

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ – ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ****Πρόγραμμα Μεταπτυχιακών Σπουδών****«Κυβερνοασφάλεια και Επιστήμη Δεδομένων»****Μεταπτυχιακή Διατριβή**

Τίτλος Διατριβής	Αναλυτική Πελατών στην Πράξη - Η Περίπτωση μιας Ψηφιακής Πλατφόρμας Παράδοσης Φαγητού Μέσω Διαδικτύου Customer Analytics in Practice – The Case of an Online Food Delivery Platform
Όνοματεπώνυμο Φοιτητή	Μαρία Ορφανίδου
Πατρώνυμο	Θεόδωρος
Αριθμός Μητρώου	ΜΠΚΕΔ21039
Επιβλέπων	Δημήτριος Αποστόλου, Καθηγητής

Ημερομηνία Παράδοσης **Οκτώβριος 2023**

Τριμελής Εξεταστική Επιτροπή

Δημήτρης Αποστόλου,
Καθηγητής

Άγγελος Πικράκης,
Επίκουρος Καθηγητής

Δρ Γρηγόρης Κορωνάκος
Διδάσκων ΠΜΣ

Ευχαριστίες

Στην ολοκλήρωση της παρούσας μεταπτυχιακής διατριβής, ιδιαίτερα σημαντική ήταν η συμβολή του Διδάσκοντα του ΠΜΣ κ. Ανδρέα Ζάρα, που προσέφερε επιστημονική και συμβουλευτική καθοδήγηση σε όλα τα στάδια εκπόνησής της. Θα ήθελα να εκφράσω την ειλικρινή μου εκτίμηση στον κ. Ανδρέα Ζάρα, για την καθοδήγηση, τα σχόλια και την εξειδίκευσή του σε όλη τη διαδικασία έρευνας και συγγραφής. Η αφοσίωσή του στην προώθηση της κριτικής σκέψης, της επεξηγηματικής προσέγγισης και της ακαδημαϊκής αριστείας έπαιξε καθοριστικό ρόλο στη διαμόρφωση της ποιότητας αυτής της μεταπτυχιακής διατριβής.

Επιπλέον θα ήθελα να ευχαριστήσω τον σύζυγο και τους γονείς μου για την στήριξη και υπομονή τους σε όλα τα στάδια της εκπόνησης της μεταπτυχιακής μου διατριβής. Η ολοκλήρωσή της θα ήταν αδύνατη χωρίς την υποστήριξη και την αγάπη τους.

Τέλος, θα ήθελα να αφιερώσω αυτήν την διατριβή στο νέο μέλος της οικογένειάς μας, το πρώτο μας παιδί. Η σκέψη του με γέμιζε με ελπίδα, δύναμη και κουράγιο σε κάθε δύσκολη και απαιτητική στιγμή. Εύχομαι να δω και εσένα να πετυχαίνεις τους στόχους σου και να πραγματοποιείς τα όνειