πειραιάς με 11% και η Θεσσαλονίκη με 7%. Οι τρεις αυτές περιοχές συνδυαστικά αντιπροσωπεύουν το 88% των παραγγελιών, τονίζοντας την ισχυρή εξάρτηση της εταιρείας από πυκνοκατοικημένα αστικά κέντρα. Αντίθετα, περιοχές όπως η Πάτρα, το Ηράκλειο, η Λάρισα και ο Βόλος συνεισφέρουν από 2% έως 3% της συνολικής δραστηριότητας, ενώ οι υπόλοιπες περιοχές παρουσιάζουν αμελητέα ποσοστά.

Η κυριαρχία της Αθήνας, του Πειραιά και της Θεσσαλονίκης αντανακλά τη σημασία των μεγάλων μητροπολιτικών κέντρων, όπου η υψηλή πληθυσμιακή πυκνότητα και η καλή υποδομή παράδοσης λειτουργούν ως καταλύτες της καταναλωτικής ζήτησης. Από την άλλη, οι χαμηλότεροι

όγκοι παραγγελιών στις μικρότερες περιοχές υποδηλώνουν είτε περιορισμένη διεύσδυση της εταιρείας, είτε προκλήσεις που σχετίζονται με την τοπική αγορά, όπως τα κόστη διανομής ή οι διαφορετικές προτιμήσεις καταναλωτών.

Οι ευκαιρίες που αναδεικνύονται από αυτήν την ανάλυση είναι πολυδιάστατες. Καταρχάς, η εταιρεία μπορεί να ενισχύσει την παρουσία της στις ήδη ισχυρές αγορές, προσφέροντας υπηρεσίες υψηλής ποιότητας, όπως ταχύτερη παράδοση ή αποκλειστικές προσφορές για να ενισχύσει την πιστότητα των πελατών. Παράλληλα, είναι κρίσιμη η διερεύνηση των αιτίων που περιορίζουν τη δραστηριότητα σε μικρότερες περιοχές. Αυτό μπορεί να περιλαμβάνει ανάλυση εμποδίων, όπως προβλήματα στη διανομή, περιορισμένη αναγνωρισμότητα του brand ή προτιμήσεις προϊόντων που δεν ικανοποιούνται από την υπάρχουσα γκάμα. Η υπερβολική εξάρτηση από μερικές μεγάλες περιοχές ενέχει κινδύνους για τη σταθερότητα των εσόδων. Δυνητικές οικονομικές, ανταγωνιστικές ή επιχειρησιακές διαταραχές σε αυτές τις αγορές θα μπορούσαν να έχουν δυσανάλογο αντίκτυπο στις πωλήσεις. Η διαφοροποίηση μέσω της επέκτασης σε μικρότερες ή αναδυόμενες αγορές μπορεί να μειώσει αυτόν τον κίνδυνο, ενώ ταυτόχρονα θα ενισχύσει τη μακροχρόνια ανάπτυξη της εταιρείας.



Εικόνα 9. (Γράφημα χάρτη & φυσαλίδων-Map & Bubble chart) Συνολικός τζίρος και αριθμός πελατών ανά περιοχή

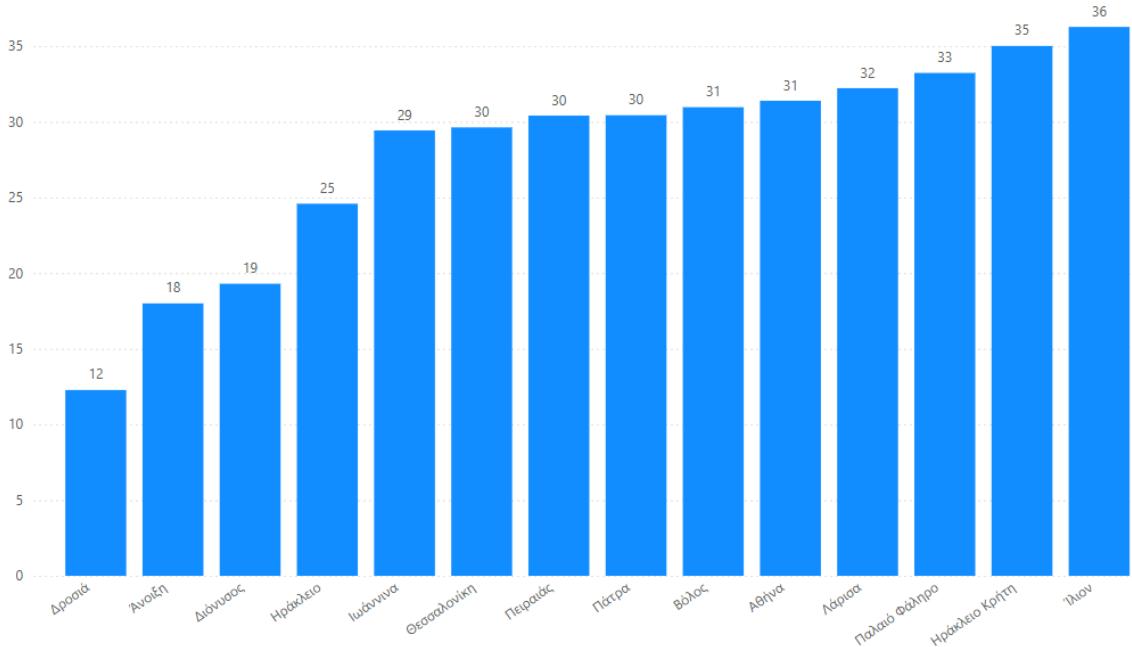
Το διάγραμμα φυσαλίδων προσφέρει μια γεωγραφική προοπτική στον συνολικό τζίρο και τη βάση πελατών της εταιρείας, επιτρέποντας μια συγκριτική ανάλυση της απόδοσης ανά περιοχή. Το μέγεθος κάθε φυσαλίδας αντιπροσωπεύει το μέγεθος του τζίρου, ενώ το χρώμα της καθορίζεται από τη συγκέντρωση των πελατών. Οι μεγαλύτερες φυσαλίδες να υποδεικνύουν περιοχές με υψηλότερη δραστηριότητα. Όπως αναμενόταν, η Αθήνα εμφανίζεται με τη μεγαλύτερη φυσαλίδα ως το κυρίαρχο κέντρο, επιβεβαιώνοντας την κεντρική σημασία της για τη συνολική απόδοση της εταιρείας. Ο Πειραιάς και η Θεσσαλονίκη ξεχωρίζουν επίσης, αν και οι φυσαλίδες τους είναι αισθητά μικρότερες, υποδεικνύοντας χαμηλότερη μεν, αλλά σημαντική συνεισφορά δε.

Περιοχές όπως η Πάτρα, το Ηράκλειο και η Λάρισα εμφανίζουν μέτρια μεγέθη φυσαλίδων, κατατάσσοντάς τις στη δευτερεύουσα κλίμακα σημαντικότητας. Στον αντίποδα, περιοχές όπως τα Ιωάννινα, ο Βόλος και μικρότερες πόλεις παρουσιάζουν ελάχιστη δραστηριότητα, με μικρές φυσαλίδες που αντικατοπτρίζουν τόσο χαμηλή βάση πελατών όσο και περιορισμένο τζίρο. Αυτή η γεωγραφική ανάλυση υπογραμμίζει δύο βασικές στρατηγικές κατευθύνσεις για την εταιρεία:

• **Ενίσχυση ισχυρών περιοχών:** Η κυριαρχία της Αθήνας προσφέρει μια εξαιρετική ευκαιρία για περαιτέρω εμβάθυνση της δραστηριότητας των πελατών μέσω βελτιωμένων υπηρεσιών, όπως ταχύτερες παραδόσεις ή εξατομικευμένες προσφορές. Παρόμοιες ενέργειες μπορούν να εφαρμοστούν σε άλλες αστικές περιοχές, όπως ο Πειραιάς και η Θεσσαλονίκη, για τη διατήρηση και την ενίσχυση της πελατειακής πίστης.

• **Διαφοροποίηση και επέκταση:** Η εξάρτηση από λίγες βασικές περιοχές αποτελεί στρατηγικό ρίσκο. Ανταγωνιστικά, οικονομικά ή λειτουργικά προβλήματα σε αυτές τις αγορές θα μπορούσαν να επηρεάσουν δυσανάλογα τα συνολικά έσοδα. Η στοχευμένη ανάπτυξη σε μικρότερες ή λιγότερο ανεπτυγμένες αγορές, όπως τα Ιωάννινα ή ο Βόλος, μπορεί να μετριάσει αυτούς τους κινδύνους, ενώ προσφέρει σημαντικές ευκαιρίες για την απόκτηση νέων πελατών.

Επιπλέον, περιοχές με μέτριες επιδόσεις, όπως η Πάτρα ή το Ηράκλειο, θα μπορούσαν να επωφεληθούν από στοχευμένες πρωθιθητικές ενέργειες ή βελτιώσεις στην εφοδιαστική αλυσίδα, αυξάνοντας το μερίδιό τους στην αγορά. Συνεργασίες με τοπικές επιχειρήσεις ή στοχευμένες καμπάνιες μπορούν να βελτιώσουν την επίγνωση του brand και να προσελκύσουν νέους πελάτες.



Εικόνα 10. Μέσος αριθμός ημερών μεταξύ παραγγελιών ανά περιοχή

Το ραβδόγραμμα (Εικόνα 5) απεικονίζει τον μέσο αριθμό ημερών μεταξύ των παραγγελιών σε διάφορες περιοχές. Χαμηλές τιμές υποδεικνύουν συχνότερες παραγγελίες, ενώ υψηλές τιμές αντικατοπτρίζουν λιγότερο τακτική αγοραστική συμπεριφορά.

Η περιοχή του Ιλίου αναδεικνύεται ως εκείνη με το μεγαλύτερο μέσο διάστημα μεταξύ παραγγελιών, στις 36 ημέρες, ακολουθούμενη από το Ηράκλειο Κρήτης με 35 ημέρες και το Παλαιό Φάληρο με 33 ημέρες. Οι περιοχές αυτές εμφανίζουν αραιότερα πρότυπα αγορών, υποδηλώνοντας ότι οι πελάτες πιθανώς βασίζονται λιγότερο στις υπηρεσίες της εταιρείας ή την αντιμετωπίζουν ως δευτερεύουσα επιλογή για τις αγορές τους. Αντίθετα, περιοχές όπως η Δροσιά (12 ημέρες), η Άνοιξη (18 ημέρες) και ο Διόνυσος (19 ημέρες) παρουσιάζουν πολύ μικρότερα διαστήματα, αναδεικνύοντας υψηλότερο επίπεδο αφοσίωσης πελατών και πιο συχνές αγορές.

Σημαντικό είναι ότι ο Πειραιάς, η Θεσσαλονίκη και η Πάτρα έχουν μέσο διάστημα παραγγελιών 30 ημερών, κατατάσσοντάς τους στη μεσαία κλίμακα συχνότητας παραγγελιών. Παρομοίως, η Αθήνα, αν και κύριος μοχλός εσόδων της εταιρείας, παρουσιάζει μέσο διάστημα 31 ημερών. Αν και οι περιοχές αυτές συμβάλλουν σημαντικά στην απόδοση της εταιρείας, υπάρχει περιθώριο περαιτέρω αύξησης της συχνότητας των παραγγελιών, με στόχο τη μεγιστοποίηση των εσόδων.

Οι περιοχές με μεγαλύτερα διαστήματα παραγγελιών, όπως το Ίλιον, το Ηράκλειο Κρήτης και το Παλαιό Φάληρο, αντιπροσωπεύουν σημαντικές ευκαιρίες ανάπτυξης. Η αντιμετώπιση πιθανών εμποδίων, όπως ζητήματα στις παραδόσεις ή στη διαθεσιμότητα προϊόντων, θα μπορούσε να ενθαρρύνει πιο συχνή δραστηριότητα πελατών. Στρατηγικές όπως στοχευμένες καμπάνιες, προγράμματα επιβράβευσης ή κίνητρα για επαναλαμβανόμενες αγορές μπορεί να είναι αποτελεσματικές στη μείωση του μέσου χρόνου μεταξύ παραγγελιών σε αυτές τις περιοχές.

Αντίθετα, οι περιοχές με συχνότερες παραγγελίες, όπως η Δροσιά, η Άνοιξη και ο Διόνυσος, χαρακτηρίζονται κυρίως από πελάτες με μεγαλύτερη οικονομική ευχέρεια, γεγονός που μπορεί να εξηγεί την τακτική αγοραστική συμπεριφορά. Η κατανόηση των παραγόντων που οδηγούν αυτή τη συμπεριφορά μπορεί να λειτουργήσει ως οδηγός για τη βελτίωση της αφοσίωσης των πελατών και σε άλλες περιοχές. Για παράδειγμα, η αναπαραγωγή της ποιότητας εξυπηρέτησης, της γκάμας προϊόντων ή των προωθητικών ενεργειών που χρησιμοποιούνται σε αυτές τις περιοχές θα μπορούσε να ενισχύσει την απόδοση και αλλού.



Εικόνα 11. Συνολικός τζίρος ανά ημέρα της εβδομάδας

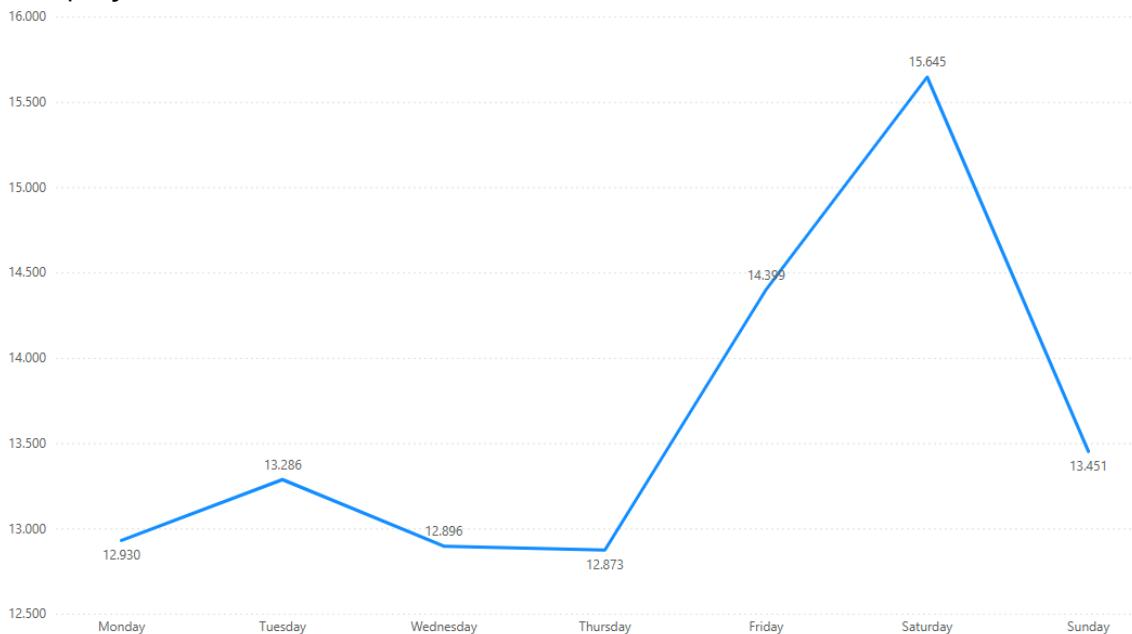
Το γραμμικό διάγραμμα (Εικόνα 6) απεικονίζει τον συνολικό τζίρο που παράγεται ανά ημέρα της εβδομάδας, καταγράφοντας την κατανομή των πωλήσεων κατά τη διάρκεια της εβδομάδας και εντοπίζοντας πρότυπα στη συμπεριφορά των πελατών. Ο οριζόντιος άξονας αναπαριστά τις ημέρες της εβδομάδας, ενώ ο κάθετος άξονας δείχνει τον συνολικό τζίρο σε ευρώ.

Τα δεδομένα αποκαλύπτουν ότι ο τζίρος ξεκινά σχετικά χαμηλά τη Δευτέρα, φτάνοντας τις €286 χιλ., και αυξάνεται ελαφρώς την Τρίτη στις €297 χιλ. Μια μικρή πτώση παρατηρείται στα μέσα της εβδομάδας, με τον τζίρο να μειώνεται στις €288 χιλ. την Τετάρτη, προτού ανακάμψει ελαφρώς στα €296 χιλ. την Πέμπτη. Ο υψηλότερος τζίρος καταγράφεται την Παρασκευή, αγγίζοντας τις €343 χιλ., σημειώνοντας μια σημαντική άνοδο στις πωλήσεις προς το τέλος της εργάσιμης εβδομάδας.

Ο τζίρος παραμένει ισχυρός το Σάββατο, στις €336 χιλ., προτού μειωθεί απότομα στα €299 χιλ. την Κυριακή.

Το πρότυπο αυτό αναδεικνύει μια σαφή αύξηση της αγοραστικής δραστηριότητας προς το τέλος της εβδομάδας, με την Παρασκευή και το Σάββατο να είναι οι πιο προσοδοφόρες ημέρες. Η υψηλή απόδοση των εν λόγω ημερών οφείλεται σε αυξημένες αγορές, που καλύπτουν τόσο βασικές ανάγκες όσο και προαιρετικές αγορές. Η πτώση της Κυριακής, παρότι εξακολουθεί να είναι ισχυρή συγκριτικά με τη Δευτέρα, υποδηλώνει μειωμένη αγοραστική δραστηριότητα κατά την ολοκλήρωση του Σαββατοκύριακου.

Αυτά τα ευρήματα υπογραμμίζουν τη σημασία της προσαρμογής των στρατηγικών μάρκετινγκ και των πρωθητικών ενεργειών σύμφωνα με τη συμπεριφορά των πελατών κατά τη διάρκεια της εβδομάδας. Για παράδειγμα, στοχευμένες καμπάνιες ή ειδικές προσφορές την Παρασκευή και το Σάββατο θα μπορούσαν να αξιοποιήσουν την αυξημένη δραστηριότητα των πελατών κατά την περίοδο αυτή. Παράλληλα, η εισαγωγή πρωθητικών ενεργειών στα μέσα της εβδομάδας, όπως την Τετάρτη, θα μπορούσε να αντιμετωπίσει την πτώση της δραστηριότητας και να ενισχύσει τις πωλήσεις.



Εικόνα 12. Συνολικός αριθμός παραγγελιών ανά ημέρα της εβδομάδας

Αντίστοιχα, το παραπάνω διάγραμμα (Εικόνα 7) παρουσιάζει τον αριθμό παραγγελιών που πραγματοποιήθηκαν σε κάθε ημέρα της εβδομάδας. Στον οριζόντιο άξονα περιλαμβάνονται οι ημέρες της εβδομάδας και στον κάθετο ο συνολικός αριθμός παραγγελιών.

Η εβδομάδα ξεκινά με 12.930 παραγγελίες τη Δευτέρα κι αυξάνεται ελαφρώς στις 13.286 την Τρίτη. Ακολουθεί πτώση στις 12.896 παραγγελίες την Τετάρτη, ενώ σημειώνεται περαιτέρω πτώση την Πέμπτη, με 12.873 παραγγελίες, οι οποίες αποτελούν το χαμηλότερο σημείο της εβδομάδας. Η δραστηριότητα αυξάνεται σημαντικά προς το τέλος της εβδομάδας, με 14.399 παραγγελίες την Παρασκευή και 15.645 παραγγελίες το Σάββατο. Την Κυριακή, ο αριθμός των παραγγελιών μειώνεται στις 13.451, παραμένοντας ωστόσο υψηλότερος από τις καθημερινές.

Το πρότυπο αυτό καταδεικνύει ότι οι πελάτες είναι πιο ενεργοί τα Σαββατοκύριακα, με ιδιαίτερη κορύφωση το Σάββατο, κάτι που επαληθεύει τη συνήθη αγοραστική συμπεριφορά για προετοιμασίες κοινωνικών δραστηριοτήτων, αναψυχής ή της επόμενης εβδομάδας.

Συγκρίνοντας το διάγραμμα με εκείνο του συνολικού τζίρου ανά ημέρα, αναδύεται μια ξεκάθαρη συσχέτιση μεταξύ του όγκου παραγγελιών και των εσόδων. Και τα δύο διαγράμματα

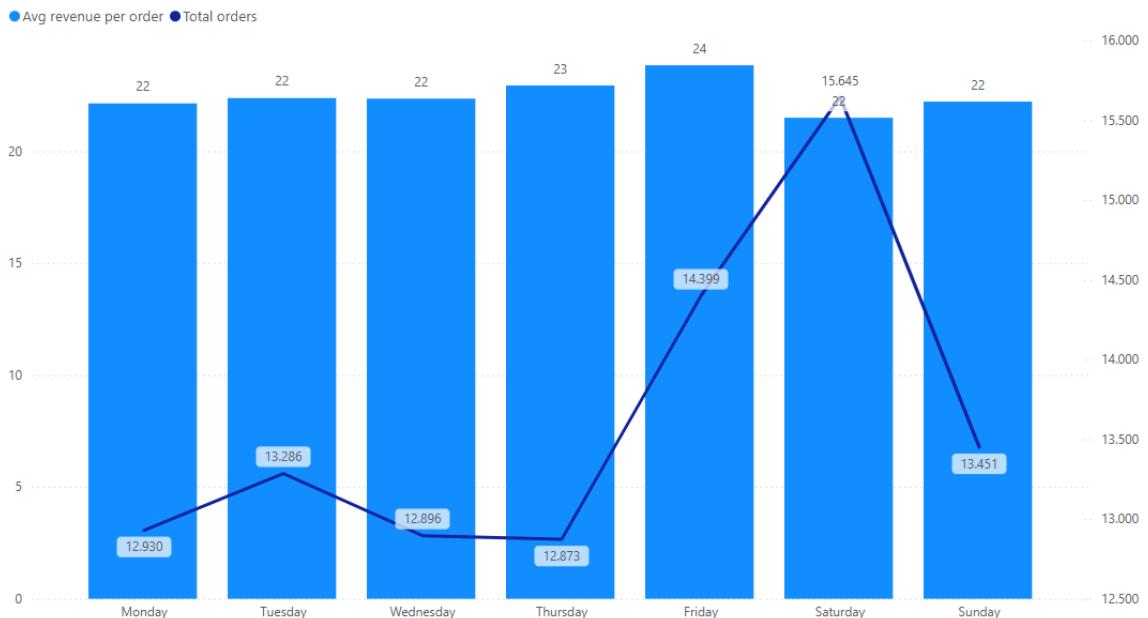
ακολουθούν παρόμοια πρότυπα, με κορυφώσεις την Παρασκευή και το Σάββατο και μειώσεις στα μέσα της εβδομάδας. Αυτό υποδεικνύει ότι η αύξηση των εσόδων τα Σαββατοκύριακα οφείλεται κυρίως στον μεγαλύτερο όγκο παραγγελιών, παρά σε αύξηση της μέσης αξίας παραγγελίας (AOV). Ωστόσο, διακρίνονται ενδιαφέρουσες διαφοροποιήσεις. Παρά τον υψηλότερο αριθμό παραγγελιών το Σάββατο σε σχέση με την Παρασκευή, ο συνολικός τζίρος την Παρασκευή είναι μεγαλύτερος. Αυτό υποδηλώνει ότι οι παραγγελίες της Παρασκευής είναι υψηλότερης αξίας κατά μέσο όρο.

Πιθανοί λόγοι για αυτό το φαινόμενο περιλαμβάνουν την προτεραιότητα των πελατών για πιο ακριβές αγορές την Παρασκευή (πχ. αλκοολούχα ποτά ή premium προϊόντα) που εξυπηρετούν κοινωνικές εκδηλώσεις του Σαββατοκύριακου. Στοχευμένες προσφορές ή καμπάνιες την Παρασκευή, που εστιάζουν σε προϊόντα υψηλότερης αξίας, μπορεί να ενθαρρύνουν τους πελάτες να ξοδέψουν ακόμα περισσότερα. Μια άλλη πιθανή εξήγηση είναι η παρουσία μαζικών ή εταιρικών παραγγελιών τις Παρασκευές, οι οποίες έχουν υψηλότερη αξία συναλλαγής συγκριτικά με τις καθημερινές αγορές.

Αντίθετα, ο υψηλός αριθμός παραγγελιών το Σάββατο φαίνεται να αντικατοπτρίζει μικρότερες, συνήθεις και μη αγορές. Αυτό έχει ως αποτέλεσμα χαμηλότερη μέση αξία παραγγελίας, παρότι η δραστηριότητα των πελατών είναι αυξημένη. Οι πελάτες ενδέχεται να ολοκληρώνουν εκκρεμότητες ή να αγοράζουν προϊόντα πρώτης ανάγκης που έχουν ζεχάσει, συμβάλλοντας στον αυξημένο αριθμό παραγγελιών, αλλά όχι απαραίτητα στον υψηλότερο τζίρο.

Αυτή η διαφοροποίηση υπογραμμίζει τη σημασία της κατανόησης τόσο του όγκου παραγγελιών όσο και της μέσης αξίας παραγγελίας για την ακριβή ερμηνεία των εσόδων. Η εταιρεία μπορεί να αξιοποιήσει αυτά τα ευρήματα, εφαρμόζοντας τις στρατηγικές που αυξάνουν το AOV της Παρασκευής σε άλλες ημέρες, όπως την Τετάρτη ή την Κυριακή, για να ενισχύσει τα έσοδα. Παρομοίως, η ενθάρρυνση μεγαλύτερων μεγεθών καλαθιού το Σάββατο μέσω προωθητικών ενεργειών ή πακέτων προσφορών θα μπορούσε να ευθυγραμμίσει τα έσοδα με τον υψηλό όγκο παραγγελιών.

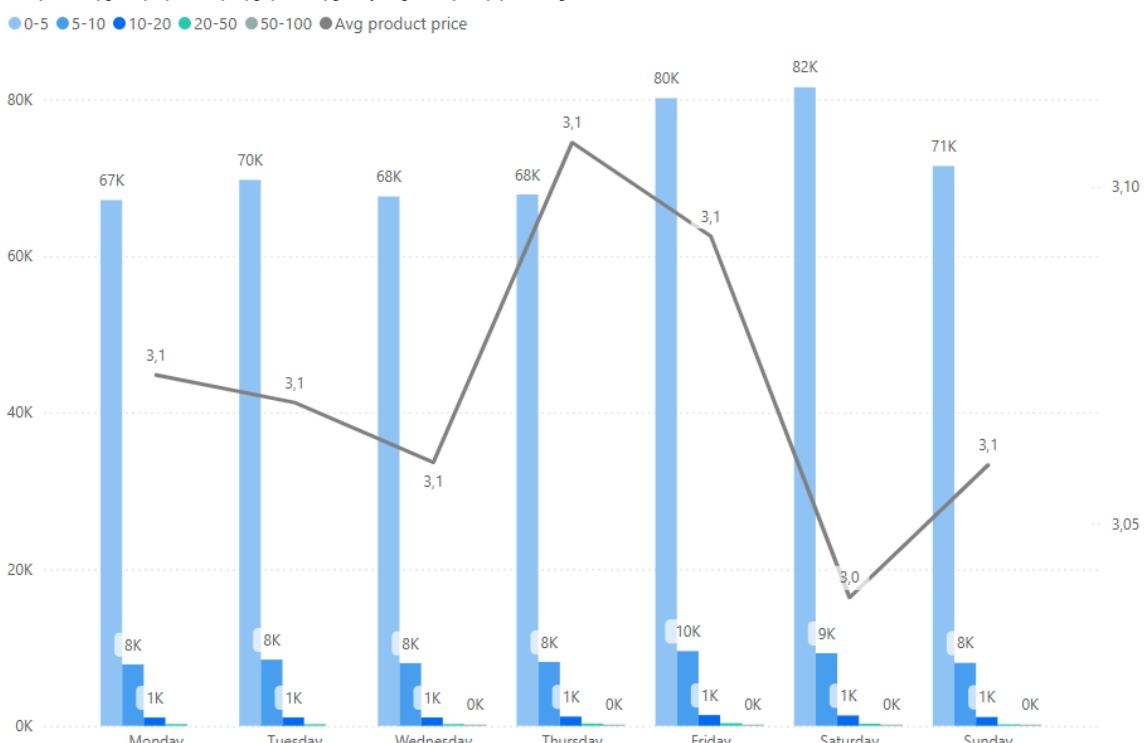
Επιπλέον, ενώ τόσο ο τζίρος όσο και ο όγκος παραγγελιών κορυφώνονται το Σάββατο, η πτώση των εσόδων την Κυριακή είναι πιο έντονη από την αντίστοιχη πτώση στον αριθμό παραγγελιών. Αυτό υποδεικνύει ότι η μέση αξία παραγγελίας την Κυριακή είναι χαμηλότερη, παρά την ισχυρή καταναλωτική δραστηριότητα. Αντίθετα, οι μέρες όπως η Πέμπτη εμφανίζουν χαμηλές τιμές και στους δύο δείκτες, καταδεικνύοντας μειωμένη δραστηριότητα των πελατών.



Εικόνα 13. Μέσος τζίρος ανά παραγγελία έναντι συνολικού αριθμού παραγγελιών ανά ημέρα της εβδομάδας

Στο σύνθετο διάγραμμα (Εικόνα 8) βλέπουμε συνδυαστικά τη σχέση μεταξύ του μέσου τζίρου ανά παραγγελία και του συνολικού αριθμού παραγγελιών ανά ημέρα της εβδομάδας, προσφέροντας μια διπλή προοπτική για τη συμπεριφορά των πελατών. Η γραμμή αντιπροσωπεύει τον συνολικό αριθμό παραγγελιών για κάθε ημέρα, ενώ οι ράβδοι υποδεικνύει τον μέσο τζίρο ανά παραγγελία σε ευρώ.

Τα δεδομένα δείχνουν έναν σταθερό μέσο τζίρο ανά παραγγελία ύψους περίπου €22 κατά τις περισσότερες ημέρες της εβδομάδας, με μικρές αυξήσεις να παρατηρούνται την Πέμπτη και την Παρασκευή, όπου φτάνει τα €23 και €24 αντίστοιχα. Παρά τον μεγαλύτερο αριθμό παραγγελιών το Σάββατο, η Παρασκευή καταγράφει τα υψηλότερα συνολικά έσοδα (€343 χιλ.), προφανώς λόγω της υψηλότερης μέσης αξίας παραγγελίας.



Εικόνα 14. Συνολικός αριθμός παραγγελιών με μέσο κόστος ανά παραγγελία, διαχωρισμένο σε ομάδες τιμών, ανά ημέρα της εβδομάδας

Το διάγραμμα της Εικόνας 9 παρέχει μια ολοκληρωμένη εικόνα του συνολικού αριθμού παραγγελιών και της μέσης τιμής ανά παραγγελία σε διαφορετικά εύρη τιμών για κάθε ημέρα της εβδομάδας. Οι ράβδοι αναπαριστούν τον συνολικό αριθμό παραγγελιών, κατηγοριοποιημένο σε εύρη τιμών (€0–5, €5–10, €10–20, €20–50, €50–100), ενώ η γραμμή δείχνει τη μέση τιμή ανά παραγγελία σε ευρώ.

Τα δεδομένα δείχνουν ότι η πλειονότητα των παραγγελιών ανήκει στο εύρος τιμών €0–5, το οποίο κυριαρχεί σταθερά σε όλες τις ημέρες της εβδομάδας. Για παράδειγμα, 67 χιλ. παραγγελίες τη Δευτέρα, 70 χιλ. την Τρίτη και 68 χιλ. την Τετάρτη προέρχονται κυρίως από αυτή την κατηγορία. Άλλα εύρη τιμών, όπως €5–10 και €10–20, συνεισφέρουν σημαντικά λιγότερο στον συνολικό όγκο παραγγελιών, ενώ τα υψηλότερα εύρη, όπως €20–50 και €50–100, έχουν αμελητέα συμμετοχή.

Η μέση τιμή ανά παραγγελία παραμένει σταθερή περίπου στα €3,1 τις περισσότερες ημέρες, με μικρές διακυμάνσεις. Μια ελαφρά πτώση παρατηρείται το Σάββατο, όπου η μέση τιμή μειώνεται στα €3, πριν ανακάμψει την Κυριακή στα €3,1. Αυτές οι τάσεις υποδηλώνουν ότι οι περισσότερες αγορές των πελατών αποτελούνται από προϊόντα χαμηλού κόστους, πιθανώς βασικά αγαθά ή προϊόντα χαμηλής αξίας που αγοράζονται συχνά.

Ενδιαφέρον παρουσιάζει το γεγονός ότι, ενώ το Σάββατο καταγράφει τον υψηλότερο συνολικό αριθμό παραγγελιών στο εύρος €0–5 (82 χιλ. παραγγελίες), η χαμηλότερη μέση τιμή υποδηλώνει ότι αυτές οι συναλλαγές αποτελούνται από μικρότερες, συνήθεις αγορές. Αντίθετα, η Παρασκευή, με 80 χιλ. συνολικές παραγγελίες, εμφανίζει ελαφρώς υψηλότερη μέση τιμή, αντανακλώντας πιθανώς αυξημένες συναλλαγές υψηλότερης αξίας σε σύγκριση με το Σάββατο. Αυτή η παρατήρηση ευθυγραμμίζεται με προηγούμενα ευρήματα, όπου η Παρασκευή ξεπέρασε το Σάββατο σε συνολικά έσοδα, παρά τον χαμηλότερο όγκο παραγγελιών.

Η μικρή συμμετοχή του υψηλότερου εύρους τιμών, ιδιαίτερα από €10–20 και άνω, υπογραμμίζει τον περιορισμένο βαθμό εμπλοκής των πελατών με premium ή υψηλής αξίας προϊόντα σε όλες τις ημέρες της εβδομάδας. Αυτή η τάση αναδεικνύει την εξάρτηση της εταιρείας από προϊόντα χαμηλού κόστους και υψηλού όγκου για την επίτευξη πωλήσεων και εσόδων.

Η εν λόγω ανάλυση τονίζει τη σημασία της κατανόησης της δυναμικής των ομάδων τιμών στη διαμόρφωση της αγοραστικής συμπεριφοράς των πελατών. Παρότι η κυριαρχία του εύρους €0–5 είναι κρίσιμη για την επιτυχία της εταιρείας, υπάρχουν ευκαιρίες για την προώθηση προϊόντων υψηλότερης αξίας, ιδιαίτερα σε ημέρες όπως η Παρασκευή, όπου οι πελάτες φαίνεται να είναι πιο διατεθειμένοι να ξοδέψουν περισσότερα. Η εισαγωγή στοχευμένων προωθητικών ενεργειών ή προσφορών σε πακέτα για τα υψηλότερα εύρη τιμών θα μπορούσε να βοηθήσει στην εξισορρόπηση του μείγματος προϊόντων και να αυξήσει τη μέση τιμή ανά παραγγελία.

	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total
Monday	253	462	572	690	719	754	750	744	738	773	1.042	1.256	1.424	1.209	889	507	148	12.930
Tuesday	265	474	553	664	689	745	698	732	771	869	1.164	1.286	1.416	1.259	990	540	171	13.286
Wednesday	252	457	597	629	718	759	736	687	674	854	1.058	1.229	1.426	1.250	872	528	170	12.896
Thursday	249	426	554	669	723	755	766	740	702	819	1.005	1.293	1.335	1.220	898	582	137	12.873
Friday	276	516	686	811	827	929	913	863	913	986	1.097	1.331	1.393	1.206	928	556	168	14.399
Saturday	228	524	784	1.018	1.050	1.114	1.025	941	1.011	1.051	1.265	1.425	1.452	1.203	842	546	166	15.645
Sunday	266	439	588	659	844	859	839	740	762	877	1.115	1.296	1.366	1.244	889	513	155	13.451
Total	1.789	3.298	4.334	5.140	5.570	5.915	5.727	5.447	5.571	6.229	7.746	9.116	9.812	8.591	6.308	3.772	1.115	95.480

Εικόνα 15. (Γράφημα θερμικού χάρτη—Heatmap chart) Αριθμός παραγγελιών ανά ώρα και ημέρα της εβδομάδας

Ο χάρτης θερμότητας (Εικόνα 10) παρέχει μια λεπτομερή εικόνα της κατανομής των παραγγελιών ανά ημέρα της εβδομάδας και ώρα της ημέρας, αναδεικνύοντας σαφή χρονικά πρότυπα στη δραστηριότητα των πελατών. Κάθε κελί αναπαριστά τον αριθμό παραγγελιών που τοποθετήθηκαν κατά τη διάρκεια μιας συγκεκριμένης ώρας και ημέρας, με πιο σκούρες αποχρώσεις να υποδεικνύουν μεγαλύτερους όγκους. Τα δεδομένα καλύπτουν τις ώρες από 08:00 έως 24:00 για όλες τις ημέρες της εβδομάδας.

Αναλύοντας τα ωριαία πρότυπα, η χρονική περίοδος 20:00–21:00 καταγράφει σταθερά τον υψηλότερο αριθμό παραγγελιών για όλες τις ημέρες της εβδομάδας. Για παράδειγμα, οι μέγιστοι όγκοι παραγγελιών για αυτή την ώρα είναι 1.424 παραγγελίες τη Δευτέρα, 1.416 την Τρίτη, 1.426 την Τετάρτη, 1.335 την Πέμπτη, 1.393 την Παρασκευή, 1.452 το Σάββατο και 1.366 την Κυριακή. Αυτοί οι αριθμοί τονίζουν την προτίμηση των πελατών για παραγγελίες τις βραδινές ώρες, πιθανότατα λόγω της ευκολίας που προσφέρει η ολοκλήρωση των καθημερινών δραστηριοτήτων. Η δραστηριότητα αρχίζει να αυξάνεται σημαντικά μετά τις 09:00, διατηρείται σε υψηλά επίπεδα καθ' όλη τη διάρκεια της ημέρας και αρχίζει να μειώνεται μετά τις 21:00, με τη λιγότερη δραστηριότητα να καταγράφεται τις πρώτες πρωινές και αργά το βράδυ ώρες. Αξιοσημείωτο

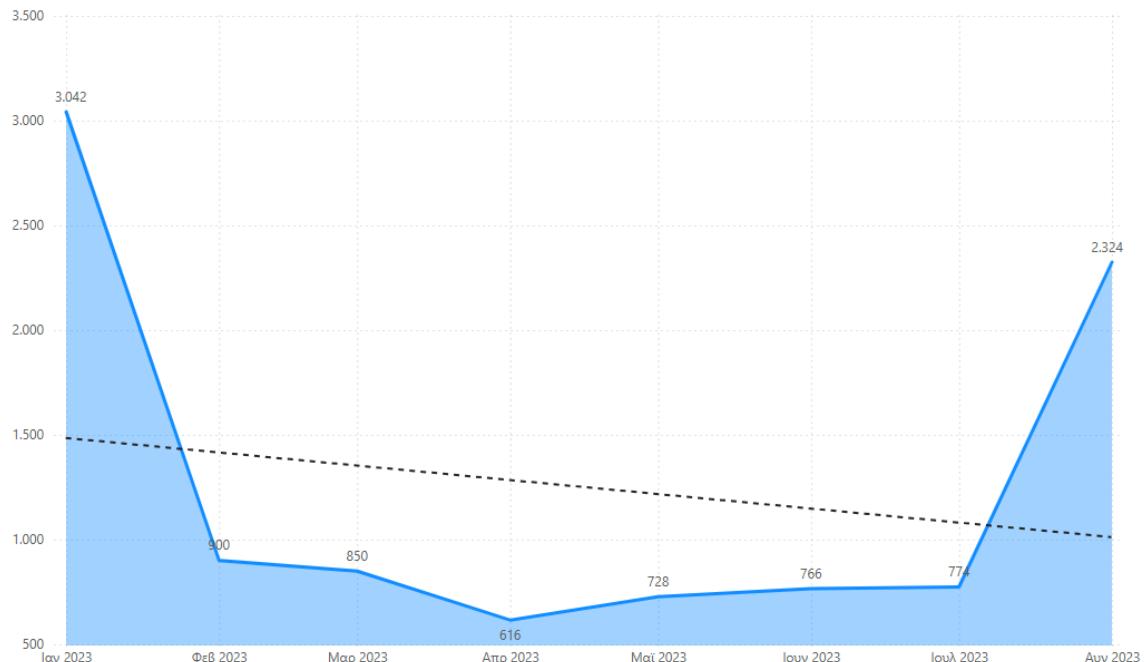
όμως, είναι το γεγονός ότι στη συγκεκριμένη χρονική περίοδο η Παρασκευή εμφανίζει λιγότερες παραγγελίες από τη Δευτέρα, την Τρίτη και την Τετάρτη, παρόλο τις ξεπερνά κατά πολύ στο γενικό σύνολο παραγγελιών.

Τα πρότυπα αυτά παρουσιάζουν αρκετές ευκαιρίες για βελτιστοποίηση λειτουργιών και στρατηγικών μάρκετινγκ. Για την αποτελεσματική διαχείριση της υψηλής ζήτησης κατά τις ώρες αιχμής, ιδιαίτερα από 20:00 έως 21:00, η εταιρεία θα μπορούσε να διαθέσει επιπλέον πόρους, όπως προσωπικό παράδοσης ή ομάδες υποστήριξης, για να διασφαλίσει την έγκαιρη εξυπηρέτηση. Επιπλέον, η αυξημένη δραστηριότητα το Σαββατοκύριακο απαιτεί ενισχυμένες εφοδιαστικές διαδικασίες για την κάλυψη των προσδοκιών των πελατών.

Για την αύξηση της εμπλοκής κατά τις πιο ήσυχες περιόδους, όπως οι ημέρες στα μέσα της εβδομάδας ή οι ώρες εκτός αιχμής, η εταιρεία θα μπορούσε να εφαρμόσει στοχευμένες προωθητικές ενέργειες, όπως για παράδειγμα, εκπτώσεις σε παραγγελίες που τοποθετούνται κατά τη διάρκεια αυτών των περιόδων. Αυτές οι πρωτοβουλίες θα μπορούσαν να ενθαρρύνουν μια πιο ισορροπημένη κατανομή των παραγγελιών και να μειώσουν την πίεση στις ώρες αιχμής.

6.3.2 Ανάλυση πελατολογίου

Σε συνέχεια της ανάλυσης πωλήσεων, ακολουθεί πιο εξειδικευμένη ανάλυση της δραστηριότητας των πελατών. Η κατανόησή της θα προσφέρει σημαντικά εφόδια στην εταιρεία προς την διατήρηση των πελατών με τη μεγαλύτερη συνεισφορά στα έσοδά της, ενώ παράλληλα θα βοηθήσει στην ενίσχυση της υφιστάμενης πελατειακής βάσης.



Εικόνα 16. Αριθμός απόκτησης νέων πελατών ανά μήνα του έτους

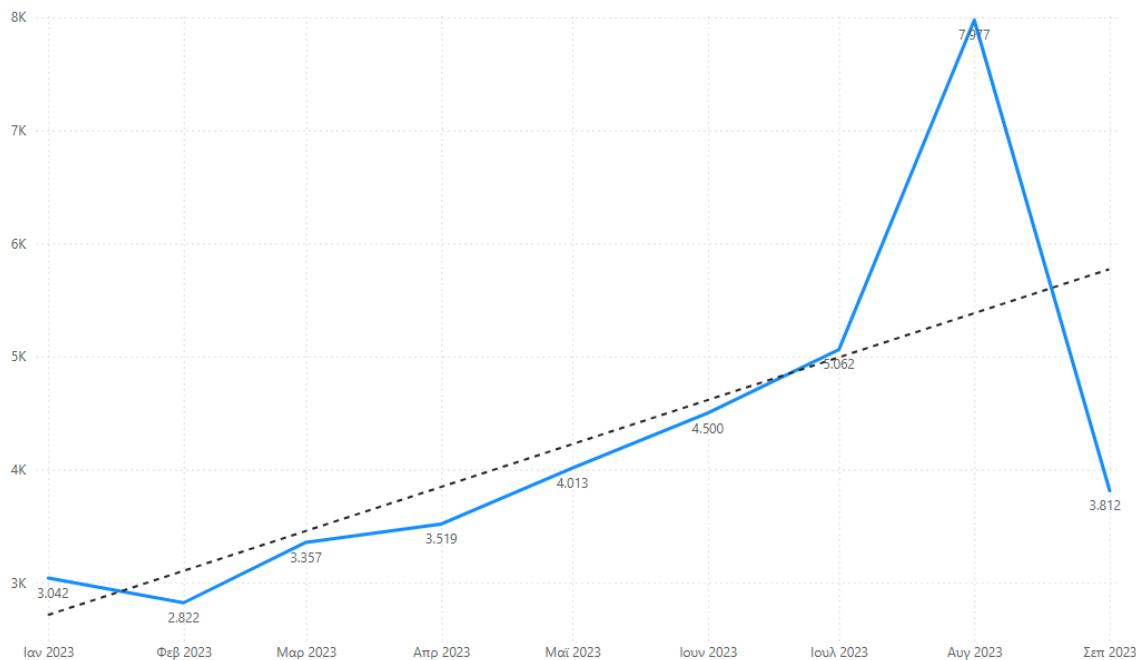
Στο γραμμικό διάγραμμα (Εικόνα 11) βλέπουμε την τάση αύξησης του πελατολογίου κατά τη διάρκεια του χρόνου, παρουσιάζοντας τον αριθμό των νέων πελατών που τοποθέτησαν παραγγελία κάθε μήνα από τον Ιανουάριο έως τον Αύγουστο 2023. Μια σημαντική κορύφωση στην απόκτηση νέων πελατών παρατηρείται τον Ιανουάριο, όπου καταγράφηκαν 3.042 νέοι πελάτες, πιθανότατα λόγω μιας επιτυχημένης προωθητικής καμπάνιας ή αυξημένου ενδιαφέροντος μετά την εορταστική περίοδο. Ωστόσο, αυτή η αρχική δυναμική δεν διατηρήθηκε,

καθώς ο αριθμός νέων πελατών μειώθηκε απότομα στις 900 τον Φεβρουάριο και συνέχισε να μειώνεται μέχρι τον Απρίλιο, φτάνοντας στο χαμηλότερο σημείο των 616 νέων πελατών.

Από τον Μάιο και μετά παρατηρείται μια σταδιακή ανάκαμψη στην ενσωμάτωση νέων πελατών, με τους αριθμούς να αυξάνονται αργά, φτάνοντας τους 766 τον Ιούνιο και τους 774 τον Ιούλιο. Αυτή η σταθερή ανάπτυξη κορυφώνεται με μια απότομη άνοδο τον Αύγουστο, όπου αποκτήθηκαν 2.324 νέοι πελάτες, αντανακλώντας μια άλλη επιτυχημένη προωθητική προσπάθεια ή εποχική ζήτηση. Παρόλα αυτά, η διακεκομμένη γραμμή τάσης στο διάγραμμα αποκαλύπτει μια γενική πτωτική πτορεία κατά την εξεταζόμενη περίοδο, υποδεικνύοντας προκλήσεις στη διατήρηση μιας σταθερής εισροής νέων πελατών.

Η απότομη αύξηση του Ιανουαρίου υποδηλώνει ότι οι στοχευμένες καμπάνιες ή οι προσπάθειες εμπλοκής πελατών μετά τις εορτές ήταν αποτελεσματικές στις αρχές του έτους. Ωστόσο, η επακόλουθη πτώση αναδεικνύει τη δυσκολία διατήρησης αυτού του επιπέδου ενδιαφέροντος, κάτι που μπορεί να οφείλεται σε μειωμένη δραστηριότητα μάρκετινγκ, κορεσμό της αγοράς ή αλλαγή στις προτεραιότητες των πελατών. Η σταδιακή ανάκαμψη από τον Μάιο και μετά, και ιδιαίτερα η κορύφωση του Αυγούστου, υποδεικνύει ότι οι ανανεωμένες προσπάθειες κατά τα μέσα και τέλη καλοκαιριού προσέλκυσαν περισσότερους πελάτες, πιθανώς μέσω καλοσχεδιασμένων στρατηγικών μάρκετινγκ ή εποχικών παραγόντων.

Από την έρευνα των δεδομένων υπογραμμίζεται η ανάγκη για μια συνεπή προσέγγιση στην απόκτηση νέων πελατών, προκειμένου να αποφευχθούν οι έντονες διακυμάνσεις που παρατηρήθηκαν στο πρώτο εξάμηνο του έτους. Η εστίαση στον εντοπισμό των παραγόντων πίσω από τις επιτυχίες του Ιανουαρίου και του Αυγούστου θα μπορούσε να προσφέρει πληροφορίες για αποτελεσματικές στρατηγικές που μπορούν να αναπαραχθούν σε άλλους μήνες για τη σταθεροποίηση των ρυθμών απόκτησης. Η αντιμετώπιση της ύφεσης στα μέσα του έτους, από τον Φεβρουάριο έως τον Απρίλιο, μέσω στοχευμένων προωθητικών ενεργειών ή κινήτρων, θα μπορούσε να ενισχύσει την εμπλοκή των πελατών κατά τις ιστορικά πιο ήσυχες περιόδους. Παράλληλα, η βελτίωση της διατήρησης πελατών μέσω προγραμμάτων αφοσίωσης ή εξατομικευμένων προσφορών θα μπορούσε να συμπληρώσει τις προσπάθειες απόκτησης, ενισχύοντας τη μακροπρόθεσμη ανάπτυξη και αφοσίωση.



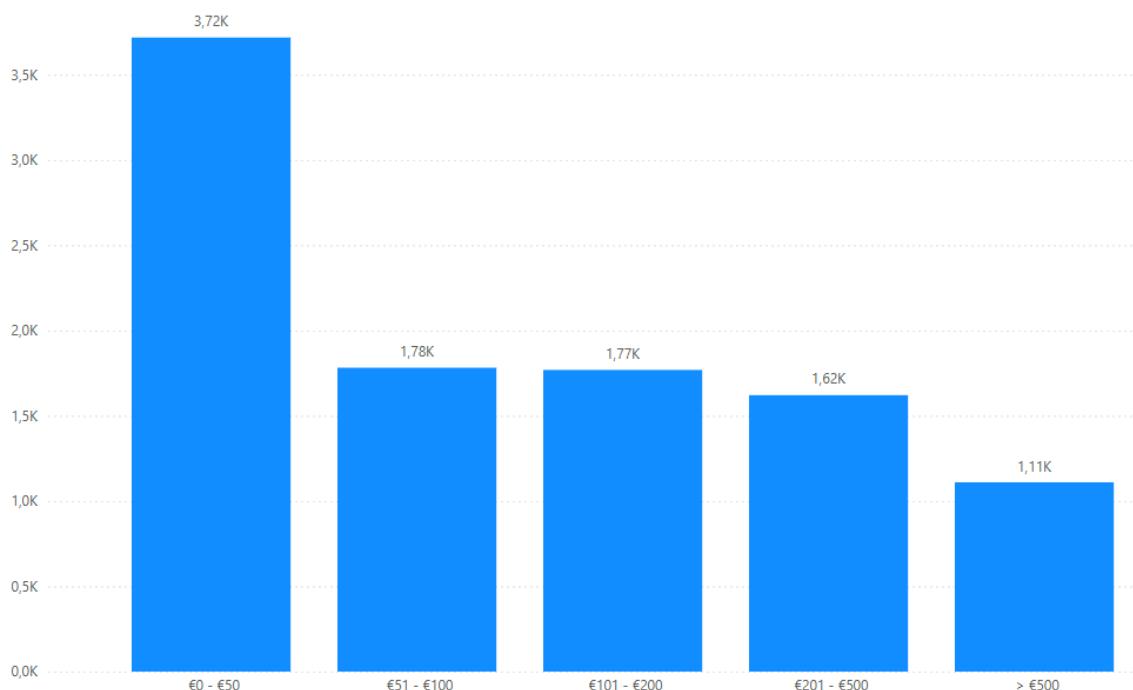
Εικόνα 17. Αριθμός μοναδικών ενεργών πελατών ανά μήνα του έτους

Το διάγραμμα της Εικόνας 12 αποτυπώνει τον αριθμό των διακριτών ενεργών πελατών ανά μήνα για την περίοδο από τον Ιανουάριο έως τον Σεπτέμβριο 2023. Ως ενεργοί πελάτες ορίζονται όσοι πραγματοποίησαν τουλάχιστον μία παραγγελία μέσα σε έναν συγκεκριμένο μήνα. Το διάγραμμα παρουσιάζει τόσο τις εποχιακές διακυμάνσεις όσο και τις μακροπρόθεσμες τάσεις, όπως αποτυπώνονται από τη συνεχή γραμμή που αντιπροσωπεύει τον αριθμό πελατών ανά μήνα και τη διακεκομένη γραμμή τάσης που δείχνει τη συνολική πρόοδο.

Τα δεδομένα ξεκινούν τον Ιανουάριο 2023 με 3.042 ενεργούς πελάτες. Ακολουθεί μια μικρή μείωση τον Φεβρουάριο, με 2.822 ενεργούς πελάτες, το χαμηλότερο σημείο της εξεταζόμενης περιόδου. Από τον Μάρτιο και μετά, ο αριθμός των ενεργών πελατών σημειώνει σταθερή ανοδική πορεία, αυξάνοντας στους 3.357 τον Μάρτιο, στους 3.519 τον Απρίλιο και στους 4.013 τον Μάιο. Η ανοδική τάση συνεχίζεται τον Ιούνιο και τον Ιούλιο, όπου ο αριθμός των ενεργών πελατών φτάνει τους 4.500 και τους 5.062 αντίστοιχα. Αυτή η σταθερή ανάπτυξη κορυφώνεται τον Αύγουστο, με 7.977 ενεργούς πελάτες, που αντιπροσωπεύουν το υψηλότερο επίπεδο δραστηριότητας κατά την αναλυόμενη περίοδο. Ωστόσο, ακολουθεί μια απότομη πτώση τον Σεπτέμβριο, όπου ο αριθμός των ενεργών πελατών μειώνεται στους 3.812, ένα επίπεδο συγκρίσιμο με αυτό του Απριλίου.

Η ανοδική τάση στους ενεργούς πελάτες κατά τους πρώτους οκτώ μήνες υποδηλώνει αυξημένη συμμετοχή πιθανώς λόγω ενός συνδυασμού προωθητικών καμπανιών, εποχιακής ζήτησης και επαναλαμβανόμενων αγορών από πιστούς πελάτες. Η κορύφωση του Αυγούστου πιθανόν να αντικατοπτρίζει την επιτυχία των στρατηγικών μάρκετινγκ ή την ενισχυμένη εποχιακή δραστηριότητα, όπως οι καλοκαιρινές αγορές ή η προετοιμασία για τις διακοπές. Αντίθετα, η απότομη πτώση τον Σεπτέμβριο δείχνει μια σημαντική μείωση της δραστηριότητας, που πιθανώς επηρεάζεται από το τέλος του καλοκαιριού και τη μετατόπιση των προτεραιοτήτων των πελατών καθώς εισέρχονται στο φθινόπωρο.

Η διακεκομένη γραμμή τάσης δείχνει μια σαφή συνολική ανοδική πορεία, υποδηλώνοντας τη βιώσιμη ανάπτυξη της εταιρείας εφόσον επιτυγχάνει να εμπλέκει πελάτες με την πάροδο του χρόνου. Αυτή η μακροπρόθεσμη τάση υποδεικνύει ότι οι στρατηγικές της εταιρείας για τη διατήρηση και την εμπλοκή της πελατειακής της βάσης αποφέρουν θετικά αποτελέσματα, παρά τις προσωρινές πτώσεις ή διακυμάνσεις που παρατηρούνται σε μεμονωμένους μήνες.



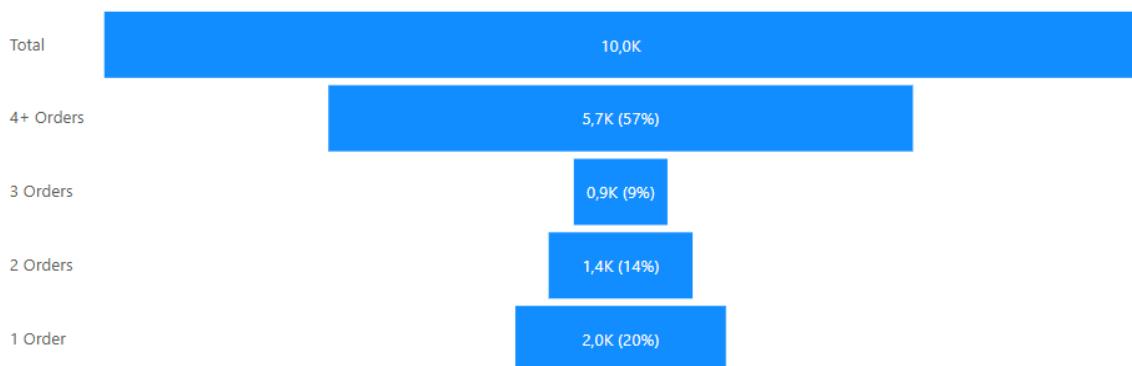
Εικόνα 18. Αριθμός μοναδικών πελατών βάσει εύρους τζίρου

Στο ραβδόγραμμα της Εικόνας 13 παρατηρούμε την κατανομή των διακριτών πελατών σε διαφορετικά εύρη δαπάνης, παρέχοντας πληροφορίες για τη συμπεριφορά των πελατών και τη συνολική συνεισφορά τους στα έσοδα της εταιρείας. Τα εύρη δαπάνης κατηγοριοποιούνται σε πέντε ομάδες: €0–€50, €51–€100, €101–€200, €201–€500 και άνω των €500. Κάθε ράβδος αντιπροσωπεύει τον αριθμό των πελατών, των οποίων η συνολική δαπάνη εμπίπτει σε ένα συγκεκριμένο εύρος.

Τα δεδομένα δείχνουν ότι η πλειονότητα των πελατών, συνολικά 3.720, ανήκει στο χαμηλότερο εύρος δαπάνης των €0–€50. Αυτή η ομάδα αποτελεί το μεγαλύτερο τμήμα, υποδηλώνοντας ότι ένα σημαντικό ποσοστό της πελατειακής βάσης αλληλεπιδρά με την εταιρεία ελάχιστα, είτε μέσω σπάνιων αγορών είτε με μικρού μεγέθους παραγγελίες. Η δεύτερη μεγαλύτερη ομάδα, με 1.780 πελάτες, αντιστοιχεί στο εύρος δαπάνης €51–€100, ενώ ακολουθεί στενά η ομάδα των 1.770 πελατών στο εύρος €101–€200. Οι δύο επόμενες ομάδες, €201–€500 και άνω των €500, περιλαμβάνουν 1.620 και 1.110 πελάτες, αντίστοιχα, εκπροσωπώντας πελάτες με υψηλότερες δαπάνες που πιθανότατα έχουν μεγαλύτερη εμπλοκή με τα προϊόντα και τις υπηρεσίες της εταιρείας.

Αυτή η κατανομή αναδεικνύει την παρουσία ενός έντονα ασύμμετρου προτύπου δαπάνης, όπου ένα μεγάλο ποσοστό πελατών συνεισφέρει σχετικά λίγο στα έσοδα της εταιρείας, ενώ μια μικρότερη αναλογία πελατών υψηλής αξίας πραγματοποιεί σημαντικές οικονομικές συνεισφορές. Οι μικρότερες ομάδες πελατών με υψηλές δαπάνες, ιδιαίτερα εκείνες στα εύρη €201–€500 και άνω των €500, αντιπροσωπεύουν ένα σημαντικό τμήμα για την εταιρεία, δεδομένου του δυναμικού τους να ενισχύουν την κερδοφορία και την αξία διάρκειας ζωής των πελατών.

Η κυριαρχία των πελατών χαμηλών δαπανών υποδηλώνει ευκαιρίες για την αύξηση της εμπλοκής σε αυτό το τμήμα. Η ενθάρρυνση μεγαλύτερων δαπανών μέσω εξατομικευμένων προωθητικών ενεργειών, προσωποποιημένων προτάσεων προϊόντων ή πακέτων προσφορών θα μπορούσε να βοηθήσει στη μετακίνηση κάποιων πελατών σε υψηλότερα εύρη δαπανών. Αντίστροφα, η διατήρηση και καλλιέργεια της πελατειακής βάσης υψηλής αξίας είναι εξίσου κρίσιμη. Προγράμματα αφοσίωσης, premium προϊόντα και αποκλειστικά κίνητρα μπορούν να χρησιμοποιηθούν για την ενίσχυση της σχέσης με αυτούς τους πελάτες και την ενθάρρυνση της συνεχούς εμπλοκής τους.



Εικόνα 19. (Γράφημα χωνιού-Funnel chart) Άθροισμα αριθμού πελατών ανά κατηγορία συνόλου παραγγελιών

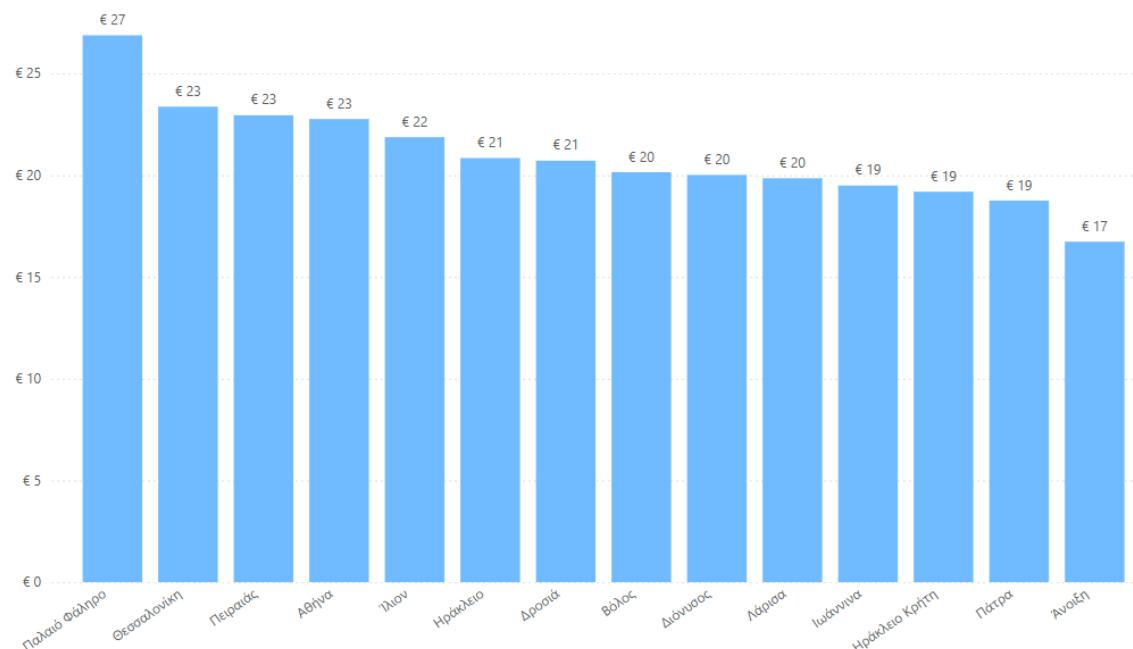
Το funnel chart (Εικόνα 14) εμφανίζει την κατανομή των πελατών βάσει του αριθμού των παραγγελιών που έχουν πραγματοποιήσει, προσφέροντας μια σαφή εικόνα για τα επίπεδα εμπλοκής τους. Ο συνολικός αριθμός πελατών ανέρχεται στους 10 χιλ., και αυτή η ομάδα

κατηγοριοποιείται σε τέσσερις διαβαθμίσεις: πελάτες που έχουν πραγματοποιήσει μία μόνο παραγγελία, δύο παραγγελίες, τρεις παραγγελίες, και τέσσερις ή περισσότερες παραγγελίες.

Το μεγαλύτερο τμήμα, που περιλαμβάνει 5,7 χιλ. πελάτες (57%), αποτελείται από εκείνους που έχουν πραγματοποιήσει τέσσερις ή περισσότερες παραγγελίες. Αυτό υποδεικνύει μια ισχυρή βάση από ιδιαίτερα εμπλεκόμενους και επαναλαμβανόμενους πελάτες, οι οποίοι πιθανότατα αποτελούν τη ραχοκοκαλία της σταθερής ροής εσόδων της εταιρείας. Η δεύτερη μεγαλύτερη ομάδα περιλαμβάνει 2 χιλ. πελάτες (20%) που έχουν πραγματοποιήσει μόνο μία παραγγελία, αντιπροσωπεύοντας ένα σημαντικό ποσοστό της πελατειακής βάσης που δεν έχει ακόμη επιδείξει συμπεριφορά επαναλαμβανόμενων αγορών. Οι υπόλοιπες ομάδες περιλαμβάνουν 1,4 χιλ. πελάτες (14%) με δύο παραγγελίες και 900 πελάτες (9%) με τρεις παραγγελίες, καταδεικνύοντας διάφορους βαθμούς επαναλαμβανόμενης δραστηριότητας.

Αυτή η κατανομή αναδεικνύει μια θετική τάση, καθώς η πλειονότητα των πελατών ανήκει στην κατηγορία υψηλής εμπλοκής με τέσσερις ή περισσότερες παραγγελίες. Ωστόσο, το σημαντικό μέγεθος της ομάδας της μίας παραγγελίας υποδηλώνει ευκαιρίες για την αύξηση της εμπλοκής και τη μετατροπή των πρώτων αγοραστών σε επαναλαμβανόμενους πελάτες. Η μετατροπή ακόμη και ενός μικρού ποσοστού αυτών των πελατών σε επαναλαμβανόμενους αγοραστές θα μπορούσε να ενισχύσει σημαντικά τα συνολικά έσοδα και την αξία διάρκειας ζωής των πελατών (CLV).

Το υψηλό ποσοστό των πελατών στην κατηγορία των τεσσάρων ή περισσότερων παραγγελιών υποδηλώνει ότι η εταιρεία επιτυγχάνει στη διατήρηση ενός πυρήνα πιστών πελατών. Στρατηγικές για τη διατήρηση αυτής της πίστης, όπως εξατομικευμένες ανταμοιβές, αποκλειστικές προσφορές ή στοχευμένη επικοινωνία, θα μπορούσαν να εξασφαλίσουν τη συνεχιζόμενη εμπλοκή αυτής της κρίσιμης ομάδας. Παράλληλα, προσπάθειες για την ενεργοποίηση των πελατών της μίας παραγγελίας θα μπορούσαν να περιλαμβάνουν καμπάνιες παρακολούθησης μάρκετινγκ, εκπτώσεις στη δεύτερη παραγγελία ή κίνητρα συνδρομής για την ενθάρρυνση επαναλαμβανόμενων αγορών.



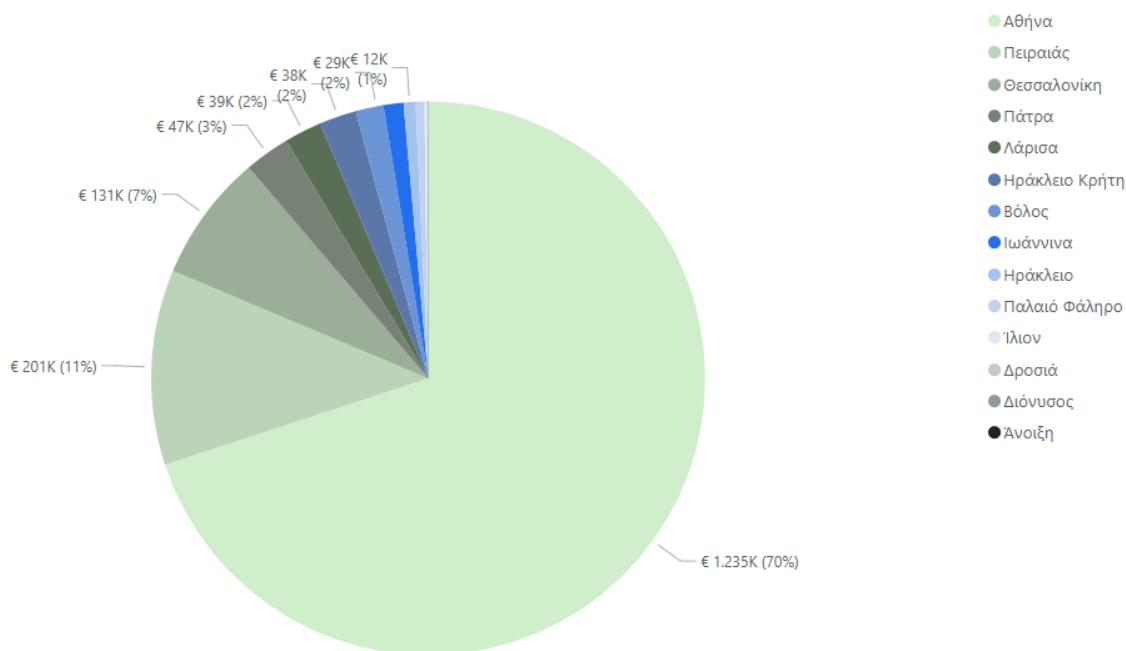
Εικόνα 20. Μέσο κόστος παραγγελίας ανά περιοχή

Το ραβδόγραμμα (Εικόνα 15) απεικονίζει τη μέση αξία παραγγελίας ανά πελάτη σε διαφορετικές περιοχές, παρέχοντας πληροφορίες για τις περιφερειακές διαφοροποιήσεις στη συμπεριφορά δαπανών. Κάθε ράβδος αναπαριστά το μέσο ποσό που δαπανάται από τους πελάτες σε μία παραγγελία, υπολογισμένο για κάθε περιοχή.

Τα δεδομένα δείχνουν ότι το Παλαιό Φάληρο έχει τη μεγαλύτερη μέση αξία παραγγελίας, στα €27, ακολουθούμενο από τη Θεσσαλονίκη, τον Πειραιά και την Αθήνα, που παρουσιάζουν μέσους όρους €23 ανά παραγγελία. Οι περιοχές αυτές, οι οποίες συγκαταλέγονται στα κύρια αστικά κέντρα, εμφανίζουν τη μεγαλύτερη δαπάνη ανά πελάτη, υποδηλώνοντας μεγαλύτερη διάθεση ή δυνατότητα για μεγαλύτερες αγορές. Το Ίλιον, το Ηράκλειο και η Δροσιά έπονται με μέσους όρους €22 και €21, αντίστοιχα. Άλλες περιοχές, όπως ο Βόλος, ο Διόνυσος και η Λάρισα, παρουσιάζουν μέτριους μέσους όρους περίπου €20 ανά παραγγελία, ενώ η Άνοιξη καταγράφει τη χαμηλότερη μέση αξία παραγγελίας, στα €17.

Οι διαφορές στη μέση αξία παραγγελίας υποδεικνύουν διαφοροποιήσεις στα προφίλ των πελατών, τη συμπεριφορά αγορών ή τις προτιμήσεις προϊόντων ανά περιοχή. Αστικές περιοχές όπως το Παλαιό Φάληρο, η Θεσσαλονίκη και η Αθήνα τείνουν να έχουν υψηλότερες μέσες αξίες παραγγελιών, πιθανότατα αντικατοπτρίζοντας μεγαλύτερη οικονομική δυνατότητα ή πρόσβαση σε ευρύτερη γκάμα προϊόντων και υπηρεσιών. Αντίθετα, περιοχές με χαμηλότερες μέσες αξίες παραγγελιών, όπως η Πάτρα, μπορεί να υποδηλώνουν μικρότερα μεγέθη καλαθιού ή χαμηλότερη αγοραστική δύναμη. Το τελευταίο ίσως δεν ισχύει για την Άνοιξη, καθώς οι κάτοικοι της συγκεκριμένης περιοχής δεν χαρακτηρίζονται από οικονομική ένδεια στο σύνολό τους.

Αυτά τα ευρήματα υποδηλώνουν την ανάγκη για προσαρμοσμένες στρατηγικές με στόχο τη βελτιστοποίηση της συμπεριφοράς δαπανών σε διαφορετικές περιοχές. Για περιοχές με υψηλή απόδοση, όπως το Παλαιό Φάληρο και η Θεσσαλονίκη, η εταιρεία θα μπορούσε να ενθαρρύνει περαιτέρω δαπάνες μέσω προσφορών σε υψηλότερης τιμής προϊόντα, αποκλειστικών εκπτώσεων ή εξατομικευμένων προωθητικών ενεργειών για πιστούς πελάτες. Αντίστοιχα, για περιοχές με χαμηλότερες μέσες αξίες παραγγελιών, όπως η Άνοιξη, στρατηγικές που στοχεύουν στην αύξηση του μεγέθους καλαθιού—όπως προωθητικές ενέργειες πακέτων, δωρεάν μεταφορικά πάνω από συγκεκριμένα ποσά ή πρωτοβουλίες cross-selling—θα μπορούσαν να ενθαρρύνουν μεγαλύτερες δαπάνες.



Εικόνα 21. Αξία του Χρόνου Ζωής του Πελάτη (Customer Lifetime Value) ανά περιοχή

Το διάγραμμα πίtaς (Εικόνα 16) αποτυπώνει την κατανομή της Αξίας Διάρκειας Ζωής Πελατών (CLV) ανά περιοχή, αναδεικνύοντας τη συνεισφορά κάθε γεωγραφικής περιοχής στα μακροπρόθεσμα έσοδα της εταιρείας. Η CLV αντιπροσωπεύει τα συνολικά έσοδα που αναμένεται να αποκομίσει η εταιρεία από έναν πελάτη σε μια συγκεκριμένη περιοχή κατά τη διάρκεια της σχέσης τους.

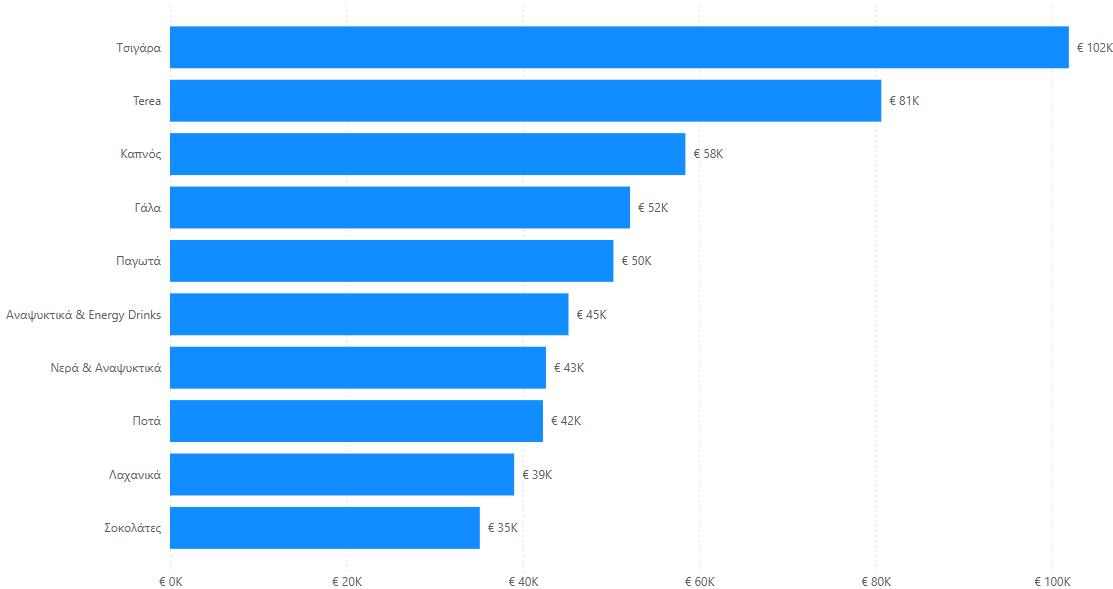
Το διάγραμμα δείχνει ότι η Αθήνα κυριαρχεί με CLV ύψους €1.235 χιλ., που αντιστοιχεί στο 70% του συνόλου. Άκολουθεί ο Πειραιάς, ο οποίος συνεισφέρει €201 χιλ. (11%), και η Θεσσαλονίκη, με €131 χιλ. (7%). Αυτά τα τρία αστικά κέντρα συνθέτουν τη συντριπτική πλειοψηφία των μακροπρόθεσμων εσόδων της εταιρείας, αντανακλώντας τις ισχυρές πελατειακές τους βάσεις. Άλλες περιοχές, όπως η Πάτρα (€47 χιλ., 3%), η Λάρισα (€39 χιλ., 2%) και το Ηράκλειο Κρήτης (€38 χιλ., 2%), συνεισφέρουν πιο μετριοπαθώς, ενώ μικρότερες περιοχές όπως η Άνοιξη και ο Διόνυσος, αντιπροσωπεύουν ένα μικρό ποσοστό του συνολικού CLV.

Αυτή η κατανομή επαληθεύει για ακόμη μία φορά τη μεγάλη εξάρτηση της εταιρείας από τις αστικές περιοχές, και ειδικότερα από την Αθήνα, για τα μακροπρόθεσμα έσοδά της. Η κυριαρχία αυτών των περιοχών αντικατοπτρίζει τις μεγαλύτερες πελατειακές βάσεις τους, τη συχνότερη τοποθέτηση παραγγελιών, καθώς και τις υψηλότερες μέσες αξίες παραγγελιών. Αντίθετα, οι μικρότερες συνεισφορές υποδεικνύουν ευκαιρίες ανάπτυξης σε λιγότερο αστικοποιημένες αγορές.

Τα δεδομένα υπογραμμίζουν τη σημασία της διατήρησης ισχυρών σχέσεων με τους πελάτες στις περιοχές υψηλής αξίας, όπως η Αθήνα, ο Πειραιάς και η Θεσσαλονίκη, καθώς αυτές οι περιοχές είναι κρίσιμες για την οικονομική ευρωστία της εταιρείας. Η ενίσχυση της αφοσίωσης των πελατών σε αυτές τις περιοχές μέσω στοχευμένων ενεργειών μάρκετινγκ θα μπορούσε να μεγιστοποιήσει την αξία διάρκειας ζωής τους. Παράλληλα, κρίνονται εξίσου σημαντικές οι προσπάθειες αύξησης της CLV σε μικρότερες περιοχές.

6.3.3 Ανάλυση προϊόντων

Κατόπιν της διεξοδικότερης ανάλυσης των πελατών, έπειτα η ανάλυση των επί μέρους προϊόντων. Ήταν λόγω μελέτη θεωρείται κρίσιμη γιατί θα αποκαλύψει τα προϊόντα με τη μεγαλύτερη ζήτηση, καθώς και τις κατηγορίες αυτών, ενώ παράλληλα, θα αποτελέσει εφαλτήριο για την περαιτέρω ανάλυση του καλαθιού αγοράς που θα ακολουθήσει.



Εικόνα 22. Κορυφαίες 10 κατηγορίες προϊόντων βάσει τζίρου

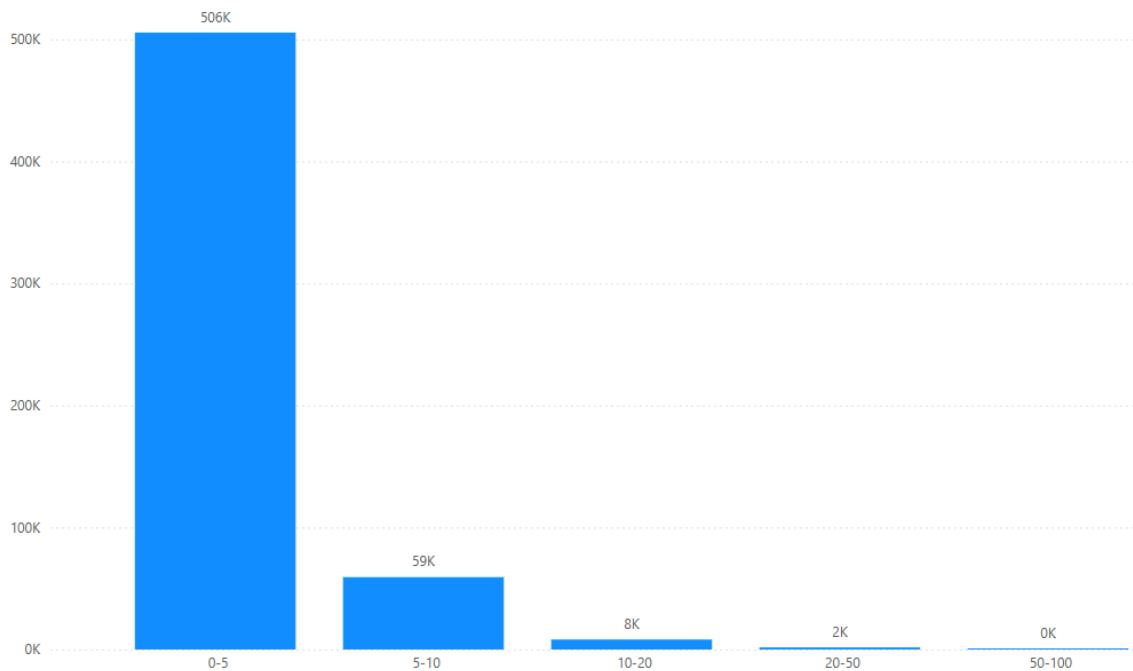
Customer Segmentation & Market Basket Analysis

Το ραβδόγραμμα της Εικόνας 17 καταγράφει τις δέκα κορυφαίες κατηγορίες προϊόντων, ταξινομημένες βάσει της συνολικής συνεισφοράς τους στα έσοδα, προσφέροντας πληροφορίες για τα πιο κερδοφόρα τμήματα προϊόντων. Κάθε ράβδος αντιπροσωπεύει τα συνολικά έσοδα που παράγονται από μια συγκεκριμένη κατηγορία, αποτυπώνοντας τη σχετική απόδοσή τους εντός του συνόλου των δεδομένων.

Τα δεδομένα δείχνουν ότι τα Τσιγάρα ηγούνται με συνολικά έσοδα ύψους €102 χιλ., καθιστώντας τα ως τη σημαντικότερη κατηγορία προϊόντων. Ακολουθεί η Tereia με €81 χιλ., ενώ στην τρίτη θέση βρίσκεται ο Καπνός, που συνεισφέρει €58 χιλ.. Το Γάλα κατατάσσεται τέταρτο με €52 χιλ. σε έσοδα. Τα Παγωτά ολοκληρώνουν την πρώτη πεντάδα με συνολικά έσοδα ύψους €50 χιλ.. Άλλες κατηγορίες στην κορυφαία δεκάδα περιλαμβάνουν τα Αναψυκτικά & Ενεργειακά Ποτά (€45 χιλ.), τα Νερά & Αναψυκτικά (€43 χιλ.), τα Ποτά (€42 χιλ.), τα Λαχανικά (€39 χιλ.) και τις Σοκολάτες (€35 χιλ.).

Η κυριαρχία των κατηγοριών που σχετίζονται με τα Τσιγάρα, την Tereia και τον Καπνό, υποδηλώνει ότι αυτά τα προϊόντα είναι βασικοί κινητήριοι μοχλοί των εσόδων της εταιρείας. Αυτό πιθανώς αντικατοπτρίζει μια πιστή πελατειακή βάση ή συχνά μοτίβα αγορών που συνδέονται με αυτά τα προϊόντα. Παράλληλα, η παρουσία βασικών καθημερινών αγαθών όπως το Γάλα, τα Νερά και τα Λαχανικά υπογραμμίζει τη συνεχή ζήτησή τους, ενώ τα προαιρετικά είδη όπως τα Παγωτά και οι Σοκολάτες αναδεικνύουν τη σημασία των προϊόντων απόλαυσης στο συνολικό μείγμα προϊόντων. Η ισχυρή απόδοση κατηγοριών όπως τα Ποτά και τα Αναψυκτικά & Ενεργειακά Ποτά φανερώνει τις προτιμήσεις των πελατών για πόσιμα είδη, είτε για καθημερινή κατανάλωση, είτε για κοινωνικές περιστάσεις.

Αυτά τα ευρήματα αναδεικνύουν διάφορες ευκαιρίες για τη βελτιστοποίηση των στρατηγικών προϊόντων της εταιρείας. Για τις κατηγορίες υψηλής απόδοσης όπως τα Τσιγάρα, ο Καπνός και η Tereia, η διασφάλιση της διαθεσιμότητας, η προσφορά προωθητικών ενεργειών και η εξασφάλιση ανταγωνιστικών τιμών θα μπορούσαν να ενισχύσουν περαιτέρω τη συνεισφορά τους στα έσοδα. Για τα προαιρετικά προϊόντα όπως τα Παγωτά και οι Σοκολάτες, στοχευμένες καμπάνιες μάρκετινγκ κατά τη διάρκεια συγκεκριμένων εποχών ή περιστάσεων θα μπορούσαν να μεγιστοποιήσουν τις δυνατότητες πωλήσεών τους. Παρομοίως, η εστίαση στα βασικά είδη όπως το Γάλα και τα Λαχανικά θα μπορούσε να ενισχύσει τη διατήρηση πελατών, τοποθετώντας την εταιρεία ως μια αξιόπιστη πηγή για καθημερινά αγαθά.

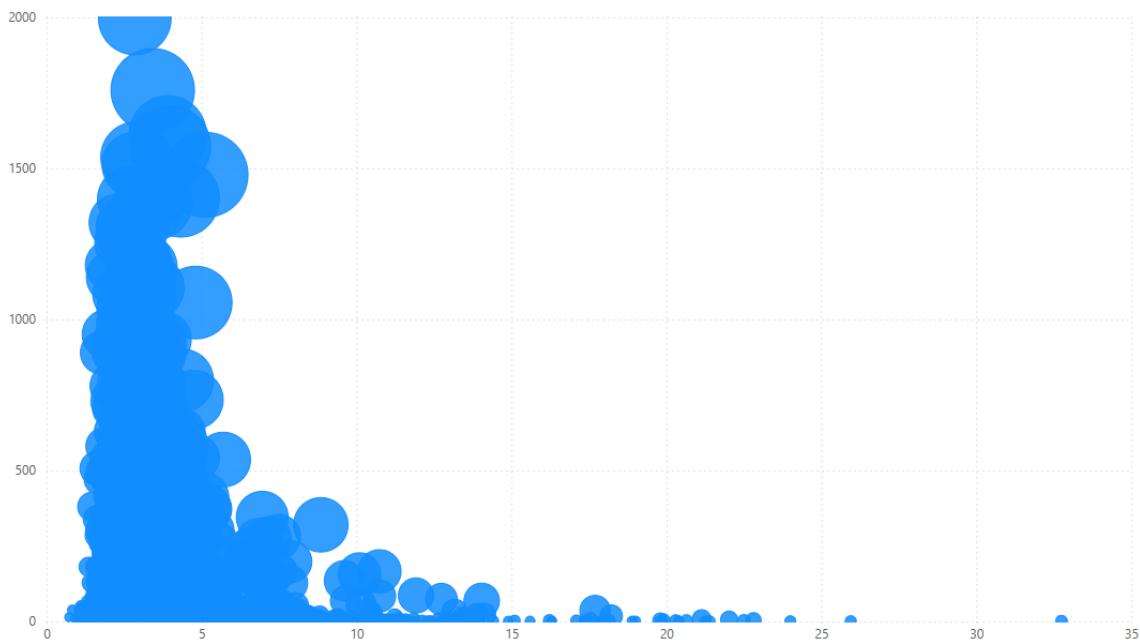


Εικόνα 23. Αριθμός προϊόντων ανά ομάδα τιμών

Στο ραβδόγραμμα (Εικόνα 18) βλέπουμε την κατανομή των προϊόντων σε διαφορετικά εύρη τιμών, προσφέροντας πληροφορίες για τη στρατηγική τιμολόγησης της εταιρείας. Κάθε ράβδος αντιπροσωπεύει τον αριθμό των προϊόντων που ανήκουν σε ένα συγκεκριμένο εύρος τιμών, επιτρέποντας μια σαφή κατανόηση του τρόπου με τον οποίο διαρθρώνεται το χαρτοφυλάκιο προϊόντων από άποψη τιμών.

Τα δεδομένα δείχνουν ότι η συντριπτική πλειοψηφία των προϊόντων, συνολικά 506 χιλ., τιμολογούνται μεταξύ €0 και €5, υποδηλώνοντας ότι το χαρτοφυλάκιο προϊόντων της εταιρείας δίνει ιδιαίτερη έμφαση στα χαμηλού κόστους είδη. Η επόμενη μεγαλύτερη ομάδα περιλαμβάνει 59 χιλ. προϊόντα στην κατηγορία τιμών €5 έως €10, ακολουθούμενη από 8 χιλ. προϊόντα στην κατηγορία €10 έως €20. Ο αριθμός των προϊόντων μειώνεται σημαντικά στις υψηλότερες κατηγορίες τιμών, με μόνο 2 χιλ. προϊόντα να τιμολογούνται μεταξύ €20 και €50, ενώ πρακτικά δεν υπάρχουν προϊόντα στην κατηγορία €50 έως €100.

Αυτή η κατανομή υποδηλώνει ότι η εταιρεία εξυπηρετεί κυρίως μια πελατειακή βάση ευαίσθητη στις τιμές, με έντονη εστίαση στα προσιτά προϊόντα. Ο περιορισμένος αριθμός ειδών υψηλότερης τιμής υποδεικνύει ότι τα πολυτελή προϊόντα δεν αποτελούν σημαντικό μέρος της τρέχουσας στρατηγικής της εταιρείας. Η κυριαρχία των προϊόντων χαμηλού κόστους ευθυγραμμίζεται καλά με την εξυπηρέτηση μιας ευρείας πελατειακής βάσης, αλλά η εταιρεία θα μπορούσε να εξετάσει τη δυνατότητα διεύρυνσης της γκάμας προϊόντων υψηλότερης τιμής, για να στοχεύσει πελάτες με μεγαλύτερη αγοραστική δύναμη. Η εισαγωγή πολυτελών ή εξειδικευμένων σειρών προϊόντων, όπως βιολογικά ή προϊόντα εξειδικευμένων αγορών, θα μπορούσε να βοηθήσει στη διαφοροποίηση του χαρτοφυλακίου και να προσελκύσει πελάτες υψηλότερης αξίας.

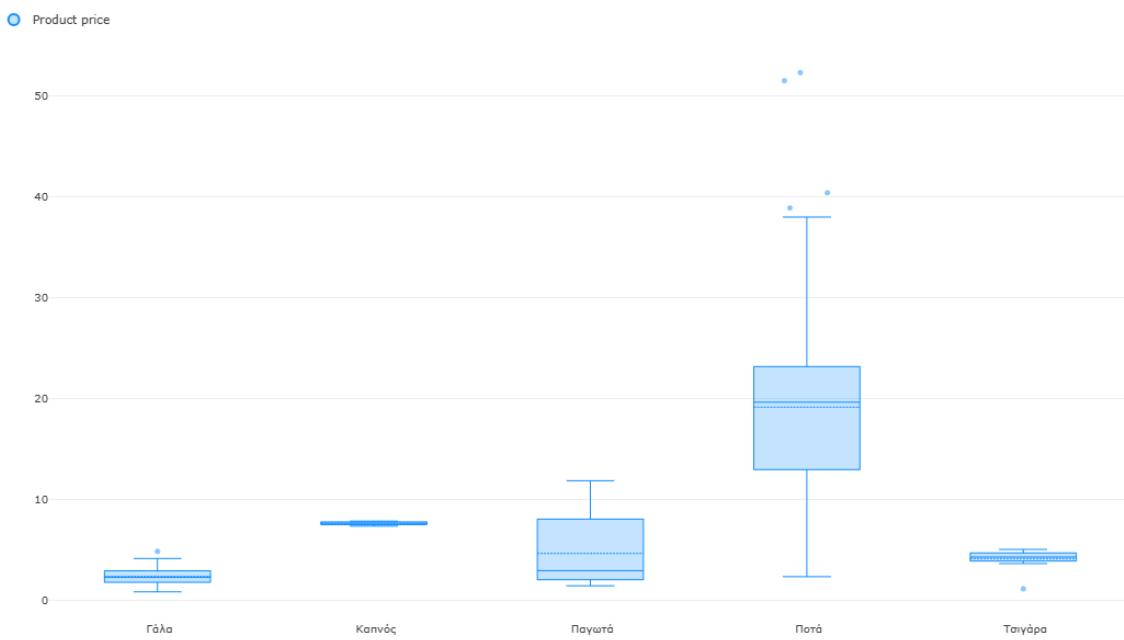


Εικόνα 24. Σύνολο πωληθέντων προϊόντων ανά πελάτη σε μέση τιμή

Το διάγραμμα φυσαλίδων της Εικόνας 19 αποτυπώνει τη σχέση μεταξύ της μέσης τιμής των προϊόντων (οριζόντιος άξονας), του αριθμού των πωλούμενων ειδών και του αριθμού των πελατών που αγοράζουν αυτά τα προϊόντα (κάθετος άξονας). Το μέγεθος κάθε φυσαλίδας αντιστοιχεί στον αριθμό των πελατών που αγοράζουν προϊόντα σε συγκεκριμένα εύρη μέσης τιμής.

Το διάγραμμα δείχνει ότι η πλειονότητα των πωλήσεων επικεντρώνεται σε προϊόντα με τιμές μεταξύ €0 και €5, όπως υποδεικνύεται από την υψηλή συγκέντρωση φυσαλίδων σε αυτό το εύρος. Αυτά τα προϊόντα αντιπροσωπεύουν τον μεγαλύτερο αριθμό πωλούμενων ειδών, αντικατοπτρίζοντας την εστίαση της εταιρείας στην προσιτή τιμή και στα προϊόντα χαμηλού κόστους αλλά υψηλού όγκου πωλήσεων. Το μέγεθος των φυσαλίδων σε αυτό το εύρος υπογραμμίζει επίσης τη σημαντική εμπλοκή των πελατών, ενισχύοντας περαιτέρω τη σημασία των οικονομικών προϊόντων στην ενίσχυση των συνολικών πωλήσεων.

Καθώς αυξάνεται η μέση τιμή, ο αριθμός των πωλούμενων ειδών και η εμπλοκή των πελατών μειώνονται απότομα. Προϊόντα στο εύρος €5 έως €10 εμφανίζουν αξιοσημείωτη μείωση τόσο στον αριθμό των πωλήσεων όσο και στη συμμετοχή των πελατών, ενώ οι υψηλότερες κατηγορίες τιμών (άνω των €20) εκπροσωπούνται σπάνια, με ελάχιστα πωλούμενα είδη και περιορισμένη δραστηριότητα των πελατών. Αυτό το πρότυπο υποδηλώνει ότι το χαρτοφυλάκιο προϊόντων και η στρατηγική τιμολόγησης της εταιρείας είναι έντονα προσανατολισμένα στα προϊόντα χαμηλότερης τιμής.



Εικόνα 25. (Θηκόγραμμα-Box plot) Εύρος τιμών των κορυφαίων 5 κατηγοριών προϊόντων

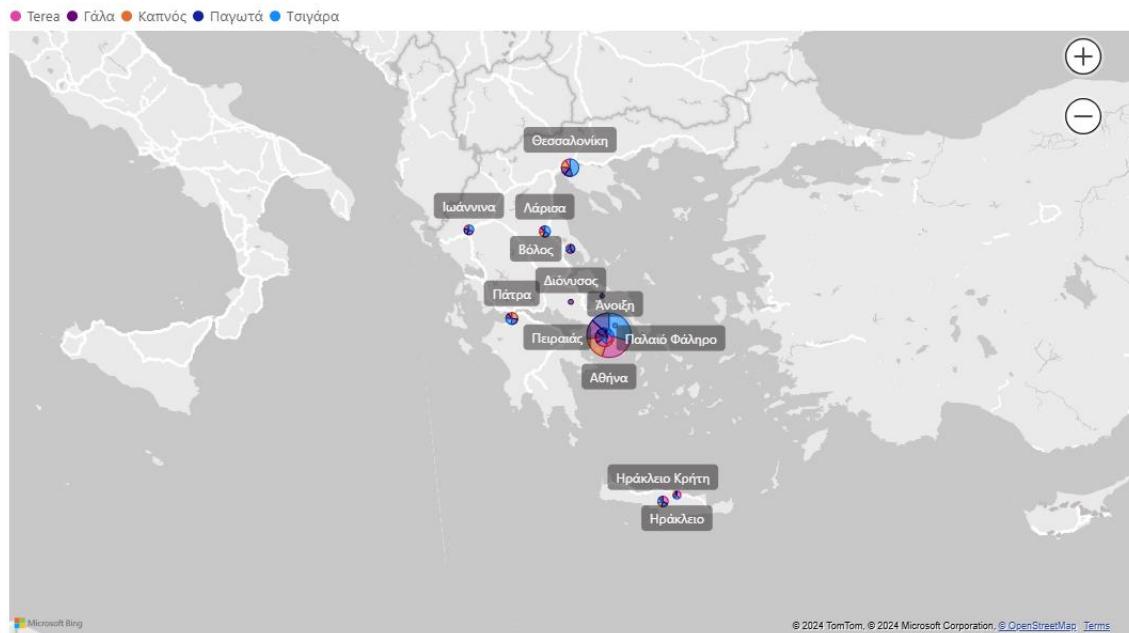
Το box plot της Εικόνας 20 δείχνει την κατανομή των τιμών για τις πέντε κορυφαίες κατηγορίες προϊόντων, ήτοι Γάλα, Καπνός, Παγωτά, Ποτά και Τσιγάρα, παρέχοντας μια λεπτομερή εικόνα για τη μεταβλητότητα, τις διάμεσες τιμές και τυχόν ακραίες τιμές (outliers) στις κατηγορίες αυτές.

Τα Ποτά παρουσιάζουν το μεγαλύτερο εύρος τιμών από τις πέντε κατηγορίες, με τιμές που κυμαίνονται από κάτω από €3 έως πάνω από €50. Το διάστημα μεταξύ των τεταρτημορίων (IQR), που αναπαρίσταται από το σκιασμένο κουτί, υποδηλώνει σημαντική μεταβλητότητα στις τιμές του Αλκοόλ, κάτι που δείχνει ότι η κατηγορία περιλαμβάνει τόσο οικονομικά όσο και λιγότερο φθηνά προϊόντα. Η διάμεση τιμή βρίσκεται λίγο κάτω από τα €20, ενώ μερικά δεδομένα εκτείνονται πάνω από τα €50.

Τα Παγωτά εμφανίζουν μέτρια μεταβλητότητα στις τιμές, με τις περισσότερες τιμές να συγκεντρώνονται κάτω από τα €10. Η διάμεση τιμή βρίσκεται λίγο κάτω από τα €5, ενώ το εύρος τιμών δείχνει ένα μείγμα από βασικά και ελαφρώς ακριβότερα προϊόντα, πιθανώς επηρεασμένα από premium brands ή προϊόντα σε συσκευασίες πολλαπλών τεμαχίων.

Το Γάλα, ο Καπνός και τα Τσιγάρα παρουσιάζουν πολύ στενότερες κατανομές τιμών. Για τα Τσιγάρα, οι τιμές συγκεντρώνονται στενά γύρω από μια διάμεση τιμή κοντά στα €4, αντικατοπτρίζοντας την ομοιομορφία τιμολόγησης αυτής της κατηγορίας, πιθανότατα λόγω ρυθμιστικών ελέγχων ή τυποποιημένων μεγεθών προϊόντων. Παρομοίως, το Γάλα έχει πολύ στενό εύρος, με τις τιμές να συγκεντρώνονται γύρω από μια διάμεση τιμή κάτω από τα €3, υποδηλώνοντας σταθερή τιμολόγηση για βασικά είδη. Ο Καπνός ακολουθεί παρόμοιο πρότυπο, με στενό εύρος τιμών και τις τιμές να συγκεντρώνονται γύρω από μια διάμεση τιμή περίπου στα €8.

Αυτά τα ευρήματα παρέχουν ευκαιρίες για στρατηγικές αποφάσεις. Για κατηγορίες με σημαντική μεταβλητότητα, όπως τα Ποτά, η εταιρεία μπορεί να βελτιστοποιήσει τη στρατηγική τιμολόγησή της, αναλύοντας τις προτιμήσεις των πελατών για ακριβότερα προϊόντα σε σχέση με οικονομικές επιλογές. Αντίθετα, σταθερές κατηγορίες όπως το Γάλα και τα Τσιγάρα τονίζουν τη σημασία της διατήρησης ανταγωνιστικών τιμών για την ενίσχυση της εμπιστοσύνης και της πιστότητας των πελατών.



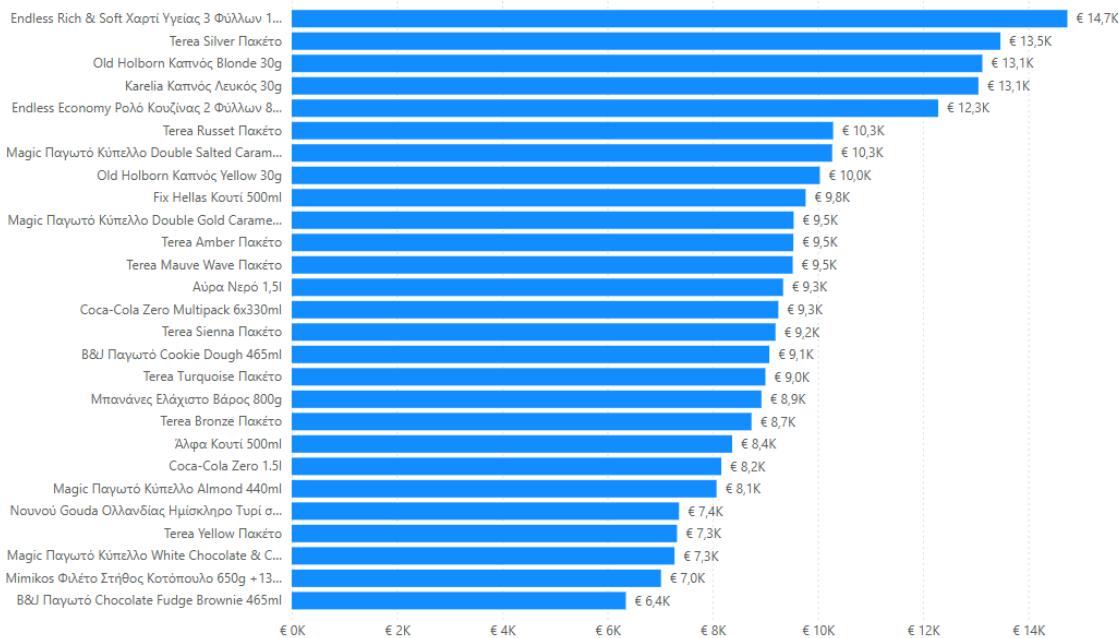
Εικόνα 26. Κορυφαίες 5 κατηγορίες προϊόντων βάσει τζίρου ανά περιοχή

Ο χάρτης της Εικόνας 21 παρέχει μια γεωγραφική ανάλυση των πέντε κορυφαίων κατηγοριών προϊόντων βάσει εσόδων, απεικονίζοντας την κατανομή τους σε διάφορες περιοχές. Κάθε φυσαλίδα αντιπροσωπεύει μια περιοχή, ενώ το μέγεθος και η διαίρεση της φυσαλίδας αντικατοπτρίζουν τη συμβολή των εσόδων κάθε κατηγορίας προϊόντων εντός της συγκεκριμένης περιοχής. Οι πέντε κορυφαίες κατηγορίες περιλαμβάνουν τα Τσιγάρα, την Τερεά, το Καπνό, το Γάλα και τα Παγωτά, που αναπαρίστανται με ξεχωριστά χρώματα.

Τα δεδομένα δείχνουν ότι η Αθήνα και οι γύρω περιοχές, όπως το Παλαιό Φάληρο και ο Πειραιάς, κυριαρχούν στις συνεισφορές εσόδων σε όλες τις κατηγορίες. Σε αυτές τις αστικές περιοχές, τα Τσιγάρα (μπλε) καταλαμβάνουν σταθερά το μεγαλύτερο μερίδιο εσόδων, ακολουθούμενα από την Τερεά (ροζ), τον Καπνό (πορτοκαλί) και το Γάλα (μοβ). Η σημαντική συμβολή των Τσιγάρων ευθυγραμμίζεται με τη συνολική ισχυρή τους απόδοση, όπως έχει ήδη φανεί σε προηγούμενα διαγράμματα.

Άλλες περιοχές, όπως η Θεσσαλονίκη, δείχνουν επίσης αξιοσημείωτες συνεισφορές, αν και το συνολικό τους μερίδιο στα έσοδα είναι μικρότερο σε σχέση με την Αθήνα. Στη Θεσσαλονίκη, τα έσοδα από τα Τσιγάρα φαίνονται πιο έντονα, υποδηλώνοντας διαφοροποιήσεις στις προτιμήσεις των καταναλωτών ή στη ζήτηση προϊόντων. Το ίδιο πρότυπο παρατηρείται και στα Ιωάννινα.

Σε μικρότερες περιοχές, όπως η Πάτρα και το Ηράκλειο Κρήτης, οι φυσαλίδες είναι σημαντικά μικρότερες, υποδεικνύοντας χαμηλότερα συνολικά έσοδα. Στις περιοχές αυτές, τα έσοδα φαίνεται να κατανέμονται πιο ομοιόμορφα στις πέντε κατηγορίες προϊόντων, χωρίς καμία κατηγορία να κυριαρχεί τόσο έντονα όσο στα μεγαλύτερα αστικά κέντρα.



Εικόνα 27. Κορυφαία 30 προϊόντα βάσει τζίρου

Το ραβδόγραμμα της Εικόνας 22 παρουσιάζει τα 30 κορυφαία προϊόντα ταξινομημένα με βάση τα έσοδα που παράγουν, παρέχοντας πληροφορίες για τα πιο σημαντικά είδη από οικονομική άποψη στο απόθεμα της εταιρείας. Κάθε ράβδος αντιπροσωπεύει ένα προϊόν, με τα αντίστοιχα έσοδά του να εμφανίζονται δίπλα.

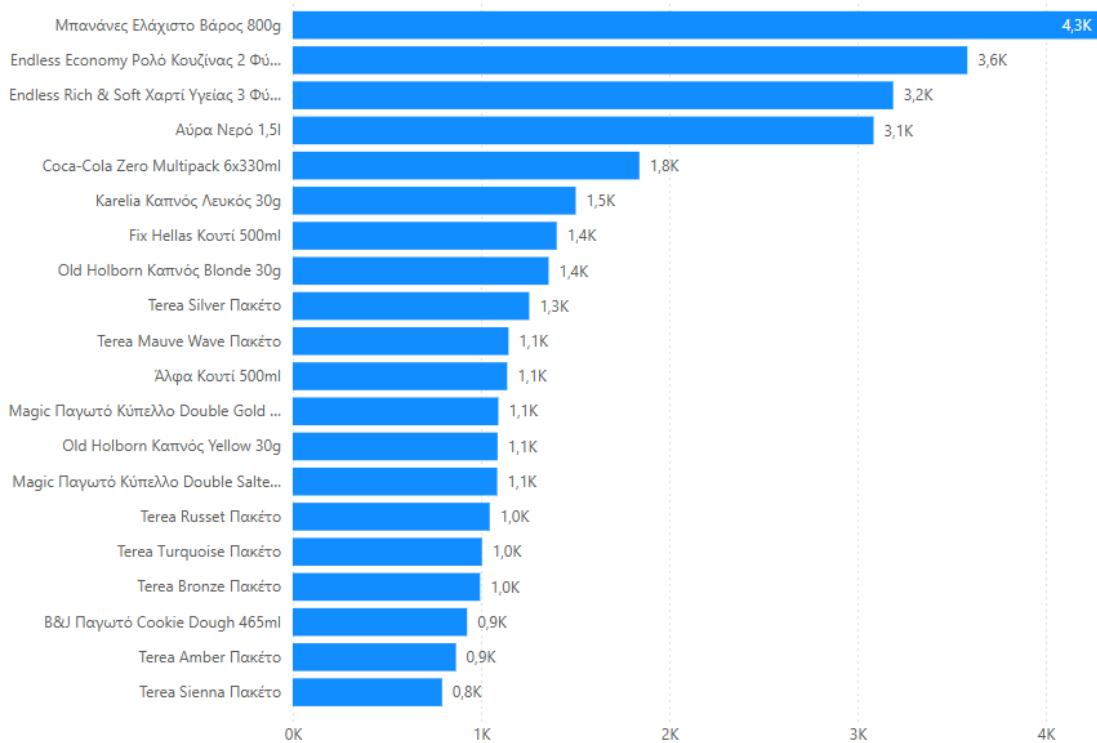
Τα δεδομένα δείχνουν ότι το "Endless Rich & Soft Χαρτί Υγείας 3 Φύλλων 12 Τεμ" ηγείται της κατάταξης, παράγοντας €14.7 χιλ. σε έσοδα. Το προϊόν αυτό ακολουθείται στενά από το "Tereia Silver Πακέτο" με €13.5 χιλ., το "Old Holborn Καπνός Blonde 30g" με €13.1 χιλ. και το "Karelia Καπνός Λευκός 30g", επίσης με €13.1 χιλ. Στην επόμενη θέση κατατάσσεται το "Endless Economy Ρολό Κουζίνας 2 Φύλλων 8 Τεμ", το οποίο συνεισφέρει €12.3 χιλ.. Αυτά τα κορυφαία προϊόντα αντικατοπτρίζουν έναν συνδυασμό κατηγοριών υψηλής ζήτησης, όπως είδη οικιακής χρήσης (χαρτί υγείας και ρολά κουζίνας) και καπνικά προϊόντα (μάρκες Tereia και Karelia). Διαπιστώνεται περαιτέρω η σημαντική συμβολή των καπνικών προϊόντων της επωνυμίας Tereia, με πολλαπλές παραλλαγές (π.χ. Silver, Russet, Amber) να εμφανίζονται εμφανώς στις κορυφαίες θέσεις.

Τα βασικά είδη οικιακής χρήσης, όπως το χαρτί υγείας και τα ρολά κουζίνας, κατατάσσονται επίσης ψηλά, αντικατοπτρίζοντας τη σταθερή εξάρτηση των πελατών από καθημερινά απαραίτητα αγαθά. Επιπλέον, τα παγωτά (π.χ. Magic Παγωτό Κύπελλο Double Salted Caramel, Double Gold Caramel, Cookie Dough) εμφανίζονται στις υψηλές θέσεις, υποδεικνύοντας την προθυμία των πελατών να αγοράσουν προϊόντα ελαφρώς υψηλότερης αξίας για απόλαυση.

Η κατώτερη βαθμίδα της λίστας των 30 κορυφαίων περιλαμβάνει προϊόντα όπως το "Mimikos Φιλέτο Στήθος Κοτόπουλο 650g" και το "B&J Παγωτό Chocolate Fudge Brownie 465ml", που παράγουν έσοδα €7 χιλ. και €6.4 χιλ., αντίστοιχα. Παρά τη χαμηλότερη κατάταξή τους, αυτά τα προϊόντα εξακολουθούν να συνεισφέρουν σημαντικά στα συνολικά έσοδα, υπογραμμίζοντας την ποικιλία των προτιμήσεων των πελατών.

Συμπερασματικά, η κυριαρχία των οικιακών και καπνικών προϊόντων υπογραμμίζει τη συνεχή ζήτηση για βασικά αγαθά, ενώ η παρουσία ελαφρώς διαφοροποιημένων προϊόντων, όπως τα παγωτά αναδεικνύει ευκαιρίες για την αξιοποίηση των προαιρετικών δαπανών. Αξιοποιώντας αυτά τα δεδομένα, η εταιρεία θα μπορούσε να εστιάσει στη βελτιστοποίηση των επιπέδων αποθεμάτων για τα προϊόντα υψηλής κατάταξης, ενώ παράλληλα θα ήταν χρήσιμο να εξερευνήσει

προωθητικές ευκαιρίες για την αύξηση των πωλήσεων σε προϊόντα που βρίσκονται στις χαμηλότερες θέσεις της κορυφαίας 30άδας.



Εικόνα 28. Κορυφαία 20 προϊόντα σε πωλήσεις ανά παραγγελία

Στην Εικόνα 23 παρουσιάζονται τα 20 προϊόντα με τη μεγαλύτερη συχνότητα παραγγελιών. Κάθε ράβδος αντιπροσωπεύει ένα προϊόν με τον αντίστοιχο αριθμό παραγγελιών. Το προϊόν με τις περισσότερες πωλήσεις είναι οι "Μπανάνες Ελάχιστο Βάρος 800g" με 4,3 χιλ. παραγγελίες, καθιστώντας το αδιαμφισβήτητο ηγέτη σε αυτή την κατηγορία. Ακολουθούν τα "Endless Economy Ρολό Κουζίνας 2 Φύλλων 8 Τεμ" με 3.6 χιλ. παραγγελίες και το "Endless Rich & Soft Χαρτί Υγείας 3 Φύλλων 10 Τεμ" με 3,2 χιλ. παραγγελίες, κάτι που υποδεικνύει σταθερή προτίμηση για βασικά είδη οικιακής χρήσης, όπως τα προϊόντα χαρτιού.

Το εμφιαλωμένο νερό, με το "Αύρα Νερό 1,5L" (Avra Water, 1,5L), κατατάσσεται τέταρτο με 3,1 χιλ. παραγγελίες. Άλλα σημαντικά προϊόντα περιλαμβάνουν το "Coca-Cola Zero Multipack 6x330ml", με 1,8 χιλ. παραγγελίες, που αντικατοπτρίζει την ισχυρή αφοσίωση των πελατών στα επίωνυμα αναψυκτικά.

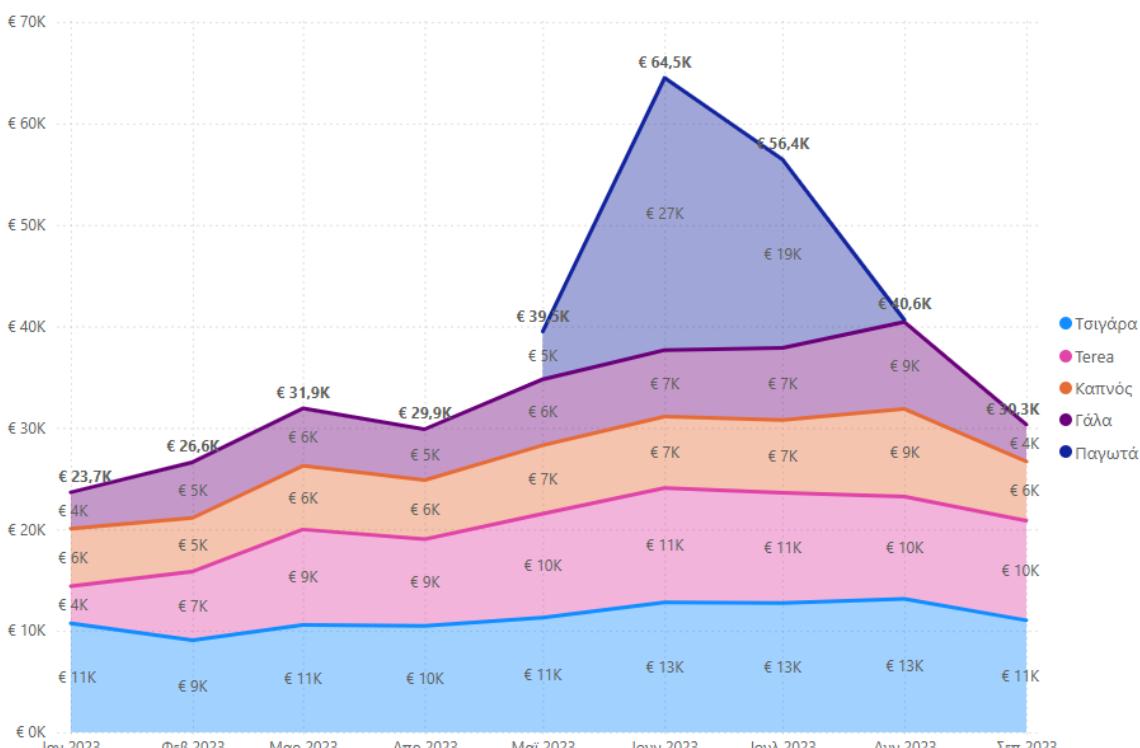
Αρκετά καπνικά προϊόντα εμφανίζονται επίσης στις υψηλές θέσεις της κατάταξης, όπως το "Karelia Καπνός Λευκός 30g" με 1.5 χιλ. παραγγελίες, καθώς και διάφορα προϊόντα Tereia και Old Holborn, το καθένα από τα οποία καταγράφει περίπου 1,0–1,4 χιλ. παραγγελίες.

Προϊόντα παγωτού κάνουν επίσης αισθητή την παρουσία τους, όπως το "Magic Παγωτό Κύπελλο Double Gold Caramel" και το "Magic Παγωτό Κύπελλο Double Salted Caramel", με περίπου 1,1 χιλ. παραγγελίες το καθένα.

Στην κατώτερη βαθμίδα της λίστας των 20 κορυφαίων περιλαμβάνονται προϊόντα όπως το "Tereia Amber Πακέτο" και το "Tereia Sienna Πακέτο", με 0,9 χιλ. και 0,8 χιλ. παραγγελίες αντίστοιχα. Παρότι αυτά τα προϊόντα συνεισφέρουν λιγότερο στον συνολικό αριθμό παραγγελιών συγκριτικά με τα προϊόντα κορυφαίας κατάταξης, εξακολουθούν να αποτελούν σημαντικό μέρος των προτιμήσεων των πελατών στις κατηγορίες τους.

Το διάγραμμα αναδεικνύει την ξεκάθαρη κυριαρχία των βασικών προϊόντων, όπως οι μπανάνες, το εμφιαλωμένο νερό και τα χαρτικά είδη, σε όρους συχνότητας παραγγελιών. Τα καπνικά προϊόντα και τα παγωτά δείχνουν επίσης ισχυρή παρουσία, αντικατοπτρίζοντας τις ποικίλες προτιμήσεις των πελατών, που καλύπτουν τόσο βασικά είδη όσο και προαιρετικές αγορές.

Από στρατηγική άποψη, η διασφάλιση της σταθερής διαθεσιμότητας αποθέματος και οι στοχευμένες πρωθητικές ενέργειες για αυτά τα προϊόντα υψηλής ζήτησης θα μπορούσαν να βελτιώσουν περαιτέρω την ικανοποίηση των πελατών και τη δημιουργία εσόδων. Επιπλέον, υπάρχουν ευκαιρίες για cross-selling ή συνδυαστικές προσφορές με αυτά τα δημοφιλή προϊόντα, με σόχο την αύξηση του μεγέθους του καλαθιού αγορών.



Εικόνα 29. Κορυφαίες 5 κατηγορίες βάσει τζίρου ανά μήνα έτους

Το area chart της Εικόνας 24 αποτυπώνει τις μηνιαίες τάσεις εσόδων για τις πέντε κορυφαίες κατηγορίες προϊόντων, επιτρέποντας μια λεπτομερή σύγκριση της συνεισφοράς τους στα συνολικά έσοδα και των εποχικών μεταβολών τους.

Η κατηγορία των Τσιγάρων καταγράφει σταθερά τα υψηλότερα έσοδα, ξεκινώντας από €11 χιλ. τον Ιανουάριο και κορυφώνοντας στις €13 χιλ. κατά τους μήνες Ιούνιο, Ιούλιο και Αύγουστο.

Η κατηγορία Tereia ακολουθεί με τα αμέσως επόμενα υψηλότερα έσοδα, ξεκινώντας από €4 χιλ. τον Ιανουάριο και φτάνοντας στην κορύφωσή της τον Ιούνιο με €11 χιλ., γεγονός που υποδηλώνει σημαντική αύξηση κατά τη συγκεκριμένη περίοδο. Ωστόσο, στη συνέχεια τα έσοδα αρχίζουν να μειώνονται ελαφρώς, υποχωρώντας στις €10 χιλ. τον Σεπτέμβριο.

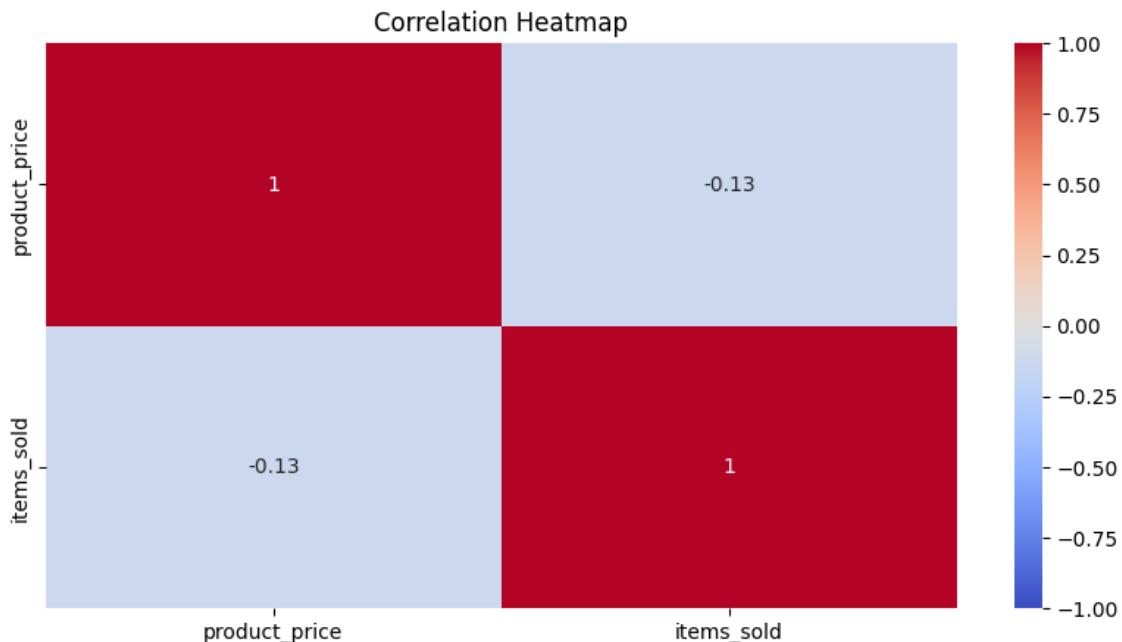
Η κατηγορία Καπνού έχει την επόμενη συμμετοχή στα έσοδα διατηρώντας αυξητική τάση εσόδων μέχρι το τέλος του καλοκαιριού, ξεκινώντας από τις €4 χιλ. τον Ιανουάριο, κορυφώνοντας στις €9 χιλ. τον Αύγουστο και καταλήγοντας στις €6 χιλ. τον Σεπτέμβριο.

Ελάχιστα πιο κάτω εμφανίζεται η κατηγορία του Γάλακτος, η οποία ξεκινά επίσης από τις €4 χιλ. τον Ιανουάριο και παρουσιάζει σταθερή αύξηση μέχρι τον Αύγουστο, φτάνοντας τις €9 χιλ. όπως ο Καπνός, πριν σημειώσει σταδιακή πτώση και επανέλθει στις €6 χιλ. τον Σεπτέμβριο.

Αντίθετα με όλα τα ανωτέρω, η κατηγορία των Παγωτών εμφανίζει έντονο εποχικό μοτίβο. Ξεκινά από τις 5 χιλ. τον Μάιο και σημειώνει απότομη αύξηση τον Ιούνιο φτάνοντας τις €27 χιλ. έσοδα, που είναι μακράν τα υψηλότερα απ' όλες τις υπόλοιπες κατηγορίες. Ωστόσο, τον Αύγουστο κλείνει η συνεισφορά τους στον συνολικό τζίρο της εταιρείας με είσπραξη €13 χιλ..

Το συνολικό μοτίβο εσόδων στις πέντε κορυφαίες κατηγορίες δείχνει έντονη ανοδική τάση από τον Ιανουάριο έως τον Αύγουστο. Η αιφνίδια πτώση κατόπιν, υποδηλώνει μια μετατόπιση των προτεραιοτήτων των καταναλωτών, πιθανώς λόγω εποχικών παραγόντων που σχετίζονται με το τέλος του καλοκαιριού. Αυτά τα ευρήματα καταδεικνύουν την ανάγκη για στρατηγικό σχεδιασμό που θα ευθυγραμμίζει το απόθεμα, το μάρκετινγκ και τις προωθητικές ενέργειες με τις εποχικές τάσεις:

Η αιχμή πωλήσεων στα παγωτά το καλοκαίρι υπογραμμίζει τη σημασία της ενίσχυσης των αποθεμάτων και των διαφημιστικών ενεργειών κατά τους θερμότερους μήνες, με στόχο τη μεγιστοποίηση των εσόδων στη συγκεκριμένη περίοδο υψηλής ζήτησης. Αντίστοιχα, η δυναμική απόδοση της Tereia υποδηλώνει ευκαιρίες για στοχευμένες καμπάνιες ή προωθητικές ενέργειες, ώστε να αξιοποιηθεί η δημοφιλία της κατά την περίοδο αυτή. Παράλληλα, η σταθερότητα των εσόδων της κατηγορίας των Τσιγάρων μπορεί να αξιοποιηθεί για μακροχρόνιες προωθητικές στρατηγικές, οι οποίες δεν επηρεάζονται έντονα από εποχικές διακυμάνσεις.



Εικόνα 30. Θερμικός χάρτης συσχέτισης μεταξύ των αριθμητικών μεταβλητών του dataset

Ο θερμικός χάρτης (Εικόνα 25) δημιουργήθηκε για την αξιολόγηση των γραμμικών σχέσεων μεταξύ των αριθμητικών μεταβλητών του συνόλου δεδομένων. Αποτυπώνει τη δύναμη και την κατεύθυνση των συσχετίσεων, με τιμές που κυμαίνονται από -1 (τέλεια αρνητική συσχέτιση) έως +1 (τέλεια θετική συσχέτιση). Όπως φαίνεται στον χάρτη, οι τιμές στη διαγώνιο είναι ίσες με 1, υποδεικνύοντας τέλεια συσχέτιση κάθε μεταβλητής με τον εαυτό της. Αυτό είναι αναμενόμενο και λειτουργεί ως σημείο αναφοράς για σύγκριση.

Εστιάζοντας στις τιμές εκτός διαγωνίου, ο συντελεστής συσχέτισης μεταξύ «product_price» και «items_sold» είναι -0.13. Αυτό υποδεικνύει μια ασθενή αρνητική γραμμική σχέση μεταξύ αυτών των μεταβλητών. Μια αρνητική συσχέτιση σημαίνει ότι καθώς η τιμή του προϊόντος αυξάνεται, ο αριθμός των πωλούμενων ειδών τείνει να μειώνεται ελαφρώς. Ωστόσο, η αδυναμία αυτής της συσχέτισης δείχνει ότι η τιμή από μόνη της δεν έχει σημαντική επίδραση στον όγκο πωλήσεων. Άλλοι παράγοντες, όπως η κατηγορία προϊόντος, οι προσφορές ή οι προτιμήσεις των πελατών, πιθανώς διαδραματίζουν πιο σημαντικό ρόλο στον καθορισμό της απόδοσης των πωλήσεων.

Η χρωματική διαβάθμιση στον χάρτη θερμότητας τονίζει περαιτέρω αυτή τη σχέση, με το κόκκινο να αντιπροσωπεύει ισχυρές θετικές συσχετίσεις (π.χ., κατά μήκος της διαγωνίου) και το μπλε να υποδηλώνει αρνητικές συσχετίσεις. Η μπλε απόχρωση εκτός διαγωνίου ευθυγραμμίζεται με την ασθενή αρνητική συσχέτιση που παρατηρείται μεταξύ «product_price» και «items_sold».

Το παραπάνω διάγραμμα, καταδεικνύει τη σημασία της διερεύνησης πρόσθετων μεταβλητών και της διεξαγωγής περαιτέρω αναλύσεων, όπως αυτές που προηγήθηκαν, προκειμένου να κατανοθούν καλύτερα οι παράγοντες που επηρεάζουν τις πωλήσεις στο σύνολο δεδομένων.

ΚΕΦΑΛΑΙΟ 7. Εφαρμογή ανάλυσης RFM

7.1 Μεθοδολογία

Το μοντέλο RFM (Recency, Frequency, and Monetary) είναι ένα ευρέως χρησιμοποιούμενο πλαίσιο για την τμηματοποίηση πελατών, επιτρέποντας στις επιχειρήσεις να εντοπίσουν τους πιο πολύτιμους πελάτες τους βάσει της αγοραστικής τους συμπεριφοράς. Οι μετρικές του RFM ορίζονται ως εξής:

Recency: Πόσο πρόσφατα πραγματοποίησε ένας πελάτης την τελευταία του αγορά.

Frequency: Πόσο συχνά ένας πελάτης πραγματοποιεί αγορές σε μια καθορισμένη χρονική περίοδο.

Monetary Value: Το συνολικό έσοδο που παράγεται από έναν πελάτη.

Ο στόχος αυτής της ανάλυσης είναι να υπολογιστούν οι μετρικές RFM για όλους τους πελάτες στο σύνολο δεδομένων, προκειμένου να δημιουργηθεί η βάση για τμηματοποίηση. Το σύνολο δεδομένων που χρησιμοποιείται από το διαθέσιμο dataset για αυτήν την ανάλυση περιλαμβάνει τις εξής στήλες:

customer_id: Μοναδικός αναγνωριστικός κωδικός για κάθε πελάτη.

order_timestamp: Ημερομηνία κάθε συναλλαγής.

product_price: Τιμή ανά μονάδα προϊόντος.

items_sold: Αριθμός μονάδων που πωλήθηκαν σε κάθε συναλλαγή.

monetary_value: Μια νέα στήλη που υπολογίζεται από το γινόμενο του *product_price* και του *items_sold*, για να καταγράψει τα έσοδα που παράγονται από κάθε συναλλαγή.

recency: Υπολογίζεται χρησιμοποιώντας μια reference date. Για τη συγκεκριμένη ανάλυση, η reference date έχει οριστεί ως η 1η Οκτωβρίου 2023, μία ημέρα μετά την τελευταία καταγεγραμμένη συναλλαγή στο σύνολο δεδομένων.

Για να υπολογίσουμε τις μετρικές RFM ακολουθούμε τα παρακάτω βήματα:

Monetary Value: Ομαδοποιούμε όλες τις συναλλαγές ανά customer_id και υπολογίζουμε το άθροισμα της στήλης monetary_value για κάθε πελάτη.

Frequency: Υπολογίζουμε τον αριθμό συναλλαγών ανά πελάτη (order_id) για να αποτυπώσουμε τη συχνότητα των αγορών τους.

Recency: Προσδιορίζουμε την πιο πρόσφατη ημερομηνία συναλλαγής για κάθε πελάτη. Αφαιρούμε την ημερομηνία συναλλαγής από τη reference date για να υπολογίσουμε τον αριθμό ημερών από την τελευταία αγορά.

7.2 Literature review

Οι Dawane et al. (2021) εφάρμοσαν την ανάλυση RFM με τον αλγόριθμο K-Means clustering για την τμηματοποίηση πελατών βάσει παρελθοντικών συναλλαγών, με σκοπό τη βελτίωση των εσόδων και της διατήρησης πελατών. Η μελέτη τόνισε ότι το μοντέλο RFM αποτελεί ένα απλό αλλά αποτελεσματικό εργαλείο για την τμηματοποίηση πελατών, ωστόσο η αποτελεσματικότητά του μπορεί να βελτιωθεί περαιτέρω με τη σωστή παραμετροποίηση. Βασικές προτάσεις για τη βελτίωση της τμηματοποίησης περιλαμβάνουν:

- Ανάλυση RFM με βάρη, όπου κάθε μεταβλητή λαμβάνει διαφορετική βαρύτητα ανάλογα με τον επιχειρηματικό στόχο (π.χ., μεγαλύτερη βαρύτητα στη νομισματική αξία για αύξηση εσόδων ή στη χρονική εγγύτητα για μέιωση του ρυθμού αποχώρησης πελατών).
- Συνδυασμός πολλαπλών μεθόδων για την εύρεση του βέλτιστου αριθμού συστάδων, αντί αυτή να βασίζεται σε μία μόνο τεχνική.
- Χρήση της ιεραρχικής συσταδοποίησης για τον προσδιορισμό του ιδανικού αριθμού συστάδων πριν από την εφαρμογή του K-Means.

Η μελέτη κατέληξε στο ότι οι στρατηγικές τμηματοποίησης πελατών θα πρέπει να προσαρμόζονται ανάλογα με τις επιχειρησιακές ανάγκες, τη διαθεσιμότητα δεδομένων και τους υπολογιστικούς πόρους, ώστε να προσφέρουν πιο αποδοτικές και εφαρμόσιμες στρατηγικές.

Οι Sarkar et al. (2024) διερεύνησαν την εφαρμογή του αλγορίθμου K-Means clustering σε συνδυασμό με την ανάλυση RFM για την αποτελεσματική τμηματοποίηση πελατών και τη βελτιστοποίηση στρατηγικών μάρκετινγκ. Η μελέτη έδειξε ότι ο αλγόριθμος K-Means, στο πλαίσιο της ανάλυσης RFM, πέτυχε ακρίβεια τμηματοποίησης 95%, όπως υποδεικνύεται από την τιμή καθαρότητας των συστάδων (cluster purity = 0,95). Αυτή η υψηλή ακρίβεια ανέδειξε την ικανότητα του αλγορίθμου να κατηγοριοποιεί με ακρίβεια τους πελάτες με βάση κοινές αγοραστικές συμπεριφορές, επιπρέποντας στις επιχειρήσεις να υλοποιήσουν στοχευμένες πρωθητικές ενέργειες και εξατομικευμένες στρατηγικές διατήρησης πελατών. Τα αποτελέσματα επιβεβαίωσαν την αποδοτικότητα του K-Means ως μεθόδου συσταδοποίησης, επικυρώνοντας τη σημασία του στην επιχειρηματική αναλυτική για τη βελτιστοποίηση της τμηματοποίησης πελατών. Η μελέτη κατέληξε στο ότι οι επιχειρήσεις που αξιοποιούν τεχνητή νοημοσύνη στην τμηματοποίηση πελατών μπορούν να αποκτήσουν ανταγωνιστικό πλεονέκτημα, κατηγοριοποιώντας με ακρίβεια το πελατολόγιό τους και εξατομικεύοντας τις στρατηγικές μάρκετινγκ.

Οι Christy et al. (2021) διερεύνησαν την τμηματοποίηση πελατών με χρήση του μοντέλου RFM με σκοπό την κατηγοριοποίηση των πελατών βάσει της αγοραστικής τους συμπεριφοράς. Αρχικά, πραγματοποίησαν ανάλυση RFM στα συναλλακτικά δεδομένα και στη συνέχεια εφάρμοσαν αλγορίθμους συσταδοποίησης, όπως ο παραδοσιακός K-Means και ο Fuzzy C-Means. Στο πλαίσιο της μελέτης, προτάθηκε μια νέα παραλαγή του αλγορίθμου K-Means, ο RM K-Means, ο οποίος βελτιώνει τη διαδικασία επιλογής αρχικών κεντροειδών μέσω του υπολογισμού τους με χρήση των αποτελεσματικών διαμέσων της κατανομής των δεδομένων. Τα ευρήματα έδειξαν ότι ο προτεινόμενος αλγόριθμος μειώνει τον αριθμό των επαναλήψεων, βελτιώνει τη συνοχή των συστάδων και απαιτεί μικρότερο χρόνο εκτέλεσης σε σύγκριση με τις κλασικές προσεγγίσεις. Η μελέτη καταλήγει ότι η αποτελεσματική τμηματοποίηση των πελατών ενισχύει τις σχέσεις με αυτούς και συμβάλλει στην αύξηση της διατήρησής τους, στοιχείο κρίσιμο για τη βιωσιμότητα και ανάπτυξη μιας επιχείρησης. Επιπλέον, πρότειναν ως μελλοντική έρευνα την ανάλυση της

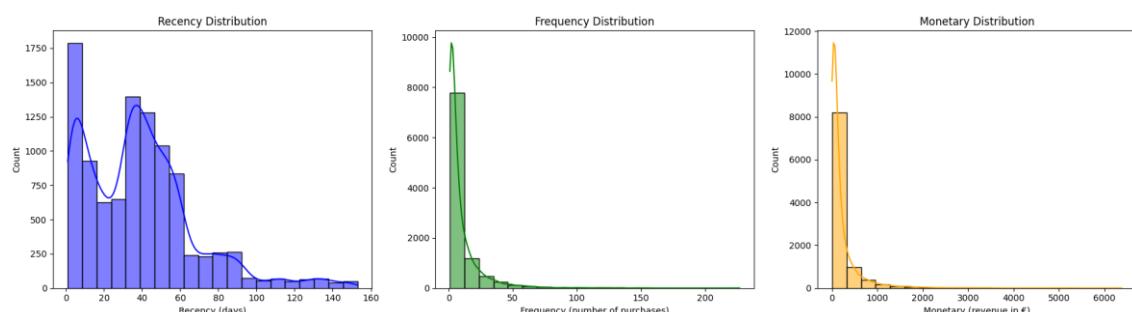
αγοραστικής συμπεριφοράς κάθε συστάδας, προκειμένου να εντοπιστούν τα προϊόντα που προτιμώνται συχνότερα και να σχεδιαστούν πιο στοχευμένες πρωθητικές στρατηγικές.

Οι Paramita et al. (2024) πραγματοποίησαν μια συγκριτική ανάλυση των αλγορίθμων συσταδοποίησης K-Means και DBSCAN για την τμηματοποίηση πελατών στον τομέα του ηλεκτρονικού εμπορίου. Η μελέτη αξιολόγησε την απόδοση των δύο αλγορίθμων χρησιμοποιώντας βασικούς δείκτες, όπως ο silhouette score, ο Davies-Bouldin και ο Calinski-Harabasz, με στόχο τον εντοπισμό της πιο αποδοτικής μεθόδου για την κατάτμηση ενός συνόλου δεδομένων πελατών e-commerce. Τα αποτελέσματα έδειξαν ότι ο DBSCAN πέτυχε υψηλότερη τιμή silhouette score (0,680) σε σύγκριση με τον K-Means (0,546), γεγονός που υποδηλώνει καλύτερο διαχωρισμό των συστάδων. Ωστόσο, ο K-Means παρήγαγε πιο ισορροπημένες και καλά καθορισμένες συστάδες, γεγονός που τον καθιστά καταλληλότερο για δομημένες στρατηγικές τμηματοποίησης, όπως η εξαπομικευμένη προώθηση και τα προγράμματα πιστότητας. Από την άλλη πλευρά, ο DBSCAN διακρίθηκε στην αναγνώριση εξειδικευμένων αγορών και στον εντοπισμό μοναδικών καταναλωτικών συμπεριφορών, καθιστώντας τον ιδανικό για τον εντοπισμό ακραίων τιμών και διαφοροποιημένων καταναλωτικών τάσεων.

Η μελέτη κατέληξε στο συμπέρασμα ότι η επιλογή του κατάλληλου αλγορίθμου θα πρέπει να εξαρτάται από τους στόχους της τμηματοποίησης πελατών—ο K-Means είναι προτιμότερος για εφαρμογές που απαιτούν σαφώς καθορισμένες και ισορροπημένες συστάδες, ενώ ο DBSCAN είναι πιο αποτελεσματικός για εταιρείες που επιδιώκουν την αναγνώριση εξειδικευμένων πελατών και της ποικιλομορφίας των καταναλωτικών προτιμήσεων. Οι συγγραφείς πρότειναν ότι η μελλοντική έρευνα θα μπορούσε να επικεντρωθεί σε υβριδικές προσεγγίσεις, οι οποίες θα συνδυάζουν τα πλεονεκτήματα και των δύο αλγορίθμων, καθώς και στην ενσωμάτωση πρόσθετων πηγών δεδομένων (π.χ. αλληλεπιδράσεις σε μέσα κοινωνικής δικτύωσης, αξιολογήσεις πελατών) για τη βελτίωση των στρατηγικών τμηματοποίησης πελατών.

7.2.1 Κατανομές μεταβλητών R,F,M

Η ανάλυση των κατανομών των μετρικών Recency, Frequency και Monetary αποτελεί ένα κρίσιμο προκαταρκτικό βήμα πριν από την αντιμετώπιση των outliers ή την εφαρμογή οποιωνδήποτε μετασχηματισμών δεδομένων. Αυτή η ανάλυση παρέχει πληροφορίες για τη δομή του συνόλου δεδομένων και βοηθά στον εντοπισμό προτύπων ή τάσεων που μπορεί να επηρεάσουν τις μετέπειτα διαδικασίες clustering και segmentation.



Εικόνα 31. Ιστογράμματα κατανομών των RFM μεταβλητών

Οι κατανομές των μετρικών Recency, Frequency, και Monetary (Εικόνα 27) παρέχουν πολύτιμες πληροφορίες για τη συμπεριφορά αγορών των πελατών στο σύνολο δεδομένων. Κάθε μετρική εμφανίζει ξεχωριστά χαρακτηριστικά που αναδεικνύουν διαφορετικά πρότυπα και τάσεις στην πελατειακή βάση.

Η ασυμμετρία (skewness) των μετρικών υπολογίστηκε για την ποσοτικοποίηση της απόκλισης από μια συμμετρική κατανομή. Το skewness είναι ένα στατιστικό μέτρο που αντικατοπτρίζει την

κατεύθυνση και το μέγεθος της ασυμμετρίας στην κατανομή. Θετικό skewness υποδηλώνει ότι η κατανομή έχει μακρύτερη ουρά προς τα δεξιά, ενώ αρνητικό skewness υποδηλώνει ότι η κατανομή έχει μακρύτερη ουρά προς τα αριστερά. Τιμές κοντά στο μηδέν υποδηλώνουν σχεδόν συμμετρικές κατανομές. Οι τιμές skewness για τις τρεις μετρικές ήταν οι εξής:

Η μέτρηση Recency (αριστερό διάγραμμα) εμφάνισε skewness 1.11, υποδεικνύοντας μέτριο θετικό skew. Αυτό σημαίνει ότι οι περισσότεροι πελάτες πραγματοποίησαν πρόσφατα αγορές, ενώ μια μικρότερη ομάδα πελατών έχει πολύ μεγαλύτερες τιμές Recency. Αυτή η συμπεριφορά είναι τυπική σε σύνολα δεδομένων πελατών, όπου ένα μεγάλο μέρος της βάσης παραμένει ενεργό σε σύντομο χρονικό διάστημα, ενώ άλλοι πελάτες είναι ανενεργοί για μεγαλύτερες περιόδους.

Η μέτρηση Frequency παρουσίασε skewness 4.39 (μεσαίο διάγραμμα), υποδεικνύοντας έντονο θετικό skew. Η κατανομή είναι συγκεντρωμένη κοντά στο χαμηλότερο άκρο, καθώς οι περισσότεροι πελάτες πραγματοποίησαν σχετικά λίγες αγορές. Ένα μικρό υποσύνολο ιδιαίτερα δραστήριων πελατών, που χαρακτηρίζονται από πολύ υψηλή συχνότητα αγορών, ευθύνεται για τη μεγάλη ουρά στην κατανομή. Αυτοί οι πελάτες υψηλής συχνότητας είναι πιθανώς πιστοί ή συνήθεις αγοραστές που συνεισφέρουν δυσανάλογα στον αριθμό συναλλαγών.

Η μέτρηση Monetary (δεξιά διάγραμμα) εμφάνισε το υψηλότερο skewness από τις τρεις μετρικές, με τιμή 5.18. Αυτή η ακραία θετική ασυμμετρία υπογραμμίζει ότι ένα μεγάλο μέρος των πελατών παράγει σχετικά χαμηλά έσοδα, ενώ ένα μικρό υποσύνολο πελατών υψηλής αξίας συνεισφέρει δυσανάλογα μεγάλα έσοδα. Αυτή η συμπεριφορά είναι συνήθης σε επιχειρηματικά δεδομένα, όπου οι πελάτες υψηλής αξίας παίζουν κρίσιμο ρόλο στην ενίσχυση των συνολικών εσόδων. Ωστόσο, η παρουσία τέτοιας έντονης ασυμμετρίας υπογραμμίζει τη σημασία της κανονικοποίησης (normalization), ώστε τα αποτελέσματα του clustering και της τμηματοποίησης να μην επηρεάζονται δυσανάλογα από ακραίες τιμές.

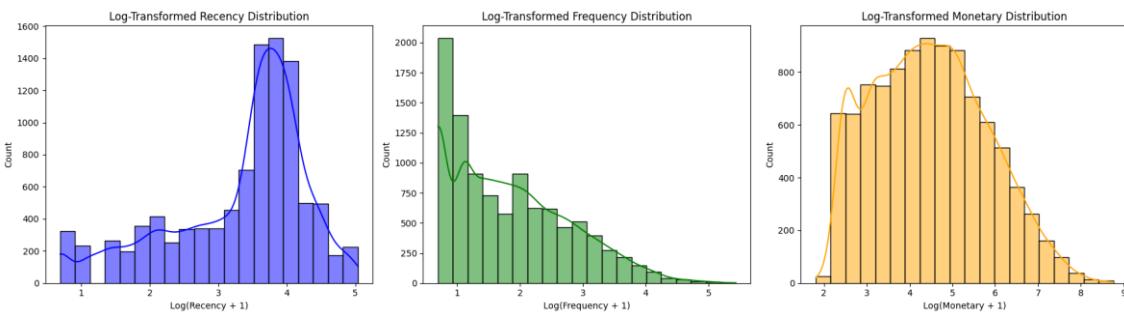
Αυτά τα ευρήματα δείχνουν ότι και οι τρεις μετρικές είναι σημαντικά ασύμμετρες, με τις Frequency και Monetary να εμφανίζουν ιδιαίτερα έντονο θετικό skewness. Για την αντιμετώπιση αυτού του προβλήματος, προτείνονται μετασχηματισμοί όπως οι λογαριθμικοί μετασχηματισμοί (logarithmic transformations) ή οι μετασχηματισμοί Box-Cox, οι οποίοι κανονικοποιούν τις κατανομές και μειώνουν την επίδραση των ακραίων τιμών. Αυτοί οι μετασχηματισμοί θα προετοιμάσουν τα δεδομένα για αξιόπιστη και σταθερή τμηματοποίηση πελατών, διασφαλίζοντας ότι όλες οι μετρικές συνεισφέρουν ισότιμα στη διαδικασία.

7.2.2 Λογαριθμικός μετασχηματισμός

Ο λογαριθμικός μετασχηματισμός είναι μία από τις πιο κοινές τεχνικές που χρησιμοποιούνται για την αντιμετώπιση της ασυμμετρίας (skewness) στα δεδομένα. Με τη συμπίεση της κλίμακας των μεγαλύτερων τιμών περισσότερο από τις μικρότερες, μειώνει αποτελεσματικά την επίδραση των ακραίων τιμών και φέρνει την κατανομή πιο κοντά στη συμμετρία. Αυτός ο μετασχηματισμός είναι ιδιαίτερα αποτελεσματικός για θετικά ασύμμετρα δεδομένα, όπως οι μετρικές Recency, Frequency, και Monetary σε αυτό το σύνολο δεδομένων.

Για την εφαρμογή του λογαριθμικού μετασχηματισμού, κάθε τιμή στο σύνολο δεδομένων μετασχηματίζεται χρησιμοποιώντας τη φυσική λογαριθμική συνάρτηση (natural logarithm). Ωστόσο, επειδή ο λογάριθμος του μηδενός ή των αρνητικών τιμών είναι απροσδιόριστος, προστίθεται μία μικρή σταθερά (π.χ. 1) σε όλες τις τιμές πριν από τον μετασχηματισμό. Αυτό διασφαλίζει ότι όλα τα σημεία δεδομένων είναι έγκυρα για τον υπολογισμό, ενώ παράλληλα διατηρείται η σχετική σειρά των τιμών.

Οι μετασχηματισμένες μετρικές στη συνέχεια οπτικοποιήθηκαν όπως προηγουμένως, με τη χρήση ιστογραμμάτων και διαγραμμάτων πυκνότητας (kernel density plots), ώστε να αξιολογηθεί ο βαθμός μείωσης της ασυμμετρίας. Ο στόχος του μετασχηματισμού ήταν η επίτευξη κατανομών πιο κοντά στη φυσιολογική κατανομή, διευκολύνοντας αμερόληπτο clustering και segmentation. Ο Python κώδικας που χρησιμοποιήθηκε για την εφαρμογή του λογαριθμικού μετασχηματισμού και την οπτικοποίηση των αποτελεσμάτων περιλαμβάνεται στο Παράρτημα.



Εικόνα 32. Ιστογράμματα κατανομών των RFM μεταβλητών κατόπιν του λογαριθμικού μετασχηματισμού

Ο μετασχηματισμός πραγματοποιήθηκε με την εφαρμογή του φυσικού λογαρίθμου (natural logarithm) σε κάθε τιμή, αφού πρώτα προστεθεί 1 για την αποφυγή απροσδιόριστων αποτελεσμάτων για μηδενικές τιμές. Μετά τον μετασχηματισμό, τα αποτελέσματα έδειξαν σημαντική βελτίωση στη συμμετρία των μετρικών, όπως περιγράφεται παρακάτω:

Η μετρική Recency, η οποία προηγουμένως εμφάνιζε υψηλή θετική ασυμμετρία με τιμή skewness 1.11, τώρα έχει skewness -0.94. Αυτό υποδηλώνει μια στροφή προς μια πιο συμμετρική κατανομή, με ελαφρώς αρνητικό skew. Η πλειοψηφία των τιμών συγκεντρώνεται κοντά στο κέντρο του μετασχηματισμένου εύρους, με λιγότερες ακραίες τιμές και στις δύο πλευρές. Αυτή η βελτίωση διασφαλίζει ότι η μετρική δεν θα επηρεάζει δυσανάλογα τα αποτελέσματα του clustering λόγω της ασυμμετρίας.

Η μετρική Frequency παρουσίασε τη μεγαλύτερη βελτίωση στη συμμετρία. Η ασυμμετρία της μειώθηκε από 4.39 σε 0.65, φέρνοντας την κατανομή κοντά στη φυσιολογική κατανομή (normality). Η μετασχηματισμένη κατανομή δείχνει μια πιο ισορροπημένη διασπορά, με την πλειοψηφία των τιμών να συγκεντρώνονται στο χαμηλό έως μεσαίο εύρος. Οι ακραίες τιμές που αντιπροσωπεύουν πελάτες με υψηλή συχνότητα αγορών είναι τώρα πιο μετριασμένες, μειώνοντας την επίδρασή τους στη διαδικασία clustering.

Η μετρική Monetary, η οποία προηγουμένως είχε τη μεγαλύτερη ασυμμετρία με τιμή skewness 5.18, τώρα παρουσιάζει skewness 0.28. Αυτό το αποτέλεσμα υποδηλώνει μια σχεδόν συμμετρική κατανομή, με την πλειοψηφία των τιμών να συγκεντρώνονται στενά κοντά στο κέντρο του μετασχηματισμένου εύρους. Οι ακραίες τιμές εσόδων, που προηγουμένως δημιουργούσαν μακριά ουρά στην αρχική κατανομή, έχουν πλέον συμπιεστεί αποτελεσματικά, καθιστώντας τη μετρική πιο ισορροπημένη και λιγότερο ευαίσθητη στην επίδραση των outliers.

Συνολικά, ο λογαριθμικός μετασχηματισμός μείωσε σημαντικά την ασυμμετρία και στις τρεις μετρικές, φέρνοντάς τες πιο κοντά στη φυσιολογική κατανομή (normality) και καθιστώντας τις πιο κατάλληλες για αλγορίθμους clustering. Αυτή η βελτίωση διασφαλίζει ότι κάθε μετρική συνεισφέρει αναλογικά στη διαδικασία τμηματοποίησης, ενισχύοντας τη σταθερότητα και την αξιοπιστία των αποτελεσμάτων.

Το επόμενο βήμα περιλαμβάνει την εξερεύνηση εναλλακτικών μετασχηματισμών, όπως ο Box-Cox μετασχηματισμός, για να αξιολογηθεί αν μπορούν να επιτευχθούν περαιτέρω βελτιώσεις.

7.2.3 Box-Cox μετασχηματισμός

Ο μετασχηματισμός Box-Cox είναι μια ευέλικτη μέθοδος για την κανονικοποίηση των κατανομών δεδομένων. Σε αντίθεση με τον λογαριθμικό μετασχηματισμό, ο οποίος έχει σταθερή λειτουργική μορφή, ο μετασχηματισμός Box-Cox προσδιορίζει την ιδανική παραμετρική τιμή, γνωστή ως λ , για τη βέλτιστη κανονικοποίηση των δεδομένων. Με τη μεταβολή του λ , ο μετασχηματισμός μπορεί να προσαρμοστεί σε ένα ευρύ φάσμα επιπέδων ασυμμετρίας (skewness), καθιστώντας τον κατάλληλο τόσο για δεδομένα με ήπια όσο και με έντονη ασυμμετρία. Ο μετασχηματισμός ορίζεται

από την παραμετρική τιμή λ , η οποία εκτιμάται από τα δεδομένα και εφαρμόζεται σε κάθε σημείο δεδομένων x , ως εξής:

- **Αν $\lambda \neq 0$:**

$$x^{(\lambda)} = \frac{x^\lambda - 1}{\lambda}$$

- **Αν $\lambda = 0$:**

$$x^{(\lambda)} = \ln(x)$$

, όπου x είναι κάθε σημείο δεδομένων και $x(\lambda)$ είναι το μετασχηματισμένο σημείο δεδομένων.

Ο στόχος του μετασχηματισμού είναι να βρεθεί η ιδανική τιμή λ , η οποία μεγιστοποιεί τη φυσιολογικότητα των μετασχηματισμένων δεδομένων. Αυτή η τιμή συνήθως προσδιορίζεται μέσω βελτιστοποίησης της λογαριθμικής συνάρτησης πιθανότητας (log-likelihood function) της φυσιολογικής κατανομής. Διαφορετικές τιμές του λ μπορούν να χρησιμοποιηθούν για την αντιμετώπιση διαφορετικών τύπων μη φυσιολογικότητας, και μερικές από τις πιο συνηθισμένες είναι:

$\lambda=2$: Μετασχηματίζει τα δεδομένα στο τετράγωνο, χρήσιμο για την αντιμετώπιση δεδομένων με θετική ασυμμετρία.

$\lambda=0.5$: Μετασχηματίζει τα δεδομένα στην τετραγωνική ρίζα, χρήσιμο για δεδομένα με θετική ασυμμετρία.

$\lambda=0$: Μετασχηματίζει τα δεδομένα σε φυσικό λογάριθμο (natural logarithm), χρήσιμο για δεδομένα με θετική ασυμμετρία.

$\lambda=-1$: Μετασχηματίζει τα δεδομένα στο αντίστροφο (inverse), χρήσιμο για δεδομένα με θετική ασυμμετρία.

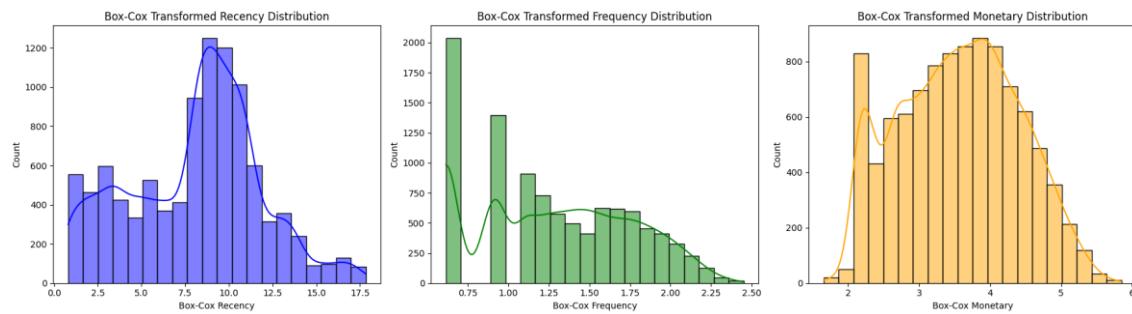
Ο μετασχηματισμός Box-Cox έχει ορισμένους περιορισμούς και παραμέτρους που πρέπει να ληφθούν υπόψη:

Εφαρμοσμότητα: Ο συγκεκριμένος μετασχηματισμός εφαρμόζεται μόνο σε δεδομένα που είναι αυστηρά θετικά, δηλαδή τα δεδομένα πρέπει να είναι μεγαλύτερα από το μηδέν. Αυτό ενδέχεται να απαιτεί προεπεξεργασία δεδομένων.

Ερμηνεία: Ο μετασχηματισμός αλλάζει την κλίμακα των δεδομένων, γεγονός που μπορεί να δυσκολέψει την ερμηνεία των αποτελεσμάτων.

Εκτίμηση Παραμέτρου: Η ιδανική τιμή της παραμέτρου λ πρέπει να εκτιμηθεί, γεγονός που μπορεί να προσθέσει υπολογιστική επιβάρυνση, αν και αυτή είναι συνήθως αμελητέα.

Παρά τους περιορισμούς αυτούς, ο Box-Cox μετασχηματισμός είναι ένα πολύτιμο εργαλείο για την προετοιμασία δεδομένων για πολλά στατιστικά μοντέλα όταν η υπόθεση φυσιολογικότητας είναι σημαντική. Η βέλτιστη επιλογή του λ μπορεί να χρειαστεί επιβεβαίωση με τη χρήση άλλων στατιστικών ελέγχων που αξιολογούν τη φυσιολογικότητα των δεδομένων, όπως ο Kolmogorov-Smirnov test ή ο Shapiro-Wilk test. (Osborne, 2010).



Εικόνα 33. Ιστογράμματα κατανομών των RFM μεταβλητών κατόπιν του Box-Cox μετασχηματισμού

Σε αντίθεση με τον λογαριθμικό μετασχηματισμό, η μέθοδος Box-Cox προσδιορίζει την ιδανική παραμετρική τιμή (λ) για κάθε μεταβλητή, προκειμένου να κανονικοποιήσει αποτελεσματικά τα δεδομένα. Οι τιμές λ υπολογίστηκαν μέσω εκτίμησης μέγιστης πιθανότητας (maximum likelihood estimation), διασφαλίζοντας ότι ο μετασχηματισμός είναι προσαρμοσμένος στα ιδιαίτερα χαρακτηριστικά κάθε μετρικής.

Μετά τον μετασχηματισμό, οι κατανομές των μετρικών Recency, Frequency, και Monetary οπτικοποιήθηκαν με ιστογράμματα, όπως φαίνεται στην Εικόνα 29. Επιπλέον, οι τιμές λ για κάθε μετρική αναφέρθηκαν, παρέχοντας πληροφορίες για τον τύπο του μετασχηματισμού που εφαρμόστηκε. Τα αποτελέσματα περιγράφονται ως εξής:

Η μετρική Recency, με ιδανικό λ 0.43, παρουσίασε σημαντική βελτίωση στη συμμετρία. Ενώ η αρχική κατανομή ήταν θετικά ασύμμετρη, ο μετασχηματισμός Box-Cox την έφερε πιο κοντά στη φυσιολογική κατανομή, με τις περισσότερες τιμές να διανέμονται κοντά στο κέντρο του μετασχηματισμένου εύρους. Η μείωση της ασυμμετρίας διασφαλίζει ότι η μετρική Recency συνεισφέρει πλέον πιο ισορροπημένα στη διαδικασία clustering, αποτρέποντας τις ακραίες τιμές από το να επηρεάζουν δυσανάλογα τα αποτελέσματα.

Η μετρική Frequency, η οποία στην αρχική της μορφή παρουσίαζε σημαντική θετική ασυμμετρία, πέτυχε ιδανικό λ -0.35. Αυτή η τιμή υποδηλώνει έναν ήπιο αντίστροφο μετασχηματισμό (inverse transformation), που συμπίζει αποτελεσματικά την κλίμακα των υψηλότερων τιμών και "απλώνει" τις χαμηλότερες τιμές. Η προκύπτουσα κατανομή είναι αισθητά πιο ισορροπημένη, με τις ακραίες τιμές που αντιπροσωπεύουν πελάτες υψηλής συχνότητας να είναι πλέον πιο μετριασμένες. Αυτός ο μετασχηματισμός διασφαλίζει ότι η επιρροή της μετρικής Frequency στο clustering είναι ανάλογη με τη σημασία της, χωρίς να κυριαρχείται από ένα μικρό υποσύνολο εξαιρετικά δραστήριων πελατών.

Η μετρική Monetary, με ιδανικό λ -0.10, υποβλήθηκε επίσης σε ήπιο αντίστροφο μετασχηματισμό. Η αρχική κατανομή, η οποία ήταν έντονα ασύμμετρη, εμφανίζει τώρα μια πιο συμμετρική μορφή, με τις περισσότερες τιμές να συγκεντρώνονται κοντά στο κέντρο του εύρους. Ο μετασχηματισμός Box-Cox συμπίεσε επιτυχώς τις ακραίες τιμές εσόδων, διασφαλίζοντας ότι οι πελάτες υψηλής αξίας δεν κυριαρχούν στη μετρική. Αυτή η κανονικοποίηση επιτρέπει στη μετρική Monetary να συνεισφέρει αποτελεσματικά στη διαδικασία τμηματοποίησης χωρίς μεροληψία.

Συνολικά, ο μετασχηματισμός Box-Cox μείωσε σημαντικά την ασυμμετρία και στις τρεις μετρικές, βελτιώνοντας την καταλληλότητά τους για clustering και segmentation. Η προσαρμοσμένη φύση του μετασχηματισμού διασφαλίζει ότι κάθε μετρική ρυθμίστηκε βέλτιστα, οδηγώντας σε πιο αξιόπιστα και αμερόληπτα αποτελέσματα. Με τη σύγκριση των επιδράσεων του Box-Cox και του λογαριθμικού μετασχηματισμού, μπορεί να ληφθεί απόφαση σχετικά με τη μέθοδο που κανονικοποιεί καλύτερα τα δεδομένα, διατηρώντας παράλληλα τη δομή τους.

7.2.4 Σύγκριση αποτελεσμάτων Λογαριθμικού και Box-Cox

Λογαριθμικός Μετασχηματισμός:

Ο λογαριθμικός μετασχηματισμός μείωσε επιτυχώς την ασυμμετρία και στις τρεις μετρικές:

Recency Skewness: Μειώθηκε από 1.11 σε -0.94.

Frequency Skewness: Μειώθηκε από 4.39 σε 0.65.

Monetary Skewness: Μειώθηκε από 5.18 σε 0.28.

Αυτή η μέθοδος έφερε τις μετρικές Frequency και Monetary πιο κοντά στη φυσιολογική κατανομή, ενώ η μετρική Recency παρέμεινε ελαφρώς αρνητικά ασύμμετρη. Ο μετασχηματισμός συμπίεσε αποτελεσματικά τις μεγάλες τιμές, αλλά παρουσίασε περιορισμούς στην επίτευξη πλήρους συμμετρίας, ειδικά για τη μετρική Recency.

Μετασχηματισμός Box-Cox:

Ο μετασχηματισμός Box-Cox βελτίωσε περαιτέρω τις κατανομές προσαρμόζοντας τον μετασχηματισμό στα συγκεκριμένα χαρακτηριστικά κάθε μετρικής:

Recency Skewness: Μειώθηκε από 1.11 σε σχεδόν συμμετρική κατανομή, με ιδανικό $\lambda = 0.43$.

Frequency Skewness: Μειώθηκε από 4.39 σε σχεδόν συμμετρική κατανομή, με $\lambda = -0.35$.

Monetary Skewness: Μειώθηκε από 5.18 σε σχεδόν συμμετρική κατανομή, με $\lambda = -0.10$.

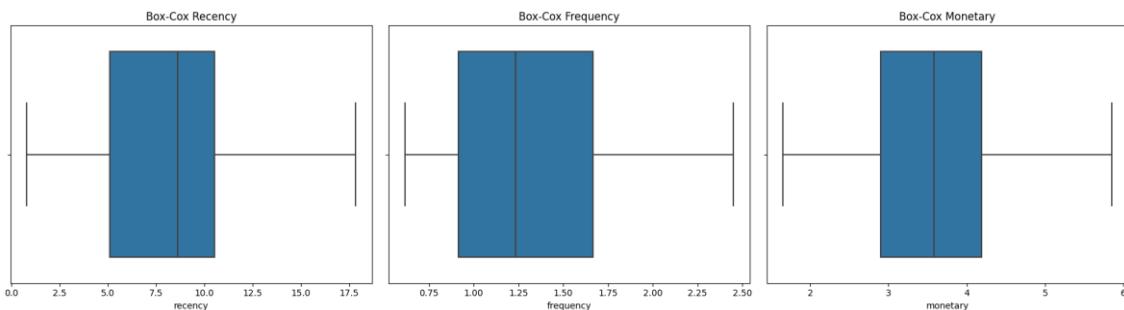
Ο μετασχηματισμός Box-Cox παρήγαγε πιο συμμετρικές κατανομές για όλες τις μετρικές σε σύγκριση με τον λογαριθμικό μετασχηματισμό. Η μετρική Recency, ειδικότερα, πέτυχε σχεδόν πλήρη κανονικοποίηση, κάτι που δεν επιτεύχθηκε πλήρως με τον λογαριθμικό μετασχηματισμό.

Ενώ και οι δύο μετασχηματισμοί μείωσαν σημαντικά την ασυμμετρία, ο μετασχηματισμός Box-Cox παρήγαγε σταθερά πιο συμμετρικές κατανομές σε όλες τις μετρικές. Η δυνατότητα του Box-Cox να προσδιορίζει προσαρμοστικά την ιδανική παραμετρική τιμή (λ) για κάθε μετρική του επιπρέπει να διαχειρίζεται ένα ευρύτερο φάσμα επιπέδων ασυμμετρίας, καθιστώντας τον πιο ευέλικτο από τη σταθερή λειτουργική μορφή του λογαριθμικού μετασχηματισμού.

Βάσει των αποτελεσμάτων, ο μετασχηματισμός Box-Cox συνιστάται για την κανονικοποίηση των μετρικών Recency, Frequency, και Monetary στο παρόν σύνολο δεδομένων. Η δυνατότητά του να επιτυγχάνει σχεδόν συμμετρικές κατανομές διασφαλίζει ότι όλες οι μετρικές συνεισφέρουν ισότιμα στο clustering, βελτιώνοντας την αξιοπιστία και την ερμηνευσιμότητα των αποτελεσμάτων τηματοποίησης. Αν και ο λογαριθμικός μετασχηματισμός είναι πιο απλός και αποδίδει καλά στις περισσότερες περιπτώσεις, η ευελιξία του Box-Cox παρέχει μια πιο ακριβή προσαρμογή για την έντονη ασυμμετρία που παρατηρείται στο σύνολο δεδομένων.

Με την εφαρμογή του μετασχηματισμού Box-Cox, το σύνολο δεδομένων προετοιμάζεται βέλτιστα για αξιόπιστο και αμερόληπτο clustering. Αυτή η επιλογή διασφαλίζει ότι οι ακραίες τιμές δεν κυριαρχούν στην ανάλυση, ενώ η υποκείμενη δομή των δεδομένων διατηρείται.

Μετά την εφαρμογή του μετασχηματισμού Box-Cox πραγματοποιήθηκε ανάλυση ακραίων τιμών (outlier analysis) για την αξιολόγηση της παρουσίας ακραίων δεδομένων. Οι outlier τιμές εντοπίστηκαν χρησιμοποιώντας z-scores, ένα στατιστικό μέτρο που ποσοτικοποιεί την απόσταση ενός σημείου δεδομένων από τον μέσο όρο σε όρους τυπικών αποκλίσεων. Τα σημεία δεδομένων με απόλυτες τιμές z-score μεγαλύτερες του 3 ταξινομήθηκαν ως outliers.



Εικόνα 34. Θηκογράμματα των RFM μεταβλητών κατόπιν του Box-Cox μετασχηματισμού

Τα αποτελέσματα της ανάλυσης ακραίων τιμών έδειξαν ότι δεν ανιχνεύθηκαν outliers σε καμία από τις τρεις μετρικές μετά την εφαρμογή του μετασχηματισμού Box-Cox. Αυτό το εύρημα αναδεικνύει την αποτελεσματικότητα του εν λόγω μετασχηματισμού στην κανονικοποίηση των δεδομένων και τη μείωση της επίδρασης των ακραίων τιμών. Τα αποτελέσματα για κάθε μετρική συνοψίζονται ως εξής:

Πριν από τον μετασχηματισμό, η μετρική Recency παρουσίαζε σημαντική θετική ασυμμετρία (positive skewness), η οποία θα μπορούσε να έχει συμβάλει στην παρουσία ακραίων τιμών. Μετά τον μετασχηματισμό Box-Cox, δεν ανιχνεύθηκαν outliers, καθώς η κατανομή της μετρικής κανονικοποιήθηκε επιτυχώς.

Η αρχική μετρική Frequency ήταν έντονα ασύμμετρη, με λίγους πελάτες να εμφανίζουν εξαιρετικά υψηλή συχνότητα αγορών. Μετά τον μετασχηματισμό, οι ακραίες τιμές συμπιέστηκαν αποτελεσματικά και δεν παρέμειναν outliers στο σύνολο δεδομένων.

Η μετρική Monetary είχε την πιο έντονη ασυμμετρία πριν από τον μετασχηματισμό, αντανακλώντας την παρουσία πελατών υψηλής αξίας. Ο μετασχηματισμός Box-Cox κανονικοποίησε επιτυχώς τη μετρική, εξαλείφοντας όλους τους outliers από την κατανομή.

Η αποτελεσματικότητα του μετασχηματισμού Box-Cox στην αντιμετώπιση της ασυμμετρίας και την απομάκρυνση των outliers υποστηρίζεται περαιτέρω από τα box plots των μετασχηματισμένων μετρικών (Εικόνα 30). Αυτά τα διαγράμματα δείχνουν συμμετρικές κατανομές χωρίς σημαντικές αποκλίσεις, επιβεβαιώνοντας ότι οι μετρικές είναι πλέον κατάλληλα προετοιμασμένες για clustering και segmentation.

Διαπιστώνουμε δηλαδή ότι ο μετασχηματισμός Box-Cox όχι μόνο κανονικοποίησε τις κατανομές των μετρικών Recency, Frequency, και Monetary, αλλά επίσης μείωσε την παρουσία outliers, διασφαλίζοντας ένα πιο ισορροπημένο και αξιόπιστο σύνολο δεδομένων για επόμενες αναλύσεις.

7.2.5 Στατιστικοί έλεγχοι κανονικότητας

Για την αξιολόγηση της αποτελεσματικότητας των μετασχηματισμών που εφαρμόστηκαν, διεξήχθησαν επιπλέον στατιστικοί έλεγχοι κανονικότητας (normality tests). Οι έλεγχοι αυτοί αξιολογούν εάν οι κατανομές των μετασχηματισμένων μετρικών αποκλίνουν σημαντικά από μια φυσιολογική κατανομή. Η διασφάλιση της κανονικότητας είναι κρίσιμη, καθώς πολλοί αλγόριθμοι clustering υποθέτουν ότι οι μεταβλητές εισόδου ακολουθούν φυσιολογική κατανομή.

- Ο έλεγχος Shapiro-Wilk είναι μια ευρέως χρησιμοποιούμενη μέθοδος για τον έλεγχο κανονικότητας:
- Ο έλεγχος Kolmogorov-Smirnov συγκρίνει την κατανομή των δεδομένων με μια φυσιολογική κατανομή:
- Ο έλεγχος D'Agostino and Pearson's συνδυάζει ασυμμετρία (skewness) και κύρτωση (kurtosis) για την αξιολόγηση της κανονικότητας.

Τα αποτελέσματα των ελέγχων κανονικότητας για τις μετασχηματισμένες ως άνω μετρικές δείχνουν εάν τα δεδομένα ακολουθούν φυσιολογική κατανομή. Κάθε έλεγχος αξιολογεί τη μηδενική υπόθεση (H_0) ότι τα δεδομένα είναι φυσιολογικά κατανεμημένα. Ένα p-value μεγαλύτερο από 0,05 υποστηρίζει τη μηδενική υπόθεση, υποδηλώνοντας ότι δεν υπάρχει σημαντική απόκλιση από τη φυσιολογική κατανομή. Αντίθετα, ένα p-value μικρότερο από 0,05 υποδηλώνει ότι τα δεδομένα αποκλίνουν σημαντικά από τη φυσιολογική κατανομή. Κατόπιν της εφαρμογής των tests, προκύπτουν τα εξής:

Recency

Shapiro-Wilk Test: Statistic = 0.9740, p-value < 0.001

Kolmogorov-Smirnov Test: Statistic = 0.1048, p-value < 0.001

D'Agostino-Pearson Test: Statistic = 153.8225, p-value < 0.001

Και οι τρεις έλεγχοι δίνουν p-values μικρότερα από 0,05, υποδεικνύοντας ότι η μετασχηματισμένη μετρική Recency αποκλίνει σημαντικά από τη φυσιολογική κατανομή. Παρότι ο μετασχηματισμός Box-Cox βελτίωσε τη συμμετρία της κατανομής, η μετρική Recency παραμένει ελαφρώς μη φυσιολογική.

Frequency:

Shapiro-Wilk Test: Statistic = 0.9401, p-value < 0.001

Kolmogorov-Smirnov Test: Statistic = 0.1206, p-value < 0.001

D'Agostino-Pearson Test: Statistic = 3699.2678, p-value < 0.001

Οι έλεγχοι δείχνουν ότι η μετρική Frequency επίσης αποκλίνει σημαντικά από τη φυσιολογική κατανομή μετά τον μετασχηματισμό. Πάλι, παρόλο που ο μετασχηματισμός Box-Cox μείωσε σημαντικά την ασυμμετρία, τα δεδομένα παραμένουν μέτρια μη φυσιολογικά.

Monetary:

Shapiro-Wilk Test: Statistic = 0.9809, p-value < 0.001

Kolmogorov-Smirnov Test: Statistic = 0.0416, p-value < 0.001

D'Agostino-Pearson Test: Statistic = 987.3736, p-value < 0.001

Παρόμοια με τις άλλες μετρικές, η μετρική Monetary αποκλίνει σημαντικά από τη φυσιολογική κατανομή βάσει και των τριών ελέγχων. Ο μετασχηματισμός Box-Cox βελτίωσε την κατανομή της, αλλά παραμένει ελαφρώς μη φυσιολογική.

Τα αποτελέσματα δείχνουν ότι καμία από τις μετρικές δεν επιτυγχάνει πλήρη κανονικότητα, ακόμη και μετά τον μετασχηματισμό Box-Cox. Αυτό είναι αναμενόμενο σε σύνολα δεδομένων πραγματικού κόσμου, καθώς η τέλεια κανονικότητα είναι σπάνια επιτεύξιμη, ιδιαίτερα σε μετρικές που επηρεάζονται από τη συμπεριφορά των πελατών, η οποία συχνά παρουσιάζει εγγενή μεταβλητότητα.

Οστόσο, ο μετασχηματισμός μείωσε επιτυχώς την ασυμμετρία και έφερε τις κατανομές πιο κοντά στη φυσιολογική κατανομή, όπως παρατηρήθηκε από προηγούμενες οπτικοποιήσεις και την ανάλυση ακραίων τιμών. Δεδομένου ότι πολλοί αλγόριθμοι clustering (π.χ., K-Means) δεν απαιτούν αυστηρά πλήρη κανονικότητα, τα μετασχηματισμένα δεδομένα εξακολουθούν να είναι κατάλληλα για τμηματοποίηση.

Αν η κανονικότητα είναι κρίσιμη απαίτηση για περαιτέρω αναλύσεις (π.χ., στη χρήση στατιστικών ελέγχων που βασίζονται σε δεδομένα με φυσιολογική κατανομή), θα μπορούσαν να εξεταστούν επιπλέον βήματα, όπως η εφαρμογή μη παραμετρικών μεθόδων, η διερεύνηση άλλων

μετασχηματισμών ή η χρήση τεχνικών όπως η ανάλυση κύριων συνιστώσων (PCA) για μείωση των διαστάσεων.

Ο τρέχων μετασχηματισμός θεωρείται επαρκής για σκοπούς clustering, καθώς επιτυγχάνει ισορροπία μεταξύ της μείωσης της ασυμμετρίας και της διατήρησης της ερμηνευσιμότητας των δεδομένων.

ΚΕΦΑΛΑΙΟ 8. Συσταδοποίηση

Η συσταδοποίηση, ή αλλιώς clustering, αποτελεί κρίσιμο βήμα στην τμηματοποίηση πελατών, καθώς επιτρέπει τον εντοπισμό διακριτών ομάδων εντός της πελατειακής βάσης, βάσει της αγοραστικής τους συμπεριφοράς. Μέσω της ανάλυσης των μετρικών RFM, το clustering μπορεί να αποκαλύψει εφαρμόσιμες πληροφορίες, όπως τον εντοπισμό πελατών υψηλής αξίας, πελατών που διατρέχουν κίνδυνο απώλειας, και πιστών αγοραστών. Για να διασφαλιστεί η αποτελεσματικότητα και η αξιοπιστία της διαδικασίας, ακολουθείται η παρακάτω μεθοδολογία:

1. Τυποποίηση Δεδομένων (Standardization)

Πριν από την εφαρμογή αλγορίθμων clustering, είναι απαραίτητο να τυποποιηθούν τα δεδομένα. Η τυποποίηση διασφαλίζει ότι όλες οι μεταβλητές συμβάλλουν εξίσου στη διαδικασία τμηματοποίησης, μετασχηματίζοντάς τις ώστε να έχουν μέσο όρο 0 και τυπική απόκλιση 1. Αυτό το βήμα είναι ιδιαίτερα σημαντικό για τις μετρικές RFM, καθώς οι κλίμακές τους διαφέρουν σημαντικά (π.χ., η Recency μετριέται σε ημέρες, ενώ η Monetary σε ευρώ). Χωρίς την τυποποίηση, οι μεταβλητές με μεγαλύτερες κλίμακες θα κυριαρχούσαν στη διαδικασία clustering, οδηγώντας σε μεροληπτικά αποτελέσματα. Η εφαρμογή της τυποποίησης αποτυπώνεται στον κώδικα που συμπεριλαμβάνεται στο Παράρτημα.

2. Προσδιορισμός Βέλτιστου Αριθμού Ομάδων

Το επόμενο βήμα περιλαμβάνει τον προσδιορισμό του βέλτιστου αριθμού clusters (k) για το σύνολο δεδομένων. Αυτό διασφαλίζει ότι τα αποτελέσματα του clustering είναι τόσο ουσιαστικά όσο και ερμηνεύσιμα. Ως προς αυτό, θα εφαρμόσουμε δύο ευρέως χρησιμοποιούμενες μεθόδους:

- *Elbow Method*: Η μέθοδος Elbow αξιολογεί το άθροισμα τετραγώνων εντός των clusters (Within-Cluster Sum of Squares, WCSS), το οποίο μετρά πόσο στενά είναι ομαδοποιημένα τα σημεία δεδομένων εντός κάθε cluster. Με τη σχεδίαση του WCSS σε σχέση με διαφορετικές τιμές του k , το "σημείο αγκώνα" (elbow point) εντοπίζεται εκεί όπου η περιθωριακή μείωση του WCSS μειώνεται σημαντικά. Αυτό το σημείο αντιπροσωπεύει τον βέλτιστο αριθμό clusters.
- *Silhouette Analysis*: Η ανάλυση Silhouette αξιολογεί πόσο καλά εντάσσονται τα σημεία δεδομένων στα clusters που έχουν ανατεθεί, σε σύγκριση με άλλα clusters. Παρέχει μια βαθμολογία που κυμαίνεται από -1 έως 1, όπου οι υψηλότερες βαθμολογίες υποδεικνύουν καλύτερα ορισμένα clusters. Ο Silhouette Score θα υπολογιστεί για διαφορετικές τιμές του k ώστε να επικυρωθούν τα αποτελέσματα της μεθόδου Elbow.

3. Εφαρμογή του αλγορίθμου K-Means Clustering με οπτικοποίηση

Ο αλγόριθμος K-Means clustering θα εφαρμοστεί για την τμηματοποίηση των πελατών. Ο K-Means είναι ένας επαναληπτικός αλγόριθμος που αναθέτει σημεία δεδομένων σε k clusters, ελαχιστοποιώντας τη διακύμανση εντός κάθε cluster. Είναι υπολογιστικά αποδοτικός και ευρέως χρησιμοποιούμενος στην τμηματοποίηση πελατών λόγω της απλότητας και της επεκτασιμότητάς του. Τα παρακάτω βήματα θα πραγματοποιηθούν:

- Αρχικοποίηση των κεντροειδών (centroids) των clusters με τη μέθοδο Forgy:

Η μέθοδος Forgy είναι μια τεχνική για την αρχικοποίηση των κεντροειδών στον αλγόριθμο K-Means. Επιλέγει ή κατανομή σημείων δεδομένων από το σύνολο δεδομένων ως αρχικά κεντροειδή. Αυτή η μέθοδος διασφαλίζει ότι τα κεντροειδή είναι πραγματικά σημεία δεδομένων, παρέχοντας μια εύλογη αρχική θέση για τη διαδικασία clustering. Η χρήση της μεθόδου Forgy μπορεί να αποτρέψει κακή αρχικοποίηση, η οποία μπορεί να οδηγήσει σε υποβελτιστικά αποτελέσματα clustering ή αργή σύγκλιση.

- Ανάθεση κάθε πελάτη στο πλησιέστερο κεντροειδές βάσει των μετρικών RFM: Οι πελάτες θα κατηγοριοποιηθούν στα clusters με βάση την εγγύτητα των μετρικών Recency, Frequency, και Monetary στα κεντροειδή.
- Επαναληπτική προσαρμογή των κεντροειδών μέχρι να επιτευχθεί σύγκλιση: Τα κεντροειδή θα αναπροσαρμόζονται επαναληπτικά, καθώς οι πελάτες ανακατανέμονται στα clusters, μέχρι η διαδικασία να συγκλίνει, δηλαδή να σταματήσουν οι αλλαγές στις αναθέσεις.

Στη συνέχεια, η οπτικοποίηση των clusters θα χρησιμεύσει για την επικύρωση του διαχωρισμού τους και την κατανόηση της κατανομής τους. Θα χρησιμοποιηθούν τεχνικές μείωσης διαστάσεων, όπως η Principal Component Analysis (PCA), με σκοπό την προβολή των δεδομένων σε 2D ή 3D χώρο για την οπτικοποίησή τους.

4. Εφαρμογή του αλγορίθμου DBSCAN με οπτικοποίηση

Για την επικύρωση των αποτελεσμάτων clustering που προέκυψαν μέσω του K-Means, θα εφαρμοστεί ένας επιπλέον αλγόριθμος clustering: DBSCAN (Density-Based Spatial Clustering of Applications with Noise), με αντίστοιχη οπτικοποίηση. Σε αντίθεση με τον K-Means, ο DBSCAN δεν απαιτεί τον προκαθορισμό του αριθμού clusters (k) και μπορεί να εντοπίσει clusters με διαφορετικά σχήματα και μεγέθη. Ο συγκεκριμένος αλγόριθμος λειτουργεί ομαδοποιώντας σημεία δεδομένων που βρίσκονται πικνά συγκεντρωμένα και χαρακτηρίζοντας τα σημεία σε περιοχές χαμηλής πυκνότητας ως θόρυβο (noise). Αυτή η ιδιότητα καθιστά τον DBSCAN ιδιαίτερα αποτελεσματικό για σύνολα δεδομένων με μη γραμμικά όρια clusters ή παρουσία outliers.

Η σύγκριση των αποτελεσμάτων μεταξύ K-Means και DBSCAN θα βοηθήσει στην αξιολόγηση της συνέπειας και της ανθεκτικότητας των αποτελεσμάτων τμηματοποίησης. Αν τα clusters που εντοπίζονται και από τις δύο μεθόδους ευθυγραμμίζονται στενά, αυτό θα ενισχύσει την εμπιστοσύνη στην ανάλυση. Επιπλέον, η ικανότητα του DBSCAN να εντοπίζει σημεία θορύβου μπορεί να προσφέρει περαιτέρω πληροφορίες για την ανώμαλη συμπεριφορά πελατών.

5. Ερμηνεία και Κατηγοριοποίηση των Clusters

Η διαδικασία clustering στοχεύει στην τμηματοποίηση της πελατειακής βάσης σε καλά ορισμένες ομάδες βάσει των μετρικών RFM. Αυτά τα τμήματα θα παρέχουν εφαρμόσιμες πληροφορίες για στοχευμένες στρατηγικές μάρκετινγκ, προγράμματα διατήρησης πελατών, και κατανομή πόρων.

Ακολουθώντας αυτήν τη δομημένη μεθοδολογία, η διαδικασία clustering θα διασφαλίσει τόσο στατιστική ακρίβεια όσο και επιχειρηματική συνάφεια, επιτρέποντας τη λήψη αποφάσεων που βασίζονται στα δεδομένα για τη διαχείριση πελατών. Αφότου δημιουργηθούν τα clusters, τα χαρακτηριστικά τους θα αναλυθούν για την εξαγωγή ουσιαστικών συμπερασμάτων. Αυτή η διαδικασία περιλαμβάνει:

- Υπολογισμό των μέσων τιμών των Recency, Frequency, και Monetary για κάθε cluster: Ο υπολογισμός αυτών των μέσων τιμών θα αποκαλύψει τις διαφορές στη συμπεριφορά των πελατών μεταξύ των clusters.
- Ερμηνεία των μοτίβων συμπεριφοράς των πελατών σε κάθε cluster: Η ανάλυση των δεδομένων θα βοηθήσει στον εντοπισμό χαρακτηριστικών, όπως πελάτες υψηλής αξίας, πελάτες που διατρέχουν κίνδυνο απώλειας, ή πιστοί πελάτες.
- Ανάθεση περιγραφικών ετικετών στα clusters: Για την ενίσχυση της επιχειρηματικής κατανόησης, τα clusters θα κατηγοριοποιηθούν με περιγραφικές ετικέτες, όπως "Loyalists" (Πιστοί πελάτες) ή "Churners" (Χαμένοι πελάτες).

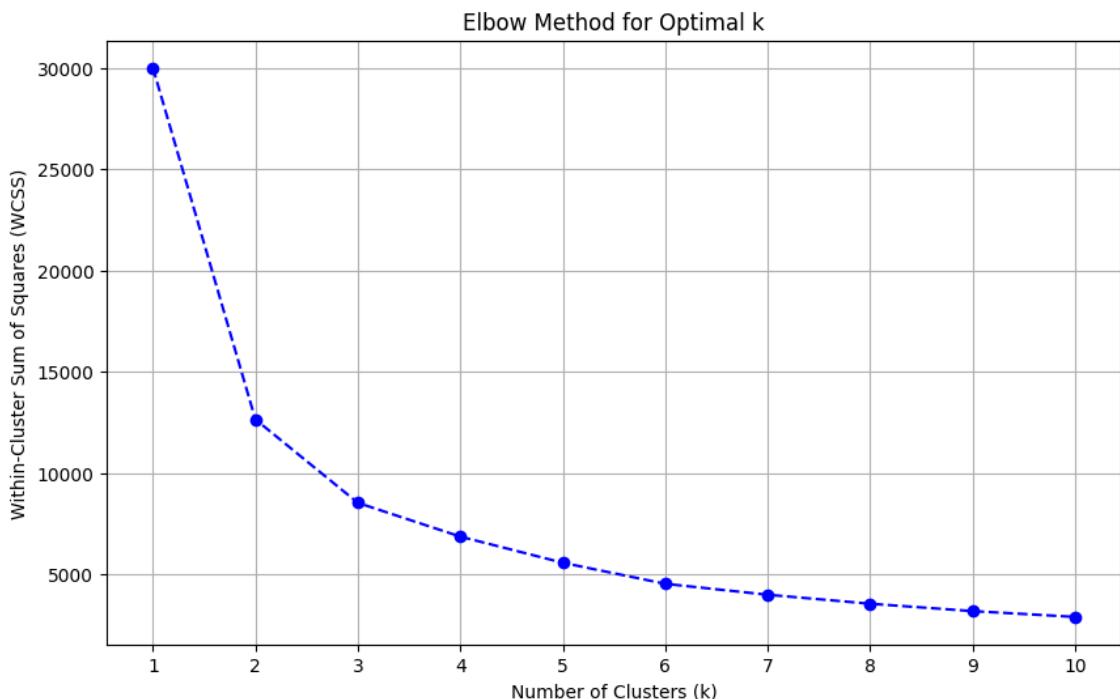
8.1 Προσδιορισμός βέλτιστου αριθμού ομάδων

8.1.1 Elbow method

Έχοντας εφαρμόσει αρχικά την τυποποίηση δεδομένων στον κώδικα, προχωρούμε στον προσδιορισμό του ιδανικού αριθμού clusters ξεκινώντας από τη μέθοδος Elbow. Αυτή είναι μια ευρέως χρησιμοποιούμενη τεχνική για τον καθορισμό του βέλτιστου αριθμού clusters (k) σε ένα σύνολο δεδομένων. Η μέθοδος αξιολογεί το Within-Cluster Sum of Squares (WCSS), το οποίο μετρά τη συμπύκνωση των clusters υπολογίζοντας το άθροισμα των τετραγώνων των αποστάσεων μεταξύ κάθε σημείου δεδομένων και του κεντροειδούς του. Χαμηλότερες τιμές WCSS υποδεικνύουν πιο σφιχτά clusters.

Η μέθοδος περιλαμβάνει τη σχεδίαση του WCSS σε σχέση με διαφορετικές τιμές του k . Αρχικά, καθώς το k αυξάνεται, το WCSS μειώνεται σημαντικά, καθώς η προσθήκη περισσότερων clusters μειώνει τη μέση απόσταση μεταξύ των σημείων δεδομένων και των κεντροειδών τους. Ωστόσο, μετά από ένα συγκεκριμένο σημείο, ο ρυθμός μείωσης επιβραδύνεται, σχηματίζοντας ένα "σχήμα αγκώνα" (elbow) στο γράφημα. Αυτό το "σημείο αγκώνα" αντιπροσωπεύει τον βέλτιστο αριθμό clusters, όπου η προσθήκη περισσότερων clusters παρέχει φθίνουσες αποδόσεις στη βελτίωση της συμπύκνωσης.

Στην παρούσα ανάλυση, η Μέθοδος Elbow εφαρμόστηκε στο τυποποιημένο σύνολο δεδομένων μετρικών RFM. Το WCSS υπολογίστηκε για τιμές k που κυμαίνονταν από 1 έως 10, και τα αποτελέσματα σχεδιάστηκαν για τον εντοπισμό του βέλτιστου k .



Εικόνα 35. Βέλτιστος αριθμός συστάδων βάσει της μεθόδου «Elbow»

Το γράφημα παραπάνω (Εικόνα 30) απεικονίζει το Within-Cluster Sum of Squares (WCSS) σε σχέση με διαφορετικές τιμές του k , που κυμαίνονται από 1 έως 10. Το WCSS αντιπροσωπεύει το άθροισμα των τετραγώνων των αποστάσεων μεταξύ κάθε σημείου δεδομένων και του κεντροειδούς του, με χαμηλότερες τιμές να υποδεικνύουν πιο συμπαγή clusters.

Η καμπύλη στο γράφημα της Μεθόδου Elbow δείχνει μια απότομη μείωση του WCSS καθώς το k αυξάνεται από 1 σε 3, ακολουθούμενη από μια πιο σταδιακή μείωση για μεγαλύτερες τιμές του k . Αυτή η συμπεριφορά αντικατοπτρίζει τη μείωση των αποδόσεων από την προσθήκη περισσότερων clusters, καθώς κάθε επιπλέον cluster συμβάλλει λιγότερο στη μείωση του WCSS.

Το "σημείο αγκώνα" (elbow point) στο γράφημα, όπου ο ρυθμός μείωσης του WCSS επιβραδύνεται σημαντικά, παρατηρείται στο $k=3$. Αυτό υποδηλώνει ότι η διαίρεση του συνόλου δεδομένων σε τρία clusters επιτυγχάνει μια ισορροπία μεταξύ συμπαγούς clustering (χαμηλό WCSS) και απλότητας (λιγότερα clusters). Πέρα από το $k=3$, η προσθήκη περισσότερων clusters οδηγεί μόνο σε οριακές βελτιώσεις του WCSS, καθιστώντας τα λιγότερο χρήσιμα για πρακτικούς σκοπούς. Βάσει της ανάλυσης της Μεθόδου Elbow, ο βέλτιστος αριθμός clusters για το σύνολο δεδομένων είναι $k=3$. Αυτή η τιμή διασφαλίζει ουσιαστική τμηματοποίηση, αποφεύγοντας περιπτή πολυπλοκότητα.

8.1.2 Silhouette Analysis

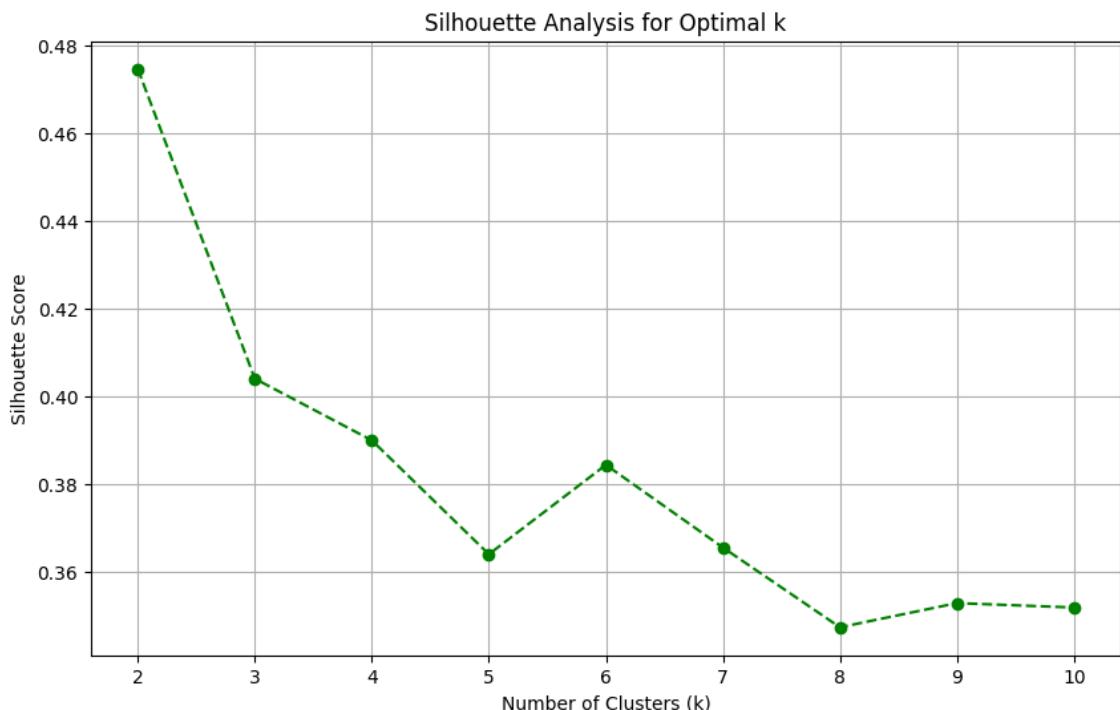
Το επόμενο βήμα περιλαμβάνει την επικύρωση αυτού του αποτελέσματος μέσω της Silhouette Analysis, για να επιβεβαιωθεί ότι τα τρία clusters παρέχουν τον καλύτερα ορισμένο διαχωρισμό των δεδομένων. Η Ανάλυση Silhouette είναι μια ευρέως χρησιμοποιούμενη μέθοδος για την αξιολόγηση της ποιότητας του clustering και τον προσδιορισμό του βέλτιστου αριθμού clusters (k). Η μέθοδος υπολογίζει τη Silhouette Score, η οποία μετρά πόσο καλά εντάσσεται κάθε σημείο δεδομένων στο cluster που του έχει ανατεθεί, σε σύγκριση με άλλα clusters. Η βαθμολογία κυμαίνεται μεταξύ -1 και 1:

Βαθμολογίες κοντά στο +1: Υποδεικνύουν ότι το σημείο δεδομένων ταιριάζει καλά με το cluster του και ταιριάζει ελάχιστα με γειτονικά clusters.

Βαθμολογίες κοντά στο 0: Υποδηλώνουν ότι το σημείο δεδομένων βρίσκεται κοντά στο όριο μεταξύ clusters.

Βαθμολογίες κοντά στο -1: Υποδεικνύουν ότι το σημείο δεδομένων μπορεί να έχει ταξινομηθεί λανθασμένα, καθώς ταιριάζει καλύτερα σε ένα γειτονικό cluster.

Η μέση Silhouette Score για όλα τα σημεία δεδομένων παρέχει ένα συνολικό μέτρο για το πόσο καλά ορισμένα είναι τα clusters. Με τον υπολογισμό της Silhouette Score για διαφορετικές τιμές του k , μπορεί να προσδιοριστεί ο βέλτιστος αριθμός clusters ως η τιμή k με τη μεγαλύτερη μέση βαθμολογία. Η ανάλυση αυτή συμπληρώνει τη Μέθοδο Elbow αξιολογώντας τον διαχωρισμό και τη συμπύκνωση των clusters, διασφαλίζοντας ότι η επιλεγμένη τιμή k παρέχει καλά ορισμένες ομάδες.



Εικόνα 36. Βέλτιστος αριθμός συστάδων βάσει της μεθόδου «Silhouette»

Η Ανάλυση Silhouette διεξήχθη για την αξιολόγηση της ποιότητας του clustering για διαφορετικές τιμές του αριθμού clusters (k). Το γράφημα της Εικόνας 31 απεικονίζει το μέσο Silhouette Score για κάθε τιμή k, που κυμαίνεται από 2 έως 10. Το σκορ μετρά πόσο καλά εντάσσονται τα σημεία δεδομένων στα clusters που τους έχουν ανατεθεί, σε σύγκριση με τα γειτονικά clusters, με υψηλότερες βαθμολογίες να υποδεικνύουν καλύτερα ορισμένα clusters.

Η υψηλότερη Silhouette Score, περίπου 0.48, παρατηρείται στο k=2. Αυτό δείχνει ότι το σύνολο δεδομένων επιτυγχάνει το καλύτερα ορισμένο clustering όταν διαιρείται σε δύο clusters. Ωστόσο, η χρήση μόνο του k=2 μπορεί να απλοποιήσει υπερβολικά την τμηματοποίηση, καθώς μπορεί να μην συλλάβει τις λεπτομερείς αποχρώσεις της συμπεριφοράς των πελατών. Η βαθμολογία μειώνεται σημαντικά καθώς το k αυξάνεται από 2 σε 3, με βαθμολογία περίπου 0.40 στο k=3. Πέρα από το k=3, οι βαθμολογίες συνεχίζουν να μειώνονται σταδιακά, φτάνοντας σε ένα σχετικά σταθερό επίπεδο γύρω στο 0.36 για k≥6.

Η Μέθοδος Elbow που εφαρμόσαμε προηγουμένως, προσδιόρισε το k=3 ως τον βέλτιστο αριθμό clusters βάσει του WCSS. Παρόλο που το k=3 έχει χαμηλότερη Silhouette Score συγκριτικά με το k=2, εξακολουθεί να αντιπροσωπεύει μια ισορροπία μεταξύ της συμπύκνωσης των clusters και της ερμηνευσιμότητας. Η ανάλυση Silhouette υποδεικνύει ότι το σύνολο δεδομένων είναι πιο συμπαγώς ομαδοποιημένο όταν k=2, αλλά το k=3 παραμένει ένας ισχυρός υποψήφιος για πρακτική τμηματοποίηση, λόγω της ερμηνευσιμότητάς του και της συμφωνίας του με τη Μέθοδο Elbow. Βάσει αυτών των αποτελεσμάτων, το k=3 είναι ένας αριθμός που μπορεί να εξεταστεί για το clustering του συγκεκριμένου dataset, καθώς προσφέρει έναν συμβιβασμό μεταξύ καλά ορισμένων clusters και επιχειρηματικής συνάφειας.

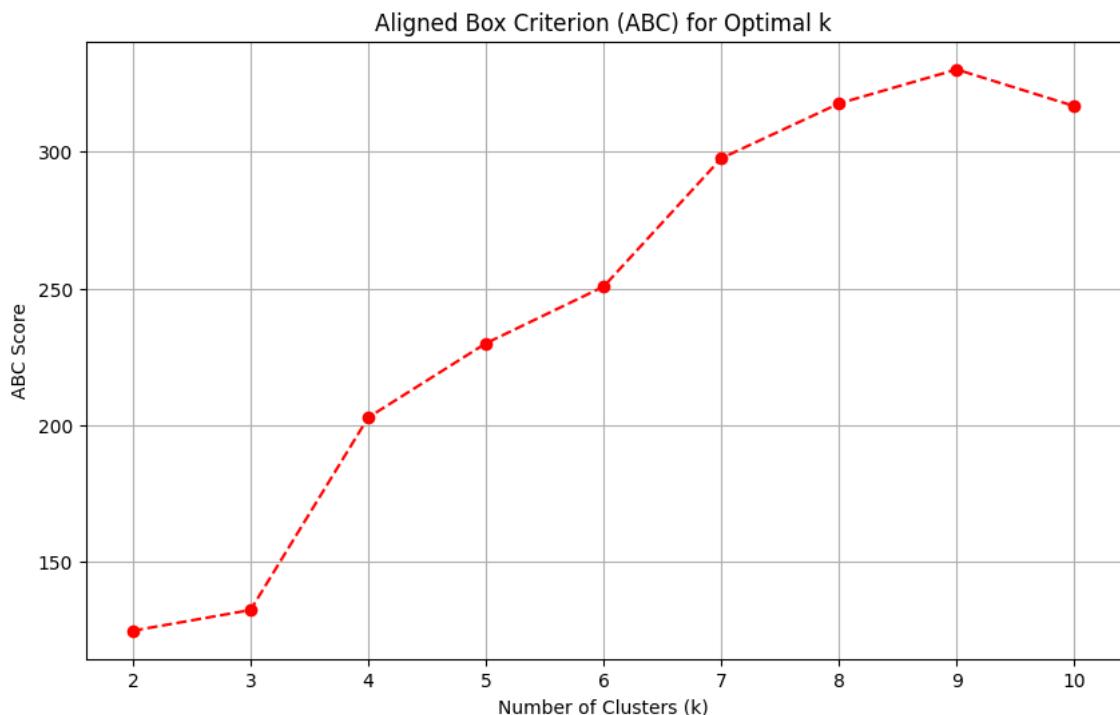
8.1.3 Αλγόριθμος Aligned Box Criterion (ABC)

Για περαιτέρω επικύρωση των αποτελεσμάτων, θα εφαρμοστεί ένας εναλλακτικός αλγόριθμος, όπως ο DBSCAN, για την επιβεβαίωση της συνέπειας των εντοπισμένων ομάδων. Στο μεταξύ όμως, για την περαιτέρω επικύρωση των αποτελεσμάτων, θα εφαρμοστεί επίσης η μέθοδος Customer Segmentation & Market Basket Analysis

Aligned Box Criterion (ABC). Αυτός ο αλγόριθμος είναι μια τεχνική επικύρωσης clustering που αξιολογεί την ποιότητα των clusters εξετάζοντας το πόσο καλά ευθυγραμμίζονται και διαχωρίζονται. Τα βασικά χαρακτηριστικά της μεθόδου ABC περιλαμβάνουν:

- **Alignment:** Μετρά πόσο στενά είναι ομαδοποιημένα τα σημεία δεδομένων εντός των clusters που τους έχουν ανατεθεί. Εξασφαλίζει ότι τα clusters είναι συμπαγή και κεντραρισμένα γύρω από τα αντίστοιχα κεντροειδή τους.
- **Separation:** Αξιολογεί τον βαθμό διαχωρισμού μεταξύ των clusters. Τα έγκυρα clusters πρέπει να έχουν σαφή όρια, με ελάχιστη επικάλυψη μεταξύ γειτονικών ομάδων.
- **Score:** Ο αλγόριθμος ABC παράγει μια συνολική βαθμολογία που συνδυάζει τις δύο παραπάνω πτυχές, επιτρέποντας μια ολοκληρωμένη αξιολόγηση της ποιότητας των clusters.

Εφαρμόζοντας τον αλγόριθμο ABC στα αποτελέσματα του clustering, μπορούμε να επικυρώσουμε την αποτελεσματικότητα του $k=3$ (ή άλλων τιμών k) και να συγκρίνουμε την απόδοσή του με άλλες πιθανές διαμορφώσεις clusters.



Εικόνα 37. Βέλτιστος αριθμός συστάδων βάσει της μεθόδου «ABC»

Το ανωτέρω γράφημα (Εικόνα 32) απεικονίζει τις βαθμολογίες ABC σε σχέση με διαφορετικές τιμές k , που κυμαίνονται από 2 έως 10. Η βαθμολογία ABC αξιολογεί την ποιότητα του clustering βάσει της ευθυγράμμισης (compactness of clusters) και του διαχωρισμού (distinct boundaries between clusters). Υψηλότερη βαθμολογία ABC υποδεικνύει καλύτερα ορισμένα clusters. Η βαθμολογία ABC παρουσιάζει σταθερή αύξηση από το $k=2$ έως το $k=9$. Αυτό δείχνει ότι η αύξηση του αριθμού των clusters βελτιώνει την ποιότητα του clustering, δημιουργώντας μικρότερες, πιο συμπαγείς ομάδες με καλύτερο διαχωρισμό.

Η υψηλότερη βαθμολογία ABC παρατηρείται στο $k=9$, υποδεικνύοντας ότι η διαμόρφωση των εννέα clusters προσφέρει την καλύτερη ισορροπία μεταξύ ευθυγράμμισης και διαχωρισμού. Ωστόσο, η βελτίωση της βαθμολογίας ABC από το $k=8$ στο $k=9$ είναι οριακή, υποδηλώνοντας ότι

οι αποδόσεις μειώνονται με την προσθήκη επιπλέον clusters. Ο ρυθμός αύξησης της βαθμολογίας ABC επιβραδύνεται σημαντικά μετά το k=7, υποδηλώνοντας ότι οι επιπλέον clusters προσφέρουν λιγότερη συνεισφορά στη βελτίωση της ποιότητας του clustering. Παρότι το k=9 επιτυγχάνει την υψηλότερη βαθμολογία ABC, η επιλογή ενός μικρότερου k, όπως το k=7, μπορεί να προσφέρει μια πιο πρακτική και ερμηνεύσιμη λύση για την τμηματοποίηση πελατών. Αυτή η επιλογή ευθυγραμμίζεται με την αρχή της ισορροπίας μεταξύ στατιστικής απόδοσης και επιχειρηματικής συνάφειας.

Η διαδικασία προσδιορισμού του βέλτιστου αριθμού clusters για την τμηματοποίηση πελατών ανέδειξ σημαντικές διαφορές μεταξύ των αποτελεσμάτων των τριών μεθόδων επικύρωσης: Μέθοδος Elbow, Ανάλυση Silhouette, και Aligned Box Criterion (ABC). Κάθε μέθοδος αξιολογεί την ποιότητα του clustering βάσει διαφορετικών κριτηρίων, γεγονός που εξηγεί τις παραπομβές αποκλίσεις.

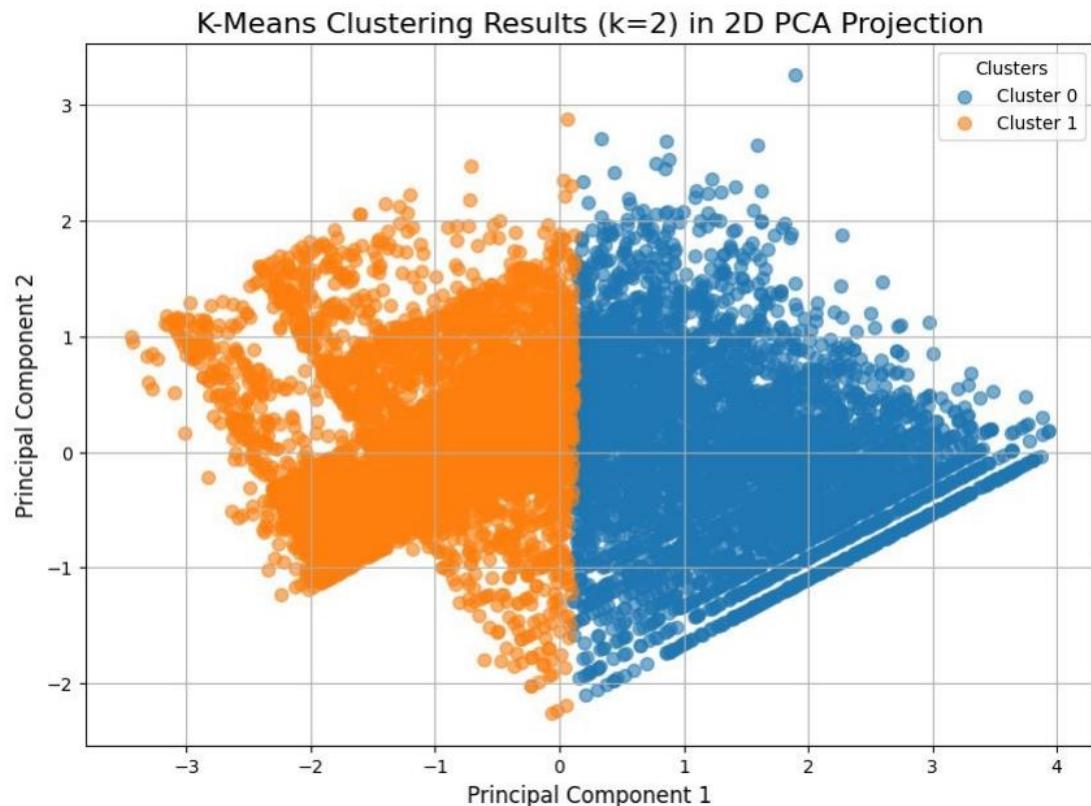
- Η Μέθοδος Elbow πρότεινε το k=3 ως τον βέλτιστο αριθμό clusters. Αυτό είναι το σημείο όπου η μείωση του WCSS αρχίζει να μειώνεται, υποδεικνύοντας μια ισορροπία μεταξύ συμπαγών clusters και απλότητας.
- Η Ανάλυση Silhouette πρότεινε το k=2 ως την καλύτερη επιλογή, καθώς πέτυχε τη μέγιστη μέση Silhouette Score. Αυτό το αποτέλεσμα αντικατοπτρίζει έναν ισχυρό διαχωρισμό μεταξύ clusters, αν και το k=3 θεωρήθηκε επίσης βιώσιμη επιλογή.
- Αντίθετα, το κριτήριο ABC υπέδειξε ότι μεγαλύτεροι αριθμοί clusters, ίδιαίτερα k=7 ή k=9, προσφέρουν την υψηλότερη ποιότητα clustering με όρους συμπύκνωσης και διαχωρισμού.

Η απόκλιση μεταξύ αυτών των μεθόδων έγκειται στα κριτήρια αξιολόγησής τους. Ενώ η μέθοδος Elbow και η ανάλυση Silhouette δίνουν έμφαση σε ευρύτερα, καλά διαχωρισμένα clusters, το κριτήριο ABC εστιάζει στη μεγιστοποίηση της ευθυγράμμισης και του διαχωρισμού, κάτι που μπορεί να ευνοεί πιο λεπτομερείς ομαδοποιήσεις.

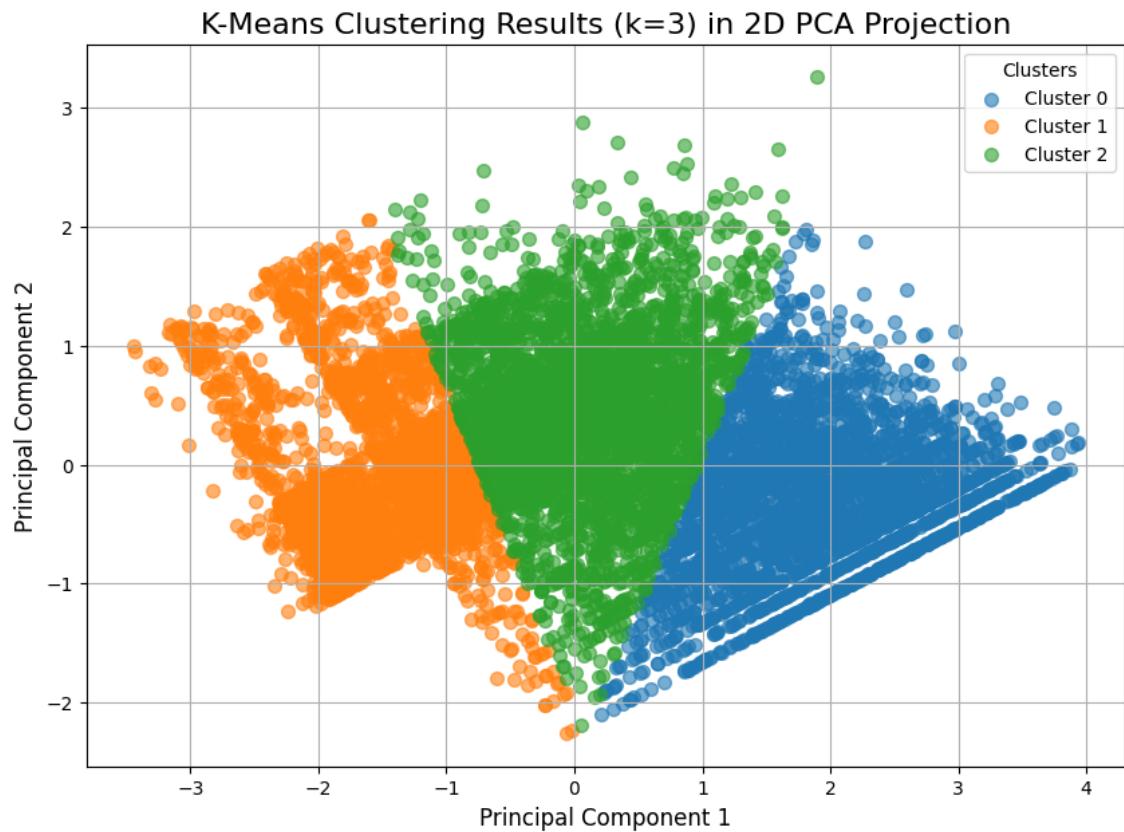
Αυτά τα ευρήματα υπογραμμίζουν τον συμβιβασμό μεταξύ της δημιουργίας στατιστικά βέλτιστων clusters και της διατήρησης πρακτικών, εφαρμόσιμων τμημάτων για επιχειρηματική χρήση. Μικρότεροι αριθμοί clusters, όπως το k=3, είναι πιο ερμηνεύσιμοι και εφαρμόσιμοι, επιτρέποντας τη διαμόρφωση σαφών στρατηγικών τμηματοποίησης πελατών. Αντίθετα, μεγαλύτεροι αριθμοί clusters, όπως το k=7, συλλαμβάνουν πιο λεπτομερείς αποχρώσεις στη συμπεριφορά των πελατών, αλλά διατρέχουν τον κίνδυνο υπερ-τμηματοποίησης, καθιστώντας δύσκολη την ανάπτυξη διακριτών στρατηγικών για κάθε ομάδα.

8.2 Εφαρμογή αλγορίθμου K-Means

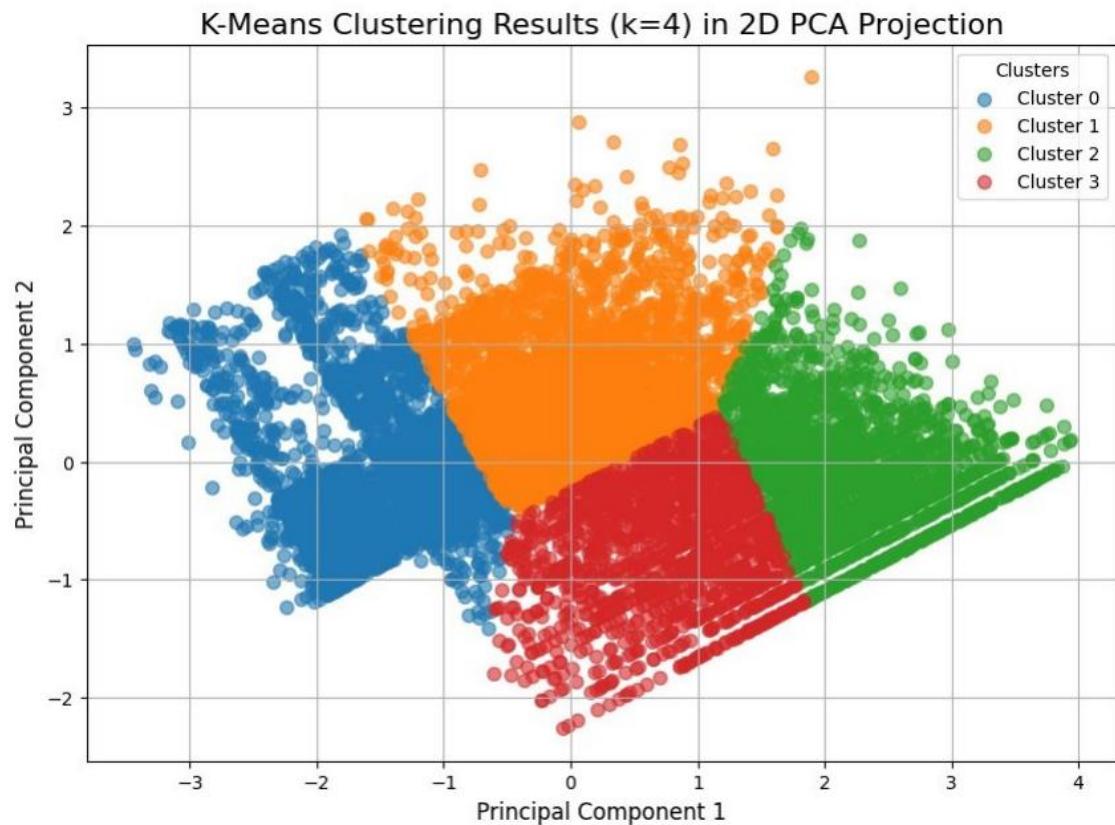
Για την αξιολόγηση της επίδρασης της μεταβολής του αριθμού των clusters στη διαμόρφωση των δεδομένων, το K-Means clustering εφαρμόστηκε για τιμές k ίσες με 2, 3, 4, 5, και 6. Τα αποτελέσματα οπτικοποιήθηκαν για τη σύγκριση των αντίστοιχων τμηματοποίησεων κι εμφανίζονται παρακάτω:



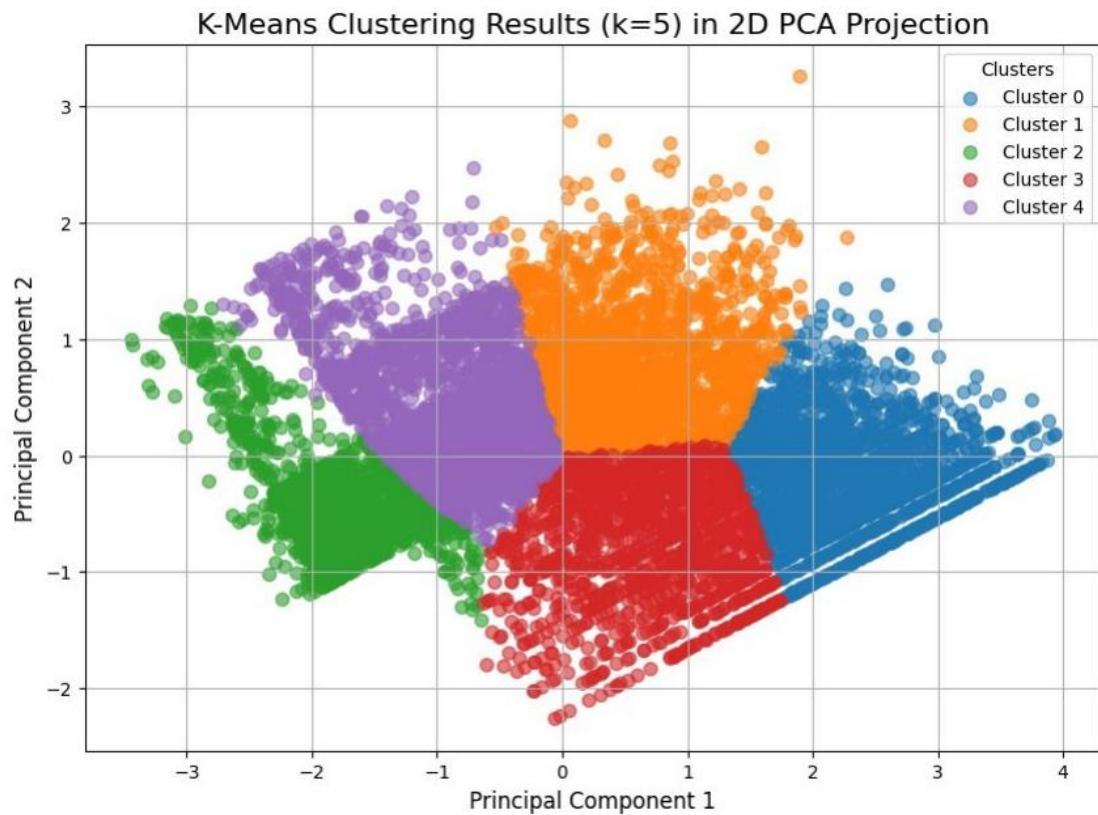
Εικόνα 38. Συσταδοποίηση βάσει αλγορίθμου K-means (2 clusters)



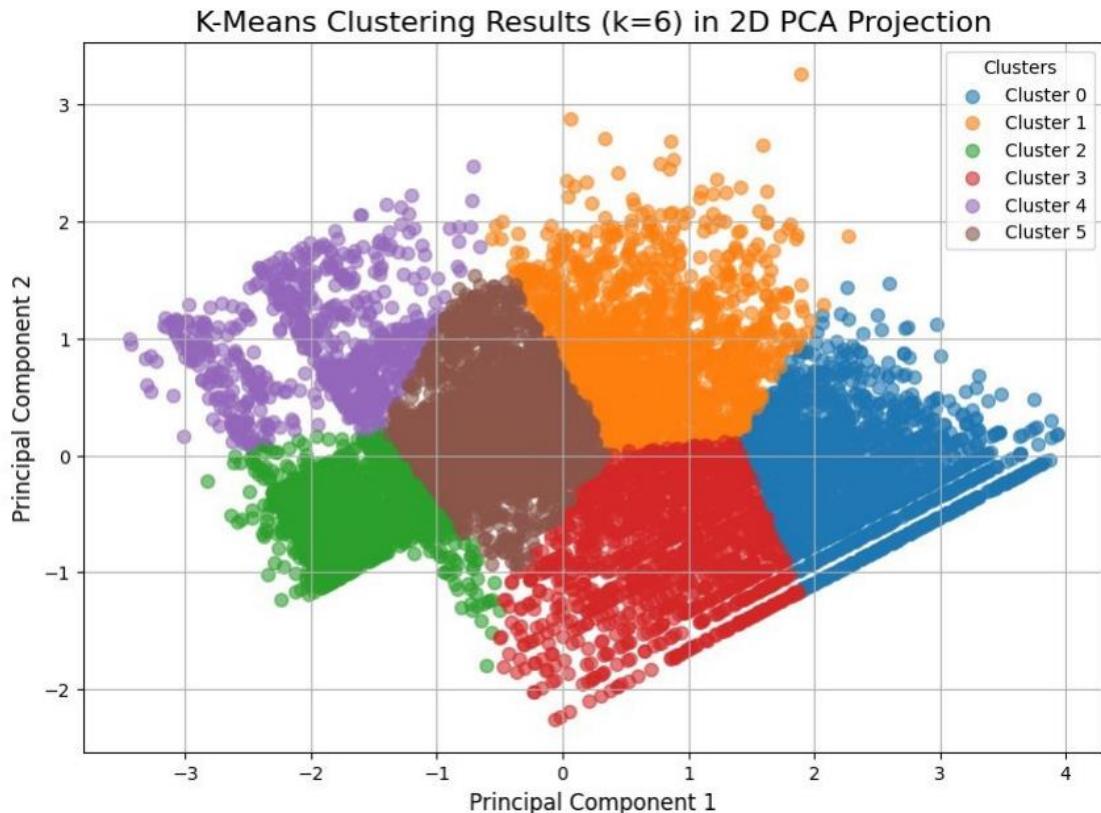
Εικόνα 39. Συσταδοποίηση βάσει αλγορίθμου K-means (3 clusters)



Εικόνα 40. Συσταδοποίηση βάσει αλγορίθμου K-means (4 clusters)



Εικόνα 41. Συσταδοποίηση βάσει αλγορίθμου K-means (5 clusters)



Εικόνα 42. Συσταδοποίηση βάσει αλγορίθμου K-means (6 clusters)

8.3 Εφαρμογή αλγορίθμου DBSCAN

Για την αντιμετώπιση αυτών των αποκλίσεων και τη διασφάλιση της πιο ανθεκτικής τμηματοποίησης, θα εφαρμοστεί ένας επιπλέον αλγόριθμος clustering: Density-Based Spatial Clustering of Applications with Noise (DBSCAN). Πρόκειται για έναν αλγόριθμο που βασίζεται στην πυκνότητα για την ομαδοποίηση σημείων που βρίσκονται κοντά το ένα στο άλλο, αναγνωρίζοντας τα outliers ως θόρυβο (noise). Δεν απαιτεί τον προκαθορισμό του αριθμού clusters (k). Ο αλγόριθμος χρησιμοποιεί δύο κύριες παραμέτρους:

eps: Η μέγιστη απόσταση μεταξύ σημείων εντός μιας περιοχής γειτονιάς.

min_samples: Ο ελάχιστος αριθμός σημείων που απαιτούνται για τον σχηματισμό μιας πυκνής περιοχής.

Τα σημεία δεδομένων κατατάσσονται ως εξής:

Core Points: Έχουν τουλάχιστον *min_samples* σημεία εντός της περιοχής *eps*.

Border Points: Βρίσκονται εντός της απόστασης *eps* ενός core point, αλλά δεν πληρούν το ελάχιστο όριο *min_samples*.

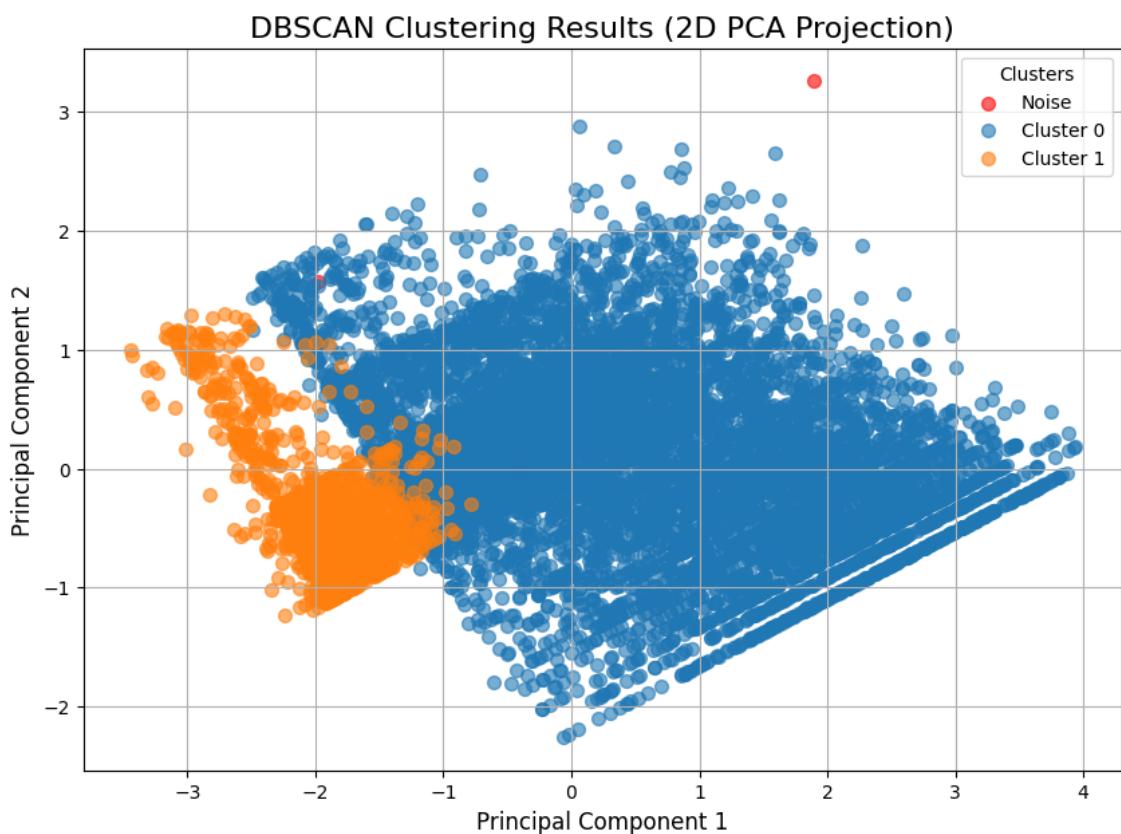
Noise Points: Αναγνωρίζονται ως outliers και δεν ανήκουν σε καμία ομάδα.

Σε αντίθεση με τον K-Means, ο DBSCAN δεν απαιτεί τον προκαθορισμό του αριθμού clusters (k). Αντίθετα, εντοπίζει clusters βάσει της πυκνότητας των σημείων δεδομένων, καθιστώντας τον ιδιαίτερα κατάλληλο για σύνολα δεδομένων με μη γραμμικές δομές clusters ή σημαντικό θόρυβο.

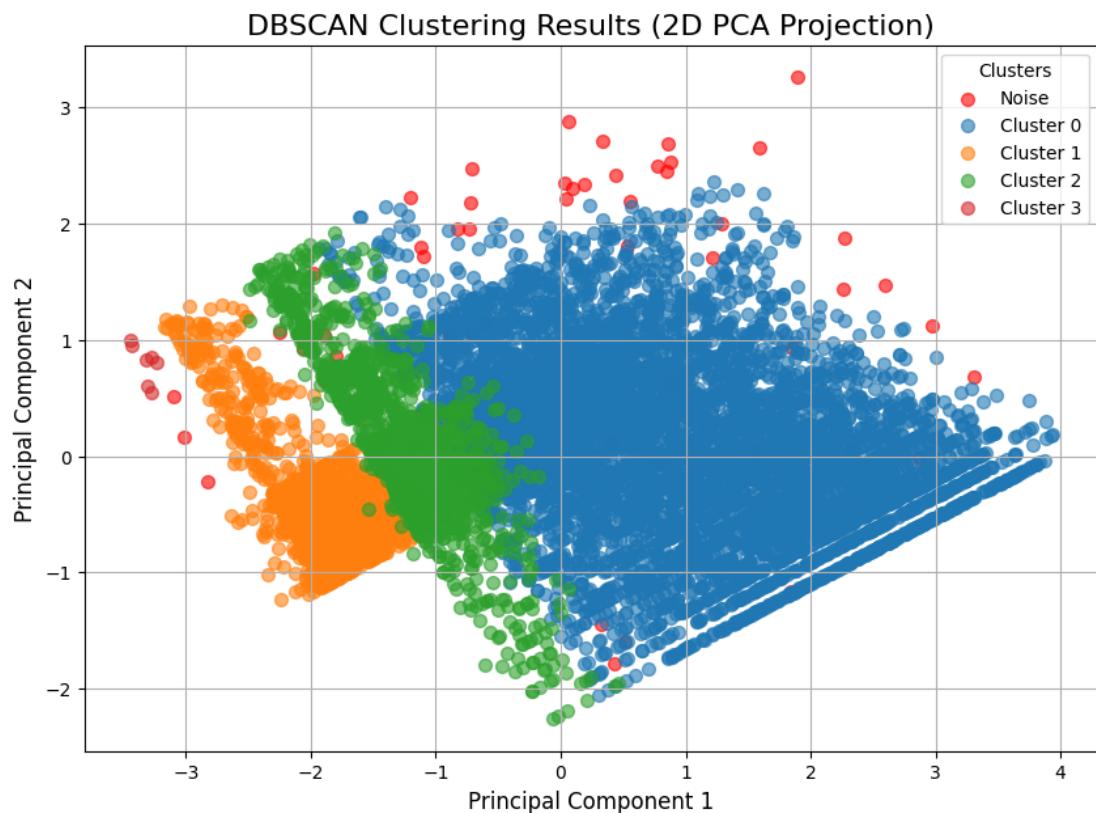
Με την εφαρμογή του DBSCAN, ο στόχος είναι η επικύρωση των αποτελεσμάτων clustering που προέκυψαν από τις προηγούμενες μεθόδους και ενδεχομένως η αποκάλυψη φυσικών ομαδοποιήσεων στο σύνολο δεδομένων χωρίς την ανάγκη προκαθορισμού του k . Αυτή η προσέγγιση θα συμπληρώσει τις προηγούμενες αναλύσεις και θα παρέχει επιπλέον εμπιστοσύνη στη στρατηγική τελικής τμηματοποίησης. Ο DBSCAN είναι αποτελεσματικός για clusters με αυθαίρετα σχήματα και είναι ανθεκτικό στα outliers, αλλά είναι ευαίσθητος στη ρύθμιση των παραμέτρων (eps και min_samples).

Το επόμενο βήμα περιλαμβάνει την εφαρμογή του DBSCAN για τον προσδιορισμό του βέλτιστου αριθμού clusters, ακολουθούμενη από συγκριτική αξιολόγηση όλων των μεθόδων για την τελική τμηματοποίηση. Αυτή η ολοκληρωμένη προσέγγιση διασφαλίζει ότι τα αποτελέσματα του clustering είναι τόσο στατιστικά έγκυρα όσο και πρακτικά συναφή.

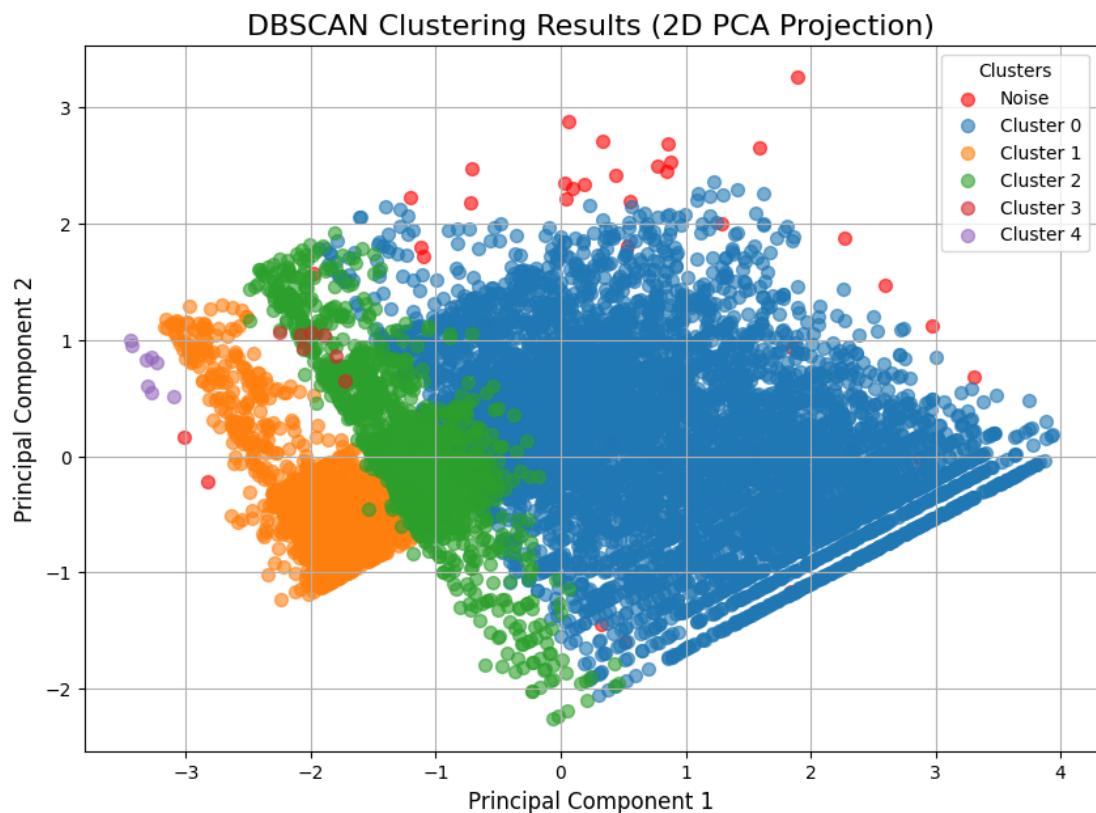
Για την αξιολόγηση της επίδρασης των διαφορετικών ρυθμίσεων παραμέτρων στη διαμόρφωση των δεδομένων, το DBSCAN clustering εφαρμόστηκε με προσαρμοσμένες τιμές eps και min_samples , με αποτέλεσμα τη δημιουργία 2, 4, 5 και 6 clusters. Τα αποτελέσματα οπτικοποιήθηκαν για τη σύγκριση των αντίστοιχων τμηματοποιήσεων.



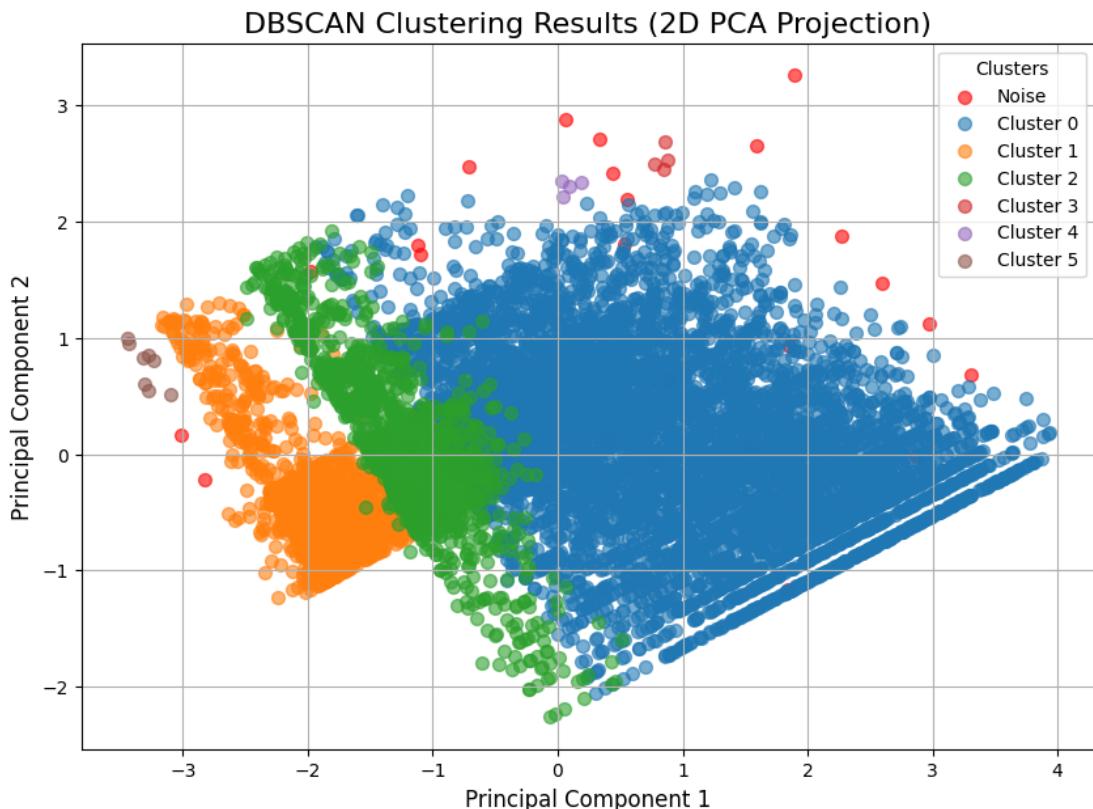
Εικόνα 43. Συσταδοποίηση βάσει αλγορίθμου DBSCAN (2 clusters)



Εικόνα 44. Συσταδοποίηση βάσει αλγορίθμου DBSCAN (4 clusters)



Εικόνα 45. Συσταδοποίηση βάσει αλγορίθμου DBSCAN (5 clusters)



Εικόνα 46. Συσταδοποίηση βάσει αλγορίθμου DBSCAN (6 clusters)

Ο DBSCAN εξετάστηκε χρησιμοποιώντας διάφορες παραμετροποιήσεις, συγκεκριμένα με ρύθμιση των eps (μέγιστη απόσταση γειτονιάς) και min_samples (ελάχιστος αριθμός σημείων που απαιτούνται για τη δημιουργία ενός cluster). Αυτή η εξερεύνηση ανέδειξε την data-driven προσέγγιση του DBSCAN:

Με μικρότερες τιμές eps , όπως 0.3, ο αλγόριθμος αναγνώρισε αρκετά μικρότερα clusters. Μεγαλύτερες τιμές eps , όπως 0.6, οδήγησαν σε δύο κύρια clusters.

Η ευαισθησία του DBSCAN στη ρύθμιση παραμέτρων επέτρεψε την καλύτερη κατανόηση της δομής των δεδομένων βάσει πυκνότητας. Παρότι ορισμένες παραμετροποιήσεις οδήγησαν σε τέσσερα ή περισσότερα clusters με σημαντική επικάλυψη στον 2D χώρο του PCA, αυτό αποκάλυψε τις πολυπλοκότητες στη δομή των δεδομένων και τα όρια μεταξύ των clusters. Επιπλέον, η ικανότητα του DBSCAN να αναγνωρίζει σημεία θορύβου (noise points) ως outliers παρείχε πολύτιμες πληροφορίες, εντοπίζοντας σημεία δεδομένων που δεν ταιριάζουν στους κύριους clusters.

8.3.1 Σύγκριση συσταδοποιήσεων K-Means και DBSCAN

Ο DBSCAN κατέδειξε σταθερά ότι τα δεδομένα δεν σχηματίζουν φυσικά τρία clusters, όπως είχε υποδειχθεί από τη Μέθοδο Elbow και την Ανάλυση Silhouette. Αυτό αποκάλυψε ότι η "επιβολή" των δεδομένων σε τρία clusters, όπως κάνει το K-Means, μπορεί να μην αντιπροσωπεύει επαρκώς τη φυσική δομή των δεδομένων. Συνολικά, η εξερεύνηση των αλγορίθμων clustering παρείχε σημαντικές πληροφορίες για τη δομή του συνόλου δεδομένων:

K-Means: Εξετάστηκε με διαφορετικούς αριθμούς clusters, από 2 έως 6. Τα αποτελέσματα έδειξαν σαφή, αλλά αυθαίρετο, διαχωρισμό των δεδομένων στους προκαθορισμένους αριθμούς

clusters, καταδεικνύοντας τους περιορισμούς του K-Means στην προσαρμογή σε φυσικές ομάδοποιήσεις.

DBSCAN: Με διάφορες παραμετροποιήσεις, εντόπισε σταθερά δύο ή περισσότερα από τρία clusters, υποδεικνύοντας μια δομή δεδομένων που δεν προσαρμόζεται φυσικά σε τρεις διακρίτες ομάδες. Επιπλέον, η αναγνώριση θορύβου και η ευελιξία στην ανίχνευση σχημάτων προσέφεραν μια πιο ανθεκτική, data-driven προσέγγιση.

Παρά τα στατιστικά πλεονεκτήματα του DBSCAN και την ικανότητά του να αποκαλύπτει την υποκείμενη δομή των δεδομένων, η τελική απόφαση να χρησιμοποιηθεί το K-Means με τρία clusters καθορίστηκε από επιχειρηματικές προτεραιότητες. Ενώ αυτή η προσέγγιση δεν είναι η πιο στατιστικά έγκυρη, η επιχειρηματική ανάγκη για τρεις διακρίτες ομάδες πελατών, στις οποίες θα εφαρμοστούν συγκεκριμένες καμπάνιες μάρκετινγκ, προτιμήθηκε. Αυτό είχε ως αποτέλεσμα την αποδοχή του μοντέλου K-Means=3, με την κατανόηση ότι αποτελεί μια εξαναγκασμένη τμηματοποίηση και όχι μια αναπαράσταση της πραγματικής δομής των δεδομένων. Είναι σημαντικό να αναγνωριστεί ότι το μοντέλο K-Means=3 πρέπει να χρησιμοποιείται με επίγνωση των περιορισμών του και ότι δεν είναι ιδανικό από στατιστική άποψη, αν και εξυπηρετεί τις επιχειρηματικές απαιτήσεις.

8.4 Αποτελέσματα συσταδοποίησης

Η ανάλυση clustering που πραγματοποιήθηκε χρησιμοποιώντας τον αλγόριθμο K-Means διαχώρισε την πελατειακή βάση σε τρεις διακρίτες ομάδες: Loyalists, Potential Loyalists, και Churners. Αυτά τα clusters αποτυπώνουν μοναδικές συμπεριφορές πελατών βάσει των τιμών Recency (R), Frequency (F), και Monetary (M), που προκύπτουν από τα συναλλακτικά δεδομένα. Κάθε ομάδα προσφέρει πολύτιμες πληροφορίες για τα επίπεδα εμπλοκής των πελατών και τη συνολική συνεισφορά τους στα έσοδα της εταιρείας.

Segment	Segment Name	Customers #	Customers %	R	F	M
1	Loyalists	2.941	29.4%	7.9	40.8	841.0
3	Potential Loyalists	3.516	35.2%	47.7	10.2	199.3
2	Churners	3.543	35.4%	57.0	1.7	27.7
Total		10.000	100.0%	19.2	32.0	657.5

Εικόνα 47. Τιμές μεταβλητών RFM κατόπιν συσταδοποίησης

1^η συστάδα: «Loyalists»

Το πρώτο cluster, που ονομάστηκε Loyalists, περιλαμβάνει 2,941 πελάτες, αντιπροσωπεύοντας το 29.4% της συνολικής πελατειακής βάσης. Η μέση τιμή Recency για αυτή την ομάδα είναι 7.9, υποδεικνύοντας ότι οι πελάτες αυτοί έχουν πραγματοποιήσει αγορές σχετικά πρόσφατα. Επιπλέον, η τιμή Frequency των 40.8 και η τιμή Monetary των 841.0 είναι σημαντικά υψηλότερες από τον συνολικό μέσο όρο αμφότερων των κατηγοριών. Αυτό αναδεικνύει τη σταθερή αγοραστική συμπεριφορά τους και τα υψηλά επίπεδα δαπανών, καθιστώντας τους το πιο πολύτιμο και αφοσιωμένο τμήμα της πελατειακής βάσης της εταιρείας.

Οι πελάτες αυτοί επιδεικνύουν πίστη και εμπιστοσύνη στην εταιρεία, γεγονός που τους καθιστά βασικούς μοχλούς εσόδων. Για να διατηρηθούν τα υψηλά επίπεδα δραστηριότητάς τους, είναι απαραίτητο να επικεντρωθεί η προσοχή στη διατήρηση της πίστης τους μέσω εξατομικευμένων ανταμοιβών, αποκλειστικών προσφορών και εξαιρετικής εξυπηρέτησης πελατών. Πρωτοβουλίες όπως προγράμματα επιβράβευσης, εξατομικευμένες προτάσεις προϊόντων και VIP εμπειρίες θα μπορούσαν να ενισχύσουν περαιτέρω την ικανοποίησή τους και να εδραιώσουν τη σχέση τους με την εταιρεία.

2^η συστάδα: «Potential Loyalists»

Customer Segmentation & Market Basket Analysis

Το δεύτερο cluster, που ονομάστηκε Potential Loyalists, αποτελείται από 3,516 πελάτες, οι οποίοι αντιπροσωπεύουν το 35.2% της πελατειακής βάσης. Όλες οι τιμές τους βρίσκονται κάτω από τον μέσο όρο, κάτι που υποδηλώνει μια μείωση της δραστηριότητας, αν και οι πελάτες αυτοί διαθέτουν τη δυναμική να εξελιχθούν σε Loyalists, αν αντιμετωπιστούν κατάλληλα.

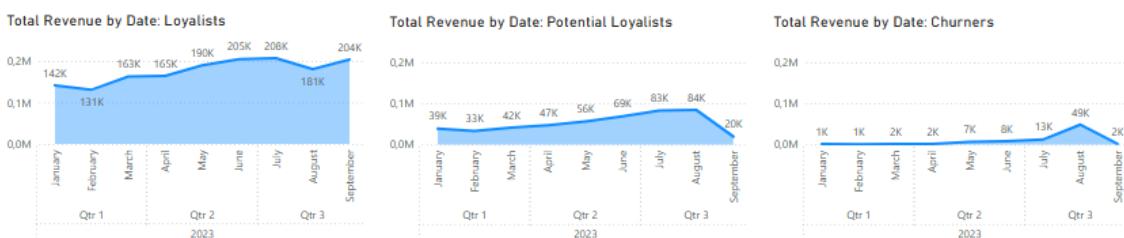
Οι πελάτες αυτοί θα μπορούσαν να επωφεληθούν από στοχευμένες στρατηγικές μάρκετινγκ που έχουν σχεδιαστεί για να τους επαναδραστηριοποιήσουν, όπως εξατομικευμένες εκπτώσεις, προσφορές περιορισμένου χρόνου και καμπάνιες μάρκετινγκ προσαρμοσμένες στις προτιμήσεις τους. Η αντιμετώπιση πιθανών εμποδίων για πιο συχνές αγορές, όπως η διαθεσιμότητα προϊόντων, οι τιμές ή η ευκολία αγορών, θα μπορούσε να τους ενθαρρύνει να αυξήσουν τη συχνότητα αγορών τους και να προσεγγίσουν την κατηγορία των Loyalists.

3^η συστάδα: «Churners»

Το τρίτο cluster, που ονομάστηκε Churners, περιλαμβάνει 3,543 πελάτες, οι οποίοι αντιπροσωπεύουν το 35.4% της πελατειακής βάσης. Αυτή η ομάδα εμφανίζει τις λιγότερο ευνοϊκές τιμές RFM, με Recency στα 57.0, που υποδεικνύει ότι δεν έχουν πραγματοποιήσει αγορές για σημαντικό χρονικό διάστημα. Οι τιμές Frequency και Monetary, που ανέρχονται σε 1.7 και 27.7 αντίστοιχα, είναι πολύ χαμηλότερες από τον συνολικό μέσο όρο, αντικατοπτρίζοντας χαμηλή εμπλοκή και ελάχιστες δαπάνες.

Οι πελάτες αυτοί διατρέχουν τον μεγαλύτερο κίνδυνο να χαθούν οριστικά, κάτι που αποτελεί κρίσιμη πρόκληση για την εταιρεία. Οι στρατηγικές που στοχεύουν στην επανενεργοποίηση αυτών των πελατών είναι ζωτικής σημασίας και θα μπορούσαν να περιλαμβάνουν καμπάνιες επαναπροσέγγισης (win-back campaigns) με δελεαστικές εκπτώσεις, εξατομικευμένες επικοινωνίες ή έρευνες για τον εντοπισμό των λόγων πίσω από τη μειωμένη εμπλοκή τους. Η διερεύνηση και αντιμετώπιση πιθανών ζητημάτων, όπως η ικανοποίηση από τα προϊόντα, οι ανησυχίες για την παράδοση ή οι προσαρμογές τιμών, μπορεί να βοηθήσει στην υπέρβαση των ενδοιασμών τους και στην αποκατάσταση της αγοραστικής τους δραστηριότητας.

Αφού ολοκληρώθηκε με επιτυχία η τμηματοποίηση της πελατειακής βάσης σε τρεις διακριτές ομάδες, το επόμενο βήμα περιλαμβάνει μια βαθύτερη ανάλυση της συμπεριφοράς και των χαρακτηριστικών κάθε ομάδας. Αυτή η φάση της ανάλυσης στοχεύει στην αποκάλυψη βασικών ευρημάτων που διαφοροποιούν αυτά τα τμήματα όσον αφορά τα πρότυπα αγορών, τη συνεισφορά στα έσοδα, καθώς και τις προτιμήσεις προϊόντων.



Εικόνα 48. Συνολικός τζίρος ανά μήνα του έτους - Συγκριτικό ανά ομάδα πελατών

Τα τρία διαγράμματα της Εικόνας 43 απεικονίζουν τη συνολική συνεισφορά εσόδων ανά μήνα του έτους για τα τρία τμήματα πελατών, καλύπτοντας την περίοδο από τον Ιανουάριο έως τον Σεπτέμβριο 2023. Διατηρώντας σταθερή την κλίμακα στον άξονα γ και στα τρία διαγράμματα, καθίσταται δυνατή η σαφής σύγκριση των προτύπων και των μεγεθών εσόδων μεταξύ των τμημάτων.

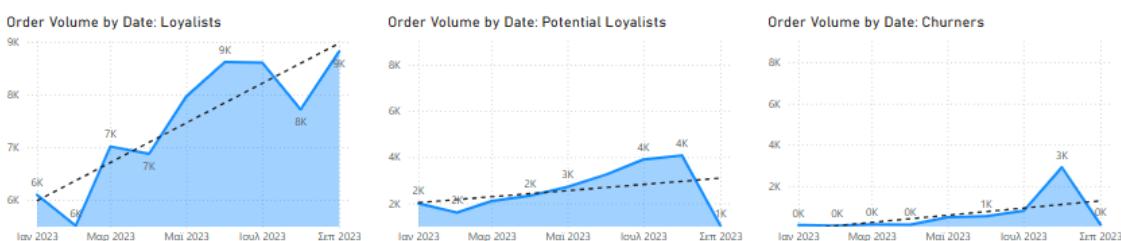
Οι Loyalists εμφανίζουν τη μεγαλύτερη και πιο σταθερή συνεισφορά εσόδων καθ' όλη τη διάρκεια της εξεταζόμενης περιόδου. Τα έσοδά τους ξεκινούν από €142 χιλ. τον Ιανουάριο και

παρουσιάζουν σταθερή αύξηση, κορυφώνοντας στα €208 χιλ. τον Ιούλιο, με μικρές διακυμάνσεις, όπως μια πτώση στα €131 χιλ. τον Φεβρουάριο. Μέχρι τον Σεπτέμβριο, το τμήμα αυτό συνεχίζει να αποφέρει €204 χιλ. σε έσοδα. Το συνεπές αυτό μοτίβο υπογραμμίζει ότι οι Loyalists αποτελούν μια ιδιαίτερα αφοσιωμένη και σταθερή ομάδα πελατών, που διαδραματίζει κρίσιμο ρόλο στη διαμόρφωση των εσόδων της εταιρείας, ακόμη και κατά τη διάρκεια περιόδων μειωμένης αγοραστικής δραστηριότητας.

Οι Potential Loyalists συνεισφέρουν σημαντικά λιγότερα έσοδα σε σύγκριση με τους Loyalists, με τα έσοδά τους να ξεκινούν από €39 χιλ. τον Ιανουάριο. Με την πάροδο του χρόνου, παρατηρείται σταδιακή αύξηση, η οποία κορυφώνεται στα €84 χιλ. τον Αύγουστο, προτού σημειωθεί απότομη πτώση στα €20 χιλ. τον Σεπτέμβριο. Αυτό το μοτίβο υποδηλώνει ότι, ενώ το τμήμα αυτό παρουσίασε αυξημένη αγοραστική δραστηριότητα κατά τους καλοκαιρινούς μήνες, η απότομη πτώση του Σεπτεμβρίου καταδεικνύει μια ασταθή αγοραστική συμπεριφορά. Η σημαντική μείωση των εσόδων τον Σεπτέμβριο αναδεικνύει την ανάγκη για στοχευμένες πρωτοβουλίες, με σκοπό τη διατήρηση της εμπλοκής αυτού του τμήματος και τη διασφάλιση μιας πιο σταθερής ροής εσόδων καθ' όλη τη διάρκεια του έτους.

Αντίθετα, οι Churners παρουσιάζουν ένα εντελώς διαφορετικό μοτίβο εσόδων, το οποίο χαρακτηρίζεται από ελάχιστη δραστηριότητα στη μεγαλύτερη διάρκεια της εξεταζόμενης περιόδου. Τα έσοδά τους ξεκινούν από μόλις €1 χιλ. τον Ιανουάριο, παραμένουν σταθερά χαμηλά και σημειώνουν αιφνίδια αύξηση στα €49 χιλ. τον Αύγουστο, πριν επιστρέψουν σε αμελητέα επίπεδα τον Σεπτέμβριο. Αυτή η περιοδική αύξηση των εσόδων κατά τη θερινή περίοδο υποδηλώνει ότι οι Churners είναι ενεργοί κυρίως κατά τους καλοκαιρινούς μήνες, πιθανότατα λόγω αγοράς εποχικών προϊόντων που παρουσιάζουν υψηλή ζήτηση αυτή την περίοδο.

Αυτή η εποχική συμπεριφορά αναδεικνύει τη σημασία της αναγνώρισης συγκεκριμένων κατηγοριών προϊόντων που συμβάλλουν στις αγορές τους και της ανάπτυξης στοχευμένων στρατηγικών επανενεργοποίησης. Για παράδειγμα, η προσφορά αποκλειστικών προωθητικών ενεργειών σε προϊόντα που σχετίζονται με το καλοκαίρι, πριν από την περίοδο αιχμής, ενδέχεται να συμβάλει στην ενεργοποίηση αυτής της ομάδας και στη μετατροπή των περιστασιακών αγοραστών σε πιο σταθερούς πελάτες.



Εικόνα 49. Συνολικός αριθμός παραγγελιών ανά μήνα του έτους - Συγκριτικό ανά ομάδα πελατών

Τα ως άνω διαγράμματα (Εικόνα 44) απεικονίζουν τον όγκο παραγγελιών ανά μήνα, όπως στην προηγούμενη ανάλυση. Παρόμοια με την ανάλυση των εσόδων, αυτά τα διαγράμματα αναδεικνύουν βασικές διαφοροποιήσεις στη συμπεριφορά των πελατών, εστιάζοντας στη συχνότητα παραγγελιών με την πάροδο του χρόνου.

Το τμήμα Loyalists εμφανίζει σταθερά υψηλό όγκο παραγγελιών, αντικατοπτρίζοντας τη συχνή και συνεχή αγοραστική τους δραστηριότητα. Οι παραγγελίες τους ξεκινούν από 6 χιλ. τον Ιανουάριο και αυξάνονται σταθερά, κορυφώνοντας στις 9 χιλ. τον Ιούνιο και τον Ιούλιο, πριν σημειώσουν ελαφρά μείωση στις 8 χιλ. τον Σεπτέμβριο. Η συνολική τάση είναι ανοδική, γεγονός που υποδηλώνει ότι αυτό το τμήμα πελατών δεν συνεισφέρει μόνο σημαντικά έσοδα, αλλά και διατηρεί υψηλή συχνότητα αγορών, ενισχύοντας τον ρόλο τους ως το πιο πολύτιμο πελατειακό τμήμα για την εταιρεία.

Οι Potential Loyalists εμφανίζουν λιγότερο έντονη αλλά σταθερή αύξηση στον όγκο παραγγελιών. Οι παραγγελίες τους ζεκινούν από 2 χιλ. τον Ιανουάριο και αυξάνονται σταδιακά, κορυφώνοντας στις 4 χιλ. τον Αύγουστο, προτού σημειώσουν απότομη πτώση στις 1 χιλ. τον Σεπτέμβριο. Το μοτίβο αυτό αντικατοπτρίζει τη συμπεριφορά των εσόδων τους, υποδηλώνοντας ότι, ενώ το τμήμα αυτό έχει δυνατότητες για υψηλότερη εμπλοκή, η δραστηριότητά του παραμένει συνδεδεμένη με συγκεκριμένες χρονικές περιόδους, όπως οι καλοκαιρινοί μήνες. Η αιφνίδια μείωση τον Σεπτέμβριο αναδεικνύει μια σημαντική ευκαιρία βελτίωσης, με στόχο τη διατήρηση της εμπλοκής τους σε όλη τη διάρκεια του έτους και τη μετατροπή τους σε πιο σταθερούς πελάτες.

Το τμήμα Churners παρουσιάζει ένα σαφώς διακριτό μοτίβο παραγγελιών, με ελάχιστη δραστηριότητα στη μεγαλύτερη διάρκεια του έτους, εκτός από μια σημαντική αύξηση τον Αύγουστο, όπου οι παραγγελίες τους φτάνουν τις 3 χιλ., πριν επιστρέψουν σε αμελητέα επίπεδα τον Σεπτέμβριο. Αυτή η τάση ενισχύει την υπόθεση ότι οι Churners καθοδηγούνται κυρίως από εποχιακή ζήτηση, πιθανότατα αγοράζοντας προϊόντα που σχετίζονται με το καλοκαίρι. Η σποραδική δραστηριότητά τους αναδεικνύει περαιτέρω την πρόκληση μετατροπής αυτού του τμήματος σε πιο αξιόπιστη πηγή παραγγελιών.



Εικόνα 50. Αριθμός ενεργών πελατών ανά μήνα του έτους - Συγκριτικό ανά ομάδα πελατών

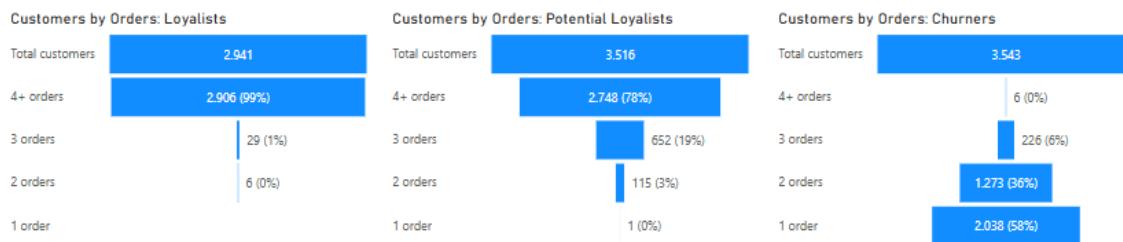
Τα διαγράμματα των ενεργών πελατών ανά μήνα της Εικόνας 45 αναπαριστούν τον αριθμό των πελατών σε κάθε τμήμα που πραγματοποίησαν τουλάχιστον μία παραγγελία κατά τη διάρκεια ενός συγκεκριμένου μήνα, παρέχοντας μια σαφέστερη εικόνα για την εξέλιξη της αγοραστικής δραστηριότητας και αναδεικνύοντας τις διαφορές στη συμπεριφορά των πελατών μεταξύ των τμημάτων.

Οι Loyalists παρουσιάζουν μια σταθερή και συνεπή αύξηση στον αριθμό των ενεργών πελατών τους, με τη δραστηριότητά τους να ακολουθεί στενά ή ακόμα και να διαμορφώνει την τάση της συνολικής πελατειακής βάσης, καθ' όλη τη διάρκεια της εξεταζόμενης περιόδου. Ζεκινώντας από 1,800 ενεργούς πελάτες τον Ιανουάριο, η βάση των ενεργών πελατών τους αυξάνεται σταδιακά, φτάνοντας τους 2,842 τον Σεπτέμβριο. Αυτή η ευθυγράμμιση με την τάση της αγοράς υποδηλώνει ότι οι Loyalists αποτελούν ένα σταθερό και αξιόπιστο πελατειακό τμήμα που διατηρεί σταθερή αλληλεπίδραση με την εταιρεία. Η σταδιακή αλλά διαρκής αύξηση των ενεργών πελατών επιβεβαιώνει τον μακροπρόθεσμο χαρακτήρα της πιστότητάς τους και τη σημαντική συμβολή τους στη δημιουργία εσόδων.

Αντίθετα, οι Potential Loyalists εμφανίζουν ένα πιο ακανόνιστο μοτίβο δραστηριότητας. Ο αριθμός των ενεργών πελατών τους ζεκινά από 1,145 τον Ιανουάριο, αυξάνεται σταδιακά στους 2,065 τον Ιούλιο και κορυφώνεται στους 2,785 τον Αύγουστο, πριν σημειώσει απότομη πτώση στους 841 τον Σεπτέμβριο. Αν και η δραστηριότητά τους εμφανίζει ανοδική τάση κατά τη διάρκεια των καλοκαιρινών μηνών, η έντονη μείωση που ακολουθεί υποδηλώνει έλλειψη διατηρήσιμης δραστηριότητας. Αυτή η μεταβλητότητα καταδεικνύει ότι, παρόλο που αυτό το τμήμα παρουσιάζει δυνατότητες ανάπτυξης, ενδεχομένως απαιτούνται πιο σταθερές και στοχευμένες στρατηγικές μάρκετινγκ για τη μετατροπή τους σε πλήρως αφοσιωμένους πελάτες.

Το τμήμα των Churners εμφανίζει μια ριζικά διαφορετική συμπεριφορά, με πολύ χαμηλά επίπεδα ενεργών πελατών στη διάρκεια του έτους, εκτός από μια απότομη κορύφωση τον Αύγουστο, όπου ο αριθμός των ενεργών πελατών εκτοξεύεται στους 2,713, πριν σημειώσει

ραγδαία πτώση στους 129 τον Σεπτέμβριο. Αυτό επιβεβαιώνει ότι οι Churners είναι κυρίως εποχιακοί αγοραστές, καθώς η εμπλοκή τους συγκεντρώνεται σε έναν και μόνο μήνα, πιθανότατα λόγω αγορών προϊόντων που σχετίζονται με το καλοκαίρι. Η ιδιαίτερα σποραδική τους αγοραστική δραστηριότητα υπογραμμίζει ακόμη περισσότερο την αναγκαιότητα στοχευμένων στρατηγικών επανενεργοποίησης, με σκοπό τη διατήρησή τους πέρα από τις εποχιακές περιόδους αιχμής.



Εικόνα 51. Άθροισμα αριθμού πελατών ανά κατηγορία συνόλου παραγγελιών - Συγκριτικό ανά ομάδα πελατών

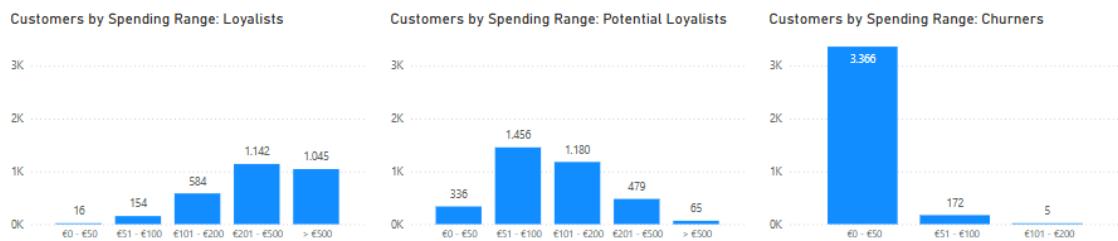
Τα funnel charts (Εικόνα 46) αποτυπώνουν την κατανομή των πελατών βάσει του αριθμού των παραγγελιών.

Μεταξύ των Loyalists, σχεδόν όλοι οι πελάτες (99%) πραγματοποίησαν τέσσερις ή περισσότερες παραγγελίες, γεγονός που υποδηλώνει ιδιαίτερα συνεπή αγοραστική συμπεριφορά. Αυτό το υψηλό επίπεδο επαναλαμβανόμενων αγορών ενισχύει την πίστη τους στην εταιρεία. Μόνο ένα αμελητέο ποσοστό των Loyalists πραγματοποίησε τρεις παραγγελίες (1%), ενώ μόλις έξι πελάτες πραγματοποίησαν δύο παραγγελίες, επιβεβαιώνοντας περαιτέρω ότι η αγοραστική τους συμπεριφορά χαρακτηρίζεται από συχνές και επαναλαμβανόμενες αγορές.

Οι Potential Loyalists από την άλλη πλευρά, εμφανίζουν μεγαλύτερη ποικιλομορφία στην κατανομή των παραγγελιών. Παρόλο που η πλειοψηφία (78%) έχει πραγματοποιήσει τέσσερις ή περισσότερες παραγγελίες, παρατηρείται μια σημαντική μείωση στη δέσμευση καθώς μειώνεται ο αριθμός των παραγγελιών. Περίπου 19% των Potential Loyalists πραγματοποίησε τρεις παραγγελίες, ενώ μικρότερα ποσοστά (3% και 0%) πραγματοποίησαν δύο ή μία παραγγελία αντίστοιχα. Αυτό το μοτίβο υποδηλώνει ότι, αν και οι περισσότεροι Potential Loyalists εμφανίζουν κάποιο επίπεδο αφοσίωσης, ένα σημαντικό ποσοστό αυτών ενδέχεται να μην είναι πλήρως δεσμευμένο σε τακτικές αγορές.

Οι Churners εμφανίζουν ένα σαφώς διαφορετικό μοτίβο, το οποίο αντικατοπτρίζει ελάχιστη αγοραστική θέληση ή δύναμη. Η πλειοψηφία (58%) πραγματοποίησε μόνο μία παραγγελία, ενώ 36% πραγματοποίησε δύο παραγγελίες. Το ποσοστό των Churners που πραγματοποίησαν τρεις παραγγελίες μειώνεται στο 6%, ενώ μόνο 0% πραγματοποίησε τέσσερις ή περισσότερες παραγγελίες. Αυτή η κατανομή υπογραμμίζει τη χαμηλή αλληλεπίδραση αυτής της ομάδας με την πλατφόρμα και επιβεβαιώνει την ανάγκη για στοχευμένες στρατηγικές επανενεργοποίησης ή μέτρα πρόληψης περαιτέρω αποστασιοποίησης.

Συγκρίνοντας τα τρία τμήματα, το funnel chart των Loyalists είναι σχεδόν πλήρες, αποδεικνύοντας τη συνεχή αγοραστική τους δραστηριότητα, ενώ οι Potential Loyalists εμφανίζουν κενά, υποδηλώνοντας πιθανή μείωση στη δέσμευσή τους. Αντίθετα, το διάγραμμα των Churners παρουσιάζει έντονη απώλεια πελατών μετά την πρώτη ή δεύτερη αγορά, αναδεικνύοντας το υψηλό ποσοστό εγκατάλειψης σε αυτό το τμήμα.



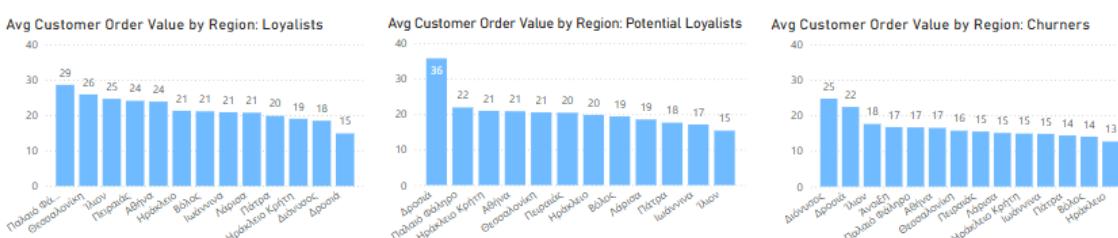
Εικόνα 52. Αριθμός μοναδικών πελατών βάσει εύρους τζίρου - Συγκριτικό ανά ομάδα πελατών

Παραπάνω, στην Εικόνα 47, βλέπουμε πώς διαμορφώνεται η κατανομή των πελατών βάσει του εύρους δαπανών τους.

Μεταξύ των Loyalists, η πλειοψηφία των πελατών κατατάσσεται στην κατηγορία €201–€500, όπου ανήκουν 1,142 πελάτες. Ακολουθεί στενά η κατηγορία >€500, που περιλαμβάνει 1,045 πελάτες. Αυτές οι δύο κατηγορίες μαζί αντιπροσωπεύουν το μεγαλύτερο μέρος των Loyalists, επιβεβαιώνοντας την τάση τους να πραγματοποιούν σημαντικές αγορές. Ένα μικρότερο, αλλά αξιόλογο, ποσοστό των Loyalists (584 πελάτες) ανήκει στην κατηγορία €101–€200, ενώ τα χαμηλότερα εύρη δαπανών (€0–€50 και €51–€100) αντιπροσωπεύουν μόνο ένα ελάχιστο ποσοστό αυτού του τμήματος.

Στους Potential Loyalists, η μεγαλύτερη συγκέντρωση πελατών παρατηρείται στο εύρος €51–€100, όπου ανήκουν 1,456 πελάτες. Ακολουθεί η κατηγορία €101–€200, με 1,180 πελάτες, γεγονός που υποδηλώνει ότι οι περισσότεροι πελάτες του συγκεκριμένου γκρουπ τείνουν να ξοδεύουν λιγότερο σε σύγκριση με τους Loyalists. Ένα μικρότερο ποσοστό (479 πελάτες) ανήκει στην κατηγορία €201–€500, ενώ μόλις 65 πελάτες βρίσκονται στην κατηγορία >€500. Ενδιαφέρον παρουσιάζει το γεγονός ότι 336 πελάτες κατατάσσονται στο χαμηλότερο εύρος δαπανών (€0–€50), γεγονός που υποδηλώνει μεγαλύτερη ποικιλομορφία στη συμπεριφορά δαπανών αυτού του τμήματος.

Οι Churners εμφανίζουν μια ριζικά διαφορετική κατανομή δαπανών, η οποία κυριαρχείται από το χαμηλότερο εύρος (€0–€50), όπου ανήκουν 3,366 πελάτες, αντιπροσωπεύοντας σχεδόν ολόκληρο το τμήμα. Μόνο 172 πελάτες ανήκουν στην κατηγορία €51–€100, ενώ ακόμη λιγότεροι εμφανίζονται στα υψηλότερα επίπεδα δαπανών. Αυτή η κατανομή αναδεικνύει τη μικρή οικονομική εμπλοκή των Churners με την πλατφόρμα, επιβεβαιώνοντας περαιτέρω τη χαμηλή αγοραστική δραστηριότητα και την έλλειψη ουσιαστικής πιστότητας σε αυτό το τμήμα.



Εικόνα 53. Μέσο κόστος παραγγελίας ανά περιοχή - Συγκριτικό ανά ομάδα πελατών

Η Εικόνα 48 δείχνει τη μέση αξία παραγγελίας ανά περιοχή για έκαστο τμήμα πελατών.

Μεταξύ των Loyalists, οι υψηλότερες μέσες τιμές παραγγελιών παρατηρούνται σε αστικές περιοχές, όπως το Παλαιό Φάληρο (€29), η Θεσσαλονίκη (€26) και το Ιλισσό (€25). Οι περιοχές αυτές χαρακτηρίζονται από ισχυρή οικονομική δραστηριότητα και υψηλή αγοραστική δύναμη, γεγονός που πιθανώς συμβάλλει στα αυξημένα ποσά ανά παραγγελία. Αντίθετα, περιοχές όπως η Δροσιά, παρά το γεγονός ότι είναι γνωστή για το υψηλό οικονομικό επίπεδο των κατοίκων της, παρουσιάζει τη χαμηλότερη μέση αξία παραγγελίας μεταξύ των Loyalists (€15). Αυτή η απόκλιση Customer Segmentation & Market Basket Analysis

μπορεί να υποδηλώνει ότι, ενώ οι Loyalists στη Δροσιά είναι σταθεροί πελάτες, τείνουν να πραγματοποιούν αγορές μικρότερης αξίας ανά παραγγελία, σε αντίθεση με άλλες περιοχές.

Για τους Potential Loyalists, η κατανομή εμφανίζει διαφοροποιήσεις. Η Δροσιά καταγράφει την υψηλότερη μέση αξία παραγγελίας, στα €36, ξεπερνώντας σημαντικά άλλες περιοχές, όπως το Παλαιό Φάληρο (€22). Αυτή η συμπεριφορά υποδηλώνει ότι οι Potential Loyalists στη Δροσιά τείνουν να πραγματοποιούν μεγαλύτερες αγορές κατά τη διάρκεια των σποραδικών τους αλληλεπιδράσεων με την πλατφόρμα, διαφοροποιούμενοι από τους Loyalists της ίδιας περιοχής, οι οποίοι πραγματοποιούν συχνότερες αλλά χαμηλότερης αξίας αγορές. Σε αυτό το σημείο αποτυπώνεται πιθανώς η λεπτή διαφοροποίηση των clusters που εξηγήσαμε νωρίτερα, καθώς είδαμε ότι βάσει της στατιστικής προσέγγισης, η τμηματοποίηση θα έπρεπε να πραγματοποιηθεί σε δύο segments αντί τριών. Σε αυτήν την περίπτωση ενδεχομένως να μην παρουσιάζονταν το φαινόμενο των πελατών της Δροσιάς.

Στο τμήμα των Churners, η περιοχή του Διονύσου ξεχωρίζει με τη υψηλότερη μέση αξία παραγγελίας (€25), ακολουθούμενη από τη Δροσιά (€22). Αυτή η τάση συνάδει με τη συμπεριφορά που παρατηρείται στους Potential Loyalists, υποδηλώνοντας ότι οι πελάτες από ευκατάστατες περιοχές όπως η Δροσιά και ο Διόνυσος μπορεί να πραγματοποιούν σποραδικές αλλά υψηλής αξίας αγορές πριν καταστούν ανενεργοί. Στις υπόλοιπες περιοχές, οι μέσες αξίες παραγγελιών είναι σημαντικά χαμηλότερες, κυμαινόμενες μεταξύ €13 και €18, γεγονός που αντικατοπτρίζει τη χαμηλή εμπλοκή των Churners στη μεγαλύτερη γεωγραφική κλίμακα.



Εικόνα 54. Μέσος αριθμός ημερών μεταξύ παραγγελιών ανά περιοχή - Συγκριτικό ανά ομάδα πελατών

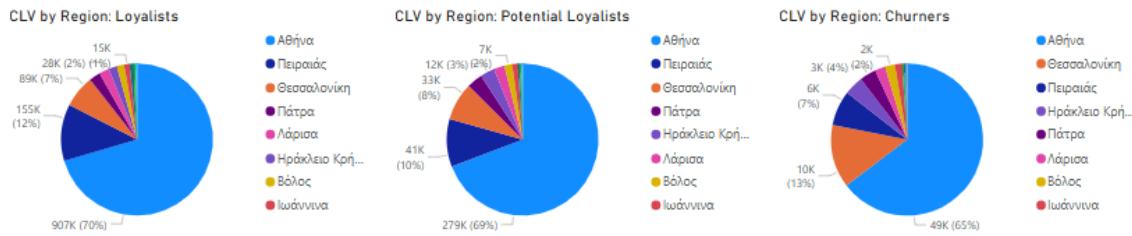
Στην Εικόνα 49 περιλαμβάνεται ανάλυση της μέσης χρονικής απόστασης μεταξύ παραγγελιών ανά περιοχή.

Μεταξύ των Loyalists, η Δροσιά εμφανίζει τον χαμηλότερο μέσο αριθμό ημερών μεταξύ παραγγελιών, με μόλις έξι ημέρες, σημαντικά χαμηλότερο από τις επόμενες πιο δραστήριες περιοχές, όπως το Ηράκλειο, το Παλαιό Φάληρο και ο Βόλος, όπου το μέσο διάστημα κυμαίνεται μεταξύ 11 και 15 ημερών. Αυτή η τάση υποδηλώνει ότι οι Loyalists στη Δροσιά είναι ιδιαίτερα συχνοί αγοραστές, παρόλο που, όπως αναφέρθηκε προηγουμένως, πραγματοποιούν μικρότερες αγορές ανά παραγγελία.

Στην ομάδα των Potential Loyalists, η Δροσιά εμφανίζει και πάλι το χαμηλότερο μέσο αριθμό ημερών μεταξύ παραγγελιών, με 18 ημέρες, τη στιγμή που οι περισσότερες άλλες περιοχές κυμαίνονται μεταξύ 29 και 42 ημερών. Αυτό που διαφοροποιεί τη Δροσιά σε αυτό το τμήμα είναι ότι οι πελάτες της εμφανίζουν τη μεγαλύτερη μέση αξία παραγγελίας, γεγονός που υποδηλώνει ότι, παρόλο που πραγματοποιούν παραγγελίες λιγότερο συχνά σε σχέση με τους Loyalists, όταν αλληλεπιδρούν με την πλατφόρμα, οι αγορές τους είναι πιο σημαντικές σε όρους αξίας. Αυτή η διπλή συμπεριφορά των πελατών της Δροσιάς, τόσο στους Loyalists όσο και στους Potential Loyalists, αναδεικνύει την ευελιξία και την εμπορική αξία τους, εφόσον αξιοποιηθούν στρατηγικά.

Στο τμήμα των Churners, οι πελάτες στον Διόνυσο εμφανίζουν μέσο αριθμό 8 ημερών μεταξύ παραγγελιών, ακολουθούμενοι από την Άνοιξη, με 18 ημέρες. Ωστόσο, το χρονικό διάστημα αιχάνεται απότομα σε άλλες περιοχές, όπως η Θεσσαλονίκη, με 38 ημέρες. Στο Ιλιον και το Παλαιό Φάληρο, ο μέσος αριθμός ημερών μεταξύ παραγγελιών φτάνει τις 64 και 67 ημέρες,

αντίστοιχα. Αυτό το μοτίβο καταδεικνύει την έλλειψη σταθερής εμπλοκής των Churners με την πλατφόρμα, επιβεβαιώνοντας την αποστασιοποίηση αυτού του τμήματος από την τακτική αγοραστική δραστηριότητα.



Εικόνα 55. Αξία του Χρόνου Ζωής του Πελάτη (Customer Lifetime Value) ανά περιοχή - Συγκριτικό ανά ομάδα πελατών

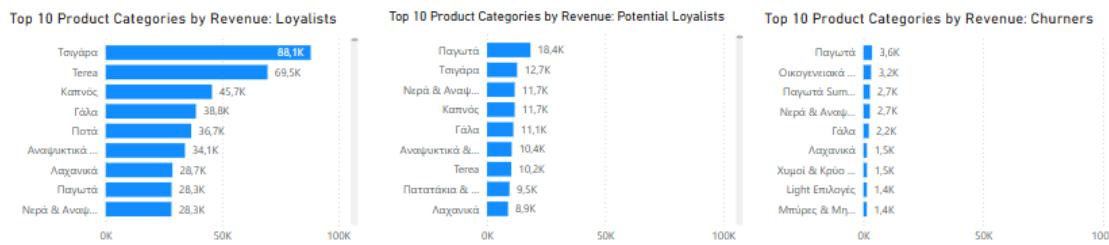
Η Εικόνα 50 περιλαμβάνει ανάλυση της Αξίας Διαρκείας Πελάτη (Customer Lifetime Value - CLV) ανά περιοχή για το κάθε τμήμα πελατών.

Στο τμήμα των Loyalists, η Αθήνα αποτελεί τη συντριπτική πλειοψηφία της συνολικής CLV, συνεισφέροντας €907 χιλ., που αντιστοιχεί στο 70% της αξίας διαρκείας αυτού του τμήματος. Αυτή η κυριαρχία αντικατοπτρίζει τη συνεπή αγοραστική συμπεριφορά των Loyalists της Αθήνας. Ο Πειραιάς ακολουθεί με €155 χιλ. (12%), ενώ η Θεσσαλονίκη και η Πάτρα συνεισφέρουν €89 χιλ. (7%) και €28 χιλ. (2%), αντίστοιχα. Οι υπόλοιπες περιοχές έχουν περιθωριακή συμβολή, με τιμές CLV κάτω του 2%.

Στο τμήμα των Potential Loyalists, η γεωγραφική κατανομή ακολουθεί παρόμοια τάση, με την Αθήνα να ηγείται, συνεισφέροντας €279 χιλ., που αντιστοιχεί στο 69% της συνολικής CLV του τμήματος. Η Θεσσαλονίκη και ο Πειραιάς διατηρούν τη σημασία τους ως βασικοί συνεισφέροντες, με €41 χιλ. (10%) και €33 χιλ. (8%), αντίστοιχα. Περιοχές όπως η Πάτρα, που είχαν περιορισμένη συνεισφορά στο τμήμα των Loyalists, εμφανίζουν παρόμοια χαμηλή αξία και σε αυτό το τμήμα, υποδηλώνοντας ότι η επίδρασή τους στη μακροπρόθεσμη κερδοφορία της εταιρείας παραμένει περιορισμένη.

Στο τμήμα των Churners, οι συνολικές τιμές CLV είναι αναμενόμενα πολύ χαμηλότερες σε σύγκριση με τα άλλα δύο τμήματα. Ωστόσο, η Αθήνα παραμένει ο μεγαλύτερος συνεισφέρων, δημιουργώντας €49 χιλ., που αντιστοιχεί στο 65% της συνολικής CLV των Churners. Η Θεσσαλονίκη και ο Πειραιάς ακολουθούν, με €10 χιλ. (13%) και €6 χιλ. (7%), αντίστοιχα. Αξιοσημείωτο είναι το γεγονός ότι περιοχές όπως το Ηράκλειο Κρήτης συνεισφέρουν ακόμα λιγότερο σε αυτό το τμήμα, με τιμές κάτω του 3%, γεγονός που υποδηλώνει έλλειψη ενδιαφέροντος για την πλατφόρμα.

Η κυριαρχία της Αθήνας σε όλα τα τμήματα υπογραμμίζει τον κρίσιμο ρόλο της στην αύξηση των εσόδων της εταιρείας και στη διατήρηση πελατών. Παρόλο που η CLV σε άλλες περιοχές είναι σχετικά μικρή, στοχευμένες προσπάθειες για την ενίσχυση της εμπλοκής και της αγοραστικής δραστηριότητας σε περιοχές όπως η Θεσσαλονίκη, ο Πειραιάς και η Πάτρα θα μπορούσαν να ξεκλειδώσουν πρόσθετη αξία. Επιπλέον, η αμελητέα συνεισφορά μικρότερων περιοχών υποδηλώνει πιθανές ευκαιρίες για τοπικές στρατηγικές μάρκετινγκ, οι οποίες θα μπορούσαν να ενισχύσουν την πελατειακή βάση και να αυξήσουν τη μακροπρόθεσμη κερδοφορία.



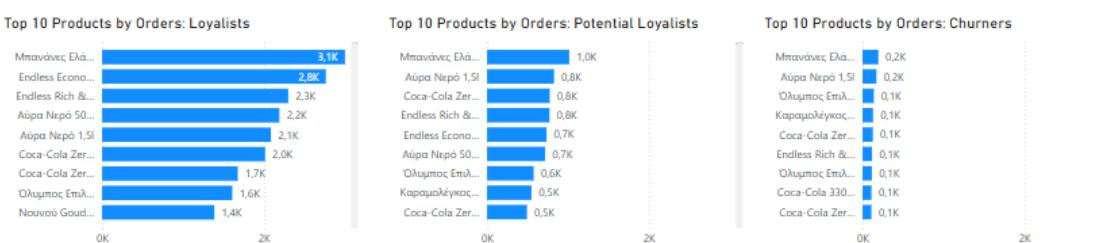
Εικόνα 56. Κορυφαίες 10 κατηγορίες προϊόντων βάσει τζίρου – Συγκριτικό ανά ομάδα πελατών

Η ανάλυση των δέκα κορυφαίων κατηγοριών προϊόντων βάσει εσόδων (Εικόνα 51) αναδεικνύει σημαντικές διαφορές τόσο στη συμπεριφορά αγορών όσο και στη συνολική συνεισφορά εσόδων μεταξύ των τμημάτων πελατών.

Στην ομάδα των Loyalists, η κατηγορία των Τσιγάρων καταλαμβάνει την πρώτη θέση, αποφέροντας €88.1 χιλ. σε έσοδα, ακολουθούμενη από τα Tereia (€69.5 χιλ.). Άλλες κορυφαίες κατηγορίες περιλαμβάνουν τον Καπνό με €45.7 χιλ., το Γάλα με €38.8 χιλ., και το Αλκοόλ με €36.7 χιλ., αντικατοπτρίζοντας μια ισορροπημένη αγοραστική συμπεριφορά που περιλαμβάνει τόσο βασικά προϊόντα όσο και προϊόντα υψηλότερης αξίας. Οι Loyalists συνεισφέρουν σημαντικά περισσότερο στα συνολικά έσοδα σε σύγκριση με τις άλλες δύο ομάδες, με κάθε μία από τις κορυφαίες κατηγορίες προϊόντων τους να αποφέρει πολλαπλάσια έσοδα σε σχέση με τις αντίστοιχες δημοφιλείς κατηγορίες των Potential Loyalists.

Οι Potential Loyalists εμφανίζουν διαφορετικές προτιμήσεις, με την κατηγορία των Παγωτών να ηγείται, αποφέροντας €18.4 χιλ., ακολουθούμενη από τα Τσιγάρα με €12.7 χιλ. και τα Νερά & Αναψυκτικά με €11.7 χιλ.. Αν και η κατηγορία των Τσιγάρων εξακολουθεί να είναι δημοφιλής, τα έσοδά της είναι σημαντικά χαμηλότερα σε σύγκριση με αυτά των Loyalists. Η ομάδα αυτή παρουσιάζει ελαφρώς μεγαλύτερη προτίμηση προς εποχικά προϊόντα, όπως τα παγωτά και τα νερά, αντικατοπτρίζοντας έναν συνδυασμό σταθερών και εποχικών αγοραστικών προτύπων.

Οι Churners, σε πλήρη αντίθεση, καταγράφουν τη χαμηλότερη συνολική συνεισφορά σε έσοδα από τις τρεις ομάδες, με την κατηγορία των Παγωτών να κυριαρχεί, αποφέροντας €9.5 χιλ. συνολικά. Αξιοσημείωτο είναι το γεγονός ότι εποχικά προϊόντα, όπως τα Παγωτά και τα Νερά & Αναψυκτικά, κατέχουν κυριαρχηθέση στις πρώτες τέσσερις κατηγορίες. Σε αντίθεση με τις άλλες ομάδες, οι Churners δεν εμφανίζουν τις κατηγορίες Καπνού ή Τσιγάρων στο top 10, γεγονός που επιβεβαιώνει την παροδική και εποχική φύση των αγορών τους. Αυτή η τάση υποδηλώνει ότι οι Churners πραγματοποιούν σποραδικές αγορές, κυρίως κατά τους καλοκαιρινούς μήνες, πριν αποστασιοποιηθούν εντελώς από την πλατφόρμα.



Εικόνα 57. Κορυφαία 20 προϊόντα σε πωλήσεις ανά παραγγελία - Συγκριτικό ανά ομάδα πελατών

Στα άνω διαγράμματα (Εικόνα 52) παρατηρούμε την ανάλυση των δέκα κορυφαίων προϊόντων βάσει αριθμού παραγγελιών.

Μεταξύ των Loyalists, το πιο συχνά παραγγελθέν προϊόν είναι οι μπανάνες (ελάχιστο βάρος 800g), με 3.1 χιλ. παραγγελίες, ακολουθούμενες από τα οικονομικά ρολά κουζίνας και το απαλό

χαρτί υγείας, και τα δύο με πάνω από 2 χιλ. παραγγελίες. Επίσης, σημαντική θέση κατέχουν το εμφιαλωμένο νερό και τα προϊόντα Coca-Cola Zero, γεγονός που υποδηλώνει προτίμηση σε βασικά και καθημερινά καταναλωτικά προϊόντα.

Οι Potential Loyalists εμφανίζουν παρόμοια πρότυπα, με τις μπανάνες να παραμένουν στην κορυφή της λίστας, με 1 χιλ. παραγγελίες, ενώ ακολουθούν το εμφιαλωμένο νερό 1.5L και η Coca-Cola Zero. Οι πελάτες αυτοί δείχνουν έντονη προτίμηση και σε άλλα βασικά προϊόντα παντοπωλείου, όπως το φρέσκο γάλα και το ψωμί, γεγονός που αντικατοπτρίζει τις τυπικές ανάγκες ενός νοικοκυριού.

Οι Churners, παρά τον σημαντικά χαμηλότερο συνολικό όγκο παραγγελιών, διατηρούν ένα παρόμοιο μοτίβο αγορών, με τις μπανάνες να βρίσκονται και πάλι στην κορυφή, με 0.2 χιλ. παραγγελίες, ακολουθούμενες από το εμφιαλωμένο νερό και το φρέσκο γάλα. Οι δέκα κορυφαίες κατηγορίες προϊόντων των Churners συμπίπτουν σε μεγάλο βαθμό με αυτές των άλλων τμημάτων. Είναι όμως αξιοσημείωτο το γεγονός ότι δεν παρατηρούνται τα παγωτά ανάμεσα στα κορυφαία προϊόντα βάσει παραγγελιών, παρόλο που είδαμε νωρίτερα ότι είναι η προτιμώμενη κατηγορία προϊόντων του συγκεκριμένου γκρουπ.

Η ελάχιστη διαφοροποίηση στις κορυφαίες κατηγορίες προϊόντων μεταξύ των τμημάτων καταδεικνύει ότι, παρά τις διαφορές στην εμπλοκή των πελατών, οι βασικές καταναλωτικές ανάγκες παραμένουν σταθερές. Αυτό τονίζει τη γενική ελκυστικότητα ορισμένων προϊόντων, ιδίως σε κατηγορίες όπως τα φρέσκα τρόφιμα και τα ποτά. Παράλληλα, αναδεικνύεται ότι η κύρια διαφοροποίηση μεταξύ των τμημάτων ενδεχομένως δεν αφορά τόσο τις προϊοντικές επιλογές τους, αλλά τη συχνότητα και τον συνολικό όγκο των παραγγελιών τους.

8.4.1 Σύγκριση τμημάτων πελατών

Η σύγκριση των Loyalists, Potential Loyalists και Churners αποκαλύπτει σημαντικά μοτίβα που προσφέρουν πολύτιμες πληροφορίες για στοχευμένες επιχειρηματικές στρατηγικές. Οι Loyalists είναι αναμφισβήτητα η πιο αφοσιωμένη και πολύτιμη ομάδα, καθώς διαμορφώνουν τη μεγαλύτερη συνεισφορά στα έσοδα και καταγράφουν τον υψηλότερο αριθμό παραγγελιών. Η δραστηριότητά τους παραμένει σταθερή και ακολουθεί μια θετική τάση σε όλους τους δείκτες, όπως έσοδα, όγκος παραγγελιών και αριθμός ενεργών πελατών ανά μήνα. Οι κορυφαίες κατηγορίες προϊόντων τους περιλαμβάνουν καπνικά προϊόντα, Tereia και τσιγάρα, τα οποία αποτελούν σημαντικό ποσοστό των συνολικών αγορών τους. Το τμήμα αυτό παρουσιάζει επίσης τον χαμηλότερο μέσο αριθμό ημερών μεταξύ παραγγελιών, γεγονός που υποδηλώνει σταθερή και προβλέψιμη καταναλωτική συμπεριφορά σε διάφορες γεωγραφικές περιοχές.

Οι Potential Loyalists, αν και λιγότερο ενεργοί από τους Loyalists, εξακολουθούν να συνεισφέρουν σημαντικά στα συνολικά έσοδα. Η δραστηριότητά τους κορυφώνεται τους καλοκαιρινούς μήνες, πριν σημειώσει αισθητή πτώση τον Σεπτέμβριο, γεγονός που υποδηλώνει λιγότερο συνεπή αγοραστική συμπεριφορά. Αξιοσημείωτο είναι ότι αυτό το τμήμα καταγράφει υψηλή μέση αξία παραγγελιάς σε κάποιες περιοχές όπως η Δροσιά, γεγονός που δείχνει ότι, αν και μπορεί να πραγματοποιούν λιγότερες αγορές, όταν το κάνουν, ξοδεύουν μεγαλύτερα ποσά ανά συναλλαγή. Οι Potential Loyalists παρουσιάζουν επίσης ορισμένες ομοιότητες με τους Loyalists στις προϊοντικές τους επιλογές, καθώς καταναλώνουν προϊόντα όπως παγωτά, καπνικά προϊόντα και ποτά, αν και σε χαμηλότερους συνολικούς όγκους.

Οι Churners διαμορφώνουν μια εντελώς διαφορετική εικόνα. Η δραστηριότητά τους συγκεντρώνεται σχεδόν αποκλειστικά τον Αύγουστο, υποδηλώνοντας ότι είναι κυρίως εποχιακοί αγοραστές, πιθανότατα προσελκυόμενοι από προϊόντα που συνδέονται με το καλοκαίρι, όπως παγωτά και εμφιαλωμένο νερό. Σε αντίθεση με τις άλλες ομάδες, οι Churners δεν περιλαμβάνουν τα καπνικά προϊόντα ή τα τσιγάρα στις κορυφαίες κατηγορίες τους, γεγονός που ενισχύει την υπόθεση ότι η αγοραστική τους συμπεριφορά είναι περιστασιακή και λιγότερο προσανατολισμένη σε επαναλαμβανόμενες αγορές. Η συνεισφορά τους στα συνολικά έσοδα και στον όγκο παραγγελιών είναι ελάχιστη, ενώ καταγράφουν τον υψηλότερο μέσο αριθμό ημερών μεταξύ παραγγελιών, επιβεβαιώνοντας τη χαμηλή και μη διατηρήσιμη εμπλοκή τους με την πλατφόρμα.

Παρά τις διαφοροποιήσεις στην αγοραστική συμπεριφορά, ορισμένα προϊόντα, όπως οι μπανάνες, το εμφιαλωμένο νερό και η Coca-Cola Zero, καταγράφουν υψηλή δημοτικότητα σε όλα τα τμήματα, γεγονός που υποδηλώνει την καθολική ελκυστικότητά τους. Ωστόσο, οι διαφορές στη συχνότητα παραγγελιών, τη μέση αξία ανά παραγγελία και τη συνεισφορά στα συνολικά έσοδα καταδεικνύουν πτοικίλα επίπεδα εμπιστοσύνης και αφοσίωσης των πελατών. Οι περιοχές όπως η Αθήνα και τα προάστια της κυριαρχούν στις συνολικές συνεισφορές εσόδων σε όλες τις ομάδες, όμως διάφορες τάσεις, όπως η υψηλή αγοραστική δύναμη των Potential Loyalists στη Δροσιά και η χαμηλή εμπλοκή των Churners στις περισσότερες περιοχές, αναδεικνύουν ευκαιρίες για τοπικές στρατηγικές προσέγγισης.

Με την ολοκλήρωση της ανάλυσης των πελατειακών τμημάτων, η οποία προσέφερε πολύτιμες πληροφορίες για τα πρότυπα και τις προτιμήσεις των πελατειακών τμημάτων, το ενδιαφέρον στρέφεται πλέον στο επόμενο κρίσιμο βήμα: την Ανάλυση Συσχέτισης Προϊόντων (Market Basket Analysis - MBA). Μέσω αυτής της μεθόδου, θα αναζητηθούν σχέσεις και συσχετίσεις μεταξύ προϊόντων που αγοράζονται συχνά μαζί, παρέχοντας εφαρμόσιμες γνώσεις για cross-selling, συνδυαστικές πρωθητικές ενέργειες και στρατηγικές διαχείρισης αποθεμάτων.

ΚΕΦΑΛΑΙΟ 9. Υλοποίηση Market Basket Analysis

9.1 Μεθοδολογία

Η Ανάλυση Συσχέτισης Προϊόντων (Market Basket Analysis - MBA) αποτελεί μία τεχνική εξόρυξης δεδομένων που χρησιμοποιείται για την ταυτοποίηση προτύπων στην αγοραστική συμπεριφορά των πελατών. Μέσω της ανάλυσης των συναλλαγών, οι επιχειρήσεις μπορούν να εντοπίσουν σχέσεις μεταξύ προϊόντων που αγοράζονται συχνά μαζί, παρέχοντας πολύτιμες πληροφορίες για στοχευμένες πρωθητικές ενέργειες, στρατηγικές συνδυαστικών πωλήσεων (cross-selling) και βελτιστοποίηση της διαχείρισης αποθέματος. Οι βασικοί δείκτες που χρησιμοποιούνται στην MBA—support, confidence και lift—ποσοτικοποιούν αυτές τις σχέσεις και αξιολογούν τη σημασία τους για την επιχείρηση.

Frequent Itemsets (Συχνά Σύνολα Προϊόντων):

Ένα frequent itemset αναφέρεται σε έναν συνδυασμό προϊόντων που εμφανίζονται συχνά μαζί σε ένα σημαντικό αριθμό συναλλαγών πελατών. Η αναγνώριση αυτών των συνδυασμών επιτρέπει στις επιχειρήσεις να αντιληφθούν τις αγοραστικές τάσεις και να βελτιστοποιήσουν τη διάταξη των προϊόντων ή τις πρωθητικές στρατηγικές τους. Για παράδειγμα, εάν ένας μεγάλος αριθμός πελατών αγοράζει συχνά ψωμί και βούτυρο μαζί, τότε ο συνδυασμός {ψωμί, βούτυρο} θεωρείται frequent itemset. Η συχνότητα εμφάνισης τέτοιων συνδυασμών αποτελεί καθοριστικό παράγοντα για τη δημιουργία association rules.

Association Rules (Κανόνες Συσχέτισης)

Ένας κανόνας συσχέτισης περιγράφει ένα πρότυπο αγορών, το οποίο δείχνει πώς η παρουσία ενός προϊόντος σε μια συναλλαγή επηρεάζει την πιθανότητα αγοράς ενός άλλου προϊόντος. Οι κανόνες αυτοί εκφράζονται στη μορφή:

$$A \rightarrow B$$

όπου A είναι το antecedent (το προϊόν που αγοράζεται πρώτο) και B είναι το consequent (το προϊόν που είναι πιθανό να αγοραστεί μαζί με το A).

Για παράδειγμα, αν η ανάλυση δείξει ότι ένας σημαντικός αριθμός πελατών που αγοράζει καφέ προχωρά επίσης στην αγορά ζάχαρης, τότε ο κανόνας:

$$\text{Καφές} \rightarrow \text{Ζάχαρη}$$

υποδεικνύει μία ισχυρή σχέση μεταξύ των δύο προϊόντων, την οποία οι επιχειρήσεις μπορούν να αξιοποιήσουν για συνδυαστικές πωλήσεις ή στοχευμένες προσφορές.

Support (Υποστήριξη - Συχνότητα Εμφάνισης)

Το support μετρά πόσο συχνά ένα προϊόν ή ένας συνδυασμός προϊόντων εμφανίζεται μέσα στο σύνολο των συναλλαγών. Ο δείκτης αυτός προσδιορίζει τη σημασία ενός συνδυασμού προϊόντων, υπολογίζοντας το ποσοστό των συναλλαγών που περιέχουν ένα συγκεκριμένο προϊόν ή συνδυασμό προϊόντων.

Μαθηματικός ορισμός:

$$\text{Support}(A) = \frac{\text{Αριθμός συναλλαγών που περιέχουν το } A}{\text{Συνολικός αριθμός συναλλαγών}}$$

Για παράδειγμα, αν ο καφές εμφανίζεται σε 500 από τις 10.000 συναλλαγές, το support για τον καφέ είναι:

$$\text{Support} = \frac{500}{10000} = 5\%$$

Αντίστοιχα, αν ο συνδυασμός καφές και ζάχαρη εμφανίζεται σε 250 συναλλαγές, το support για τον συνδυασμό είναι:

$$\text{Support} = \frac{250}{10000} = 2.5\%$$

Υψηλές τιμές support υποδηλώνουν ότι ένα προϊόν ή ένας συνδυασμός προϊόντων εμφανίζεται συχνά, καθιστώντας τον σημαντικό για περαιτέρω ανάλυση.

Confidence (Εμπιστοσύνη - Πιθανότητα Αγοράς Συνοδευτικού Προϊόντος)

Η confidence μετρά την πιθανότητα ένας πελάτης που αγοράζει ένα προϊόν A να αγοράσει επίσης ένα προϊόν B. Ο δείκτης αυτός αξιολογεί τη δύναμη ενός κανόνα συσχέτισης και υπολογίζεται ως εξής:

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

Για παράδειγμα, αν ο καφές εμφανίζεται σε 500 συναλλαγές και ο συνδυασμός καφές και ζάχαρη εμφανίζεται σε 250 συναλλαγές, τότε η confidence για τον κανόνα Καφές → Ζάχαρη είναι:

$$\text{Confidence} = \frac{250}{500} = 50\%$$

Αυτό σημαίνει ότι το 50% των πελατών που αγοράζουν καφέ αγοράζουν επίσης ζάχαρη. Όσο υψηλότερη είναι η confidence, τόσο πιο ισχυρή θεωρείται η σχέση μεταξύ των προϊόντων.

Lift (Συσχέτιση Προϊόντων - Αύξηση Πιθανοτήτων Αγοράς)

Το lift μετρά κατά πόσο η πιθανότητα αγοράς ενός προϊόντος αυξάνεται όταν ένα άλλο προϊόν έχει ήδη αγοραστεί, συγκριτικά με το εάν τα προϊόντα αυτά αγοράζονταν ανεξάρτητα.

Μαθηματικός ορισμός:

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

- Αν $Lift = 1$, τα προϊόντα είναι ανεξάρτητα (δηλαδή η αγορά του A δεν επηρεάζει την πιθανότητα αγοράς του B).
- Αν $Lift > 1$, η αγορά του A αυξάνει την πιθανότητα αγοράς του B (ισχυρή συσχέτιση).
- Αν $Lift < 1$, η αγορά του A μειώνει την πιθανότητα αγοράς του B (αρνητική συσχέτιση).

Για παράδειγμα, αν:

- $\text{Confidence}(\text{Καφές} \rightarrow \text{Ζάχαρη}) = 50\%$
- $\text{Support}(\text{Ζάχαρη}) = 30\%$

τότε το lift είναι:

$$Lift = \frac{50\%}{30\%} = 1.67$$

Αυτό σημαίνει ότι η αγορά καφέ αυξάνει κατά 67% την πιθανότητα ο πελάτης να αγοράσει και ζάχαρη.

Η κατανόηση των δεικτών support, confidence και lift επιτρέπει στις επιχειρήσεις να λαμβάνουν τεκμηριωμένες αποφάσεις σε τομείς όπως:

- Product Bundling: Ανάπτυξη πακέτων προϊόντων (π.χ. καφές + ζάχαρη).
- Cross-Selling: Προτάσεις συμπληρωματικών προϊόντων (smart recommendations).
- Στοχευμένες Προσφορές: Εκπτώσεις σε προϊόντα υψηλής lift για αύξηση πωλήσεων.
- Διαχείριση Αποθεμάτων: Βελτιστοποίηση διαθεσιμότητας προϊόντων με ισχυρές συσχετίσεις.
- Αξιοποιώντας αυτά τα ευρήματα, οι επιχειρήσεις μπορούν να αυξήσουν τα έσοδα, να βελτιώσουν την εμπειρία των πελατών και να εφαρμόσουν αποτελεσματικότερες στρατηγικές μάρκετινγκ. (GeeksforGeeks, n.d.)

Το σύνολο δεδομένων που χρησιμοποιήθηκε για αυτή την ανάλυση αποτελείται από 574,855 συναλλαγές, καλύπτοντας 10,000 μοναδικούς πελάτες και 17 χαρακτηριστικά, συμπεριλαμβανομένων των order IDs, product names, product categories, αριθμού προϊόντων που πωλήθηκαν και cluster labels. Κάθε συναλλαγή αντιπροσωπεύει μια αγορά πελάτη, η οποία μπορεί να περιλαμβάνει ένα ή περισσότερα προϊόντα. Ο στόχος της MBA ήταν να παραχθούν frequent itemsets και association rules που να πληρούν τα καθορισμένα thresholds για support, confidence και lift.

9.2 Εφαρμογή του Apriori αλγορίθμου

Ο αλγόριθμος Apriori αποτελεί μία θεμελιώδη μέθοδο στην εξόρυξη δεδομένων (data mining), η οποία χρησιμοποιείται για την ανακάλυψη συχνών συνόλων αντικειμένων (frequent itemsets) και τη δημιουργία κανόνων συσχέτισης (association rules). Η σημασία του έγκειται στην ικανότητά του να αναγνωρίζει σχέσεις μεταξύ αντικειμένων σε μεγάλα σύνολα δεδομένων, γεγονός που τον καθιστά εξαιρετικά χρήσιμο στην ανάλυση καλαθιού αγορών.

Ο αλγόριθμος Apriori λειτουργεί μέσω μίας διαδοχικής διαδικασίας που περιλαμβάνει τα στάδια:

Εντοπισμός Συχνών Συνόλων Αντικειμένων: Ο αλγόριθμος αρχικά αναζητά μεμονωμένα αντικείμενα (1-itemsets) και μετρά τη συχνότητά τους. Χρησιμοποιείται ένα κατώφλι υποστήριξης (support threshold), ώστε να καθοριστεί ποια αντικείμενα θεωρούνται συχνά.

Δημιουργία Υποψήφιων Ομάδων Αντικειμένων: Στη συνέχεια, παράγονται συνδυασμοί αντικειμένων δύο ή περισσότερων στοιχείων (2-itemsets, 3-itemsets κ.ο.κ.), με βάση τα συχνά αντικείμενα της προηγούμενης φάσης. Η διαδικασία συνεχίζεται μέχρι να μην υπάρχουν πλέον υποψήφια σύνολα που πληρούν το όριο υποστήριξης.

Απομάκρυνση Ασυνήθιστων Συνδυασμών: Ο αλγόριθμος χρησιμοποιεί την ιδιότητα του Apriori (Apriori Property), η οποία δηλώνει ότι αν ένα σύνολο αντικειμένων είναι σπάνιο, τότε και όλα τα υπερσύνολά του θα είναι επίσης σπάνια. Αυτό μειώνει σημαντικά τον αριθμό των συνδυασμών που χρειάζεται να αξιολογηθούν.

Εξαγωγή Κανόνων Συσχέτισης: Αφού εντοπιστούν τα συχνά σύνολα αντικειμένων, δημιουργούνται κανόνες συσχέτισης που περιγράφουν πώς σχετίζονται μεταξύ τους, χρησιμοποιώντας δείκτες αξιολόγησης, όπως:

Υποστήριξη (Support): Δείχνει πόσο συχνά ένα αντικείμενο ή ένας συνδυασμός αντικειμένων εμφανίζεται στο σύνολο δεδομένων.

Εμπιστοσύνη (Confidence): Εκτιμά την πιθανότητα ότι ένα αντικείμενο Υ αγοράζεται όταν αγοράζεται ένα αντικείμενο X.

Ανύψωση (Lift): Καθορίζει πόσο ισχυρή είναι η σχέση μεταξύ δύο αντικειμένων, συγκρίνοντας τη συχνότητα της συσχέτισης με τη συχνότητα που θα αναμενόταν να εμφανίζονται ανεξάρτητα (GeeksforGeeks, 2025).

Ο αλγόριθμος Apriori παρουσιάζει αρκετά πλεονεκτήματα που τον καθιστούν χρήσιμο εργαλείο στην εξόρυξη δεδομένων και ειδικότερα στην ανάλυση καλαθιού αγορών:

Απλότητα και Ευκολία Κατανόησης: Ο αλγόριθμος είναι διαισθητικός, με σαφή και διαδοχικά βήματα.

Αποτελεσματικός σε Δεδομένα Λιανικής: Βρίσκει συσχετίσεις που μπορούν να αξιοποιηθούν σε πρωθητικές ενέργειες και cross-selling στρατηγικές.

Εφαρμογή σε Διάφορους Τομείς: Χρησιμοποιείται όχι μόνο στο λιανικό εμπόριο, αλλά και σε τομείς όπως η ιατρική διάγνωση, η ανίχνευση απάτης και η ανάλυση κοινωνικών δικτύων.

Ωστόσο, έχει και ορισμένους περιορισμούς:

Υψηλή Υπολογιστική Πολυτλοκότητα: Σε μεγάλα σύνολα δεδομένων, ο αριθμός των υποψήφιων συνόλων αντικειμένων αυξάνεται εκθετικά, καθιστώντας την εκτέλεσή του χρονοβόρα και απαιτητική σε μνήμη.

Αναποτελεσματικός για Σπάνιες Συσχετίσεις: Αν ένα προϊόν αγοράζεται σπάνια, ενδέχεται να μην εντοπιστεί από τον αλγόριθμο.

Απαιτεί Χειροκίνητο Καθορισμό Κατωφλίων: Η επιλογή ελάχιστης υποστήριξης και εμπιστοσύνης πρέπει να γίνεται προσεκτικά, καθώς επηρεάζει άμεσα την ποιότητα των κανόνων που παράγονται.

Παρά τις αδυναμίες του, ο αλγόριθμος Apriori παραφένει μία ευρέως χρησιμοποιούμενη μέθοδος στην ανάλυση συσχετίσεων, συμβάλλοντας σημαντικά στη βελτιστοποίηση επιχειρηματικών στρατηγικών και στη βελτίωση της εμπειρίας των πελατών.

Στην παρούσα ανάλυση, οι αρχικές παράμετροι που τέθηκαν περιελάμβαναν support threshold 0.05 και confidence threshold 0.10. Αν και αυτά τα κατώφλια είναι συνήθως κατάλληλα για μικρότερα σύνολα δεδομένων ή για συχνότερους συνδυασμούς προϊόντων, το αποτέλεσμα ήταν μόλις 2 association rules. Το περιορισμένο πλήθος κανόνων δεν επέτρεπε την εξαγωγή

ουσιαστικών επιχειρηματικών συμπερασμάτων, καθιστώντας αναγκαία τη ρύθμιση των thresholds για τη βελτίωση των αποτελεσμάτων.

Για την αντιμετώπιση αυτού του ζητήματος, το support threshold μειώθηκε αρχικά σε 0.01, προκειμένου να συμπεριληφθούν λιγότερο συχνά itemsets διατηρώντας παράλληλα διαχειρίσιμο υπολογιστικό φόρτο. Παρότι αυτή η αλλαγή αύξησε ελαφρώς τον αριθμό των κανόνων, τα αποτελέσματα παρέμεναν περιορισμένα. Με περαιτέρω μείωση του support threshold σε 0.005 και του confidence threshold σε 0.01, προέκυψαν περισσότεροι κανόνες, αλλά η εκτέλεση του Apriori οδήγησε σε MemoryError, λόγω των υπολογιστικών απαιτήσεων του αλγορίθμου σε ένα τόσο μεγάλο σύνολο δεδομένων.

Για να ξεπεραστούν οι υπολογιστικοί περιορισμοί και να ενισχυθεί η παραγωγή κανόνων, το σύνολο δεδομένων μειώθηκε σε μέγεθος, εστιάζοντας στα 100 πιο συχνά αγοραζόμενα προϊόντα. Με το φίλτραρισμά του και την εφαρμογή του Apriori, αυτή η προσαρμογή απέδωσε 44 association rules. Ωστόσο, όταν η ανάλυση επεκτάθηκε στα 300 πιο συχνά αγοραζόμενα προϊόντα, προέκυψαν μόλις 20 association rules, γεγονός που υποδηλώνει ότι η ενσωμάτωση λιγότερο δημοφιλών προϊόντων μείωσε το συνολικό support, περιορίζοντας τον αριθμό των frequent itemsets.

Με περαιτέρω προσαρμογές, το support threshold μειώθηκε σε 0.002, διατηρώντας την ανάλυση στα 100 κορυφαία προϊόντα. Αυτός ο συνδυασμός απέδωσε 142 association rules, επιτυχάνοντας ισορροπία μεταξύ υπολογιστικής αποδοτικότητας και ποικιλομορφίας κανόνων. Ωστόσο, δεν κατέστη δυνατός ο εντοπισμός κανόνων με πολλαπλά antecedents, οι οποίοι θα μπορούσαν να αποκαλύψουν πιο σύνθετα αγοραστικά μοτίβα. Όταν επιχειρήθηκε περαιτέρω μείωση του support threshold σε 0.001, οι υπολογιστικοί περιορισμοί, όπως τα MemoryErrors, συνέχισαν να αποτελούν εμπόδιο, επιβεβαιώνοντας τα όρια του Apriori σε μεγάλα σύνολα δεδομένων.

Τελικά, το support threshold ορίστηκε στο 0.001 και το confidence threshold στο 0.005, οδηγώντας στη δημιουργία 454 association rules. Από αυτά, 250 ταξινομήθηκαν ως high-lift rules, γεγονός που υποδηλώνει ισχυρές συσχετίσεις μεταξύ προϊόντων. Τα επίπεδα confidence κυμάνθηκαν από 0.015 έως 0.400, με μέση τιμή 0.083, αντικατοπτρίζοντας διαφορετικά επίπεδα προβλεπτικής ισχύος μεταξύ των κανόνων. Τα επίπεδα support κυμάνθηκαν από 0.001 έως 0.0105, διασφαλίζοντας ότι οι κανόνες εφαρμόζονται σε ένα λογικό ποσοστό των συναλλαγών. Η τελική ανάλυση περιλάμβανε 82 μοναδικά προϊόντα, τα οποία κάλυπταν ευρεία γκάμα κατηγοριών, όπως είδη παντοπωλείου, ποτά, σνακ και φρέσκα προϊόντα.

Count	support	confidence	lift	rule
238	0.002497397040794524349	0.2497959305145578	34.480313093709426	Tiborg Pink Grapefruit 330ml → Καρέλα Κομψό Απόσκοτο, Lime & Λαζαρές 330ml Tiborg Pink Grapefruit 330ml → Καρέλα Κομψό Απόσκοτο + Σwan φλαγκόνι Extra Slim 120τμ
100	0.00103424650631435538	0.352941130705882	24.480313093709426	Karelia Kompsi Apskso 300g, Rizla Xpertsos Extra Slim 120τμ
100	0.00103424650631435538	0.4	39.959692307692308	Swan φλαγκόνι Extra Slim 120τμ + Karelia Kompsi Apskso 300g, Rizla Xpertsos Απόσκοτο
100	0.00103424650631435538	0.2956497130477	24.480313093709426	Rizla Xpertsos Απόσκοτο 120τμ + Karelia Kompsi Apskso 300g
100	0.00103424650631435538	0.2566037735849056	35.678699707680043	Karelia Kompsi Apskso 300g, Swan φλαγκόνι Extra Slim 120τμ + Karelia Kompsi Apskso 300g, Rizla Xpertsos Απόσκοτο 120τμ
179	0.0013545801846798934	0.05912915972992766	3.13102915972992766	Karelia Kompsi Apskso 300g, Rizla Xpertsos Απόσκοτο 120τμ
304	0.00113101904857078113	0.24879807692307693	35.41747744212387	Ruffles Φιλοξενη Άλατη 90g + Ruffles Φιλοξενη Ριγέν 90g
100	0.0013545801846798934	0.13102915972992766	3.13102915972992766	Ruffles Φιλοξενη Άλατη 90g + Ruffles Φιλοξενη Ριγέν 90g
232	0.0024356789588570832	0.18566393247943595	24.4795863805264512	Bio Delicious Αγγειοπλαστική Κουτιού Bio delicious Βάρος 900g → Bio Delicious Ντιζάιτες Bio delicious Βιολογικές Συσκευασμένες Ελαχύτο Βάρος 900g
232	0.0024356789588570832	0.19410319410319413	24.4795863805264512	Bio Delicious Ντιζάιτες Bio delicious Βιολογικές Ελαχύτο Βάρος 900g → Bio Delicious Αγγειοπλαστική Κουτιού Bio delicious Βιολογικές Συσκευασμένες Ελαχύτο Βάρος 900g
128	0.00234116450767630763	0.0298415961644335	14.26153461262624	Rizla Xpertsos Ταζάτσα → Swan φλαγκόνι Extra Slim 120τμ, Old Holborn Kompsi, Blonde 30g
128	0.00234116450767630763	0.0298415961644335	14.26153461262624	Old Holborn Kompsi, Blonde 30g, Rizla Xpertsos Ταζάτσα → Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα → Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra Slim 120τμ
128	0.00234116450767630763	0.06661307692307692	31.2757352176941176471	Old Holborn Kompsi, Blonde 30g + Rizla Xpertsos Ταζάτσα + Swan φλαγκόνι Extra Slim 120τμ + Swan φλαγκόνι Extra

Η παρούσα αναφορά (Εικόνα 53) συνοψίζει τους 50 κορυφαίους κανόνες που προέκυψαν από την εφαρμογή του αλγορίθμου Apriori, εστιάζοντας σε εκείνους με τις υψηλότερες τιμές lift. Η ανάλυση αποκάλυψε ότι τα προϊόντα με συμπληρωματικά ή συνηθισμένα αγοραστικά πρότυπα παρουσίαζαν ισχυρές συσχετίσεις. Ο δείκτης lift, ο οποίος υποδεικνύει πόσο πιο πιθανό είναι να αγοραστεί το consequent όταν έχει αγοραστεί το antecedent σε σύγκριση με την τυχαία πιθανότητα, χρησιμοποιήθηκε ως το βασικό κριτήριο κατάταξης των κανόνων. Οι τιμές lift στους 50 κορυφαίους κανόνες κυμάνθηκαν από περίπου 8.8 έως πάνω από 34, υποδεικνύοντας ισχυρές σχέσεις μεταξύ προϊόντων.

Για παράδειγμα, μία από τις ισχυρότερες συσχετίσεις ($lift = 34.48$) παρατηρήθηκε μεταξύ των προϊόντων Tuborg Pink Grapefruit 330ml και Tuborg Σόδα Ανανάς, Lime & Δυόσμος 330ml. Ο κανόνας αυτός ήταν αμφίδρομος, καθώς και τα δύο προϊόντα εμφανίζονται συχνά μαζί στις συναλλαγές. Αυτό υποδηλώνει ότι οι πελάτες που αγοράζουν το ένα από αυτά τα ποτά είναι πολύ πιθανό να αγοράσουν και το άλλο, γεγονός που καθιστά τα προϊόντα αυτά ιδανικούς υποψήφιους για bundling ή κοινές προωθητικές ενέργειες.

Ένας άλλος αξιοσημείωτος κανόνας αφορούσε τα προϊόντα Karelia Καπνός Λευκός 30g, Rizla Χαρτάκια Ασημένια, και Swan Φιλτράκια Extra Slim 120τεμ, με $lift = 19.96$. Ο κανόνας αυτός αναδεικνύει τη συνηθισμένη φύση των αγορών στην κατηγορία των καπνικών προϊόντων, όπου οι πελάτες συχνά αγοράζουν χαρτάκια και φίλτρα μαζί με καπνό. Μια τόσο ισχυρή συσχέτιση προσφέρει ευκαιρίες για στοχευμένες προωθητικές ενέργειες, όπως εκπτώσεις σε πακέτα προϊόντων ή προγράμματα επιβράβευσης για συχνούς αγοραστές.

Ένα σημαντικό μοτίβο αναδείχθηκε και στην κατηγορία των φρέσκων προϊόντων. Για παράδειγμα, ο κανόνας που περιλαμβάνει τα Μπανάνες Ελάχιστο Βάρος 800g, Δάμα Αγγουράκια Κνωσού Συσκευασμένα Ελάχιστο Βάρος 600g, και Δάμα Ντοματίνια Βελανίδια Συσκευασμένα Ελάχιστο Βάρος 250g παρουσίασε $lift = 13.14$. Το εύρημα αυτό υποδηλώνει ότι οι πελάτες που αγοράζουν αυτά τα φρούτα και λαχανικά είναι πιθανότατα υγιεινής διατροφής και προτιμούν φρέσκα, υψηλής ποιότητας προϊόντα. Στρατηγικές μάρκετινγκ θα μπορούσαν να εστιάσουν σε προτάσεις συνταγών ή πακέτα γευμάτων που περιλαμβάνουν αυτά τα προϊόντα για την ενίσχυση των πωλήσεων.

Η ανάλυση αποκάλυψε επίσης ισχυρές συσχετίσεις στις κατηγορίες των σνακ και γαλακτοκομικών προϊόντων. Για παράδειγμα, τα Ruffles Πατατάκια Ρίγανη 90g και Ruffles Πατατάκια Αλάτι 90g παρουσίασαν $lift = 15.41$, υποδεικνύοντας ότι οι πελάτες συχνά αγοράζουν πολλαπλές γεύσεις της ίδιας μάρκας σε μία συναλλαγή. Προωθητικές ενέργειες που βασίζονται σε "mix-and-match" εκπτώσεις για σνακ θα μπορούσαν να αξιοποιήσουν αυτή τη συμπεριφορά αγορών.

Τα επόμενα βήματα αυτής της ανάλυσης περιλαμβάνουν την εφαρμογή της MBA ξεχωριστά στα πελατειακά τμήματα, όπως οι Loyalists, οι Potential Loyalists, και οι Churners, για την εξαγωγή εξατομικευμένων πληροφοριών ανά πελατειακή ομάδα.

Top 50 Rules for Loyalists						
count	support	confidence	lift	rule	rhs	
102	0.0010725470848170235	0.053134962805516036	19.971674968290426	Rizla Φιλτράκια Extra Slim 120τμ → Karelia Καπνός Αποκά 30g → Rizla Χαρτάκια Ασημένια		
102	0.0010725470848170235	0.401259894516129	19.971674968290422	Karelia Καπνός Αποκά 30g → Rizla Φιλτράκια Ultra Slim 120τμ → Rizla Χαρτάκια Extra Slim 120τμ		
126	0.0013299583851731099	0.50819672131131754	19.963382222851816	Rizla Φιλτράκια Ultra Slim 120τμ → Old Holborn Καπνός, Blonde 30g → Rizla Χαρτάκια Γαλάζια		
126	0.0013299583851731099	0.7311277531036477	19.9719304170338	Old Holborn Καπνός, Blonde 30g → Rizla Φιλτράκια Ultra Slim 120τμ → Rizla Χαρτάκια Γαλάζια		
126	0.0013299583851731099	0.60672477962781587	15.524518552293192	Old Holborn Καπνός, Blonde 30g → Rizla Φιλτράκια Ultra Slim 120τμ → Rizla Χαρτάκια Γαλάζια		
126	0.0013299583851731099	0.60672477962781587	15.524518552293192	Old Holborn Καπνός, Blonde 30g → Rizla Φιλτράκια Ultra Slim 120τμ → Rizla Χαρτάκια Γαλάζια		
131	0.0011722860616856579	0.309178749613527	15.3169975499365	Karelia Καπνός Αποκά 30g → Rizla Φιλτράκια Ultra Slim 120τμ		
131	0.0011722860616856579	0.0880125259107733	15.3169975499365	Rizla Φιλτράκια Ultra Slim 120τμ → Karelia Καπνός Αποκά 30g → Rizla Χαρτάκια Γαλάζια		
131	0.0011722860616856579	0.257375757575757	15.3169975499365	Rizla Φιλτράκια Ultra Slim 120τμ → Rizla Χαρτάκια Γαλάζια		
905	0.00013619031274547	0.257375757575757	15.3169975499365	Karelia Καπνός Αποκά 30g → Rizla Φιλτράκια Ultra Slim 120τμ → Rizla Χαρτάκια Γαλάζια		
102	0.0010725470848170235	0.0861728950617283	15.3169975499365	Rizla Φιλτράκια Ultra Slim 120τμ → Karelia Καπνός Αποκά 30g → Rizla Χαρτάκια Γαλάζια		
409	0.0004901881192880093	0.21503981522210414	12.232934425057224	Rizla Φιλτράκια Extra Slim 120τμ → Rizla Χαρτάκια Γαλάζια		
131	0.0011722860616856579	0.309178749613527	15.3169975499365	Rizla Φιλτράκια Ultra Slim 120τμ → Rizla Χαρτάκια Γαλάζια		
131	0.0011722860616856579	0.309178749613527	15.3169975499365	Rizla Φιλτράκια Ultra Slim 120τμ → Rizla Χαρτάκια Γαλάζια		
961	0.000577558024797289	0.22776392352452204	11.293612979719627	Rizla Φιλτράκια Ultra Slim 120τμ → Rizla Χαρτάκια Γαλάζια		
561	0.005877558024797289	0.291179596174827	11.293612979719627	Rizla Φιλτράκια Ultra Slim 120τμ → Rizla Χαρτάκια Γαλάζια		
206	0.002616545113303874	0.1629031258064512	11.054513140594848	Magic Παγωτό Κύπελλο Double Salted Caramel 440ml → Magic Παγωτό Κύπελλο Double Gold Caramel 440ml		
270	0.0026131524033916942	0.15384615384615386	10.8666666666666657	Old Holborn Καπνός, Yellow 30g → Rizla Φιλτράκια Ultra Slim 120τμ		
372	0.003904071388713954	0.1517884455527847	10.866077204211693	Rizla Φιλτράκια Ultra Slim 120τμ → Rizla Χαρτάκια Ultra Slim 120τμ		
593	0.0002073091988736	0.17728444644917966	10.8419013109717	Rizla Φιλτράκια Ultra Slim 120τμ → Rizla Χαρτάκια Ultra Slim 120τμ		
593	0.0002073091988736	0.2499065973856076	10.471243901310971	Δίαιο Ντοματίνια Βελανίδια Συσκευασμένα Ελάχιστο Βάρος 250g → Διαιό Αγγουράκια Καπνού Συσκευασμένα Ελάχιστο Βάρος 600g		
593	0.0002073091988736	0.2499065973856076	10.471243901310971	Δίαιο Αγγουράκια Καπνού Συσκευασμένα Ελάχιστο Βάρος 600g → Δίαιο Ντοματίνια Βελανίδια Συσκευασμένα Ελάχιστο Βάρος 250g		
180	0.0002513213101681755	0.1401778857505138	10.2697942386613	Bio Delicous Ντοματίνια Βελανίδια Βιολογικά Συσκευασμένα Σε Τούτα Ελάχιστο Βάρος 250g → Bio Delicous Ντοματίνια Βελανίδια Βιολογικά Συσκευασμένα Ελάχιστο Βάρος 600g		
180	0.0002513213101681755	0.1401778857505138	10.2697942386613	Bio Delicous Ντοματίνια Βελανίδια Βιολογικά Συσκευασμένα Σε Τούτα Ελάχιστο Βάρος 250g → Bio Delicous Ντοματίνια Βελανίδια Βιολογικά Συσκευασμένα Ελάχιστο Βάρος 600g		
180	0.0002513213101681755	0.1401778857505138	10.2697942386613	Rizla Φιλτράκια Ultra Slim 120τμ → Rizla Χαρτάκια Γαλάζια		
102	0.0010725470848170235	0.08420808360277893	9.901869108879506	Rizla Φιλτράκια Ultra Slim 120τμ → Rizla Χαρτάκια Γαλάζια		
194	0.00031101101913466901	0.115802884691358027	9.751894172984668	Karelia Καπνός Αποκά 30g → Swan Φιλτράκια Extra Slim 120τμ, Rizla Χαρτάκια Ασημένια		
194	0.00031101101913466901	0.115802884691358027	9.751894172984668	Rizla Φιλτράκια Ultra Slim 120τμ → Old Holborn Καπνός, Yellow 30g		
137	0.00014732130936548115	0.085815283842794766	9.673281922616193	Rizla Χαρτάκια Ultra Slim 120τμ → Fix Hellas Kouti 500ml → Fix Hellas Kouti 330ml		
284	0.00291680859731325	0.14771519659934256	9.343573091310333	Swan Φιλτράκια Μεταλλικά Swan Φιλτράκια Ultra Slim 120τμ → Rizla Χαρτάκια Γαλάζια		
131	0.0011722860616856579	0.2335766423356642	9.251381403915683	Rizla Φιλτράκια Ultra Slim 120τμ → Rizla Χαρτάκια Γαλάζια		
423	0.0004404403311424477	0.1759708581138487	8.712803043093945	Rizla Φιλτράκια Ultra Slim 120τμ → Rizla Χαρτάκια Γαλάζια		
423	0.0004404403311424477	0.2199784601487777	8.712803043093945	Swan Φιλτράκια Extra Slim 120τμ → Rizla Χαρτάκια Γαλάζια		
464	0.004869537505959286	0.2223104797926452	8.651651531391414	Old Holborn Καπνός, Yellow 30g → Rizla Χαρτάκια Γαλάζια		
249	0.0026170148869553537	0.18484848484848485	8.440021181347866	Rizla Φιλτράκια Ultra Slim 120τμ → Old Holborn Καπνός, Yellow 30g		
968	0.0038611695013412847	0.209790209705098	8.12961510473815	Rizla Φιλτράκια Ultra Slim 120τμ → Old Holborn Καπνός, Yellow 30g		
968	0.0038611695013412847	0.4961259316209477	8.12961510473815	Rizla Φιλτράκια Ultra Slim 120τμ → Old Holborn Καπνός, Yellow 30g		

Εικόνα 59. Κορυφαίοι 50 κανόνες συσχέτισης μεταξύ αγορασμένων προϊόντων για την ομάδα πελατών "Loyalists", βάσει του Apriori αλγορίθμου

Η MBA που πραγματοποιήθηκε για το τμήμα των Loyalists (Εικόνα 54) εντόπισε 50 κανόνες με υψηλές τιμές lift. Το τμήμα αυτό περιλαμβάνει πελάτες με υψηλή αγοραστική συχνότητα και αξία συναλλαγών, καθιστώντας τους καθοριστικούς για τη διατήρηση σταθερών εσόδων της επιχείρησης.

Ο ισχυρότερος κανόνας που εντοπίστηκε, με lift = 19.98, αφορά την αγορά των Swan Φιλτράκια Extra Slim 120τεμ, η οποία συσχετίζεται με τα Karelia Καπνός Λευκός 30g και Rizla Χαρτάκια Ασημένια, με confidence 53.13%. Αυτό σημαίνει ότι πάνω από το 50% των πελατών που αγοράζουν τα φίλτρα Swan, αγοράζουν επίσης τα συγκεκριμένα καπνικά προϊόντα. Η υψηλή τιμή lift δείχνει ότι τα προϊόντα αυτά αγοράζονται μαζί πολύ συχνότερα από ότι θα συνέβαινε τυχαία, αντικατοπτρίζοντας ένα καθιερωμένο αγοραστικό μοτίβο στην κατηγορία των καπνικών προϊόντων και των σχετικών ειδών.

Ένας ακόμη σημαντικός κανόνας, με lift = 19.69, εντοπίστηκε μεταξύ των Rizla Χαρτάκια Γαλάζια, τα οποία παρουσιάζουν ισχυρή συσχέτιση με τα Rizla Φιλτράκια Ultra Slim 120τεμ και Old Holborn Καπνός Blonde 30g, με confidence 51.53%. Αυτό επιβεβαιώνει ότι οι Loyalists επιδεικνύουν σταθερή αγοραστική συμπεριφορά όσον αφορά τα προϊόντα καπνού, προτιμώντας να αγοράζουν χαρτάκια, φίλτρα και καπνό σε συνδυασμό.

Επιπλέον, η MBA ανέδειξε σημαντικές σχέσεις και σε άλλες κατηγορίες προϊόντων. Για παράδειγμα, ο κανόνας Magic Παγωτό Κύπελλο Double Gold Caramel 440ml → Magic Παγωτό Κύπελλο Double Salted Caramel 440ml, με lift = 11.05, υποδηλώνει ότι οι πελάτες που αγοράζουν μία γεύση του συγκεκριμένου brand παγωτού είναι ιδιαίτερα πιθανό να αγοράσουν και μία εναλλακτική γεύση της ίδιας μάρκας. Αυτό δημιουργεί ευκαιρίες για διασταυρούμενες πρωθητικές ενέργειες (cross-promotion) στα προϊόντα επιδορπίων, ενθαρρύνοντας την αγορά πολλαπλών επιλογών μέσα από προσφορές ή συνδυαστικές εκπτώσεις.

Η ανάλυση αποκάλυψε επίσης ισχυρές συσχέτισης στην κατηγορία των φρέσκων προϊόντων. Ο κανόνας Δάμα Ντοματίνια Βελανίδια Συσκευασμένα Ελάχιστο Βάρος 250g → Δάμα Αγγουράκια Κνωσσού Συσκευασμένα Ελάχιστο Βάρος 600g, με lift = 10.47, καταδεικνύει προτίμηση των πελατών στην αγορά συγκεκριμένων συνδυασμών φρούτων και λαχανικών. Αυτό το εύρημα υποδηλώνει ότι η ομαδοποίηση αυτών των προϊόντων σε συνδυαστικά πακέτα ή η εφαρμογή προσφορών με εκπτώσεις σε αγορές πολλαπλών ειδών θα μπορούσε να ενισχύσει την ικανοποίηση των πελατών και να αυξήσει τις συνολικές πωλήσεις στην κατηγορία των φρέσκων τροφίμων.

Εικόνα 60. Κορυφαίοι 50 κανόνες συσχέτισης μεταξύ αγορασμένων προϊόντων για την ομάδα πελατών "Potential Loyalists", βάσει του Apriori αλγορίθμου

Οι 50 κορυφαίοι κανόνες που παρήχθηκαν για το τμήμα των Potential Loyalists (Εικόνα 55), παρουσιάζουν τιμές lift που κυμαίνονται από 10.35 έως 29.37. Ο ισχυρότερος κανόνας, με lift = 29.37, αναδεικνύει μία ισχυρή σχέση μεταξύ των Tuborg Pink Grapefruit 330ml και Tuborg Σόδα Ανανάς, Lime & Δυόσμος 330ml. Ο κανόνας αυτός είναι αμφίδρομος, όπως αποδεικνύεται από τις υψηλές confidence τιμές των 29.21% και 38.06% αντίστοιχα. Τα αποτελέσματα αυτά υποδηλώνουν μια σταθερή προτίμηση των πελατών για τη συνδυαστική αγορά αυτών των συμπληρωματικών γεύσεων αναψυκτικών. Αυτή η ισχυρή συσχέτιση μπορεί να αξιοποιηθεί μέσω cross-promotions, όπως πακέτα προσφορών (bundles) ή “buy one, get one” πρωθητικές ενέργειες, με στόχο την αύξηση των πωλήσεων.

Ένας ακόμη σημαντικός κανόνας, με lift = 24.91, αφορά τη σχέση μεταξύ των Swan Φιλτράκια Extra Slim 120τεμ, Old Holborn Καπνός Blonde 30g και Rizla Χαρτάκια Γαλάζια. Οι confidence τιμές (14.86% και 51.61%) καταδεικνύουν ότι οι Potential Loyalists συχνά αγοράζουν προϊόντα καπνού και σχετικά αξεσουάρ μαζί. Αυτό το αγοραστικό μοτίβο αντικατοπτρίζει συνήθειες που παρατηρούνται και σε άλλα τμήματα, γεγονός που υπογραμμίζει τη δυνατότητα δημιουργίας συνδυαστικών πακέτων που θα ενισχύσουν την ευκολία των πελατών και θα βελτιώσουν τη συνολική εμπειρία αγορών.

Στην κατηγορία των φρέσκων προϊόντων, ένας κανόνας με lift = 17.99 αποκάλυψε ισχυρή συσχέτιση μεταξύ των Δάμα Ντοματίνια Βελανίδια Συσκευασμένα Ελάχιστο Βάρος 250g, Φρεσκούλης Σαλάτα Ιταλική 200g και Δάμα Αγγουράκια Κνωσσού Συσκευασμένα Ελάχιστο Βάρος 600g. Οι confidence τιμές (34.63% και 53.33%) υποδεικνύουν ότι οι πελάτες που αγοράζουν αυτά τα προϊόντα τείνουν να προτιμούν συνδυασμούς λαχανικών και έτοιμες σαλάτες, πιθανότατα λόγω προτίμησης σε υγιεινές διατροφικές επιλογές. Αυτό το εύρημα μπορεί να αξιοποιηθεί στη διαμόρφωση στρατηγικών τοποθέτησης προϊόντων (product placement) και στη δημιουργία στοχευμένων πρωθήσεων για συνδυαστικά πακέτα φρέσκων προϊόντων, ενισχύοντας την αξία για τους καταναλωτές και την εταιρεία.

Επιπλέον, εντοπίστηκαν σημαντικές συσχετίσεις στην κατηγορία των επιδορπίων. Για παράδειγμα, ο κανόνας B&J Παγωτό Chocolatey Love A-fair 465ml → B&J Sundae Cookie Vermont-ster 427ml, με lift = 11.82, υποδεικνύει ότι οι πελάτες που αγοράζουν μία γεύση παγωτού είναι ιδιαίτερα πιθανό να αγοράσουν και μία ακόμη παραλλαγή της ίδιας μάρκας. Παρόμοιοι κανόνες σε αυτήν την κατηγορία υποδηλώνουν ευκαιρίες για την προώθηση πολυσυσκευασιών πολλαπλών γεύσεων (multi-flavor packs), ενθαρρύνοντας μεγαλύτερες αγορές και ενισχύοντας τη συνολική καταναλωτική εμπειρία.

Top 50 Rules for Churners					
count	support	confidence	lift	rule	note
109	0.0011418783899514702	1.0	42.204819277108435	Φρεσκόλιο Ζακέτα Πταλακί 200g → Endless Rich & Soft Χαρτί Υγείας 3 Φύλλων 10τεμ → Endless Economy Ρολό Κουζίνας 2 Φύλλων 800g	
109	0.0011418783899514702	0.04811000000000001	0.9378	Φρεσκόλιο Ζακέτα Πταλακί 200g → Endless Rich & Soft Χαρτί Υγείας 3 Φύλλων 10τεμ	
218	0.022837567799029405	0.18181818181818182	37.465206421711233	Scandal Παγωτό Κομπότο Chocolate Orgy 140ml → Scandal Παγωτό Κομπότο Caramel Secrets 140ml	Alexandros Papandreou Μπιφτέκια Σάρδινες Με Πατάτες Baby & Μπιέκον 350g → Scandal Παγωτό Κομπότο Chocolate Orgy 140ml, Scandal Παγωτό Κομπότο Caramel Secrets 140ml
136	0.0011418783899514702	0.47050215254171762	10.455206421711233	Nτομάτα Βελανίδια Συκούσια Ελάχιστο Βάρος 600g → Δάμα Αγγειούκια Κουτσούρα Ελάχιστο Βάρος 250g	Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 250g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 250g
136	0.0011418783899514702	0.34041538401538463	30.620023797063378	Ντομάτα Βελανίδια Συκούσια Ελάχιστο Βάρος 600g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 250g	Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 250g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 250g
327	0.022837567799029405	0.15584415584415587	28.732740942627267	Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 250g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 250g	Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 250g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 250g
109	0.0011418783899514702	0.6812000000000001	28.689521809513811	Χρήση Αιγαίνων Καρπούζια & Ατσάλι	Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 250g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 250g
109	0.0011418783899514702	0.05194805194805195	25.996289424860595	Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 600g → Χρήση Αιγαίνων Καρπούζια & Ατσάλι	Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 600g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 250g
136	0.0011418783899514702	0.06493506493506493	25.274123121212927	Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 600g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 250g	Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 600g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 250g
327	0.00346351398944107	0.75	25.021428571428571	Μπανιέρες Ελάχιστο Βάρος 100g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 250g	Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 250g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 250g
163	0.0011418783899514702	0.14285714285714285	23.021428571428571	Tuborg Ζέρν Αναψυκτικό Λεμονί & Λαζαρός 330ml → Tuborg Pink Grapefruit 330ml	Tuborg Ζέρν Αναψυκτικό Λεμονί & Λαζαρός 330ml → Tuborg Pink Grapefruit 330ml
109	0.0011418783899514702	0.03418801418801419	23.951218751218752	Endless Rich & Soft Χαρτί Υγείας 3 Φύλλων 10τεμ → Endless Rich & Soft Χαρτί Υγείας 3 Φύλλων 800g	Endless Rich & Soft Χαρτί Υγείας 3 Φύλλων 10τεμ → Endless Rich & Soft Χαρτί Υγείας 3 Φύλλων 800g
109	0.0011418783899514702	0.8	23.951218751218752	Φρεσκόλιο Ζακέτα Πταλακί 200g → Endless Rich & Soft Χαρτί Υγείας 3 Φύλλων 800g	Endless Rich & Soft Χαρτί Υγείας 3 Φύλλων 10τεμ → Endless Rich & Soft Χαρτί Υγείας 3 Φύλλων 800g
218	0.022837567799029405	0.13896103619610398	22.7408751246751247	Alexandros Papandreou Μπιφτέκια Σάρδινες Με Πατάτες Baby & Μπιέκον 350g → Scandal Παγωτό Κομπότο Chocolate Orgy 140ml	Scandal Παγωτό Κομπότο Chocolate Orgy 140ml → Scandal Παγωτό Κομπότο Caramel Secrets 140ml
109	0.0011418783899514702	0.3809523809523809	22.2412698326983269	Δάμα Ηπειρώτικη Βελανίδια Συκούσια Ελάχιστο Βάρος 250g → Φρεσκόλιο Ζακέτα Πταλακί 200g	Δάμα Ηπειρώτικη Βελανίδια Συκούσια Ελάχιστο Βάρος 250g → Δάμα Ηπειρώτικη Βελανίδια Συκούσια Ελάχιστο Βάρος 600g
136	0.0011418783899514702	0.03836363636363865	21.385830385830388	Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 1.5kg → Ντομάτα Ελάχιστο Βάρος 600g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 1.5kg	Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 1.5kg → Ντομάτα Ελάχιστο Βάρος 600g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 1.5kg
136	0.0011418783899514702	0.03836363636363865	21.385830385830388	Ντομάτα Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 600g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 600g	Ντομάτα Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 600g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 600g
109	0.0011418783899514702	0.25	19.903409090909095	Mπανιέρες Ελάχιστο Βάρος 100g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 600g	Mπανιέρες Ελάχιστο Βάρος 100g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 600g
109	0.0011418783899514702	0.26710000000000003	19.64492175526264962	Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 600g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 250g	Alexandros Papandreou Μπιφτέκια Σάρδινες Με Πατάτες Baby & Μπιέκον 350g → Scandal Παγωτό Κομπότο Chocolate Orgy 140ml
517	0.005423922552269405	0.41181818181818181	19.64492175526264962	Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 600g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 250g	Alexandros Papandreou Μπιφτέκια Σάρδινες Με Πατάτες Baby & Μπιέκον 350g → Scandal Παγωτό Κομπότο Chocolate Orgy 140ml
109	0.0011418783899514702	0.17150000000000003	19.581200000000003	Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 600g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 250g	Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 600g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 250g
109	0.0011418783899514702	0.37410000000000003	19.3746259794026	Φρεσκόλιο Ζακέτα Πταλακί 200g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 600g	Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 600g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 250g
109	0.0011418783899514702	0.01184100000000001	19.3746259794026	Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 600g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 250g	Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 600g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 250g
109	0.0011418783899514702	0.05194805194805195	16.5430933270566	Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 600g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 600g	Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 600g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 600g
109	0.0011418783899514702	0.03836363636363865	16.5430933270566	Mπανιέρες Ελάχιστο Βάρος 800g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 600g	Mπανιέρες Ελάχιστο Βάρος 800g → Δάμα Ηπειρώτικη Ελάχιστη Ισχυρότητα Συκούσια Ελάχιστο Βάρος 600g
218	0.022837567799029405	0.0816326530812425	15.050483351232523	Alexandros Papandreou Μπιφτέκια Σάρδινες Με Πατάτες Baby & Μπιέκον 350g → Scandal Παγωτό Κομπότο Caramel Secrets 140ml	Scandal Παγωτό Κομπότο Caramel Secrets 140ml → Alexandros Papandreou Μπιφτέκια Σάρδινες Με Πατάτες Baby & Μπιέκον 350g, Scandal Παγωτό Κομπότο Chocolate Orgy 140ml
327	0.00346351398944107	0.27	15.472155745157923	Ruffles Πατάτες Άρκτης 100g → Ruffles Πατάτες Άρκτης 100g	Ruffles Πατάτες Άρκτης 100g → Ruffles Πατάτες Άρκτης 100g
436	0.0045507513559865801	0.2191780212170086	14.2161633668484302	B&G Πατάτες Κορνέτ Βρούνι 100ml → B&G Πατάτες Κορνέτ Βρούνι 100ml	B&G Πατάτες Κορνέτ Βρούνι 100ml → B&G Πατάτες Κορνέτ Βρούνι 100ml
872	0.009135027119611762	0.41558441558441556	13.8646481793252372	Δάμα Ηπειρώτικη Κυνούδια Λούσια Ελάχιστο Βάρος 600g → Δάμα Ηπειρώτικη Κυνούδια Λούσια Ελάχιστο Βάρος 250g	Δάμα Ηπειρώτικη Κυνούδια Λούσια Ελάχιστο Βάρος 600g → Δάμα Ηπειρώτικη Κυνούδια Λούσια Ελάχιστο Βάρος 250g
872	0.009135027119611762	0.30476150476150475	13.8646481793252372	Δάμα Ηπειρώτικη Κυνούδια Λούσια Ελάχιστο Βάρος 600g → Δάμα Ηπειρώτικη Κυνούδια Λούσια Ελάχιστο Βάρος 600g	Δάμα Ηπειρώτικη Κυνούδια Λούσια Ελάχιστο Βάρος 600g → Δάμα Ηπειρώτικη Κυνούδια Λούσια Ελάχιστο Βάρος 600g
354	0.03371110473119611762	0.16881116881116881	13.4412632817237374	Δάμα Ηπειρώτικη Κυνούδια Λούσια Ελάχιστο Βάρος 600g → Δάμα Ηπειρώτικη Κυνούδια Λούσια Ελάχιστο Βάρος 600g	Δάμα Ηπειρώτικη Κυνούδια Λούσια Ελάχιστο Βάρος 600g → Δάμα Ηπειρώτικη Κυνούδια Λούσια Ελάχιστο Βάρος 600g
272	0.0285469597487277	0.272715574488288	13.309275015717526	B&G Νερού Έλαφη Λούσια Ελάχιστο Βάρος 600g → B&G Νερού Έλαφη Λούσια Ελάχιστο Βάρος 600g	B&G Νερού Έλαφη Λούσια Ελάχιστο Βάρος 600g → B&G Νερού Έλαφη Λούσια Ελάχιστο Βάρος 600g
190	0.001988271024150727	0.15227391304347872	13.00159666808595	Days Mini Κρούσσος Καρδιά 103g → Days Mini Κρούσσος Μιλεύειλε 103g → Days Mini Κρούσσος Καρδιά 103g	Days Mini Κρούσσος Καρδιά 103g → Days Mini Κρούσσος Μιλεύειλε 103g → Days Mini Κρούσσος Καρδιά 103g
190	0.001988271024150727	0.17073103103707318	13.00159666808595	Days Mini Κρούσσος Καρδιά 103g → Days Mini Κρούσσος Μιλεύειλε 103g → Days Mini Κρούσσος Καρδιά 103g	Days Mini Κρούσσος Καρδιά 103g → Days Mini Κρούσσος Μιλεύειλε 103g → Days Mini Κρούσσος Καρδιά 103g

Εικόνα 61. Κορυφαίοι 50 κανόνες συσχέτισης μεταξύ αγορασμένων προϊόντων για την ομάδα πελατών "Churners", βάσει του Apriori αλγορίθμου

Στην Εικόνα 56 αποτυπώνονται οι 50 κορυφαίοι κανόνες για την ομάδα των Churners, με τιμές lift που κυμαίνονται από 13.00 έως 42.20. Ο ισχυρότερος κανόνας που εντοπίστηκε στην ανάλυση, με lift = 42.20, αφορά τη σχέση μεταξύ των προϊόντων Φρεσκούλης Σαλάτα Ιταλική 200g και Endless Rich & Soft Χαρτί Υγείας 3 Φύλλων 10τεμ, τα οποία συνδέονται με το Endless Economy Ρολό Κουζίνας 2 Φύλλων 800g ως consequent. Ο κανόνας αυτός παρουσίασε confidence 100%, γεγονός που σημαίνει ότι κάθε συναλλαγή που περιέχει τα antecedent προϊόντα προϊόντα από πάντα θα επιλέγουν συγκεκριμένους συνδυασμούς lift = 42.20. Οι ισχυρή συσχέτιση up-to-date περιλαμβάνει την ανάλυση την ομάδα πελατών Churners, με τιμές lift = 42.20, αφορά τη σχέση μεταξύ των προϊόντων Φρεσκούλης Σαλάτα Ιταλική 200g και Endless Rich & Soft Χαρτί Υγείας 3 Φύλλων 10τεμ, τα οποία συνδέονται με το Endless Economy Ρολό Κουζίνας 2 Φύλλων 800g ως consequent. Ο κανόνας αυτός παρουσίασε confidence 100%, γεγονός που σημαίνει ότι κάθε συναλλαγή που περιέχει τα antecedent προϊόντα προϊόντα από πάντα θα επιλέγουν συγκεκριμένους συνδυασμούς lift = 42.20.

Ένας ακόμη αξιοσημείωτο εύρημα αφορούσε τις κατεψυγμένες λιχουδιές και τα έτοιμα γεύματα. Ένα

Επιπλέον, η ανάλυση ανέδειξε ισχυρές συσχετίσεις στην κατηγορία των ποτών. Για παράδειγμα, ο κανόνας που συνδέει το Tuborg Pink Grapefruit 330ml με το Tuborg Σόδα Ανανάς, Lime & Δυόσμος 330ml εμφάνισε lift = 25.02. Η ισχυρή αυτή σχέση υποδηλώνει ότι οι πελάτες προτιμούν να αγοράζουν συμπληρωματικές γεύσεις αναψυκτικών μαζί, γεγονός που δημιουργεί ευκαιρίες για cross-selling και προωθητικές στρατηγικές πακέτων.

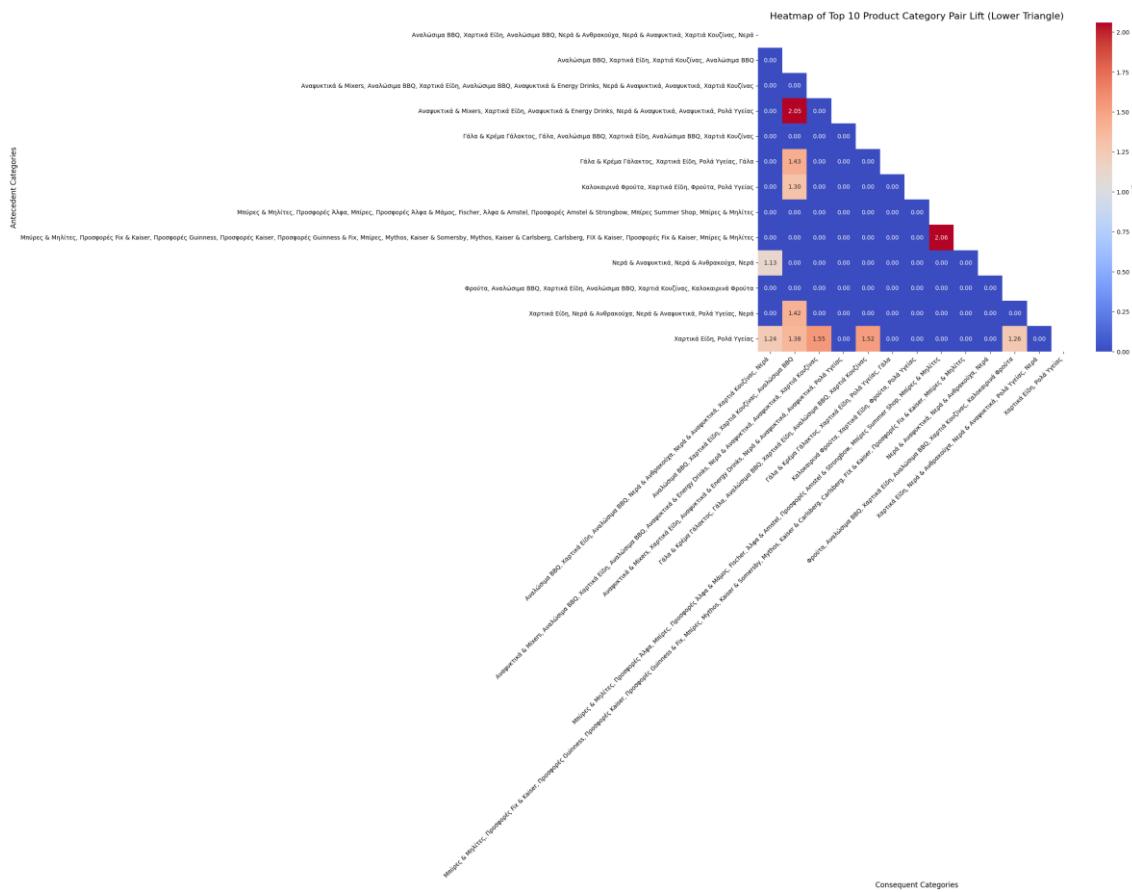
9.2.1 Σύγκριση αποτελεσμάτων MBA μεταξύ τμημάτων πελατών

Η Ανάλυση Market Basket Analysis που πραγματοποιήθηκε ξεχωριστά για τα τμήματα Loyalists, Potential Loyalists και Churners ανέδειξε διακριτές αγοραστικές συμπεριφορές μεταξύ των τριών πελατειακών ομάδων.

Το τμήμα των Loyalists παρουσίασε ισχυρά επαναλαμβανόμενα αγοραστικά μοτίβα, ιδίως στην κατηγορία των καπνικών προϊόντων, όπου χαρτάκια, φίλτρα και καπνός εμφανίζονταν συχνά μαζί στις συναλλαγές. Οι αγοραστικές συνήθειες αυτής της ομάδας χαρακτηρίζονται από ευρεία γκάμα προϊόντων, με τις τιμές lift να κυμαίνονται σε μέτρια επίπεδα, ενώ ο ισχυρότερος κανόνας εμφάνισε lift = 19.98. Αυτά τα μοτίβα καταδεικνύουν πολυδιάστατη εμπλοκή με διάφορες κατηγορίες προϊόντων, γεγονός που υποδηλώνει ευκαιρίες για στρατηγικές cross-selling και προγράμματα επιβράβευσης, ειδικά σε συγκεκριμένες κατηγορίες προϊόντων με υψηλή συσχέτιση.

Στον αντίποδα, το τμήμα των Potential Loyalists εμφάνισε μεικτά αγοραστικά πρότυπα, συνδυάζοντας προϊόντα που σχετίζονται τόσο με υγιεινές διατροφικές επιλογές όσο και με προϊόντα «πολυτελείας». Ισχυρές συσχετίσεις αναδειχθηκαν στις κατηγορίες φρέσκων προϊόντων, επιδορπίων και ποτών, με τον κορυφαίο κανόνα να καταγράφει lift = 29.37. Αυτή η διπτή αγοραστική συμπεριφορά, που περιλαμβάνει τόσο πρακτικά όσο και λιγότερο καθημερινά προϊόντα, υπογραμμίζει τη δυνατότητα μετατροπής των Potential Loyalists σε Loyalists, μέσω εξαπομικευμένων προσφορών και στρατηγικών που συνδυάζουν διαφορετικές κατηγορίες προϊόντων.

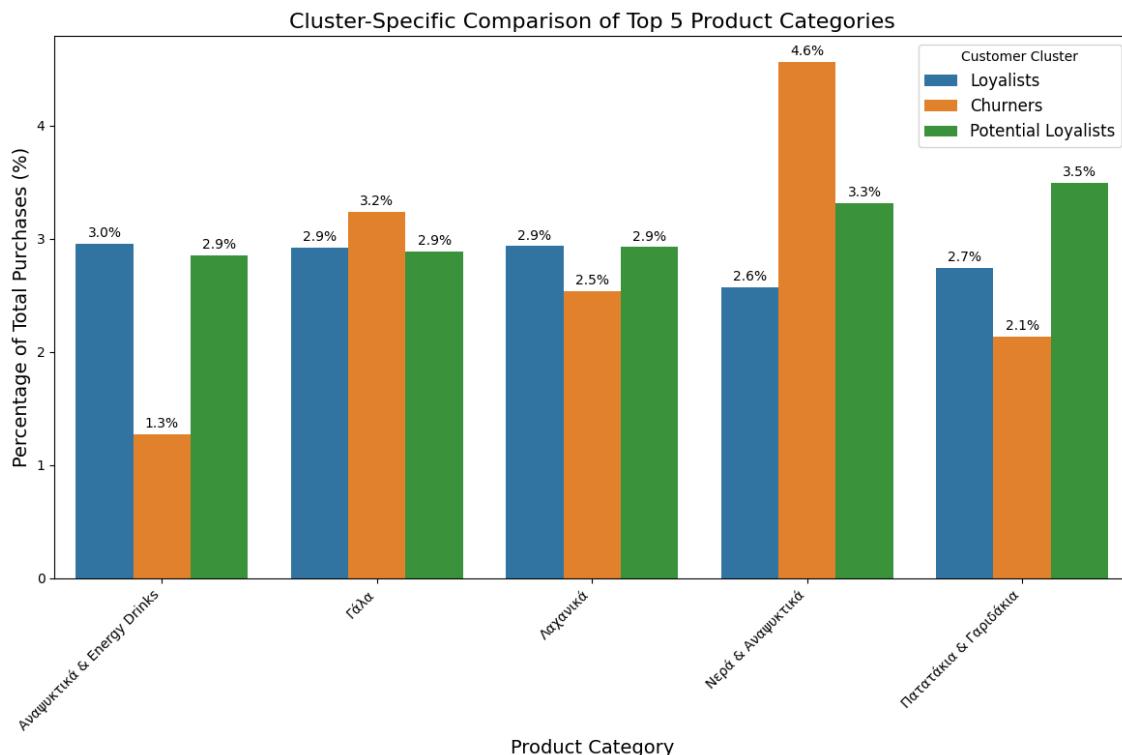
Το τμήμα των Churners παρουσίασε τις υψηλότερες τιμές lift συνολικά, με τον ισχυρότερο κανόνα να καταγράφει lift = 42.20. Η ομάδα αυτή εμφάνισε συγκεντρωτικά αγοραστικά μοτίβα, κυρίως στις κατηγορίες φρέσκων προϊόντων, κατεψυγμένων γευμάτων και ειδών οικιακής χρήσης. Η ακαμψία στις αγοραστικές τους επιλογές, όπως υποδηλώνουν τα υψηλά επίπεδα confidence, υποδεικνύει ότι οι στοχευμένες στρατηγικές επανενεργοποίησης μέσω συνδυαστικών προσφορών (bundled offers) ή εκπτώσεων σε προϊόντα υψηλής συχνότητας αγορών θα μπορούσαν να είναι αποτελεσματικές για την αύξηση της πιθανότητας επαναλαμβανόμενων αγορών. Επιπλέον, οι ισχυρές συσχετίσεις σε βασικά καταναλωτικά αγαθά υποδηλώνουν ότι οι πελάτες αυτοί δίνουν προτεραιότητα στην πρακτικότητα και την αξία, γεγονός που προσφέρει ευκαιρίες για στοχευμένες καμπάνιες που θα δώσουν έμφαση στην άνεση και την οικονομική ωφέλεια.



Εικόνα 62. Θερμικός χάρτης συσχέτισης κορυφαίων 10 κατηγοριών προϊόντων βάσει του Apriori αλγορίθμου

Η Θερμική απεικόνιση (Εικόνα 57) του lift μεταξύ ζευγών κατηγοριών προϊόντων παρέχει μια αναλυτική οπτική των ισχυρότερων συσχετίσεων μεταξύ κατηγοριών, αναδεικνύοντας αγοραστικά μοτίβα των πελατών. Εστιάζοντας στις δέκα κορυφαίες κατηγορίες προϊόντων, προκύπτουν σημαντικές σχέσεις που μπορούν να αξιοποιηθούν για στρατηγικές μάρκετινγκ και πρωθιμικές ενέργειες. Για παράδειγμα, η συσχέτιση μεταξύ των κατηγοριών "Χαρτικά Είδη" (Paper Goods) και "Ρολά Υγείας" (Toilet Rolls) παρουσιάζει υψηλή τιμή lift = 2.05, γεγονός που υποδηλώνει ισχυρή σχέση μεταξύ αυτών των κατηγοριών. Αυτό το εύρημα δείχνει ότι οι πελάτες συχνά αγοράζουν αυτά τα προϊόντα μαζί, πιθανότατα λόγω της συμπληρωματικής τους χρήσης σε οικιακές ανάγκες. Αντίστοιχα, η σχέση μεταξύ των "Αναψυκτικά BBQ" (BBQ Drinks) και "Καλοκαιρινά Φρούτα" (Summer Fruits) παρουσιάζει lift = 1.30, γεγονός που υποδεικνύει εποχιακά πρότυπα αγορών, τα οποία είναι ιδιαίτερα σημαντικά για πρωθιμικές καμπάνιες κατά τους καλοκαιρινούς μήνες.

Τα αποτελέσματα αυτά προσφέρουν εφαρμόσιμες επιχειρηματικές ευκαιρίες, όπως συνδυαστικές προσφορές (bundling) για συμπληρωματικά προϊόντα, ώστε να ενισχυθεί το cross-selling ή στοχευμένες προωθήσεις εποχικών προϊόντων, με σκοπό την αύξηση της μέσης αξίας παραγγελίας κατά περιόδους αυξημένης ζήτησης.



Εικόνα 63. Ραβδόγραμμα σύγκρισης κορυφαίων 5 κατηγοριών σε πωλήσεις προϊόντων ανά ομάδα πελατών

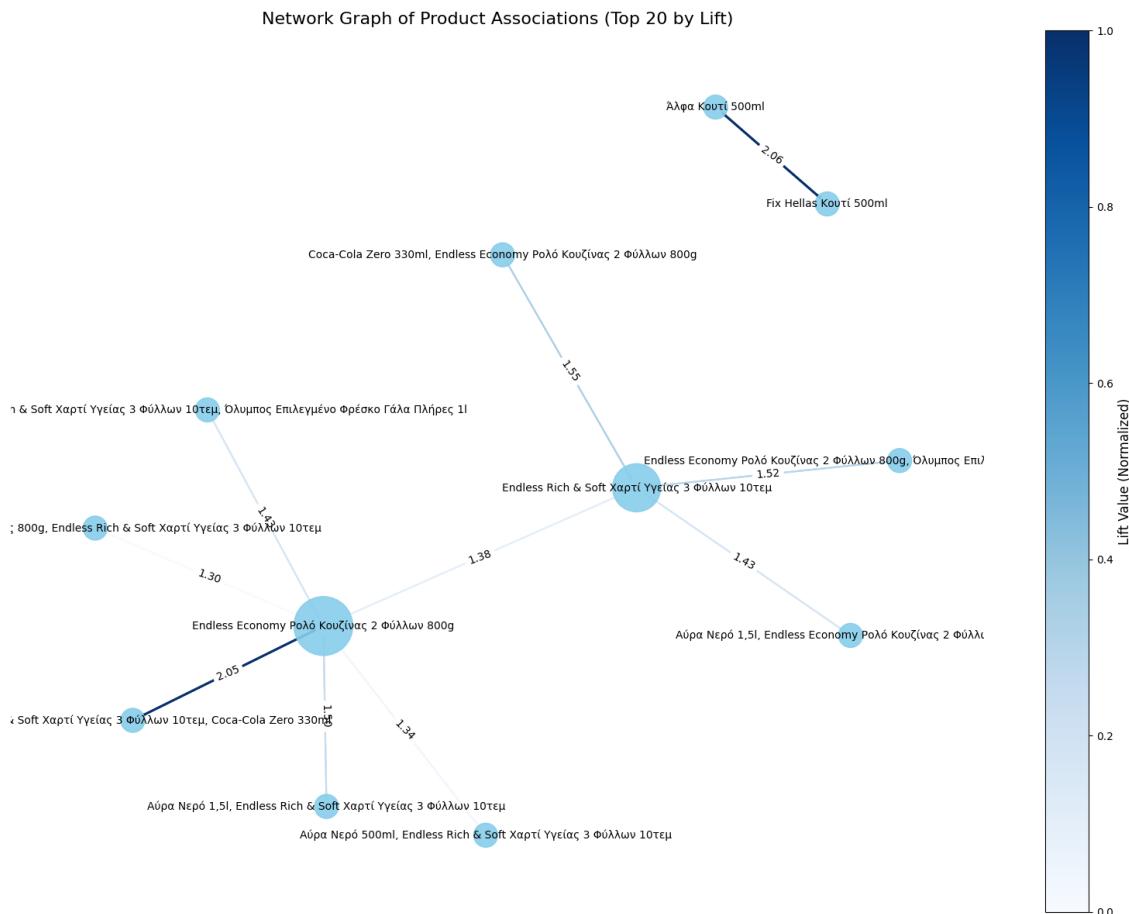
Το ραβδόγραμμα της Εικόνας 58 παρουσιάζει μία σύγκριση μεταξύ των τριών πελατειακών τμημάτων, εστιάζοντας στις πέντε κορυφαίες κατηγορίες προϊόντων βάσει του ποσοστιαίου μεριδίου τους στις συνολικές αγορές κάθε τμήματος.

Μεταξύ των Loyalists, η κατανομή των συνολικών αγορών εμφανίζεται σχετικά ισορροπημένη στις κορυφαίες κατηγορίες προϊόντων. Για παράδειγμα, οι κατηγορίες "Νερά & Αναψυκτικά" και "Πατατάκια & Γαριδάκια" παρουσιάζουν σταθερή δραστηριότητα, με κάθε μία να συνεισφέρει περίπου 2.6–2.7% των συνολικών αγορών αυτής της ομάδας. Αυτή η ισορροπημένη καταναλωτική συμπεριφορά υποδηλώνει ότι οι Loyalists είναι σταθεροί αγοραστές, οι οποίοι πραγματοποιούν συχνές αγορές διαφορετικών τύπων προϊόντων, γεγονός που δημιουργεί ευκαιρίες για εξατομικευμένα προγράμματα επιβράβευσης και cross-selling στρατηγικές.

Αντίθετα, οι Churners εμφανίζουν σαφή εστίαση σε συγκεκριμένες κατηγορίες προϊόντων. Η κατηγορία "Νερά & Αναψυκτικά" αντιπροσωπεύει 4.6% των συνολικών αγορών τους, ποσοστό που αποτελεί τη μεγαλύτερη αναλογία που παρατηρείται σε οποιοδήποτε τμήμα. Αυτή η έντονη προτίμηση υπογραμμίζει μία σημαντική ευκαιρία για στοχευμένες ενέργειες επανενεργοποίησης, όπως εκπτώσεις ή προωθητικές ενέργειες σε προϊόντα της κατηγορίας των ποτών. Αντιθέτως, άλλες κατηγορίες προϊόντων καταλαμβάνουν μικρότερα ποσοστά, με την κατηγορία "Αναψυκτικά & Energy Drinks" να έχει τη χαμηλότερη συνεισφορά στο 1.3%.

Οι Potential Loyalists παρουσιάζουν αγοραστικά μοτίβα που γεφυρώνουν τη συμπεριφορά των άλλων δύο τμημάτων. Το ποσοστό των συνολικών αγορών τους για την κατηγορία "Νερά & Αναψυκτικά" ανέρχεται σε 3.3%, τοποθετώντας τους μεταξύ των Loyalists και Churners. Επιπλέον, αυτό το τμήμα εμφανίζει υψηλότερη αναλογία αγορών στην κατηγορία "Πατατάκια & Γαριδάκια" (3.5%), ξεπερνώντας τόσο τους Loyalists όσο και τους Churners. Αυτή η αυξημένη προτίμηση στα σνακ υποδηλώνει πιθανές ευκαιρίες για ενίσχυση της εμπιστοσύνης και της

αφοσίωσης αυτής της ομάδας, μέσω προωθητικών ενεργειών που σχετίζονται με τα εν λόγω προϊόντα.



Εικόνα 64. Γράφημα δικτύου με τις συσχετίσεις των κορυφών 20 κατηγοριών προϊόντων

Το network γράφημα (Εικόνα 59) απεικονίζει τις 20 ισχυρότερες συσχετίσεις προϊόντων, καταταγμένες βάσει των τιμών lift, προσφέροντας πολύτιμες πληροφορίες για τη δύναμη των σχέσεων μεταξύ διαφορετικών προϊόντων. Κάθε κόμβος στο γράφημα αντιπροσωπεύει ένα συγκεκριμένο προϊόν, ενώ οι συνδέσεις (edges) μεταξύ των κόμβων απεικονίζουν τους κανόνες συσχέτισης. Το πάχος και η ένταση του χρώματος των συνδέσεων αντικατοπτρίζουν τις τιμές lift, όπου υψηλότερες τιμές lift υποδηλώνουν ισχυρότερες τάσεις συν-αγοράς.

Μία αξιοσημείωτη παρατήρηση αφορά το απομονωμένο ζεύγος κόμβων που αντιπροσωπεύει τις μπύρες Άλφα Koutí 500ml και Fix Hellas Koutí 500ml. Το συγκεκριμένο ζεύγος καταγράφει την υψηλότερη τιμή lift στο γράφημα, με 2.06, υποδεικνύοντας μια ισχυρή συσχέτιση μεταξύ αυτών των δύο εμπορικών σημάτων μπύρας. Ωστόσο, η φυσική απομόνωσή τους από τους υπόλοιπους κόμβους υποδηλώνει ότι η συμπεριφορά συν-αγοράς τους διαφέρει από τις κύριες ομάδες προϊόντων, πιθανώς αντανακλώντας εξειδικευμένες καταναλωτικές προτιμήσεις.

Η δεύτερη υψηλότερη τιμή lift, 2.05, παρατηρείται μεταξύ του Endless Economy Ρολό Κουζίνας 2 Φύλλων 800g και ενός συνδυασμένου κόμβου, που περιλαμβάνει τα Endless Rich & Soft Χαρτί Υγείας 3 Φύλλων 10τεμ και Coca Cola Zero 330ml. Αυτά τα προϊόντα αποτελούν μέρος της κεντρικής δομής του γράφου, με το Endless Economy Ρολό Κουζίνας 2 Φύλλων 800g να ξεχωρίζει

ως ο μεγαλύτερος κόμβος. Αυτή η προεξέχουσα θέση υποδηλώνει τον κεντρικό ρόλο του προϊόντος στα μοτίβα συν-αγοράς, ειδικά στην κατηγορία των οικιακών ειδών πρώτης ανάγκης.

Ο δεύτερος μεγαλύτερος κόμβος, το Endless Rich & Soft Χαρτί Υγείας 3 Φύλλων 10τεμ, διαδραματίζει επίσης κρίσιμο ρόλο στο δίκτυο. Εμφανίζει lift = 1.55 με έναν συνδυασμένο κόμβο που περιλαμβάνει τα Coca Cola Zero 330ml και Endless Economy Ρολό Κουζίνας 2 Φύλλων 800g. Αυτή η σχέση υποδηλώνει μέτρια ισχυρή συσχέτιση μεταξύ των χαρτικών προϊόντων και των ποτών, κάτι που θα μπορούσε να διερευνηθεί περαιτέρω για πιθανές ευκαιρίες προωθητικών ενεργειών και συνδυαστικών πωλήσεων.

Οι υπόλοιπες συνδέσεις στο γράφημα παρουσιάζουν τιμές lift μεταξύ 1.52 και 1.30, αντικατοπτρίζοντας μέτρια μοτίβα συν-αγοράς. Οι συσχετίσεις αυτές αφορούν κυρίως συνδυασμούς βασικών καταναλωτικών αγαθών, ποτών και ειδών ευκολίας. Για παράδειγμα, οι κόμβοι Άυρα Νερό 1.5L και Endless Rich & Soft Χαρτί Υγείας 3 Φύλλων 10τεμ συνδέονται μέσω σταθερών, αν και λιγότερο έντονων, αγοραστικών σχέσεων.

Συνολικά, το γράφημα αναδεικνύει δύο βασικά ευρήματα:

- Η ισχυρή αλλά απομονωμένη συσχέτιση μεταξύ των εμπορικών σημάτων μπύρας (Άλφα Κουτί 500ml και Fix Hellas Κουτί 500ml) υποδηλώνει πιθανές ευκαιρίες μάρκετινγκ που στοχεύουν συγκεκριμένα καταναλωτικά τμήματα.
- Ο κεντρικός ρόλος των χαρτικών ειδών πρώτης ανάγκης, ειδικά του Endless Economy Ρολό Κουζίνας 2 Φύλλων 800g, υποδεικνύει ευκαιρίες για cross-promotional στρατηγικές, συνδυάζοντας ποτά και άλλα βασικά καταναλωτικά προϊόντα.

9.3 Εφαρμογή Frequent Pattern (FP) Growth αλγορίθμου

Ο αλγόριθμος FP-Growth (Frequent Pattern Growth) αποτελεί μία αποδοτική και κλιμακούμενη μέθοδο εξόρυξης συχνών προτύπων (frequent patterns) από μεγάλα σύνολα δεδομένων. Εισήχθη από τον Han et al. και διαφοροποιείται από τον αλγόριθμο Apriori, καθώς δεν απαιτεί δημιουργία υποψήφιων συνόλων (candidate generation), γεγονός που βελτιώνει την απόδοσή του. Ο αλγόριθμος βασίζεται στη χρήση μίας συμπιεσμένης δομής δεδομένων, του FP-Tree (Frequent Pattern Tree), η οποία αποθηκεύει τη συχνότητα εμφάνισης στοιχείων σε ένα δέντρο προθεμάτων (prefix tree). Μέσω αυτής της προσέγγισης, μειώνει τον αριθμό των επαναλαμβανόμενων σαρώσεων της βάσης δεδομένων, καθιστώντας την εξόρυξη συχνών συνδυασμών πιο αποδοτική. Τα βασικά στάδια λειτουργίας του είναι:

Συμπίεση Βάσης Δεδομένων: Δημιουργείται το FP-tree, το οποίο αποθηκεύει τις πληροφορίες των συχνών στοιχείων με συμπιεσμένη μορφή.

Διαίρεση Βάσης Δεδομένων: Διαχωρίζει τα δεδομένα σε υποσύνολα (conditional databases), που σχετίζονται με κάθε συχνό πρότυπο.

Ανεύρεση Συχνών Προτύπων: Η εξόρυξη πραγματοποιείται με αναδρομική (recursive) αναζήτηση μικρών μοτίβων, τα οποία ενώνονται για να σχηματίσουν μεγαλύτερα σύνολα αντικειμένων.

Αυτή η διαίρει και βασίλευε (divide-and-conquer) προσέγγιση επιτρέπει την ανακάλυψη συχνών συνόλων αντικειμένων πιο αποδοτικά από τον Apriori, αποφεύγοντας την επανειλημμένη δημιουργία υποψήφιων συνδυασμών (Javatpoint.com, n.d.).

Τα πλεονεκτήματα και οι περιορισμοί του FP-Growth καταγράφονται συνοπτικά ως εξής:

- Υψηλή Απόδοση & Κλιμάκωση:** Είναι πιο αποδοτικός από τον Apriori, καθώς δεν απαιτεί επαναλαμβανόμενη δημιουργία και έλεγχο υποψήφιων συνδυασμών, κάτι που μειώνει δραστικά τον υπολογιστικό φόρτο.
- Συμπιεσμένη Αναπαράσταση Δεδομένων:** Ο FP-tree καταλαμβάνει λιγότερη μνήμη και αποθηκεύει μόνο κρίσιμες πληροφορίες για τις συχνές σχέσεις των στοιχείων.
- Λειτουργεί Αποτελεσματικά σε Μεγάλα Δεδομένα:** Η προσέγγιση διαίρει και βασίλευε επιτρέπει την ανάλυση μεγάλων συνόλων δεδομένων πιο αποδοτικά.

✗ Μεγάλη Απαίτηση Μνήμης: Η κατασκευή του FP-tree απαιτεί αποθήκευση ολόκληρου του δέντρου στη μνήμη RAM, κάτι που μπορεί να είναι αδύνατο σε πολύ μεγάλα δεδομένα.

✗ Δυσκολία στην Παράλληλη Επεξεργασία: Σε αντίθεση με τον Apriori, ο FP-Growth δεν υποστηρίζει εύκολα παραλληλισμό, γεγονός που δυσχεραίνει την επεξεργασία κατανεμημένων δεδομένων.

✗ Ευαισθησία στη Δομή των Δεδομένων: Η απόδοση εξαρτάται από τη δομή της βάσης δεδομένων, καθώς μια πολύ ανισομερής κατανομή δεδομένων μπορεί να δημιουργήσει ένα μη ισορροπημένο FP-tree.

Παρά τους περιορισμούς του, ο FP-Growth αποτελεί μία από τις πιο αποδοτικές μεθόδους εξόρυξης συχνών προτύπων και χρησιμοποιείται ευρέως σε εφαρμογές όπως market basket analysis, ανίχνευση απάτης, και ανάλυση κοινωνικών δικτύων.

Εφαρμόζοντας τον αλγόριθμο FP-Growth, συνοψίζονται τα ευρήματα που προέκυψαν και πραγματοποιείται σύγκριση με τα αποτελέσματα της προηγούμενης ανάλυσης μέσω Apriori. Ο αλγόριθμος FP-Growth, εφαρμόστηκε στο πλήρες σύνολο συναλλαγών, χρησιμοποιώντας ελάχιστο όριο support 0.001, και παρήγαγε συνολικά 370 association rules. Μεταξύ αυτών, 14 κανόνες εμφάνισαν τιμή lift μεγαλύτερη από 100, γεγονός που υποδηλώνει ισχυρές και αξιοσημείωτες συσχετίσεις. Αυτοί οι κανόνες υψηλού lift αναδεικνύουν ζεύγη ή ομάδες προϊόντων που αγοράζονται μαζί με συχνότητα σημαντικά υψηλότερη από την τυχαία πιθανότητα, υποδεικνύοντας κρίσιμες εμπορικές ευκαιρίες για cross-selling και στοχευμένες πρωθητικές ενέργειες.

count	support	confidence	lift	rule
110	0.0001527037718894	0.371981352097034	201.14524070765576	Vitamin Well Βιταμινούχο Ρόφημα Antioxidant 500ml → Vitamin Well Βιταμινούχο Ρόφημα Reload 500ml
96	0.000054461667384893	0.39506172839386723	201.14524070765574	Δέλτα Advance Επιδόρπιο Γιαουρτιού Μήλο & Μπανάνα 2x145g → Δέλτα Advance Επιδόρπιο Γιαουρτιού Μήλο & Μπανάνα 2x145g
96	0.000054461667384893	0.39506172839386723	117.87654320987653	Δέλτα Advance Επιδόρπιο Γιαουρτιού Μήλο & Μπανάνα 2x145g → Δέλτα Advance Επιδόρπιο Γιαουρτιού Μήλο & Μπανάνα 2x145g
116	0.00012419141181399246	0.4027777777777778	116.53703703703704	Oriental Express Noodles 3 - Ήπιες Ταράτσες 87g → Oriental Express Noodles 3 - Αγανάκι Oriental 87g
153	0.0016024982823628	0.2688927984609884	113.10081069355584	Gourmet Gold Mousse για Τάτσες Ψηφία του Οικείου 85g → Gourmet Gold Mousse για Τάτσες με Βούνο 85g, Gourmet Gold Mousse για Τάτσες Κοτόπουλο 85g → Gourmet Gold Mousse για Τάτσες Κοτόπουλο 85g
109	0.00011416033331487224	0.2486609334841629	107.51585776556532	Felix Agali Υγρή Τροφή για Τάτσες Κοτόπουλο Με Ζελάτε Σε Φρέσκα Λάβα 85g → Felix Agali Υγρή Τροφή για Τάτσες Κοτόπουλο Με Ζελάτε Σε Φρέσκα Λάβα 85g
109	0.00011416033331487224	0.2477168949771124	107.51585776556532	Felix Agali Υγρή Τροφή για Τάτσες Κοτόπουλο Με Ζελάτε Σε Φρέσκα Λάβα 85g → Felix Agali Υγρή Τροφή για Τάτσες Μούριας ή Ζελάτε 85g
153	0.00016024982823628	0.642851428571428	104.03898305808473	Gourmet Gold Mousse για Τάτσες Κοτόπουλο 85g → Gourmet Gold Mousse για Τάτσες με Βούνο 85g
124	0.00012987012987012987	0.3084977114747861	102.4622996130459	Oriental Express Noodles 3 - Ήπιες Ταράτσες 87g → Oriental Express Noodles 3 - Αγανάκι Ταράτσες 87g
153	0.00016024982823628	0.4309827984609884	98.125	Gourmet Gold Mousse για Τάτσες Ψηφία του Οικείου 85g → Gourmet Gold Mousse για Τάτσες Κοτόπουλο 85g → Gourmet Gold Mousse για Τάτσες Κοτόπουλο 85g
153	0.00016024982823628	0.478125	98.15000000000001	Gourmet Gold Mousse για Τάτσες Ψηφία του Οικείου 85g → Gourmet Gold Mousse για Τάτσες Κοτόπουλο 85g
153	0.00016024982823628	0.3290327550645104	98.175	Gourmet Gold Mousse για Τάτσες Κοτόπουλο 85g → Gourmet Gold Mousse για Τάτσες με Βούνο 85g, Gourmet Gold Mousse για Τάτσες Ψηφία του Οικείου 85g
153	0.00016024982823628	0.3290327550645104	97.88839135910108	Oriental Express Noodles 3 - Ήπιες Ταράτσες 87g → Oriental Express Noodles 3 - Αγανάκι Ταράτσες 87g
136	0.00042428382065433599	0.3383084577114248	97.888391374551078	Gourmet Gold Mousse για Τάτσες Κοτόπουλο 85g → Gourmet Gold Mousse για Τάτσες με Βούνο 85g
97	0.000105191564306612	0.266483164851054	97.11391661773342	Davidoff Slims Άσπρο 5 Πάκετο 10 → Κρέπ Ροζέ Πορτοφόλι Flowers Κρέπ 75ml
320	0.000351487224549644	0.19227224549644	97.01188525810969	Kleen Poi - Έργαλες Flowers Κρέπ 100ml → Κρέπ Ροζέ Πορτοφόλι Άσπρο 5 Πάκετο 10 → Davidoff Slims Άσπρο 5 Πάκετο 10
320	0.000351487224549644	0.19227224549644	97.01188525810969	Gourmet Gold Mousse για Τάτσες με Βούνο 85g → Gourmet Gold Mousse για Τάτσες Κοτόπουλο 85g
238	0.000249766862170088	0.5118275989892432	95.8883905222027	Gourmet Gold Mousse για Τάτσες Κοτόπουλο 85g → Gourmet Gold Mousse για Τάτσες Ψηφία του Οικείου 85g
101	0.000105191564306612	0.266483164851054	95.8883905222027	Gourmet Gold Mousse για Τάτσες Κοτόπουλο 85g → Gourmet Gold Mousse για Τάτσες με Βούνο 85g
227	0.0023774612484289905	0.4881720430107523	79.00112994350283	Gourmet Gold Mousse για Τάτσες Κοτόπουλο 85g → Gourmet Gold Mousse για Τάτσες με Βούνο 85g
227	0.0023774612484289905	0.3847457627118644	79.00112994350282	Gourmet Gold Mousse για Τάτσες με Βούνο 85g → Gourmet Gold Mousse για Τάτσες Κοτόπουλο 85g
101	0.000105191564306612	0.266483164851054	79.00112994350282	Gourmet Gold Mousse για Τάτσες Κοτόπουλο 85g → Gourmet Gold Mousse για Τάτσες Ψηφία του Οικείου 85g
101	0.000105191564306612	0.266483164851054	68.79159881368785	Aligida Cornetto Soft Caramel 140ml → Aligida Cornetto Soft Cookies 140ml
162	0.000169690406367825	0.35249317647058826	60.75124025513821	Tuborg 3.6% Αναψυκτικό Λεμόνι & Δρύινος 330ml → Tuborg Pink Grapefruit 330ml
104	0.000105191564306612	0.18259590238263	59.137718253855285	Τuborg 3.6% Αναψυκτικό Λεμόνι & Δρύινος 330ml → Φρεσκάκι Αναψυκτικό Λεμόνι & Δρύινος 330ml
104	0.0001089233472978634	0.2962962962962963	48.19483871090902	Φρεσκάκι Αναψυκτικό Βάρος 40g → Φρεσκάκι Μανιτάρια 40g
104	0.0001089233472978634	0.17717206132879046	48.19483871090916	Φρεσκάκι Μανιτάρια 40g → Φρεσκάκι Αναψυκτικό Βάρος 40g
104	0.0001089233472978634	0.17717206132879046	48.19483871090916	Φρεσκάκι Μανιτάρια 40g → Φρεσκάκι Αναψυκτικό Βάρος 40g
104	0.00011018012438207	0.303994301994302	46.2831710764301	Φρεσκάκι Αναψυκτικό Βάρος 40g → Φρεσκάκι Καραμελώδες Φόλες Καθημερινά 100g
173	0.001790795984499371	0.225598667546171	43.0793667546171	Schweppes Orangeade A Hint Of Orange Blossom 330ml → Schweppes Lemonade With A Hint Of Bergamot & Hibiscus 330ml
97	0.000105191564306612	0.2029567198177678	41.044645315214097	Schweppes Lemonade With A Hint Of Bergamot & Hibiscus 330ml → Βιβλί Πηγών Chocolate Fudge Brownie 100ml
97	0.000105191564306612	0.18871593103073902	41.04464531214096	Βιβλί Πηγών Chocolate Fudge Brownie 100ml → Βιβλί Πηγών Cookie Dough 100ml
322	0.000133012491411814	0.18526974646475983	31.157718253855285	King Cone Πηγών Λιμόνια 250ml → King Cone Πηγών Βανίλια 250ml
120	0.0001260748070420616	0.18337672817764166	29.84027527893238	Scandal Πετσέτα Καπέλο Χαροκόπειο 350gr → Scandal Πετσέτα Καπέλο Χαροκόπειο 360ml
120	0.000105191564306612	0.18259590238263	28.70841450505695	Scandal Πετσέτα Καπέλο Χαροκόπειο 350gr → Scandal Πετσέτα Baby & Μπικίνι 350gr
96	0.0001054461667384893	0.18259590238263	28.70841450505695	Lay's Πατάτες Άσπρο 50g → Lay's Πατάτες Ρύζι 50g
103	0.0001077959497276918	0.1943396264150474	28.41582598275788	Alexandros Papandreaou Μηνιαίκα 350gr → Alexandros Papandreaou Μηνιαίκα Σύρος Μηνιαίκα Baby & Μπικίνι 350gr
107	0.000126053540003788	0.17174959871589088	27.936374251095845	Φρεσκάκι Καραμελώδες Φόλες Καθημερινά 100g → Φρεσκάκι Καραμελώδες Φόλες Καθημερινά 100g
107	0.000126053540003788	0.182282797886712998	27.936374251095845	Φρεσκάκι Καραμελώδες Φόλες Καθημερινά 100g → Φρεσκάκι Καραμελώδες Φόλες Καθημερινά 100g
133	0.00013720150816925012	0.2578740117480315	26.277279641005382	Swan Φτεράκια Extra Slim 54 Πάκετο 10 → Rizla Χαρτάκια Micron

Εικόνα 65. Κορυφαίοι 50 κανόνες συσχέτισης μεταξύ αγορασμένων προϊόντων σε όλες τις ομάδες πελατών, βάσει του FP-Growth αλγορίθμου

Στον ανωτέρω πίνακα (Εικόνα 60) αποτυπώνονται οι 50 κορυφαίοι κανόνες συσχέτισης επί του συνόλου των προϊόντων. Ο κανόνας με την υψηλότερη τιμή lift, 201.15, παρατηρείται μεταξύ του Vitamin Well Βιταμινούχο Ρόφημα Antioxidant 500ml ως antecedent και του Vitamin Well Βιταμινούχο Ρόφημα Reload 500ml ως consequent (και αντίστροφα). Αυτή η ισχυρή σχέση αντικατοπτρίζει σαφή προτίμηση των πελατών για αυτά τα συμπληρωματικά ροφήματα, υποδηλώνοντας ότι οι καταναλωτές συχνά αγοράζουν αυτές τις συγκεκριμένες παραλαγές μαζί.

Παρόμοιες υψηλές lift καταγράφηκαν σε ζεύγη προϊόντων γιασουρτιού και noodles. Για παράδειγμα, ο συνδυασμός "Δέλτα Advance Επιδόρπιο Γιαουρτιού Μήλο & Μπανάνα 2x145g" και "Δέλτα Advance Επιδόρπιο Γιαουρτιού Με Δημητριακά 2x145g" παρουσίασε lift = 117.88, ενώ

η συσχέτιση μεταξύ "Oriental Express Noodles 3 Λαχανικά Oriental 87g" και "Oriental Express Noodles 3 Ψητές Γαρίδες 87g" εμφάνισε lift = 116.54. Αυτά τα ευρήματα υποδεικνύουν μία σαφή τάση των καταναλωτών να αγοράζουν συγκεκριμένες ποικιλίες προϊόντων μαζί, εντός αυτών των κατηγοριών.

Επιπλέον, η ανάλυση ανέδειξ ισχυρές συσχετίσεις μεταξύ προϊόντων τροφής για γάτες. Για παράδειγμα, ο κανόνας που περιλαμβάνει το "Gourmet Gold Mousse για Γάτες με Βοδινό 85g" ως antecedent και το "Gourmet Gold Mousse για Γάτες Κοτόπουλο 85g, Ψάρια του Ωκεανού 85g" ως consequent εμφάνισε lift = 113.10. Αυτό υποδηλώνει ότι οι καταναλωτές με κατοικίδια τείνουν να αγοράζουν πολλαπλές γεύσεις της ίδιας μάρκας μαζί, γεγονός που θα μπορούσε να αξιοποιηθεί μέσω πακέτων προσφορών ή loyalty programs για προϊόντα κατοικίδιων.

Συνολικά, σε όλους τους 370 κανόνες, οι τιμές support κυμάνθηκαν από 0.001 έως 0.007, με διάμεση τιμή 0.00137, γεγονός που δείχνει ότι, παρόλο που οι κανόνες καλύπτουν μικρό ποσοστό των συνολικών συναλλαγών, η παρουσία τους δεν κρίνεται ασήμαντη. Οι confidence τιμές κυμάνθηκαν από 2.2% έως 67.4%, με διάμεση τιμή 11.99%, αντικατοπτρίζοντας το βαθμό αξιοπιστίας των κανόνων στην πρόβλεψη του consequent, δεδομένου του antecedent. Παρόλο που οι συνολικές τιμές confidence είναι σχετικά χαμηλές, οι υψηλές τιμές lift υπογραμμίζουν τη δυνατότητα αξιοποίησης των συσχετίσεων για στρατηγικές cross-selling και bundling.

Συνολικά, η ανάλυση ανέδειξε 131 μοναδικά προϊόντα που συμμετείχαν στους παραχθέντες κανόνες, υποδεικνύοντας μία ποικιλία συχνά συν-αγοραζόμενων προϊόντων. Είναι αξιοσημείωτο ότι οι κανόνες κάλυψαν ευρύ φάσμα κατηγοριών, ενισχύοντας τη σημασία της διακατηγορικής ανάλυσης (cross-category analysis) για την ταυτοποίηση πιθανών ευκαιριών δημιουργίας συνδυαστικών πακέτων και πρωθητικών ενεργειών.

Εικόνα 66. Κορυφαίοι 50 κανόνες συσχέτισης μεταξύ αγορασμένων προϊόντων για την ομάδα πρελατών "The Ovalists". Βάσει του EP-Growth αλγορίθμου.

Η εφαρμογή του αλγορίθμου FP-Growth στο τμήμα των Loyalists ανέδειξ διακριτά αγοραστικά μοτίβα σε σύγκριση με τους γενικούς κανόνες που προέκυψαν από το πλήρες σύνολο δεδομένων. Αυτό το αποτέλεσμα είναι αναμενόμενο, καθώς η ομαδοποίηση των πελατών σε διακριτές συμπεριφορικές ομάδες επιτρέπει στον αλγόριθμο να εστιάσει σε τάσεις που είναι μοναδικές για κάθε τμήμα. Οι κανόνες που προκύπτουν από την ανάλυση ανά cluster δεν αποτελούν απλώς υποσύνολα των συνολικών κανόνων, αλλά ανεξάρτητα αποτελέσματα, διαμορφωμένα από τις ειδικές αγοραστικές δυναμικές του κάθε τμήματος.

Η ανάλυση των 50 κορυφαίων κανόνων για τους Loyalists (Εικόνα 61) αποκάλυψε ισχυρές συσχετίσεις με υψηλές τιμές lift, γεγονός που υποδηλώνει ισχυρές και αξιοποιήσιμες σχέσεις μεταξύ προϊόντων εντός αυτού του πελατειακού τμήματος.

Η υψηλότερη τιμή lift, 166.93, παρατηρείται μεταξύ των προϊόντων "Vitamin Well Βιταμινούχο Ρόφημα Reload 500ml" και "Vitamin Well Βιταμινούχο Ρόφημα Antioxidant 500ml". Τα προϊόντα αυτά απευθύνονται πιθανότατα σε πελάτες που ενδιαφέρονται για υγιεινές επιλογές, υποδεικνύοντας δυνατότητες για bundling σε αυτή την κατηγορία προϊόντων.

Αντίστοιχα, παρατηρείται lift = 141.33 για τη σχέση μεταξύ των "Oriental Express Noodles 3' Λαχανικά Oriental 87g" και "Oriental Express Noodles 3' Ψητές Γαρίδες 87g", γεγονός που καταδεικνύει ισχυρή καταναλωτική πίστη (brand loyalty) στην κατηγορία των έτοιμων γευμάτων προς μαγείρεμα.

Επιπλέον, η ανάλυση ανέδειξε ισχυρές συσχετίσεις σε προϊόντα κατοικίδιων. Για παράδειγμα, το "Gourmet Gold Mousse για Γάτες με Βοδινό 85g" παρουσιάζει ισχυρή συσχέτιση με το "Gourmet Gold Mousse για Γάτες Ψάρια του Ωκεανού 85g", με lift = 112.89. Αυτό το εύρημα υποδηλώνει ότι οι Loyalists συχνά αγοράζουν διαφορετικές γεύσεις της ίδιας σειράς προϊόντων, γεγονός που προσφέρει ευκαιρίες για cross-selling μέσω πακέτων προϊόντων ή προγραμμάτων επιβράβευσης πιστών πελατών.

Μία ακόμη αξιοσημείωτη συσχέτιση αφορά τα προϊόντα καπνού. Ο συνδυασμός "Tereia Silver Πακέτο" και "Tereia Willow Πακέτο" όταν αγοράζεται με "Coca-Cola Zero Multipack 2x1.5L" εμφανίζει lift = 124.43, υποδηλώνοντας συμπληρωματική αγοραστική συμπεριφορά μεταξύ των προϊόντων καπνού και των ποτών. Αυτό το εύρημα μπορεί να αξιοποιηθεί στη διαμόρφωση στρατηγικών τοποθέτησης προϊόντων εντός καταστημάτων, με στόχο την αύξηση των συνολικών πωλήσεων μέσω προσεκτικά επιλεγμένων πρωθητικών ενεργειών.

Εικόνα 67. Κορυφαίοι 50 κανόνες συσχέτισης μεταξύ αγορασμένων προϊόντων για την ομάδα πελατών "Potential Loyalists", βάσει του FP-Growth αλγορίθμου

Η εφαρμογή του αλγορίθμου FP-Growth στο τμήμα των Potential Loyalists αποκάλυψε πολύτιμες πληροφορίες σχετικά με τις μοναδικές αγοραστικές συμπεριφορές και προτιμήσεις αυτής της πελατειακής ομάδας. Σε αντίθεση με τους γενικούς κανόνες που προέκυψαν από το πλήρες σύνολο δεδομένων, οι κανόνες που δημιουργήθηκαν για αυτό το τμήμα αντανακλούν συγκεκριμένα μοτίβα, τα οποία διαφοροποιούν αυτήν από τις υπόλοιπες ομάδες.

Μεταξύ των σημαντικότερων ευρημάτων στους 50 κορυφαίους κανόνες (Εικόνα 62), αναδεικνύεται η ισχυρή συσχέτιση μεταξύ premium ποτών. Ο κανόνας που συνδέει το "The Locals Pom Royal Vodka Blended with Pomegranate & Ginger 250ml" με το "The Locals Madras Vodka Customer Segmentation & Market Basket Analysis 102

"Blended with Cranberry & Orange 250ml" καταγράφει την υψηλότερη τιμή lift, 359.57. Αυτό υποδηλώνει έντονη συμπληρωματικότητα μεταξύ αυτών των προϊόντων, αποκαλύπτοντας ότι οι Potential Loyalists δείχνουν προτίμηση σε premium αλκοολούχα ποτά. Αυτά τα ευρήματα προσφέρουν ευκαιρίες για συνδυαστικές πρωθητικές ενέργειες (bundled promotions) που επικεντρώνονται σε αυτή την κατηγορία προϊόντων.

Ένα ακόμη αξιοσημείωτο μοτίβο αφορά τα προϊόντα που σχετίζονται με την υγεία και τη φυσική κατάσταση. Εντοπίστηκε υψηλή τιμή lift, 127.30, μεταξύ του "Arla Protein Shake Ρόφημα Γάλακτος Σοκολάτα BCAA 250ml" και της "Lab Nutrition XL-High Protein Bar 33% Σοκολάτα 90g". Αυτή η σχέση υποδηλώνει ότι οι Potential Loyalists τείνουν να συμπεριλαμβάνουν προϊόντα υψηλής πρωτεΐνης και fitness-oriented προϊόντα στις αγορές τους, ενισχύοντας τη δυνατότητα διασταυρούμενων προωθήσεων (cross-marketing) μεταξύ αυτών των ειδών.

Παρόμοια τάση καταγράφεται και στην κατηγορία των προϊόντων για κατοικίδια. Πολλοί κανόνες αναδεικνύουν ισχυρές συσχετίσεις μεταξύ διαφορετικών γεύσεων της σειράς Gourmet Gold Mousse για γάτες. Για παράδειγμα, ο κανόνας που συνδέει το "Gourmet Gold Mousse for Cats with Beef 85g" με το "Gourmet Gold Mousse for Cats Ocean Fish 85g" εμφανίζει lift = 99.17, γεγονός που αντικατοπτρίζει έντονη προτίμηση των καταναλωτών σε συγκεκριμένα brands και παραλλαγές προϊόντων. Αυτό δημιουργεί ευκαιρίες για στοχευμένες εκπτώσεις και loyalty programs εντός της κατηγορίας των προϊόντων κατοικιδίων.

Επιπλέον, η κατηγορία των έτοιμων γευμάτων και των ethnic τροφίμων κατέχει σημαντική θέση στα αποτελέσματα. Για παράδειγμα, η ισχυρή συσχέτιση μεταξύ των "Oriental Express Noodles 3' Λαχανικά Oriental 87g" και "Oriental Express Noodles 3' Ψητές Γαρίδες 87g", με lift = 100.64, υποδεικνύει ότι τα convenience foods έχουν ιδιαίτερη δημοτικότητα εντός αυτού του τμήματος. Αυτό το εύρημα επιβεβαιώνει την ανάγκη για βελτιστοποιημένη τοποθέτηση αυτών των προϊόντων τόσο σε φυσικά καταστήματα όσο και σε ψηφιακές πλατφόρμες, ώστε να μεγιστοποιηθεί η έκθεση και να ενισχυθεί η πιθανότητα συνδυαστικών αγορών.

Εικόνα 68. Κορυφαίοι 50 κανόνες συσχέτισης μεταξύ αγορασμένων προϊόντων για την ομάδα πλειαράντων "Churners". Βάσει του FP-Growth αλγορίθμου.

Αναφορικά με το τμήμα των Churners, η εφαρμογή του αλγορίθμου FP-Growth έφερε επίσης ισχυρές συσχετίσεις ανάμεσα στους 50 κορυφαίους κανόνες (Εικόνα 64). Ένα σημαντικό εύρημα αφορά τη συσχέτιση μεταξύ του "Alexandros Papandreou Μουσακάς 350g" και του "Scandal Παγωτό Κύπελλο Chocolate Orgy 140ml" καθώς και των αντίστοιχων παραλλαγών τους. Ο κανόνας που συνδέει το "Alexandros Papandreou Μουσακάς 350g" και το "Alexandros Papandreou Μπιφτέκια Σχάρας Με Πατάτες Baby & Μπέικον 350gr" με τα "Scandal Παγωτό Κύπελλο Chocolate Orgy 140ml" και "Scandal Παγωτό Κύπελλο Caramel Secrets 140ml" εμφανίζει εξαιρετικά υψηλό lift = 266.52. Αυτό υποδηλώνει ότι οι πελάτες αυτού του τμήματος αγοράζουν συχνά αυτά τα προϊόντα μαζί, αντανακλώντας μία προτίμηση για έτοιμα γεύματα και

επιδόρπια. Τέτοιοι συνδυασμοί προσφέρουν ευκαιρίες για στοχευμένες εκστρατείες επανενεργοποίησης, μέσω εξατομικευμένων πακέτων και εκπτώσεων σε αυτά τα είδη.

Παρομοίως, η ισχυρή συσχέτιση μεταξύ του "Krambals Μπρουσκέτες με Ντομάτα & Μοτσαρέλα 70g" και του "Krambals Μπρουσκέτες με Μανιτάρια & Βούτυρο 70g", με lift = 145.11, αναδεικνύει μια κοινή προτίμηση για φαγητό τύπου σνακ εντός αυτής της ομάδας. Αυτό υπογραμίζει τη σημασία της προώθησης premium snack προϊόντων μέσω εξατομικευμένων προτάσεων ή περιορισμένων προσφορών.

Η ανάλυση ανέδειξε επίσης ισχυρές συσχέτισεις μεταξύ προϊόντων ποτών, ιδιαίτερα στις διάφορες παραλλαγές του "Ολυμπος Τσάι του Βουνού". Ο κανόνας που συνδέει το "Ολυμπος Τσάι του Βουνού Λεμόνι 500ml" με το "Ολυμπος Τσάι του Βουνού Ροδάκινο 500ml" εμφανίζει lift = 132.15, υποδεικνύοντας μια τάση αγοράς συμπληρωματικών γεύσεων. Αυτή η πληροφορία θα μπορούσε να αξιοποιηθεί μέσω πολυγευστικών πρωθήσεων (multi-flavor promotions) ή εκπτώσεων, με σκοπό την αύξηση των επαναλαμβανόμενων αγορών.

Στην κατηγορία των αλμυρών προϊόντων, η ισχυρή συσχέτιση μεταξύ των "Tsakiris Πατατάκια Ρίγανη 120g" και "Tsakiris Πατατάκια Αλάτι 120g", με lift = 125.86, καταδεικνύει μια σαφή προτίμηση των Churners για παραδοσιακά ελληνικά πατατάκια. Αυτή η πληροφορία προσφέρει ευκαιρίες επανενεργοποίησης των Churners μέσω πακέτων που συνδυάζουν διαφορετικές γεύσεις ή αποκλειστικών εκπτώσεων σε ελληνικά σνακ προϊόντα.

9.3.1 Σύγκριση αποτελεσμάτων αλγορίθμων Apriori και FP-Growth

Η σύγκριση μεταξύ των τιμών lift που παράγονται από τους αλγορίθμους FP-Growth και Apriori αναδεικνύει σημαντικές διαφορές στην ένταση των συσχέτισεων που ανιχνεύονται από κάθε μέθοδο. Συγκεκριμένα, οι τιμές lift που προκύπτουν από τον FP-Growth είναι σταθερά υψηλότερες σε σύγκριση με αυτές του Apriori. Για παράδειγμα, ο ισχυρότερος κανόνας που ανιχνεύθηκε μέσω του FP-Growth εμφανίζει lift = 201.14 για έναν συνδυασμό ποτών της Vitamin Well, ενώ ο ισχυρότερος κανόνας του Apriori φτάνει μόλις στο 34.48 για έναν παρόμοιο συνδυασμό προϊόντων Tuborg.

Αυτό το μοτίβο υψηλότερων lift τιμών στον FP-Growth παρατηρείται σε πολλαπλά ζεύγη προϊόντων, με τις τιμές του FP-Growth να είναι σημαντικά υψηλότερες από αυτές του Apriori. Τα υψηλά αυτά lift μπορεί να υποδηλώνουν ισχυρότερες συσχέτισεις, αλλά ταυτόχρονα μπορεί να οφείλονται και σε πιθανή υπερπροσαρμογή του αλγορίθμου (overfitting), λόγω της ικανότητάς του να εξερευνά αποδοτικότερα μεγαλύτερα σύνολα συχνών itemsets.

Σε ό,τι αφορά την τιμή support, οι δύο αλγόριθμοι εμφανίζουν παρόμοια επίπεδα υποστήριξης για κοινούς κανόνες, γεγονός που υποδηλώνει ότι και οι δύο μέθοδοι εντοπίζουν παρόμοια επίπεδα συχνότητας για τις συσχέτισεις προϊόντων. Ωστόσο, οι confidence τιμές ακολουθούν παρόμοια τάση με το lift, με τον FP-Growth να παράγει γενικά υψηλότερα confidence scores. Αυτό ενισχύει περαιτέρω την τάση του FP-Growth να αναδεικνύει ισχυρότερες συσχέτισεις, πιθανώς λόγω της δυνατότητάς του να χειρίζεται μεγαλύτερα itemsets χωρίς να εφαρμόζει επιθετικό pruning, όπως ο Apriori.

Από την άποψη της αξιοπιστίας, ο Apriori μπορεί να θεωρηθεί πιο ανθεκτικός για την ερμηνεία των τιμών lift σε επιχειρηματικά πλαίσια. Οι χαμηλότερες τιμές lift είναι λιγότερο πιθανό να προέρχονται από υπερπροσαρμογή, καθώς ο Apriori εφαρμόζει αυστηρότερα κριτήρια pruning με βάση τα κατώφλια minimum support και confidence.

Αντίθετα, οι υψηλότερες τιμές lift του FP-Growth μπορούν να προσφέρουν πολύτιμες πληροφορίες, ειδικά όταν αναλύονται εξειδικευμένα ή πιο εξειδικευμένα σύνολα προϊόντων, αλλά μπορεί περιστασιακά να υπερτονίζουν τη δύναμη των συσχέτισεων. Με βάση αυτά τα χαρακτηριστικά:

- Ο Apriori είναι προτιμότερος για πιο συντηρητικά και γενικευμένα επιχειρηματικά insights, καθώς οι αποδιδόμενες συσχέτισεις είναι πιο αξιόπιστες και εύκολα ερμηνεύσιμες.

- Ο FP-Growth είναι πιο κατάλληλος για την εξερεύνηση πιο λεπτομερών και πολύπλοκων συσχετίσεων, ιδιαίτερα σε μεγαλύτερα σύνολα δεδομένων.

Και οι δύο αλγόριθμοι αποτελούν ισχυρά εργαλεία για την αποκάλυψη κρυφών συσχετίσεων και προτύπων στις συναλλαγές πελατών. Ωστόσο, διαφέρουν στη διαχείριση μεγάλων όγκων δεδομένων:

- Ο Apriori λειτουργεί μέσω επαναληπτικής διαδικασίας (iterative process), γεγονός που μπορεί να καταστήσει τον αλγόριθμο υπολογιστικά δαπανηρό, ιδιαίτερα όταν αναλύονται σύνολα δεδομένων με μεγάλο αριθμό μοναδικών προϊόντων ή συναλλαγών.
- Ο FP-Growth, από την άλλη πλευρά, χρησιμοποιεί μια δομή δέντρου (tree-based structure) για τη συμπίεση και ανάλυση των συχνών itemsets, βελτιώνοντας σημαντικά την υπολογιστική απόδοση. Αυτό τον καθιστά ιδανικό για την ανάλυση μεγάλων συνόλων δεδομένων, χωρίς τους περιορισμούς μνήμης που αντιμετωπίζει ο Apriori. Ωστόσο, η συμπιεσμένη αναπαράσταση και η ελάχιστη εφαρμογή pruning μπορεί περιστασιακά να οδηγήσουν σε αποτέλεσμα πιο δύσκολα στην ερμηνεία ή λιγότερο αξιόπιστα για τη λήψη στρατηγικών αποφάσεων.

Σε γενικές γραμμές, όταν δεν υπάρχουν περιορισμοί στη μνήμη και την υπολογιστική ισχύ, η συνδυαστική χρήση των δύο αλγορίθμων μπορεί να προσφέρει μια ολοκληρωμένη ανάλυση της αγοραστικής συμπεριφοράς των πελατών:

Ο συντηρητικός χαρακτήρας του Apriori μπορεί να συμπληρώσει τα λεπτομερή ευρήματα του FP-Growth, επιτρέποντας στις επιχειρήσεις να εξάγουν τόσο ευρεία όσο και εξειδικευμένα actionable insights. Ταυτόχρονα, οι γενικευμένες και αξιόπιστες συσχετίσεις του Apriori μπορούν να χρησιμοποιηθούν ως θεμέλιο για στρατηγικές λήψης αποφάσεων, ενώ οι πιο λεπτομερείς και εξειδικευμένες ανακαλύψεις του FP-Growth μπορούν να αξιοποιηθούν για καινοτόμες και πιο στοχευμένες στρατηγικές προώθησης προϊόντων.

Η συνδυαστική εφαρμογή των δύο μεθόδων επιτρέπει στις επιχειρήσεις να εντοπίσουν κρυμμένα μοτίβα, να βελτιστοποιήσουν στρατηγικές διασταυρούμενων πωλήσεων (cross-selling), να εξατομικεύσουν πρωθητικές ενέργειες και να αυξήσουν τη συνολική αξία των συναλλαγών.

ΣΥΜΠΕΡΑΣΜΑΤΑ

Η παρούσα διατριβή ανέλυσε δεδομένα συναλλαγών από μια διαδικτυακή πλατφόρμα παράδοσης ειδών σούπερ μάρκετ, με στόχο την εξαγωγή χρήσιμων επιχειρησιακών πληροφοριών μέσω της Τμηματοποίησης Πελατών και της Ανάλυσης Καλαθιού Αγορών. Τα ευρήματα ανέδειξαν σημαντικά μοτίβα στη συμπεριφορά των πελατών, τις τάσεις των πωλήσεων και τις προτιμήσεις προϊόντων, τα οποία μπορούν να αξιοποιηθούν για τη βελτίωση της στρατηγικής της εταιρείας.

Από τη διερευνητική ανάλυση των δεδομένων, παρατηρήθηκαν έντονες εποχικές διακυμάνσεις στα έσοδα, με χαμηλή απόδοση στις αρχές του έτους και κορύφωση κατά τους καλοκαιρινούς μήνες, ακολουθούμενη από απότομη πτώση τον Σεπτέμβριο. Οι περισσότερες πωλήσεις προέρχονται από τις μεγάλες πόλεις, με την Αθήνα, τον Πειραιά και τη Θεσσαλονίκη να συγκεντρώνουν το μεγαλύτερο ποσοστό των παραγγελιών, ενώ μικρότερες περιοχές εμφανίζουν περιορισμένη δραστηριότητα. Οι συνήθειες αγορών διαφέρουν ανάλογα με την περιοχή, καθώς σε ορισμένες οι πελάτες πραγματοποιούν παραγγελίες πιο συχνά, ενώ σε άλλες η αγοραστική συμπεριφορά είναι πιο περιστασιακή. Η ζήτηση κορυφώνεται τις Παρασκευές και τα Σάββατα, γεγονός που συνδέεται με τις καταναλωτικές συνήθειες του σαββατοκύριακου, ενώ η αξία των παραγγελιών είναι υψηλότερη την Παρασκευή, πιθανώς λόγω μεγαλύτερων ή πιο ακριβών αγορών.

Η ανάλυση της πελατειακής βάσης επέτρεψε τη δημιουργία τριών διακριτών ομάδων: των Πιστών Πελατών (Loyalists), που πραγματοποιούν συχνές και υψηλής αξίας αγορές, των Δυνητικά Πιστών Πελατών (Potential Loyalists), οι οποίοι έχουν μέτρια αγοραστική δραστηριότητα και μπορούν να ενισχύσουν τη σχέση τους με την εταιρεία, και των Πελατών σε Κίνδυνο Αποχώρησης (Churners), οι οποίοι έχουν ασταθείς ή σποραδικές συναλλαγές. Αυτή η ταξινόμηση επιτρέπει τη στοχευμένη προσέγγιση των πελατών με προσαρμοσμένες πρωθητικές ενέργειες, προγράμματα επιβράβευσης και στρατηγικές επαναδραστηριοποίησης.

Η Ανάλυση Καλαθιού Αγορών ανέδειξε ισχυρές συσχετίσεις μεταξύ προϊόντων, προσφέροντας πολύτιμες πληροφορίες για την τοποθέτηση και την προώθησή τους. Οι πιο συχνοί συνδυασμοί αγορών μπορούν να αξιοποιηθούν για τη δημιουργία πακέτων προσφορών ή για τη βελτιστοποίηση των προτάσεων προϊόντων στους πελάτες. Παράλληλα, η ενσωμάτωση των ευρημάτων της τμηματοποίησης στην ανάλυση καλαθιού επιτρέπει πιο στοχευμένες στρατηγικές, προσαρμοσμένες στις ανάγκες κάθε πελατειακής ομάδας.

Η διατριβή ολοκληρώνεται με την ανάδειξη συγκεκριμένων επιχειρησιακών στρατηγικών που μπορούν να συμβάλουν στη βελτίωση της απόδοσης της εταιρείας. Η διατήρηση των πιστών πελατών, η αύξηση της αγοραστικής συχνότητας στις λιγότερο ενεργές ομάδες, η αξιοποίηση της εποχικότητας για πρωθητικές ενέργειες και η διεύρυνση της γεωγραφικής εμβέλειας αποτελούν βασικές κατευθύνσεις για τη μελλοντική ανάπτυξη. Επιπλέον, η αξιοποίηση των κανόνων συσχέτισης προϊόντων μπορεί να βελτιώσει την αγοραστική εμπειρία, αυξάνοντας τόσο τη μέση αξία παραγγελίας όσο και τη συνολική ικανοποίηση των πελατών.

Η αποτελεσματική αξιοποίηση των δεδομένων μπορεί να προσφέρει στην εταιρεία ανταγωνιστικό πλεονέκτημα, επιτρέποντάς της να προσαρμόζεται δυναμικά στις μεταβαλλόμενες ανάγκες των καταναλωτών και να μεγιστοποιεί την κερδοφορία της. Οι επιχειρήσεις που επενδύουν στην αναλυτική δεδομένων αποκτούν τη δυνατότητα να λαμβάνουν πιο τεκμηριωμένες αποφάσεις, ενισχύοντας τη στρατηγική τους με γνώμονα την πελατοκεντρική ανάπτυξη.

ΠΑΡΑΡΤΗΜΑ

Αρχικά παρατίθεται ο κώδικας SQL που χρησιμοποιήθηκε για την υλοποίηση της διαδικασίας tou imputation.

Υλοποίηση μετασχηματισμού και imputation με SQL

```
--Dataset info
SELECT      COLUMN_NAME,
            DATA_TYPE,
            CHARACTER_MAXIMUM_LENGTH,
            NUMERIC_PRECISION,
            NUMERIC_SCALE,
            IS_NULLABLE
FROM        INFORMATION_SCHEMA.COLUMNS
WHERE       TABLE_NAME = 'Dataset' AND
            TABLE_SCHEMA = 'dbo'
ORDER BY    ORDINAL_POSITION
```

--Backup orginal dataset

```
SELECT      *
```

```
INTO      Dataset_Backup
FROM      [Dissertation].[dbo].[Dataset]
```

```
--Change 'product_price' data type from decimal to float
ALTER TABLE [Dissertation].[dbo].[Dataset]
ALTER COLUMN product_price FLOAT
```

```
/* IMPUTATION PROCESS */
```

```
--Initial dataset
SELECT      *
FROM        [Dissertation].[dbo].[Dataset]
```

```
--Identify all columns that contains null values
SELECT      *
FROM        [Dissertation].[dbo].[Dataset]
WHERE       product_name IS NULL --135
OR         product_category IS NULL --5
OR         f0_ IS NULL --13
```

```
--Identify 'product_name' null values for non-null 'product_category' values
SELECT      *
FROM        [Dissertation].[dbo].[Dataset]
WHERE       TRIM(product_category) IN (
    N'Healthy Αλμυρά Snacks ',
    N'Make A Wish ',
    N'Αρνί ',
    N'Αρτοσκευάσματα ',
    N'Αυγά ',
    N'Γλυκά ',
    N'Είδη επάλειψης ',
    N'Ετοιμες Σαλάτες ',
    N'Ζύμες, Πίτες & Πίτσες ',
    N'Κρέατα & Πουλερικά ',
    N'Κρουασάν & Τσουρέκια ',
    N'Πατατάκια & Αλμυρά Snacks ',
    N'Πατατάκια & Γαριδάκια ',
    N'Προϊόντα Κρεάτων & Πουλερικών ',
    N'Σοκολάτες',
    N'Χοιρινό ',
    N'Ψωμί '
)
```

```

AND      product_name IS NULL
ORDER BY product_category,
         product_price
--130

```

```

--Identify 'product_name' null values for null 'product_category' values
SELECT      *
FROM        [Dissertation].[dbo].[Dataset]
WHERE       product_name IS NULL
AND        product_category IS NULL
--5

```

```

--1a. Identify null product_name values for product_category 'Healthy Αλμυρά Snacks'
SELECT      *
FROM        [Dissertation].[dbo].[Dataset]
WHERE       TRIM(product_category) = N'Healthy Αλμυρά Snacks'
AND        product_name IS NULL --(1.6)

```

```

--1b. Finding mode for non-null product_name values for product_category 'Healthy Αλμυρά Snacks'
SELECT      product_name,
            COUNT(*) AS ProductCount
FROM        [Dissertation].[dbo].[Dataset]
WHERE       TRIM(product_category) = N'Healthy Αλμυρά Snacks'
AND        product_name IS NOT NULL
AND        product_price = '1.6'
GROUP BY   product_name
ORDER BY   2 DESC
--'Nestle Fitness Multigrain Chips Δημητριακών Θαλασσινό Αλάτι 75g'

```

```

--2a. Identify null product_name values for product_category 'Make A Wish'
SELECT      *
FROM        [Dissertation].[dbo].[Dataset]
WHERE       TRIM(product_category) = N'Make A Wish'
AND        product_name IS NULL

```

```

--2b. Finding mode for non-null product_name values for product_category 'Make A Wish'
SELECT      product_name, COUNT(*) AS ProductCount
FROM        [Dissertation].[dbo].[Dataset]
WHERE       TRIM(product_category) = N'Make A Wish'
GROUP BY   product_name
--NoName

```

--3a. Identify null product_name values for product_category 'Αρνί'

```
SELECT      *
FROM      [Dissertation].[dbo].[Dataset]
WHERE      TRIM(product_category) = N'Αρνί'
AND        product_name IS NULL
```

--3b. Finding mode for non-null product_name values for product_category 'Αρνί'

```
SELECT      product_name, COUNT(*) AS ProductCount
FROM      [Dissertation].[dbo].[Dataset]
WHERE      TRIM(product_category) = N'Αρνί'
AND        product_name IS NOT NULL
GROUP BY  product_name
--NoName
```

--4a. Identify null product_name values for product_category 'Αρτοσκευάσματα'

```
SELECT      *
FROM      [Dissertation].[dbo].[Dataset]
WHERE      TRIM(product_category) = N'Αρτοσκευάσματα'
AND        product_name IS NULL --(1.6)
```

--4b. Finding mode for non-null product_name values for product_category 'Αρτοσκευάσματα'

```
SELECT      product_name, COUNT(*) AS ProductCount
FROM      [Dissertation].[dbo].[Dataset]
WHERE      TRIM(product_category) = N'Αρτοσκευάσματα'
AND        product_name IS NOT NULL
AND        product_price = '1.6'
GROUP BY  product_name
--ELITE Φρυγανιές Σικάλεως 180g
```

--5a. Identify null product_name values for product_category 'Αυγά'

```
SELECT      *
FROM      [Dissertation].[dbo].[Dataset]
WHERE      TRIM(product_category) = N'Αυγά'
AND        product_name IS NULL --(3, 4, 7.25)
ORDER BY   product_price
```

--5b. Finding mode for non-null product_name values for product_category 'Αυγά'

```
SELECT      product_name,
           product_price,
```

```

        COUNT(*) AS ProductCount
FROM      [Dissertation].[dbo].[Dataset]
WHERE     TRIM(product_category) = N'Αυγά'
AND       product_name IS NOT NULL
--AND      product_price = '7.25'
GROUP BY  product_name,
          product_price
ORDER BY   3 DESC
-- 3='Χρυσά Αυγά Αχυρώνα με Βιταμίνη D 53-63g 6τεμ -0.5€', 4='Χρυσά Αυγά Μεσαία 10τεμ σε Πλαστική Θήκη', 7.25='NoName'

```

--6a. Identify null product_name values for product_category 'Γλυκά'

```

SELECT      *
FROM      [Dissertation].[dbo].[Dataset]
WHERE     TRIM(product_category) = N'Γλυκά'
AND       product_name IS NULL --(6.75)
ORDER BY   product_price

```

--6b. Finding mode for non-null product_name values for product_category 'Γλυκά'

```

SELECT      product_name, COUNT(*) AS ProductCount
FROM      [Dissertation].[dbo].[Dataset]
WHERE     TRIM(product_category) = N'Γλυκά'
AND       product_name IS NOT NULL
AND       product_price = '6.75'
GROUP BY  product_name
--'Μακεδονικός Χαλβάς Κακάο 400g'

```

--7a. Identify null product_name values for product_category 'Είδη επάλειψης'

```

SELECT      *
FROM      [Dissertation].[dbo].[Dataset]
WHERE     TRIM(product_category) = N'Είδη επάλειψης'
AND       product_name IS NULL --(6.75)
ORDER BY   product_price

```

--7b. Finding mode for non-null product_name values for product_category 'Είδη επάλειψης'

```

SELECT      product_name, COUNT(*) AS ProductCount
FROM      [Dissertation].[dbo].[Dataset]
WHERE     TRIM(product_category) = N'Είδη επάλειψης'
AND       product_name IS NOT NULL
AND       product_price = '6.75'
GROUP BY  product_name
--'Μακεδονικός Χαλβάς Κακάο 400g'

```

--8a. Identify null product_name values for product_category "Ετοιμες Σαλάτες"

```
SELECT      *
FROM      [Dissertation].[dbo].[Dataset]
WHERE      TRIM(product_category) = N'Ετοιμες Σαλάτες'
AND        product_name IS NULL --(3.5)
ORDER BY    product_price
```

--8b. Finding mode for non-null product_name values for product_category "Ετοιμες Σαλάτες"

```
SELECT      product_name, COUNT(*) AS ProductCount
FROM      [Dissertation].[dbo].[Dataset]
WHERE      TRIM(product_category) = N'Ετοιμες Σαλάτες'
AND        product_name IS NOT NULL
AND        product_price = '3.5'
GROUP BY   product_name
--'NoName'
```

--9a. Identify null product_name values for product_category 'Ζύμες, Πίτες & Πίτσες'

```
SELECT      *
FROM      [Dissertation].[dbo].[Dataset]
WHERE      TRIM(product_category) = N'Ζύμες, Πίτες & Πίτσες'
AND        product_name IS NULL --(5.8)
ORDER BY    product_price
```

--9b. Finding mode for non-null product_name values for product_category 'Ζύμες, Πίτες & Πίτσες'

```
SELECT      product_name, COUNT(*) AS ProductCount
FROM      [Dissertation].[dbo].[Dataset]
WHERE      TRIM(product_category) = N'Ζύμες, Πίτες & Πίτσες'
AND        product_name IS NOT NULL
AND        product_price = '5.88'
GROUP BY   product_name
--'Alfa Κασερόπιτα Τσαλακωτή Πηλίου 850g'
```

--10a. Identify null product_name values for product_category 'Κρέατα & Πουλερικά'

```
SELECT      *
FROM      [Dissertation].[dbo].[Dataset]
WHERE      TRIM(product_category) = N'Κρέατα & Πουλερικά'
AND        product_name IS NULL --(28.05)
ORDER BY    product_price
```

--10b. Finding mode for non-null product_name values for product_category 'Κρέατα & Πουλερικά'

```

SELECT      product_name, COUNT(*) AS ProductCount
FROM        [Dissertation].[dbo].[Dataset]
WHERE       TRIM(product_category) = N'Κρέατα & Πουλερικά'
AND         product_name IS NOT NULL
AND         product_price = '28.05'
GROUP BY    product_name
--'NoName'

```

--11a. Identify null product_name values for product_category 'Κρουασάν & Τσουρέκια'

```

SELECT      *
FROM        [Dissertation].[dbo].[Dataset]
WHERE       TRIM(product_category) = N'Κρουασάν & Τσουρέκια'
AND         product_name IS NULL --(5.1,6.1)
ORDER BY    product_price

```

--11b. Finding mode for non-null product_name values for product_category 'Κρουασάν & Τσουρέκια'

```

SELECT      product_name, COUNT(*) AS ProductCount
FROM        [Dissertation].[dbo].[Dataset]
WHERE       TRIM(product_category) = N'Κρουασάν & Τσουρέκια'
AND         product_name IS NOT NULL
AND         product_price IN ('5.1','6.1')
GROUP BY    product_name
--'NoName'

```

--12a. Identify null product_name values for product_category 'Πατατάκια & Αλμυρά Snacks'

```

SELECT      *
FROM        [Dissertation].[dbo].[Dataset]
WHERE       TRIM(product_category) = N'Πατατάκια & Αλμυρά Snacks'
AND         product_name IS NULL --(1.7)
ORDER BY    product_price

```

--12b. Finding mode for non-null product_name values for product_category 'Πατατάκια & Αλμυρά Snacks'

```

SELECT      product_name, COUNT(*) AS ProductCount
FROM        [Dissertation].[dbo].[Dataset]
WHERE       TRIM(product_category) = N'Πατατάκια & Αλμυρά Snacks'
AND         product_name IS NOT NULL
AND         product_price IN ('1.7')
GROUP BY    product_name
ORDER BY    2 DESC
--'Lay's Πατατάκια στο Φούρνο Γιαούρτι & Μυρωδικά 105g'

```

--13a. Identify null product_name values for product_category 'Πατατάκια & Γαριδάκια'

```
SELECT      *
FROM      [Dissertation].[dbo].[Dataset]
WHERE      TRIM(product_category) = N'Πατατάκια & Γαριδάκια'
AND        product_name IS NULL --(1.7)
ORDER BY    product_price
```

--13b. Finding mode for non-null product_name values for product_category 'Πατατάκια & Γαριδάκια'

```
SELECT      product_name, COUNT(*) AS ProductCount
FROM      [Dissertation].[dbo].[Dataset]
WHERE      TRIM(product_category) = N'Πατατάκια & Γαριδάκια'
AND        product_name IS NOT NULL
AND        product_price IN ('1.7')
GROUP BY   product_name
ORDER BY 2 DESC
--'Tasty Πιτσίνια 110g'
```

--14a. Identify null product_name values for product_category 'Προϊόντα Κρεάτων & Πουλερικών'

```
SELECT      *
FROM      [Dissertation].[dbo].[Dataset]
WHERE      TRIM(product_category) = N'Προϊόντα Κρεάτων & Πουλερικών'
AND        product_name IS NULL --(4.62,5.39,6.16,6.94,9.91)
ORDER BY    product_price
```

--14b. Finding mode for non-null product_name values for product_category 'Προϊόντα Κρεάτων & Πουλερικών'

```
SELECT      product_name, COUNT(*) AS ProductCount
FROM      [Dissertation].[dbo].[Dataset]
WHERE      TRIM(product_category) = N'Προϊόντα Κρεάτων & Πουλερικών'
AND        product_name IS NOT NULL
AND        product_price IN ('4.62','5.39','6.16','6.94','9.91')
GROUP BY   product_name
ORDER BY 2 DESC
--'NoName'
```

--15a. Identify null product_name values for product_category 'Σοκολάτες'

```
SELECT      *
FROM      [Dissertation].[dbo].[Dataset]
WHERE      TRIM(product_category) = N'Σοκολάτες'
AND        product_name IS NULL --(0.5,1.8,3.6)
ORDER BY    product_price
```

--15b. Finding mode for non-null product_name values for product_category 'Σοκολάτες'

```
SELECT      product_name, COUNT(*) AS ProductCount
FROM       [Dissertation].[dbo].[Dataset]
WHERE      TRIM(product_category) = N'Σοκολάτες'
AND        product_name IS NOT NULL
AND        product_price IN ('0.5','1.8','3.6')
GROUP BY   product_name
ORDER BY   2 DESC
--0.5='Πιασιλίδης Kiss Σοκολάτα Γάλακτος 27.5g',1.8="Iov Break Σοκολάτα Γάλακτος με Φουντούκι 85g',3.6='NoName'
```

--16a. Identify null product_name values for product_category 'Χοιρινό'

```
SELECT      *
FROM       [Dissertation].[dbo].[Dataset]
WHERE      TRIM(product_category) = N'Χοιρινό'
AND        product_name IS NULL --(7.3)
ORDER BY   product_price
```

--16b. Finding mode for non-null product_name values for product_category 'Χοιρινό'

```
SELECT      product_name, COUNT(*) AS ProductCount
FROM       [Dissertation].[dbo].[Dataset]
WHERE      TRIM(product_category) = N'Χοιρινό'
AND        product_name IS NOT NULL
AND        product_price IN ('7.3')
GROUP BY   product_name
ORDER BY   2 DESC
--'NoName'
```

--17a. Identify null product_name values for product_category 'Ψωμί'

```
SELECT      *
FROM       [Dissertation].[dbo].[Dataset]
WHERE      TRIM(product_category) = N'Ψωμί'
AND        product_name IS NULL --()
ORDER BY   product_price
```

--17b. Finding mode for non-null product_name values for product_category 'Ψωμί'

```
SELECT      product_name, COUNT(*) AS ProductCount
FROM       [Dissertation].[dbo].[Dataset]
WHERE      TRIM(product_category) = N'Ψωμί'
AND        product_name IS NOT NULL
AND        product_price IN ('1.6')
GROUP BY   product_name
ORDER BY   2 DESC
```

--'Παπαδοπούλου Γεύση2 Ψωμί Τοστ Σίτου 700g'

--Final dataset with non-null values

```
DROP TABLE IF EXISTS #FilledNull
SELECT      order_id,
            customer_id,
            order_timestamp,
CASE WHEN product_name IS NULL AND TRIM(product_category) = N'Healthy Αλμυρά Snacks'      AND
product_price = '1.6'                      THEN N'Nestle Fitness Multigrain Chips Δημητριακών Θαλασσινό
Αλάτι 75g'
WHEN product_name IS NULL AND TRIM(product_category) = N'Make A Wish'                         AND
product_price = '2'                           THEN N'NoName'
WHEN product_name IS NULL AND TRIM(product_category) = N'Αρνί'                            AND
product_price IN ('16.2','43.52')
WHEN product_name IS NULL AND TRIM(product_category) = N'Αροσκευάσματα'                   AND
product_price = '1.6'                         THEN N'ELITE Φρυγανίές Σικάλεως 180g'
WHEN product_name IS NULL AND TRIM(product_category) = N'Αυγά' THEN
CASE
WHEN product_price = '3' THEN N'Χρυσά Αυγά Αχυρώνα με Βιταμίνη D 53-63g 6τεμ -0.5€'
WHEN product_price = '4' THEN N'Χρυσά Αυγά Μεσαία 10τεμ σε Πλαστική Θήκη'
WHEN product_price = '7.25' THEN N'NoName'
ELSE product_name
END
WHEN product_name IS NULL AND TRIM(product_category) = N'Γλυκά'                         AND
product_price = '6.75'                       THEN N'Μακεδονικός Χαλβάς Κακάο 400g'
WHEN product_name IS NULL AND TRIM(product_category) = N'Είδη επάλειψης'                  AND
product_price = '6.75'                       THEN N'Μακεδονικός Χαλβάς Κακάο 400g' --wrong value found
WHEN product_name IS NULL AND TRIM(product_category) = N'Ετοιμες Σαλάτες'                 AND
product_price = '3.5'
WHEN product_name IS NULL AND TRIM(product_category) = N'Ζύμες, Πίτες & Πίτσες'           AND
product_price = '5.8'
WHEN product_name IS NULL AND TRIM(product_category) = N'Κρέατα & Πουλερικά'             AND
product_price = '28.05'
WHEN product_name IS NULL AND TRIM(product_category) = N'Κρουασάν & Τσουρέκια'          AND
product_price IN ('5.1','6.1')
WHEN product_name IS NULL AND TRIM(product_category) = N'Πατατάκια & Αλμυρά Snacks'    AND
product_price = '1.7'                         THEN N'Lay's Πατατάκια στο Φούρνο Γιαούρτι & Μυρωδικά 105g'
WHEN product_name IS NULL AND TRIM(product_category) = N'Πατατάκια & Γαριδάκια'          AND
product_price = '1.7'                         THEN N'Tasty Πιτσίνια 110g'
WHEN product_name IS NULL AND TRIM(product_category) = N'Προϊόντα Κρεάτων & Πουλερικών' AND
product_price IN ('4.62','5.39','6.16','6.94','9.91') THEN N'NoName'
WHEN product_name IS NULL AND TRIM(product_category) = N'Σοκολάτες' THEN
CASE
WHEN product_price = '0.5' THEN N'Παυλίδης Kiss Σοκολάτα Γάλακτος 27.5g'
WHEN product_price = '1.8' THEN N'Iov Break Σοκολάτα Γάλακτος με Φουντούκι 85g'
WHEN product_price = '3.6' THEN N'NoName'
ELSE product_name
END
```

```

WHEN product_name IS NULL AND TRIM(product_category) = N'Χοιρινό'          AND
product_price IN ('7.3')           THEN N'NoName'

WHEN product_name IS NULL AND TRIM(product_category) = N'Ψωμί'          AND
product_price IN ('1.6')           THEN N'Παπαδοπούλου Γεύση2 Ψωμί Τοστ Σίτου 700g'

WHEN product_name IS NULL AND product_category IS NULL THEN N'NoName'
ELSE product_name END AS product_name,
CASE WHEN product_category IS NULL THEN N'NoCategory' ELSE product_category END AS
product_category,
product_price,
items_sold,
geographical_region_of_user,
user_last_order_date,
CASE
WHEN f0_ IS NULL THEN (SELECT MIN(order_timestamp)   FROM
[Dissertation].[dbo].[Dataset] WHERE customer_id = 'c4b990d9f8cc112c80ee56e7c8a0ebf8')
ELSE f0_
END AS f0_
INTO #FilledNull
FROM [Dissertation].[dbo].[Dataset]

```

--Check if 'f0_' values are older than 'order_timestamp' values

```

SELECT *
FROM [Dissertation].[dbo].[Dataset]
WHERE f0_ > order_timestamp

```

--Control: Check columns for remaining null values (should be blank)

```

SELECT *
FROM #FilledNull
WHERE product_name IS NULL
OR product_category IS NULL
OR f0_ IS NULL

```

--Change the datetime columns format to remove milliseconds and form final dataset

```

DROP TABLE IF EXISTS [Dissertation].[dbo].[ProcessedDataset]
SELECT order_id,
customer_id,
FORMAT(order_timestamp, 'yyyy-MM-dd HH:mm:ss') as order_timestamp,
product_name,
product_category,
product_price,
items_sold,
geographical_region_of_user,
FORMAT(user_last_order_date, 'yyyy-MM-dd HH:mm:ss') as user_last_order_date,
FORMAT(f0_, 'yyyy-MM-dd HH:mm:ss') as f0_

```

```
INTO      [Dissertation].[dbo].[ProcessedDataset]
FROM      #FilledNull
```

```
--Identify rows where product_name is manually assigned as 'NoName'
SELECT      *
FROM      [Dissertation].[dbo].[ProcessedDataset]
WHERE      product_name='NoName'
ORDER BY    product_category
--91
```

```
--Identify possible free giveaways
SELECT      *
FROM      [Dissertation].[dbo].[ProcessedDataset]
WHERE      product_price = 0.0
--489, Πρώτο Θέμα
```

```
--Create a table with the first order dates per customer
DROP TABLEIF EXISTS #CustomerFirstOrders
CREATE TABLE #CustomerFirstOrders
(
    customer_id NVARCHAR(50),
    first_order_date DATE
)
```

```
--Get the first (minimum) order date for each customer
INSERT INTO#CustomerFirstOrders
SELECT      customer_id,
            MIN(CAST(order_timestamp AS DATE)) AS first_order_date
FROM      [Dissertation].[dbo].[ProcessedDataset]
GROUP BY  customer_id
```

```
--Count new customers by their first order month
DROP TABLEIF EXISTS [Dissertation].[dbo].[NewCustomers]
SELECT      FORMAT(first_order_date, 'yyyy-MM') AS first_order_month,
            COUNT(customer_id) AS new_customer_count
INTO      [Dissertation].[dbo].[NewCustomers]
FROM      #CustomerFirstOrders
GROUP BY  FORMAT(first_order_date, 'yyyy-MM')
ORDER BY  first_order_month
```

Υλοποίηση ανάλυσης με Python

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import skew
from scipy.stats import boxcox
from scipy.stats import zscore
from scipy.stats import shapiro, kstest, normaltest
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
from sklearn.metrics import silhouette_score
from sklearn.cluster import DBSCAN
from mlxtend.frequent_patterns import apriori, association_rules
from mlxtend.frequent_patterns import fpgrowth
import networkx as nx
from matplotlib import cm

In [2]:
# Load the dataset
file_path = 'c:\\\\Users\\\\User\\\\Desktop\\\\Dissertation\\\\Dataset\\\\customer segmentation & market
basket analysis_efm_ProCESSEDdataset.xlsx'
df = pd.read_excel(file_path, sheet_name="Dissertation.dbo.Dataset")

In []:
df.info()
In []:
df.head()
In []:
# Get descriptive statistics for numerical columns
print(df.describe())
In []:
# Print the dataframe
print(df)
In []:
# Display count of null values per column
null_summary = df.isnull().sum()
print(null_summary)
In []:
# Check how many 'NoName' values exist in 'product_name'
noname = (df['product_name'] == 'NoName').sum()
print(f"NoName' product names after imputation: {noname}")
In []:

```

```

# Number of unique values in categorical columns
print("Unique values in key columns:")
for col in ['product_name', 'product_category', 'customer_id', 'geographical_region_of_user']:
    print(f'{col}: {df[col].nunique()}')


In []:
# Heatmap to visualize correlation between numerical variables
plt.figure(figsize=(10, 5))
sns.heatmap(df[['product_price', 'items_sold']].corr(), annot=True, cmap='coolwarm', vmin=-1, vmax=1)
plt.title('Correlation Heatmap')
plt.show()


In []:
# Drop the redundant 'f0_' column
sales_data = df.drop(columns=['f0_'])

# Verify that zero-priced products match 'Πρώτο Θέμα'
zero_price_products = sales_data[sales_data['product_price'] == 0]
mismatched_zero_price = zero_price_products[
    (zero_price_products['product_name'] != 'Πρώτο Θέμα') |
    (zero_price_products['product_category'] != 'Πρώτο Θέμα')
]

if mismatched_zero_price.empty:
    print("All zero-priced products match 'Πρώτο Θέμα'.")
else:
    print("There are mismatched zero-price entries:")
    print(mismatched_zero_price)

# Display the dataset info after dropping the column
print(sales_data.info())
RFM model implementation

In []:
# Convert 'order_timestamp' to datetime if not already
sales_data['order_timestamp'] = pd.to_datetime(sales_data['order_timestamp'])

# Create a reference date for recency calculations
reference_date = sales_data['order_timestamp'].max()

# Calculate RFM metrics
rfm_metrics = sales_data.groupby('customer_id').agg(
    recency=('order_timestamp', lambda x: (reference_date - x.max()).days),
    frequency=('order_id', 'nunique'),
    monetary=('product_price', 'sum')
)

```

```

).reset_index()

# Preview the RFM metrics
print(rfm_metrics.head())

# Save the RFM metrics dataset to a new CSV file
output_file = "RFM metrics_dataset.csv"
rfm_metrics.to_csv(output_file, index=False, float_format='%.2f', decimal=',')
In [ ]:

# Box plot for RFM metrics
fig, axes = plt.subplots(1, 3, figsize=(15, 5), sharey=False)

# Box plots
metrics = ['recency', 'frequency', 'monetary']
titles = ['Recency', 'Frequency', 'Monetary']
for i, metric in enumerate(metrics):
    axes[i].boxplot(rfm_metrics[metric], vert=False)
    axes[i].set_title(titles[i])

plt.tight_layout()
plt.show()
In [ ]:

# Calculate z-scores for RFM metrics
rfm_z_scores = rfm_metrics[['recency', 'frequency', 'monetary']].apply(
    lambda x: (x - x.mean()) / x.std()
)

# Identify outliers
outliers = (rfm_z_scores.abs() > 3).any(axis=1)
print(f"Number of outliers detected: {outliers.sum()}")
In [ ]:

# Calculate RFM metrics directly from 'df'

# Convert 'order_timestamp' to datetime if it is not already in datetime format
df['order_timestamp'] = pd.to_datetime(df['order_timestamp'])

# Create a reference date for calculating Recency (e.g., the day after the last transaction in the dataset)
reference_date = df['order_timestamp'].max() + pd.Timedelta(days=1)

# Calculate RFM metrics
rfm_metrics = df.groupby('customer_id').agg(
    recency=('order_timestamp', lambda x: (reference_date - x.max()).days),
)

```

```

frequency=('order_id', 'nunique'),
monetary=('product_price', lambda x: (x * df.loc[x.index, 'items_sold']).sum())
).reset_index()

# Plot histograms and kernel density plots for the calculated RFM metrics
fig, axes = plt.subplots(1, 3, figsize=(18, 5))

# Plot Recency
sns.histplot(rfm_metrics['recency'], kde=True, ax=axes[0], color='blue', bins=20)
axes[0].set_title('Recency Distribution')
axes[0].set_xlabel('Recency (days)')

# Plot Frequency
sns.histplot(rfm_metrics['frequency'], kde=True, ax=axes[1], color='green', bins=20)
axes[1].set_title('Frequency Distribution')
axes[1].set_xlabel('Frequency (number of purchases)')

# Plot Monetary
sns.histplot(rfm_metrics['monetary'], kde=True, ax=axes[2], color='orange', bins=20)
axes[2].set_title('Monetary Distribution')
axes[2].set_xlabel('Monetary (revenue in €)')

plt.tight_layout()
plt.show()

In []:
# Calculate skewness for each RFM metric
recency_skewness = skew(rfm_metrics['recency'])
frequency_skewness = skew(rfm_metrics['frequency'])
monetary_skewness = skew(rfm_metrics['monetary'])

# Print skewness values
print(f'Recency Skewness: {recency_skewness:.2f}')
print(f'Frequency Skewness: {frequency_skewness:.2f}')
print(f'Monetary Skewness: {monetary_skewness:.2f}')

In []:
# Add 1 to all values to avoid issues with log(0)
rfm_metrics_log = rfm_metrics.copy()
rfm_metrics_log['recency'] = np.log1p(rfm_metrics['recency'])
rfm_metrics_log['frequency'] = np.log1p(rfm_metrics['frequency'])
rfm_metrics_log['monetary'] = np.log1p(rfm_metrics['monetary'])

# Visualize the distributions after logarithmic transformation
fig, axes = plt.subplots(1, 3, figsize=(18, 5))

```

```

# Plot Log-Transformed Recency
sns.histplot(rfm_metrics_log['recency'], kde=True, ax=axes[0], color='blue', bins=20)
axes[0].set_title('Log-Transformed Recency Distribution')
axes[0].set_xlabel('Log(Recency + 1)')

# Plot Log-Transformed Frequency
sns.histplot(rfm_metrics_log['frequency'], kde=True, ax=axes[1], color='green', bins=20)
axes[1].set_title('Log-Transformed Frequency Distribution')
axes[1].set_xlabel('Log(Frequency + 1)')

# Plot Log-Transformed Monetary
sns.histplot(rfm_metrics_log['monetary'], kde=True, ax=axes[2], color='orange', bins=20)
axes[2].set_title('Log-Transformed Monetary Distribution')
axes[2].set_xlabel('Log(Monetary + 1)')

plt.tight_layout()
plt.show()
In []:
# Calculate skewness_log for each RFM metric
recency_skewness_log = skew(rfm_metrics_log['recency'])
frequency_skewness_log = skew(rfm_metrics_log['frequency'])
monetary_skewness_log = skew(rfm_metrics_log['monetary'])

# Print skewness_log values
print(f'Recency Skewness: {recency_skewness_log:.2f}')
print(f'Frequency Skewness: {frequency_skewness_log:.2f}')
print(f'Monetary Skewness: {monetary_skewness_log:.2f}')
In []:
# Copy the original RFM metrics for Box-Cox transformation
rfm_metrics_boxcox = rfm_metrics.copy()

# Add a small constant to ensure all values are positive
rfm_metrics_boxcox['recency'], recency_lambda = boxcox(rfm_metrics['recency'] + 1)
rfm_metrics_boxcox['frequency'], frequency_lambda = boxcox(rfm_metrics['frequency'] + 1)
rfm_metrics_boxcox['monetary'], monetary_lambda = boxcox(rfm_metrics['monetary'] + 1)

# Visualize the distributions after Box-Cox transformation
fig, axes = plt.subplots(1, 3, figsize=(18, 5))

# Plot Box-Cox Transformed Recency
sns.histplot(rfm_metrics_boxcox['recency'], kde=True, ax=axes[0], color='blue', bins=20)
axes[0].set_title('Box-Cox Transformed Recency Distribution')

```

```

axes[0].set_xlabel('Box-Cox Recency')

# Plot Box-Cox Transformed Frequency
sns.histplot(rfm_metrics_boxcox['frequency'], kde=True, ax=axes[1], color='green', bins=20)
axes[1].set_title('Box-Cox Transformed Frequency Distribution')
axes[1].set_xlabel('Box-Cox Frequency')

# Plot Box-Cox Transformed Monetary
sns.histplot(rfm_metrics_boxcox['monetary'], kde=True, ax=axes[2], color='orange', bins=20)
axes[2].set_title('Box-Cox Transformed Monetary Distribution')
axes[2].set_xlabel('Box-Cox Monetary')

plt.tight_layout()
plt.show()

# Display the Box-Cox transformation parameters (lambdas)
print(f'Recency Lambda: {recency_lambda:.2f}')
print(f'Frequency Lambda: {frequency_lambda:.2f}')
print(f'Monetary Lambda: {monetary_lambda:.2f}')
In [ ]:

# Calculate z-scores for the Box-Cox transformed metrics
rfm_z_scores = rfm_metrics_boxcox[['recency', 'frequency', 'monetary']].apply(zscore)

# Identify outliers (absolute z-score > 3)
outliers = (rfm_z_scores.abs() > 3)

# Count the number of outliers in each metric
outlier_summary = outliers.sum()

# Visualize the Box-Cox transformed metrics with outliers using box plots
fig, axes = plt.subplots(1, 3, figsize=(18, 5))

# Box plots for transformed metrics
metrics = ['recency', 'frequency', 'monetary']
titles = ['Box-Cox Recency', 'Box-Cox Frequency', 'Box-Cox Monetary']
for i, metric in enumerate(metrics):
    sns.boxplot(x=rfm_metrics_boxcox[metric], ax=axes[i])
    axes[i].set_title(titles[i])

plt.tight_layout()
plt.show()

# Display the summary of outliers

```

```

print("Outlier Summary:")
print(outlier_summary)
In []:
# Perform normality tests on the Box-Cox transformed metrics
normality_results = {}

for metric in ['recency', 'frequency', 'monetary']:
    # Shapiro-Wilk Test
    shapiro_stat, shapiro_p = shapiro(rfm_metrics_boxcox[metric])

    # Kolmogorov-Smirnov Test
    ks_stat, ks_p = kstest(rfm_metrics_boxcox[metric], 'norm',
                           args=(rfm_metrics_boxcox[metric].mean(), rfm_metrics_boxcox[metric].std()))

    # D'Agostino and Pearson's Test
    dagostino_stat, dagostino_p = normaltest(rfm_metrics_boxcox[metric])

    # Store results
    normality_results[metric] = {
        'Shapiro-Wilk': {'statistic': shapiro_stat, 'p-value': shapiro_p},
        'Kolmogorov-Smirnov': {'statistic': ks_stat, 'p-value': ks_p},
        'D'Agostino-Pearson': {'statistic': dagostino_stat, 'p-value': dagostino_p}
    }

# Print the results
for metric, results in normality_results.items():
    print(f"\nNormality Test Results for {metric.capitalize()}:")
    for test, values in results.items():
        print(f" {test}: Statistic = {values['statistic']:.4f}, p-value = {values['p-value']:.4f}")
In []:
# Step 1: Standardize the Box-Cox transformed RFM metrics
scaler = StandardScaler()
rfm_scaled = scaler.fit_transform(rfm_metrics_boxcox[['recency', 'frequency', 'monetary']])

# Step 2: Implement K-Means using the Forgy method for initialization
# (Random initialization from the dataset)
kmeans = KMeans(n_clusters=3, init='random', random_state=42) # n_clusters can be adjusted as needed
kmeans.fit(rfm_scaled)

# Step 3: Add the cluster labels to the original dataset
rfm_metrics_boxcox['cluster'] = kmeans.labels_

```

```

# Step 4: Summarize the clustering results
# Inverse transform cluster centers to original scale
cluster_centers = scaler.inverse_transform(kmeans.cluster_centers_)
cluster_summary = pd.DataFrame(cluster_centers, columns=['recency', 'frequency', 'monetary'])
cluster_summary['cluster'] = range(1, len(cluster_summary) + 1)

# Output the results
print("Cluster Summary:")
print(cluster_summary)

# Save the dataset with clusters
rfm_metrics_boxcox.to_csv("rfm_with_clusters.csv", index=False)
In []:
kmeans = KMeans(n_clusters=3, init='random', random_state=42)
kmeans.fit(rfm_scaled)
kmeans_labels = kmeans.labels_

# Step 2: Reduce dimensionality to 2D using PCA ---Find source documentation for PCA
pca = PCA(n_components=2)
rfm_2d = pca.fit_transform(rfm_scaled)

# Step 3: Create a scatter plot of the clusters
plt.figure(figsize=(10, 7))
unique_labels = set(kmeans_labels)

for label in unique_labels:
    plt.scatter(
        rfm_2d[kmeans_labels == label, 0],
        rfm_2d[kmeans_labels == label, 1],
        s=50,
        label=f'Cluster {label}',
        alpha=0.6
    )

# Step 4: Customize the plot
plt.title('K-Means Clustering Results (k=3) in 2D PCA Projection', fontsize=16)
plt.xlabel('Principal Component 1', fontsize=12)
plt.ylabel('Principal Component 2', fontsize=12)
plt.legend(title='Clusters', fontsize=10, loc='best')
plt.grid(True)
plt.show()
In []:
# Elbow Method to Determine Optimal Number of Clusters
Customer Segmentation & Market Basket Analysis

```

```
wcss = [] # Within-Cluster Sum of Squares
k_values = range(1, 11) # Test k from 1 to 10

for k in k_values:
    kmeans = KMeans(n_clusters=k, init='random', random_state=42)
    kmeans.fit(rfm_scaled)
    wcss.append(kmeans.inertia_) # Append WCSS for each k

# Plot the Elbow Method graph
plt.figure(figsize=(10, 6))
plt.plot(k_values, wcss, marker='o', linestyle='--', color='b')
plt.title('Elbow Method for Optimal K')
plt.xlabel('Number of Clusters (k)')
plt.ylabel('Within-Cluster Sum of Squares (WCSS)')
plt.xticks(k_values)
plt.grid()
plt.show()
In [ ]:
# Silhouette Analysis for Optimal Number of Clusters
silhouette_scores = [] # Store silhouette scores for each k
k_values = range(2, 11) # Test k from 2 to 10

for k in k_values:
    kmeans = KMeans(n_clusters=k, init='random', random_state=42) # Use Forgy method
    cluster_labels = kmeans.fit_predict(rfm_scaled) # Predict cluster labels
    silhouette_avg = silhouette_score(rfm_scaled, cluster_labels) # Calculate Silhouette Score
    silhouette_scores.append(silhouette_avg)

# Plot the Silhouette Scores
plt.figure(figsize=(10, 6))
plt.plot(k_values, silhouette_scores, marker='o', linestyle='--', color='g')
plt.title('Silhouette Analysis for Optimal k')
plt.xlabel('Number of Clusters (k)')
plt.ylabel('Silhouette Score')
plt.xticks(k_values)
plt.grid()
plt.show()
In [ ]:
# Aligned Box Criterion (ABC) Implementation
def aligned_box_criterion(X, labels):
    """
    Aligned Box Criterion Implementation:
    X: Dataset
    
```

```

labels: Cluster labels
Returns: ABC Score
"""

unique_labels = np.unique(labels)
total_score = 0

for cluster in unique_labels:
    cluster_points = X[labels == cluster]
    box_min = cluster_points.min(axis=0)
    box_max = cluster_points.max(axis=0)
    box_volume = np.prod(box_max - box_min)

    cluster_score = len(cluster_points) / box_volume # Density of points within the box
    total_score += cluster_score

return total_score / len(unique_labels)

# Step 1: Run K-Means for k=3
kmeans_k3 = KMeans(n_clusters=3, init='random', random_state=42)
kmeans_k3.fit(rfm_scaled) # rfm_scaled contains the standardized RFM data

# Step 2: Calculate ABC Score for k=3
labels_k3 = kmeans_k3.labels_ # Get cluster labels
abc_score_k3 = aligned_box_criterion(rfm_scaled, labels_k3)

# Step 3: Print the ABC Score for k=3
print(f"Aligned Box Criterion (ABC) Score for k=3: {abc_score_k3:.4f}")

Implement DBSCAN algorithm
In []:
# Step 1: Implement DBSCAN

# Parameters: eps (neighborhood size) and min_samples (minimum points to form a cluster)
eps_value = 0.3 # Set the desired eps value
min_samples_value = 5 # Set the desired min_samples value
dbscan = DBSCAN(eps=eps_value, min_samples=min_samples_value) # Adjust these values as needed for better results
dbscan.fit(rfm_scaled) # rfm_scaled contains the standardized RFM metrics

# Step 2: Extract cluster labels
dbscan_labels = dbscan.labels_

# Step 3: Count the number of clusters (ignoring noise points labeled as -1)
n_clusters_dbscan = len(set(dbscan_labels)) - (1 if -1 in dbscan_labels else 0)

```

```

n_noise_points = list(dbSCAN_labels).count(-1)

# Step 4: Compute silhouette score (only if more than 1 cluster is formed)
if n_clusters_dbSCAN > 1:
    silhouette_avg_dbSCAN = silhouette_score(rfm_scaled, dbSCAN_labels)
else:
    silhouette_avg_dbSCAN = "Not Applicable (only one cluster formed)"

# Step 5: Append DBSCAN cluster labels to the dataset
rfm_metrics_boxcox['dbSCAN_cluster'] = dbSCAN_labels

# Display the DBSCAN results summary with parameters
dbSCAN_results = {
    "Eps (Neighborhood Size)": eps_value,
    "Min Samples": min_samples_value,
    "Number of Clusters (excluding noise)": n_clusters_dbSCAN,
    "Number of Noise Points": n_noise_points,
    "Silhouette Score": silhouette_avg_dbSCAN,
}
print("DBSCAN Clustering Results:")
print(pd.DataFrame([dbSCAN_results]).to_string())

# Save the dataset with DBSCAN cluster labels for further analysis
rfm_metrics_boxcox.to_csv("rfm_with_dbSCAN_clusters.csv", index=False)
In []:
# Step 1: Reduce dimensionality to 2D using PCA
pca = PCA(n_components=2)
rfm_2d = pca.fit_transform(rfm_scaled)

# Step 2: Create a scatter plot of the clusters
plt.figure(figsize=(10, 7))
unique_labels = np.unique(dbSCAN_labels)

for label in unique_labels:
    if label == -1: # Noise points
        plt.scatter(
            rfm_2d[dbSCAN_labels == label, 0],
            rfm_2d[dbSCAN_labels == label, 1],
            s=50,
            c='red',
            label='Noise',
            alpha=0.6
        )

```

```

else:
    plt.scatter(
        rfm_2d[dbSCAN_labels == label, 0],
        rfm_2d[dbSCAN_labels == label, 1],
        s=50,
        label=f'Cluster {label}',
        alpha=0.6
    )

# Step 3: Customize plot
plt.title('DBSCAN Clustering Results (2D PCA Projection)', fontsize=16)
plt.xlabel('Principal Component 1', fontsize=12)
plt.ylabel('Principal Component 2', fontsize=12)
plt.legend(title='Clusters', fontsize=10, loc='best')
plt.grid(True)
plt.show()
In []:
# Define a function to visualize DBSCAN results
def visualize_dbSCAN_clusters(rfm_scaled, labels, eps, min_samples):
    from sklearn.decomposition import PCA
    pca = PCA(n_components=2)
    rfm_2d = pca.fit_transform(rfm_scaled)

    unique_labels = np.unique(labels)
    plt.figure(figsize=(10, 7))

    for label in unique_labels:
        if label == -1: # Noise
            plt.scatter(rfm_2d[labels == label, 0], rfm_2d[labels == label, 1],
                        s=50, c='red', label='Noise', alpha=0.6)
        else:
            plt.scatter(rfm_2d[labels == label, 0], rfm_2d[labels == label, 1],
                        s=50, label=f'Cluster {label}', alpha=0.6)

    plt.title(f'DBSCAN Clustering Results (eps={eps}, min_samples={min_samples})', fontsize=16)
    plt.xlabel('Principal Component 1', fontsize=12)
    plt.ylabel('Principal Component 2', fontsize=12)
    plt.legend(title='Clusters', fontsize=10, loc='best')
    plt.grid(True)
    plt.show()

# Grid search for eps and min_samples
eps_values = [0.3, 0.4, 0.5, 0.6] # Test smaller eps values
Customer Segmentation & Market Basket Analysis

```

```

min_samples_values = [3, 5, 7] # Test stricter min_samples

for eps in eps_values:
    for min_samples in min_samples_values:
        dbSCAN = DBSCAN(eps=eps, min_samples=min_samples)
        dbSCAN.fit(rfm_scaled)
        labels = dbSCAN.labels_

        # Count clusters (excluding noise)
        n_clusters = len(set(labels)) - (1 if -1 in labels else 0)
        n_noise = list(labels).count(-1)

        print(f"eps={eps}, min_samples={min_samples}: {n_clusters} clusters, {n_noise} noise points")

    # Visualize if 3 clusters are formed
    if n_clusters == 3:
        visualize_dbSCAN_clusters(rfm_scaled, labels, eps, min_samples)

After having examined K-means and DBSCAN results, we decided to keep K-means output. Now we will analyze the 3 created clusters

In []:
# Step 1: Standardize the RFM metrics
scaler = StandardScaler()
rfm_scaled = scaler.fit_transform(rfm_metrics_boxcox[['recency', 'frequency', 'monetary']])

# Step 2: Implement K-Means using the Forgy method for initialization
kmeans = KMeans(n_clusters=3, init='random', random_state=42) # Adjust n_clusters if needed
kmeans.fit(rfm_scaled)

# Step 3: Add the cluster labels to the original dataset
rfm_metrics_boxcox['cluster'] = kmeans.labels_

# Step 4: Summarize the clustering results
# Inverse transform cluster centers to the original scale
cluster_centers = scaler.inverse_transform(kmeans.cluster_centers_)
cluster_summary = pd.DataFrame(cluster_centers, columns=['recency', 'frequency', 'monetary'])
cluster_summary['cluster'] = range(1, len(cluster_summary) + 1)

# Step 5: Add customer count and percentage
cluster_counts = rfm_metrics_boxcox['cluster'].value_counts().sort_index()
cluster_summary['customer_count'] = cluster_counts.values
cluster_summary['percentage'] = (cluster_counts.values / cluster_counts.sum()) * 100

```

```

# Step 6: Apply conditional formatting
# Calculate averages
average_recency = cluster_summary['recency'].mean()
average_frequency = cluster_summary['frequency'].mean()
average_monetary = cluster_summary['monetary'].mean()

# Add conditional formatting (as an additional column for display)
cluster_summary['recency_color'] = cluster_summary['recency'].apply(
    lambda x: 'green' if x < average_recency else 'red'
)
cluster_summary['frequency_color'] = cluster_summary['frequency'].apply(
    lambda x: 'green' if x > average_frequency else 'red'
)
cluster_summary['monetary_color'] = cluster_summary['monetary'].apply(
    lambda x: 'green' if x > average_monetary else 'red'
)

# Round RFM variables to 2 decimal places for readability
cluster_summary[['recency', 'frequency', 'monetary']] = cluster_summary[['recency', 'frequency',
'monetary']].round(2)

# Step 7: Save or display the cluster summary table
cluster_summary_file = "cluster_summary.csv"
cluster_summary.to_csv(cluster_summary_file, index=False)

# Print confirmation
print(f"Cluster summary saved to {cluster_summary_file}")
In []:
# Adjust cluster numbering to start from 1 instead of 0
rfm_metrics_boxcox['cluster_label'] = rfm_metrics_boxcox['cluster'] + 1

# Round RFM values to two decimal places for better readability
rfm_metrics_boxcox[['recency', 'frequency', 'monetary']] = rfm_metrics_boxcox[['recency',
'frequency', 'monetary']].round(2)

# Merge the corrected RFM metrics and clusters back into the original dataset
final_dataset = pd.merge(
    df,
    rfm_metrics_boxcox[['customer_id', 'recency', 'frequency', 'monetary', 'cluster_label']],
    on='customer_id',
    how='left'
)

```

```

# Save the CSV file with a comma as the decimal separator
output_file = "full_dataset_with_clusters_corrected.csv"
final_dataset.to_csv(output_file, index=False, float_format='%.2f', decimal=',')

# Print confirmation and preview of the corrected dataset
print(f"Updated dataset saved with corrected cluster labels and rounded RFM values to {output_file}")
print(final_dataset.head().to_string())
In []:
# Retain the original RFM values
rfm_metrics_original = rfm_metrics.copy()
rfm_metrics_original.columns = ['customer_id', 'recency_original', 'frequency_original', 'monetary_original']

# Add cluster_label column, starting from 1 instead of 0
rfm_metrics_boxcox['cluster_label'] = rfm_metrics_boxcox['cluster'] + 1

# Merge the original RFM metrics with the Box-Cox transformed data
rfm_combined = pd.merge(
    rfm_metrics_original,
    rfm_metrics_boxcox[['customer_id', 'recency', 'frequency', 'monetary', 'cluster_label']],
    on='customer_id',
    how='left'
)
# Merge the combined RFM data with the original dataset
final_dataset_corrected = pd.merge(
    df,
    rfm_combined,
    on='customer_id',
    how='left'
)
# Save the corrected dataset
output_file = "full_dataset_with_original_and_transformed_RFMs.csv"
final_dataset_corrected.to_csv(output_file, index=False, float_format='%.2f', decimal=',')
# Print confirmation and preview of the corrected dataset
print(f"Final dataset saved to {output_file}")
print(final_dataset_corrected.head().to_string())
In []:
# Adjust pandas display settings to show all columns in one line
pd.set_option('display.width', None) # Disable width limitation

```

```

pd.set_option('display.max_columns', None) # Show all columns
pd.set_option('display.max_rows', None) # Ensure all rows are displayed
pd.set_option('display.max_colwidth', None) # Prevent column text from being truncated

# Read the dataset
file_path = 'c:\\Users\\User\\Desktop\\Dissertation\\RFM_FinalDataset.csv'
fd = pd.read_csv(file_path)

# Display the first 5 rows of the dataframe, showing all columns in one line
print(fd.head().to_string(index=False)) # Use to_string() to display in one line without index
In []:
fd.info()
In [48]:
# Create a transaction dataset with 'order_id' and 'product_name'
transactions = fd.groupby('order_id')['product_name'].apply(list).reset_index(name='products')

# One-hot encode the transaction data
from mlxtend.preprocessing import TransactionEncoder

te = TransactionEncoder()
te_ary = te.fit(transactions['products']).transform(transactions['products'])
df_encoded = pd.DataFrame(te_ary, columns=te.columns_)
Apriori algorithm
The following code cell is commented out because it cannot run due to memory restriction
In [49]:
# frequent_itemsets_apriori = apriori(df_encoded, min_support=0.003, use_colnames=True)

# # Check if itemsets are generated
# if not frequent_itemsets_apriori.empty:
#     # Generate association rules using num_itemsets
#     rules_apriori = association_rules(frequent_itemsets_apriori, metric="confidence",
#                                         min_threshold=0.1, num_itemsets=len(frequent_itemsets_apriori))
#     print(rules_apriori.head().to_string())
# else:
#     print("No frequent itemsets found. Try lowering the support threshold.")
We reduce the dataset size by focusing on top 100 frequently purchased products to save
memory:
In []:
top_products = fd['product_name'].value_counts().head(100).index
filtered_data = fd[fd['product_name'].isin(top_products)]

# Create transactions and re-encode

```

```

transactions_filtered = filtered_data.groupby('order_id')['product_name'].apply(list).reset_index(name='products')
te_ary = te.fit(transactions_filtered['products']).transform(transactions_filtered['products'])
df_encoded_filtered = pd.DataFrame(te_ary, columns=te.columns_)

# Run Apriori on the reduced dataset
frequent_itemsets_filtered = apriori(df_encoded_filtered, min_support=0.001,
use_colnames=True)
rules_filtered = association_rules(frequent_itemsets_filtered, min_threshold=0.005, num_itemsets=len(frequent_itemsets_filtered))
print(rules_filtered.to_string())
Explore findings:
In []:
# Display top 10 itemsets with higher support
print(frequent_itemsets_filtered.sort_values(by='support', ascending=False).head(10))
In []:
#Display all itemsets with support > 0.02
high_support_itemsets = frequent_itemsets_filtered[frequent_itemsets_filtered['support'] > 0.02]
print(high_support_itemsets)
In []:
# How many transactions contain more than one product
transactions['num_products'] = transactions['products'].apply(len)
print(transactions['num_products'].value_counts())
Check the generated rules
In []:
print(rules_filtered.to_string())
Descriptive stats on generated rules
In []:
print(f"Number of rules: {len(rules_filtered)}")
In []:
# Display generated rules with lift > 2
high_lift_rules = rules_filtered[rules_filtered['lift'] > 2]
print(f"Number of high-lift rules: {len(high_lift_rules)}")
print(high_lift_rules.to_string())
In []:
# Coverage Across Product Categories:
unique_products = set()
for rule in rules_filtered['antecedents']:
    unique_products.update(rule)
for rule in rules_filtered['consequents']:
    unique_products.update(rule)
print(f"Number of unique products in rules: {len(unique_products)}")
In []:

```

```

# Higher support means the rule applies to a larger portion of transactions. Should have a balance of high-support and niche (low-support) rules
print(rules_filtered['support'].describe())
In [ ]:
# Confidence indicates the probability that the consequent is purchased when the antecedent is purchased
print(rules_filtered['confidence'].describe())
In [ ]:
# Display rules with Lift > 2
high_lift_rules = rules_filtered[rules_filtered['lift'] > 2]
print(high_lift_rules.to_string())
In [ ]:
rules_filtered['antecedent_category'] = rules_filtered['antecedents'].apply(lambda x: ', '.join(list(x)))
print(rules_filtered[['antecedent_category', 'consequents', 'support', 'confidence', 'lift']].head(10).to_string())
Visualize top 50 generated rules by descending lift in a table format
In [ ]:
# Extract relevant columns
table_data = rules_filtered[['antecedents', 'consequents', 'support', 'confidence', 'lift']]

# Convert sets of items into strings for readability
table_data['antecedents'] = table_data['antecedents'].apply(lambda x: ', '.join(list(x)))
table_data['consequents'] = table_data['consequents'].apply(lambda x: ', '.join(list(x)))

# Set rule depiction
table_data['rule'] = table_data['antecedents'] + ' → ' + table_data['consequents']

# Calculate the number of times a rule has been applied (absolute support count)
table_data['count'] = (table_data['support'] * len(fd['order_id'].unique())).astype(int)

# Change column order to display the count first
table_data = table_data[['count', 'support', 'confidence', 'lift', 'rule']]

# Sort by lift in descending order
table_data = table_data.sort_values(by='lift', ascending=False)

# Filter the top 50 rules by lift
top_50_rules = table_data.sort_values(by='lift', ascending=False).head(50)

# Convert to string format for display
fig, ax = plt.subplots(figsize=(12, 8))
ax.axis('tight')
ax.axis('off')

```

```

# Create a table
table = ax.table(cellText=top_50_rules.values,
                 colLabels=top_50_rules.columns,
                 cellLoc='center',
                 loc='center')

# Adjust table font size
table.auto_set_font_size(False)
table.set_fontsize(10)
table.auto_set_column_width(col=list(range(len(top_50_rules.columns)))))

plt.show()

Let's analyze the dataset to ensure it includes enough multi-item transactions:
In []:
# Filter rules where the antecedents contain more than one item
multi_antecedent_rules = rules_filtered[rules_filtered['antecedents'].apply(lambda x: len(x) > 1)]
print(multi_antecedent_rules[['antecedents', 'consequents', 'support', 'confidence', 'lift']].to_string())

Market Basket Analysis (MBA) separately for each customer cluster:
In []:
# Function to filter top products, apply Apriori, and generate rules
def cluster_mba_analysis(cluster_data, top_n=100, min_support=0.001, min_confidence=0.005):
    # Filter top N products
    top_products = cluster_data['product_name'].value_counts().head(top_n).index
    filtered_data = cluster_data[cluster_data['product_name'].isin(top_products)]

    # Create transactions and re-encode
    transactions_filtered = filtered_data.groupby('order_id')['product_name'].apply(list).reset_index(name='products')
    te_ary = te.fit(transactions_filtered['products']).transform(transactions_filtered['products'])
    df_encoded_filtered = pd.DataFrame(te_ary, columns=te.columns_)

    # Run Apriori on the reduced dataset
    frequent_itemsets_filtered = apriori(df_encoded_filtered, min_support=min_support, use_colnames=True)
    rules_filtered = association_rules(frequent_itemsets_filtered, metric="confidence", min_threshold=min_confidence, num_itemsets=len(frequent_itemsets_filtered))

    return rules_filtered

# Separate data by clusters
loyalists_data = fd[fd['cluster_label'] == 1]

```

```

potential_loyalists_data = fd[fd['cluster_label'] == 3]
churners_data = fd[fd['cluster_label'] == 2]

# Apply MBA for each cluster
print("Running MBA for Loyalists...")
loyalists_rules = cluster_mba_analysis(loyalists_data)
print("Loyalists Rules:\n", loyalists_rules)

print("Running MBA for Potential Loyalists...")
potential_loyalists_rules = cluster_mba_analysis(potential_loyalists_data)
print("Potential Loyalists Rules:\n", potential_loyalists_rules)

print("Running MBA for Churners...")
churners_rules = cluster_mba_analysis(churners_data)
print("Churners Rules:\n", churners_rules)
In []:
# Function to process rules for visualization
def visualize_cluster_rules(rules_filtered, cluster_name):
    # Extract relevant columns
    table_data = rules_filtered[['antecedents', 'consequents', 'support', 'confidence', 'lift']]

    # Convert sets of items into strings for readability
    table_data['antecedents'] = table_data['antecedents'].apply(lambda x: ', '.join(list(x)))
    table_data['consequents'] = table_data['consequents'].apply(lambda x: ', '.join(list(x)))

    # Set rule depiction
    table_data['rule'] = table_data['antecedents'] + ' → ' + table_data['consequents']

    # Calculate the number of times a rule has been applied (absolute support count)
    table_data['count'] = (table_data['support'] * len(fd['order_id'].unique())).astype(int)

    # Change column order to display the count first
    table_data = table_data[['count', 'support', 'confidence', 'lift', 'rule']]

    # Sort by lift in descending order
    table_data = table_data.sort_values(by='lift', ascending=False)

    # Filter the top 50 rules by lift
    top_50_rules = table_data.head(50)

    # Visualize the table
    fig, ax = plt.subplots(figsize=(12, 8))
    ax.axis('tight')

```

```

ax.axis('off')

# Create a table
table = ax.table(cellText=top_50_rules.values,
                  colLabels=top_50_rules.columns,
                  cellLoc='center',
                  loc='center')

# Adjust table font size
table.auto_set_font_size(False)
table.set_fontsize(10)
table.auto_set_column_width(col=list(range(len(top_50_rules.columns)))))

# Add a title above the table
plt.suptitle(f"Top 50 Rules for {cluster_name} (Apriori)", fontsize=14, y=1.05) # Adjust 'y' for vertical placement
plt.show()

# Visualize rules for each cluster
print("Visualizing rules for Loyalists...")
visualize_cluster_rules(loyalists_rules, "Loyalists")

print("Visualizing rules for Potential Loyalists...")
visualize_cluster_rules(potential_loyals_rules, "Potential Loyalists")

print("Visualizing rules for Churners...")
visualize_cluster_rules(churners_rules, "Churners")
We apply analysis to the top 10 products for computational reasons
In []:
# Step 1: Filter to the top 10 products
top_products = fd['product_name'].value_counts().head(10).index
filtered_data = fd[fd['product_name'].isin(top_products)]=

# Step 2: Create transactions for filtered data
transactions
filtered_data.groupby('order_id')['product_name'].apply(list).reset_index(name='products')

# Step 3: One-hot encode the filtered transactions
te = TransactionEncoder()
te_ary = te.fit(transactions['products']).transform(transactions['products'])
df_encoded = pd.DataFrame(te_ary, columns=te.columns_)

# Step 4: Apply Apriori to generate frequent itemsets

```

```

frequent_itemsets = apriori(df_encoded, min_support=0.001, use_colnames=True)

# Step 5: Generate association rules
rules      = association_rules(frequent_itemsets,      metric="lift",      min_threshold=1,
num_itemsets=len(frequent_itemsets))

# Step 6: Map antecedents and consequents to product categories
product_categories = filtered_data[['product_name', 'product_category']].drop_duplicates()

rules['antecedent_categories'] = rules['antecedents'].apply(
    lambda x: ', '.join(set(product_categories.loc[product_categories['product_name'].isin(x),
'product_category'])))
)
rules['consequent_categories'] = rules['consequents'].apply(
    lambda x: ', '.join(set(product_categories.loc[product_categories['product_name'].isin(x),
'product_category'])))
)

# Step 7: Group by categories and calculate the mean lift for each category pair
category_lift           = rules.groupby(['antecedent_categories',
'consequent_categories'])['lift'].mean().unstack(fill_value=0)

# Step 8: Create a mask for the upper triangle
mask = np.triu(np.ones_like(category_lift, dtype=bool))

# Step 9: Plot the heatmap
plt.figure(figsize=(12, 10))
sns.heatmap(category_lift,      mask=mask,      annot=True,      fmt=".2f",      cmap='coolwarm',
cbar_kws={'label': 'Lift'})
plt.title('Heatmap of Top 10 Product Category Pair Lift (Lower Triangle)', fontsize=16)
plt.xlabel('Consequent Categories', fontsize=12)
plt.ylabel('Antecedent Categories', fontsize=12)
plt.xticks(rotation=45, ha='right')
plt.yticks(rotation=0)
plt.tight_layout()
plt.show()
In [ ]:

# Grouping by cluster and product category, then counting the purchases
category_counts           = fd.groupby(['cluster_label',
'product_category'])['order_id'].count().reset_index()
category_counts.columns = ['Customer Cluster', 'Product Category', 'Count']

# Calculating the total purchases per cluster
total_counts = category_counts.groupby('Customer Cluster')['Count'].sum().reset_index()

```

```

total_counts.columns = ['Customer Cluster', 'Total Purchases']

# Merging to calculate percentage of total purchases for each category within each cluster
category_counts = pd.merge(category_counts, total_counts, on='Customer Cluster')
category_counts['Percentage of Total Purchases'] = (category_counts['Count'] / category_counts['Total Purchases']) * 100

# Filtering top 5 product categories across all clusters
top_categories = category_counts.groupby('Product Category')['Count'].sum().nlargest(5).index
filtered_data = category_counts[category_counts['Product Category'].isin(top_categories)]

# Rename clusters for clarity
cluster_mapping = {1: 'Loyalists', 3: 'Potential Loyalists', 2: 'Churners'}
filtered_data['Customer Cluster'] = filtered_data['Customer Cluster'].map(cluster_mapping)

# Create the bar plot
plt.figure(figsize=(12, 8))
ax = sns.barplot(x='Product Category', y='Percentage of Total Purchases', hue='Customer Cluster', data=filtered_data)

# Add data labels
for p in ax.patches:
    ax.annotate(f'{p.get_height():.1f}%', (p.get_x() + p.get_width() / 2., p.get_height()),
                ha='center', va='baseline', fontsize=10, color='black', xytext=(0, 5),
                textcoords='offset points')

# Set plot title and labels
plt.title('Cluster-Specific Comparison of Top 5 Product Categories', fontsize=16)
plt.xlabel('Product Category', fontsize=14)
plt.ylabel('Percentage of Total Purchases (%)', fontsize=14)
plt.xticks(rotation=45, ha='right')
plt.legend(title='Customer Cluster', fontsize=12)
plt.tight_layout()

# Show the plot
plt.show()
In []:
# Filter the rules to select top N associations by lift
top_n = 20 # You can adjust this value
filtered_rules = rules.sort_values(by='lift', ascending=False).head(top_n)

```

```

# Create a directed graph
G = nx.DiGraph()

# Add edges to the graph based on the filtered rules
for _, rule in filtered_rules.iterrows():
    antecedents = ', '.join(list(rule['antecedents']))
    consequents = ', '.join(list(rule['consequents']))
    lift = rule['lift']
    G.add_edge(antecedents, consequents, weight=lift)

# Set graph layout with increased spacing
pos = nx.spring_layout(G, seed=42, k=1) # Adjust k for spacing

# Calculate node degrees and edge weights for scaling
node_sizes = [len(list(G.neighbors(n))) * 500 for n in G.nodes()]
edge_weights = [d['weight'] for _, _, d in G.edges(data=True)]

# Normalize edge weights for color mapping
norm_weights = np.array(edge_weights)
norm_weights = (norm_weights - norm_weights.min()) / (norm_weights.max() - norm_weights.min())

# Draw nodes
plt.figure(figsize=(16, 12))
nx.draw_networkx_nodes(G, pos, node_size=node_sizes, node_color='skyblue', alpha=0.9)

# Draw edges with color mapping
edges = nx.draw_networkx_edges(
    G, pos, edge_color=norm_weights, edge_cmap=cm.Blues, width=[w for w in edge_weights],
    arrowsize=2
)

# Draw labels with optimized font size
nx.draw_networkx_labels(G, pos, font_size=10, font_weight='normal')

# Add edge labels for lift values
edge_labels = {(u, v): f'{d["weight"]:.2f}' for u, v, d in G.edges(data=True)}
nx.draw_networkx_edge_labels(G, pos, edge_labels=edge_labels, font_size=10, label_pos=0.5)

# Add a colorbar for edge weights
sm = plt.cm.ScalarMappable(cmap=cm.Blues, norm=plt.Normalize(vmin=norm_weights.min(),
vmax=norm_weights.max()))
sm.set_array([])
```

```

cbar = plt.colorbar(sm)
cbar.set_label('Lift Value (Normalized)', fontsize=12)

# Add a title and display the graph
plt.title('Network Graph of Product Associations (Top 20 by Lift)', fontsize=16)
plt.axis('off')
plt.tight_layout()
plt.show()

FP-Growth implementation

In []:
# Ensure dataset is in transactional format (assuming 'fd' is the dataset)
# Transform the dataset to a transactional format suitable for FP-Growth
basket = fd.groupby(['order_id', 'product_name'])['items_sold'].sum().unstack().fillna(0)
basket = basket.applymap(lambda x: 1 if x > 0 else 0)

# Apply FP-Growth algorithm to generate frequent itemsets
min_support = 0.001 # Adjust this value if needed
frequent_itemsets = fpgrowth(basket, min_support=min_support, use_colnames=True)

# Generate association rules from frequent itemsets
rules = association_rules(frequent_itemsets, metric="lift", min_threshold=1,
                           num_itemsets=len(frequent_itemsets))

# Sort the rules by lift in descending order
sorted_rules = rules.sort_values(by='lift', ascending=False)

# Display the top 20 rules for a quick preview
print(sorted_rules.head(20).to_string())

# Save the rules to a CSV file (optional)
sorted_rules.to_csv('fp_growth_rules.csv', index=False)

In []:
print(f"Number of rules: {len(sorted_rules)}")
In []:
# Display generated rules with lift > 2
high_lift_rules = sorted_rules[sorted_rules['lift'] > 100]
print(f"Number of high-lift rules (>100): {len(high_lift_rules)}")
print(high_lift_rules.to_string())

In []:
# Coverage Across Product Categories:
unique_products = set()
for rule in sorted_rules['antecedents']:
    unique_products.update(rule)

```

```

for rule in sorted_rules['consequents']:
    unique_products.update(rule)
print(f"Number of unique products in rules: {len(unique_products)}")
In [ ]:
# Higher support means the rule applies to a larger portion of transactions. Should have a balance of high-support and niche (low-support) rules
print(sorted_rules['support'].describe())
In [ ]:
# Confidence indicates the probability that the consequent is purchased when the antecedent is purchased
print(sorted_rules['confidence'].describe())
In [ ]:
# Extract relevant columns from the FP-Growth rules
fp_growth_table_data = rules[['antecedents', 'consequents', 'support', 'confidence', 'lift']]

# Convert sets of items into strings for readability
fp_growth_table_data['antecedents'] = fp_growth_table_data['antecedents'].apply(lambda x: ', '.join(list(x)))
fp_growth_table_data['consequents'] = fp_growth_table_data['consequents'].apply(lambda x: ', '.join(list(x)))

# Set rule depiction
fp_growth_table_data['rule'] = fp_growth_table_data['antecedents'] + ' → ' + fp_growth_table_data['consequents']

# Calculate the number of times a rule has been applied (absolute support count)
fp_growth_table_data['count'] = (fp_growth_table_data['support'] * len(fd['order_id'].unique())).astype(int)

# Change column order to display the count first
fp_growth_table_data = fp_growth_table_data[['count', 'support', 'confidence', 'lift', 'rule']]

# Sort by lift in descending order
fp_growth_table_data = fp_growth_table_data.sort_values(by='lift', ascending=False)

# Filter the top 50 rules by lift
fp_growth_top_50_rules = fp_growth_table_data.head(50)

# Convert to string format for display
fig, ax = plt.subplots(figsize=(12, 8))
ax.axis('tight')
ax.axis('off')

# Create a table

```

```

table = ax.table(cellText=fp_growth_top_50_rules.values,
                 colLabels=fp_growth_top_50_rules.columns,
                 cellLoc='center',
                 loc='center')

# Adjust table font size
table.auto_set_font_size(False)
table.set_fontsize(10)
table.auto_set_column_width(col=list(range(len(fp_growth_top_50_rules.columns)))))

plt.show()
In []:
print(fp_growth_top_50_rules.sort_values(by='lift',
                                         ascending=False).head(20).to_string(index=False))
In []:
# Function to preprocess and apply FP-Growth for a specific cluster
def apply_fp_growth_for_cluster(cluster_data, min_support=0.001):
    # Ensure dataset is in transactional format for the cluster
    basket_cluster = cluster_data.groupby(['order_id',
                                           'product_name'])['items_sold'].sum().unstack().fillna(0)
    basket_cluster = basket_cluster.applymap(lambda x: 1 if x > 0 else 0)

    # Apply FP-Growth algorithm
    frequent_itemsets_cluster = fpgrowth(basket_cluster, min_support=min_support,
                                           use_colnames=True)

    # Generate association rules
    rules_cluster = association_rules(frequent_itemsets_cluster, metric="lift", min_threshold=1,
                                       num_itemsets=len(frequent_itemsets))

    return rules_cluster

# Apply FP-Growth to each cluster
loyalists_data = fd[fd['cluster_label'] == 1]
potential_loyals_data = fd[fd['cluster_label'] == 3]
churners_data = fd[fd['cluster_label'] == 2]

loyalists_rules = apply_fp_growth_for_cluster(loyalists_data)
potential_loyals_rules = apply_fp_growth_for_cluster(potential_loyals_data)
churners_rules = apply_fp_growth_for_cluster(churners_data)

# Function to visualize the top 50 rules for each cluster
def visualize_fp_growth_cluster_rules(rules_filtered, cluster_name):
    # Extract relevant columns

```

```

fp_growth_table_data = rules_filtered[['antecedents', 'consequents', 'support', 'confidence', 'lift']]
fp_growth_table_data['antecedents'] = fp_growth_table_data['antecedents'].apply(lambda x: ', '.join(list(x)))
fp_growth_table_data['consequents'] = fp_growth_table_data['consequents'].apply(lambda x: ', '.join(list(x)))
fp_growth_table_data['rule'] = fp_growth_table_data['antecedents'] + ' → ' + fp_growth_table_data['consequents']
fp_growth_table_data['count'] = (fp_growth_table_data['support'] * len(fd['order_id'].unique())).astype(int)
fp_growth_table_data = fp_growth_table_data[['count', 'support', 'confidence', 'lift', 'rule']]
fp_growth_table_data = fp_growth_table_data.sort_values(by='lift', ascending=False)
top_50_rules = fp_growth_table_data.head(50)

# Visualize the table
fig, ax = plt.subplots(figsize=(12, 8))
ax.axis('tight')
ax.axis('off')
table = ax.table(cellText=top_50_rules.values,
                  colLabels=top_50_rules.columns,
                  cellLoc='center',
                  loc='center')
table.auto_set_font_size(False)
table.set_fontsize(10)
table.auto_set_column_width(col=list(range(len(top_50_rules.columns))))
plt.suptitle(f"Top 50 Rules for {cluster_name} (FP-Growth)", fontsize=14, y=1.05)
plt.show()

# Visualize FP-Growth rules for each cluster
print("Visualizing FP-Growth rules for Loyalists...")
visualize_fp_growth_cluster_rules(loyalists_rules, "Loyalists")

print("Visualizing FP-Growth rules for Potential Loyalists...")
visualize_fp_growth_cluster_rules(potential_loyals_rules, "Potential Loyalists")

print("Visualizing FP-Growth rules for Churners...")
visualize_fp_growth_cluster_rules(churners_rules, "Churners")
In []:
# Generate heatmap data from FP-Growth rules
# category_lift = rules.groupby(['antecedent_categories', 'consequent_categories'])['lift'].mean().unstack(fill_value=0)

# Create a mask for the upper triangle
mask = np.triu(np.ones_like(category_lift, dtype=bool))

```

```

# Set up the matplotlib figure
plt.figure(figsize=(12, 10))

# Draw the heatmap with the mask
sns.heatmap(
    category_lift,
    mask=mask,
    cmap='coolwarm',
    annot=True,
    fmt=".2f",
    cbar_kws={'label': 'Lift'},
    linewidths=0.5
)

# Customize the plot
plt.title('Heatmap of Top 10 Product Category Pair Lift', fontsize=16)
plt.xlabel('Consequent Categories', fontsize=12)
plt.ylabel('Antecedent Categories', fontsize=12)
plt.xticks(rotation=45, ha='right')
plt.yticks(rotation=0)
plt.tight_layout()

# Show the plot
plt.show()
In []:
# Grouping by cluster and product category, then counting the purchases
category_counts_fp = fd.groupby(['cluster_label',
'product_category'])['order_id'].count().reset_index()
category_counts_fp.columns = ['Customer Cluster', 'Product Category', 'Count']

# Calculating the total purchases per cluster
total_counts_fp = category_counts_fp.groupby('Customer Cluster')['Count'].sum().reset_index()
total_counts_fp.columns = ['Customer Cluster', 'Total Purchases']

# Merging to calculate percentage of total purchases for each category within each cluster
category_counts_fp = pd.merge(category_counts_fp, total_counts_fp, on='Customer Cluster')
category_counts_fp['Percentage of Total Purchases'] = (category_counts_fp['Count'] / category_counts_fp['Total Purchases']) * 100

# Filtering top 5 product categories across all clusters
top_categories_fp = category_counts_fp.groupby('Product Category')['Count'].sum().nlargest(5).index

```

```

filtered_data_fp           = category_counts_fp[category_counts_fp['Product
Category'].isin(top_categories_fp)]


# Rename clusters for clarity
cluster_mapping_fp = {1: 'Loyalists', 3: 'Potential Loyalists', 2: 'Churners'}
filtered_data_fp['Customer Cluster'] = filtered_data_fp['Customer
Cluster'].map(cluster_mapping_fp)


# Create the bar plot
plt.figure(figsize=(12, 8))
ax_fp = sns.barplot(x='Product Category', y='Percentage of Total Purchases', hue='Customer
Cluster', data=filtered_data_fp)

# Add data labels
for p in ax_fp.patches:
    ax_fp.annotate(f'{p.get_height():.1f}%', 
                  (p.get_x() + p.get_width() / 2., p.get_height()),
                  ha='center', va='baseline', fontsize=10, color='black', xytext=(0, 5),
                  textcoords='offset points')

# Set plot title and labels
plt.title('Cluster-Specific Comparison of Top 5 Product Categories (FP-Growth)', fontsize=16)
plt.xlabel('Product Category', fontsize=14)
plt.ylabel('Percentage of Total Purchases (%)', fontsize=14)
plt.xticks(rotation=45, ha='right')
plt.legend(title='Customer Cluster', fontsize=12)
plt.tight_layout()

# Show the plot
plt.show()

```

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Ajah, I. A., & Nweke, H. F. (2019). Big data and business analytics: Trends, platforms, success factors and applications. *Big Data and Cognitive Computing*, 3(2), 32.
- Akerkar, R. (2019). *Artificial Intelligence for Business*. Springer.
- Akshara, R., & Jain, A. (2024). Data to decisions: Optimizing e-commerce sales potential with analytics. *International Research Journal on Advanced Engineering Hub (IRJAEH)*, 2(04), 1087–1093. doi:<https://doi.org/10.47392/IRJAEH.2024.0150>
- Aldino, A. A., Pratiwi, E. D., Sintaro, S., & Putra, A. D. (2021). 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)* (pp. 29-34). IEEE.
- Avagianou, A., Chatzichristos, G., Herod, A., & Gialis, S. (2024). Precarity and Agency in Youthspaces of Work: The Case of Food Delivery Platform Workers in Athens, Greece. *Environment and Planning A: Economy and Space*, 0308518X241282536.
- Ballestar, M. T., Grau-Carles, P., & Sainz, J. (2018). Customer segmentation in e-commerce: Applications to the cashback business model. *Journal of Business Research*, 88, 407-414.
- Basal, M., Saraç, E., & Kadir, Ö. (2024). Dynamic pricing strategies using artificial intelligence algorithm. *Open Journal of Applied Sciences*, 14(8), 1963-1978.
- Bharadiya, J. P. (2023). Machine learning and AI in business intelligence: Trends and opportunities. *International Journal of Computer (IJC)*, 48(1), 123-134.
- Bibri, S. E. (2018). The IoT for smart sustainable cities of the future: An analytical framework for sensor-based big data applications for environmental sustainability. *Sustainable Cities and Society*, 38, 230-253.
- Bory, P., Natale, S., & Katzenbach, C. (2024). Strong and weak AI narratives: An analytical framework. *AI & SOCIETY*, 1-11.
- Brijs, B. (2016). *Business analysis for business intelligence*. CRC Press.
- Canhoto, A. I., & Clear, F. (2020). Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential. *Business Horizons*, 63(2), 183-193.
- Chaudhary, K., & Alam, M. (2022). *Big data analytics: Applications in business and marketing*. Auerbach Publications.
- Christy, A. J. (2021). RFM ranking – An effective approach to customer segmentation. *Journal of King Saud University - Computer and Information Sciences*, 33(10), 1251–1257. doi:<https://doi.org/10.1016/j.jksuci.2018.09.004>
- Christy, A. J., Umamakeswari, A., Priyatharsini, L., & Neyaa, A. (2021). RFM ranking–An effective approach to customer segmentation. *Journal of King Saud University-Computer and Information Sciences*, 33(10), 1251-1257.
- Davenport, H. T. (2014). How strategists use “big data” to support internal business decisions, discovery and production. *Strategy & Leadership*, 42(4), 45-50.
- Dawane, V. W. (2021). RFM analysis using K-means clustering to improve revenue and customer retention. (p. Proceedings of the International Conference on Smart Data Intelligence (ICSMDI 2021)). SSRN. doi:<https://doi.org/10.2139/ssrn.3852887>
- Deekshith, A. (2023). Scalable Machine Learning: Techniques for Managing Data Volume and Velocity in AI Applications. *International Scientific Journal for Research*, 5(5).
- Doğan, O., Ayçin, E., & Bulut, Z. (2018). Customer segmentation by using RFM model and clustering methods: A case study in retail industry. *International Journal of Contemporary Economics and Administrative Sciences*, 8.

- Emmanuel, I., & Stanier, C. (2016). Defining big data. *Proceedings of the International Conference on Big Data and Advanced Wireless Technologies*, (pp. 1-6).
- Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2024). Artificial intelligence and business value: A literature review. *Information Systems Frontiers*, 24(5), 1709-1734.
- Flowers, J. C. (2019, March). Strong and Weak AI: Deweyan Considerations. *AAAI Spring Symposium: Towards Conscious AI Systems*, 2287(7).
- Fragkos, N., Liapakis, A., Ntaliani, M., Ntalianis, F., & Costopoulou, C. (2024). A Sentiment Analysis Approach for Exploring Customer Reviews of Online Food Delivery Services: A Greek Case. *Digital*, 4(3), 698-709.
- Ganou, E., Karavasilis, G., Vrana, V., & Kehris, E. (2022). Online food ordering and delivery applications: An empirical study of the factors affecting intention to reuse. *Transcending Borders in Tourism Through Innovation and Cultural Heritage: 8th International Conference, IACuDiT, Hydra, Greece, 2021* (pp. 961-975). Cham: Springer International Publishing.
- GeeksforGeeks. (2025). Retrieved 2 9, 2025, from GeeksforGeeks: <https://www.geeksforgeeks.org/apriori-algorithm/>
- Gil de Zúñiga, H., Goyanes, M., & Durotoye, T. (2024). A scholarly definition of artificial intelligence (AI): Advancing AI as a conceptual framework in communication research. *Political Communication*, 41(2), 317-334.
- Günther, W. A., Mehrizi, M. H., Huysman, M., & Feldberg, F. (n.d.). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*, 26(3), 191-209.
- Gupta, I., Singh, A. K., Lee, C. N., & Buyya, R. (2017). Secure data storage and sharing techniques for data protection in cloud environments: A systematic review, analysis, and future directions. *IEEE Access*, 10, 71247-71277.
- Hallikainen, H., Savimäki, E., & Laukkanen, T. (2020). Fostering B2B sales with customer big data analytics. *Industrial Marketing Management*, 86, 90-98.
- Hamilton, R. H., & Sodeman, W. A. (2020). The questions we ask: Opportunities and challenges for using big data analytics to strategically manage human capital resources. *Business Horizons*, 63(1), 85-95.
- Hamzehi, M., & Hosseini, S. (2022). Business intelligence using machine learning algorithms. *Multimedia Tools and Applications*, 81(23), 33233-33251.
- Hoque, M. D. (2024). Optimizing decision-making through customer-centric market basket analysis. *Journal of Operational and Strategic Analysis*, 2(2), 72–83. doi:<https://doi.org/10.56578/josa020201>
- Hoque, M. D., Islam, M. S., & Mohtasim, S. A. (2024). Optimizing decision-making through customer-centric Market Basket Analysis. *J. Oper. Strateg. Anal.*, 2(2), 72-83.
- Hossain, M., Sattar, A. S., & Paul, M. K. (2019). Market basket analysis using Apriori and FP-Growth algorithm. *2019 22nd International Conference on Computer and Information Technology (ICCIT)* (pp. 1-6). IEEE.
- Javatpoint.com. (n.d.). *FP Growth Algorithm in Data Mining*. Retrieved 2 9, 2025, from Javatpoint: [https://www.javatpoint.com/fp-growth-algorithm-in-data-mining#:~:text=The%20FP%2DGrowth%20Algorithm%20proposed,tree%20\(FP%2Dtree\).](https://www.javatpoint.com/fp-growth-algorithm-in-data-mining#:~:text=The%20FP%2DGrowth%20Algorithm%20proposed,tree%20(FP%2Dtree).)
- Jin, D. H., & Kim, H. J. (2018). Integrated understanding of big data, big data analysis, and business intelligence: A case study of logistics. *Sustainability*, 10(10), 3778.
- Kalusivalingam, A. K., Sharma, A., Patel, N., & Singh, V. (2020). Enhancing Customer Segmentation through AI: Leveraging K-Means Clustering and Neural Network Classifiers. *International Journal of AI and ML*, 1(3).

- Kansal, M., Tanwar, K., Pandey, A. K., Kumar, V., Singh, P., & Upadhyay, S. (2023). Implementing Market Basket Analysis Using Eclat and Apriori Algorithm on Grocery Product Marketing Strategy. *2023 3rd International Conference on Advancement in Electronics & Communication Engineering (AECE)* (pp. 465-469). IEEE.
- Kansal, T., Bahuguna, S., Singh, V., & Choudhury, T. (2018). Customer segmentation using K-means clustering. *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)* (pp. 135-139). IEEE.
- Katsikas, S. K., & Gritzalis, S. (2017). Digitalization in Greece: State of play, barriers, challenges, solutions. In *Beyond Bureaucracy: Towards Sustainable Governance Informatisation* (pp. 355-375).
- Khan, W. A., Chung, S. H., Awan, M. U., & Wen, X. (2020). Machine learning facilitated business intelligence (Part I) Neural networks learning algorithms and applications. *Industrial Management & Data Systems*, 120(1), 164-195.
- Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, 3(1), 2053951716631130.
- Kitsios, F., & Kamariotou, M. (2021). Artificial intelligence and business strategy towards digital transformation: A research agenda. *Sustainability*, 13(4), 2025.
- Kumar, P., Manisha, K. N., & Nivetha, M. (2024). Enhancing Retail Performance Through Market Basket Analysis. *2024 Second International Conference on Advances in Information Technology (ICAIT)*. 1, pp. 1-7. IEEE.
- Kurniawan, F., Umayah, B., Hammad, J., Nugroho, S. M., & Hariadi, M. (2018). Market Basket Analysis to identify customer behaviours by way of transaction data. *Knowledge Engineering and Data Science*, 1(1), 20.
- Langer, A., & Mukherjee, A. (23). Data Strategy for Exponential Growth. In *Developing a Path to Data Dominance: Strategies for Digital Data-Centric Enterprises* (pp. 65-112). Cham: Springer International Publishing.
- Laursen, G. H., & Thorlund, J. (2016). *Business analytics for managers: Taking business intelligence beyond reporting*. John Wiley & Sons.
- Lee, I., & Shin, Y. J. (2020). Machine learning for enterprises: Applications, algorithm selection, and challenges. *Business Horizons*, 63(2), 157-170.
- Liu, B. (2021). "Weak AI" is Likely to Never Become "Strong AI", So What is its Greatest Value for Us? *arXiv preprint*. doi:arXiv:2103.15294
- Mohammed, A. B., Al-Okaily, M., Qasim, D., & Al-Majali, M. K. (2024). Towards an understanding of business intelligence and analytics usage: Evidence from the banking industry. *International Journal of Information Management Data Insights*, 4(1), 100215.
- Ng, K., Steinhubl, S. R., DeFilippi, C., Dey, S., & Stewart, W. F. (2016). Early detection of heart failure using electronic health records: Practical implications for time before diagnosis, data diversity, data quantity, and data density. *Circulation: Cardiovascular Quality and Outcomes*, 9(6), 649-658.
- Niu, Y., Ying, L., Yang, J., Bao, M., & Sivaparthipan, C. B. (2021). Organizational business intelligence and decision making using big data analytics. *Information Processing & Management*, 58(6), 102725.
- Omol, E. J. (2024). Apriori algorithm and market basket analysis to uncover consumer buying patterns: Case of a Kenyan supermarket. *Buana Information Technology and Computer Sciences (BIT and CS)*, 5(2), 51-63.
- Osborne, J. W. (2010). Improving normality of data using transformations. *Practical Assessment, Research, and Evaluation*, 15(4), 1-9.
- Oussous, A., Benjelloun, F. Z., Lahcen, A. A., & Belfkih, S. (2018). Big Data technologies: A survey. *Journal of King Saud University-Computer and Information Sciences*, 30(4), 431-448.

- Paramita, A. S. (2024). Comparison of K-Means and DBSCAN algorithms for customer segmentation in e-commerce. *Journal of Digital Market and Digital Currency*, 1(1), 43–62. doi:<https://doi.org/10.47738/jdmdc.v1i1.3>
- Park, M. H., Lee, J. S., & Doo, I. C. (2020). A Study of the Demand Forecasting Model for Publishing Business using Business Analysis. *International Journal of Computing and Digital Systems*, 9(5), 1-12.
- Phillips-Wren, G., Daly, M., & Burstein, F. (2021). Reconciling business intelligence, analytics and decision support systems: More data, deeper insight. *Decision Support Systems*, 146, 113560.
- Polydoropoulou, A., Tsirimpa, A., Karakikes, I., Tsouros, I., & Pagoni, I. (2022). Mode choice modeling for sustainable last-mile delivery: The Greek perspective. *Sustainability*, 14(15), 8976.
- Russell, S., & Norvig, P. (2021). *Τεχνητή Νοημοσύνη: Μια Σύγχρονη Προσέγγιση*. (M. Π. Καναβός, Trans.) Αθήνα: Κλειδάριθμος.
- Samanta-Rounti, I., Kyriazopoulos, P., & Pitticas, N. (2016). Can e-marketplaces improve the logistics strategy of traditional firms? The case of firms operating in Greece.
- Sarkar, M. P. (2024). Optimizing marketing strategies with RFM method and K-means clustering-based AI customer segmentation analysis. *Journal of Business and Management Studies*, 6(2), 54–60. doi:<https://doi.org/10.32996/jbms.2024.6.2.5>
- Siddiqa, A., Karim, A., & Gani, A. (2017). Big data storage technologies: A survey. *Frontiers of Information Technology & Electronic Engineering*, 1040-1070.
- Skalkos, D., Kalyva, Z. C., & Kosma, I. S. (2023). The Impact of the COVID-19 Pandemic on College Students' Food Choice Motives in Greece. *Sustainability*, 15(13), 9865.
- Strohbach, M., Daubert, J., Ravkin, H., & Lischka, M. (2016). Big data storage. In *New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe* (pp. 119-141).
- Tavera Romero, C. A., Ortiz, J. H., Khalaf, O. I., & Ríos Prado, A. (2021). Business intelligence: Business evolution after Industry 4.0. *Sustainability*, 13(18), 10026.
- Ünvan, Y. A. (2021). Market basket analysis with association rules. *Communications in Statistics- Theory and Methods*, 50(7), 1615-1628.
- Wang, G., Gunasekaran, A., Ngai, E. W., & Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics*, 176, 98-110.
- Whig, P. (2019). Artificial intelligence and machine learning in business. *International Journal on Integrated Education*, 2(2), 334128.
- Widjaja, A. E. (2024). Analysis of Apriori and FP-Growth algorithms for market basket insights: A case study of The Bread Basket bakery sales. *Journal of Digital Market and Digital Currency*, 1(1), 44–59.
- Zhang, Y., Xu, S., Zhang, L., & Yang, M. (2021). Big data and human resource management research: An integrative review and new directions for future research. *Journal of Business Research*, 133, 34-50.
- Βασιλάκη, Α. (2017). Μελέτη των ηλεκτρονικών καταστημάτων παράδοσης φαγητού στην Ελλάδα (Online Food Delivery).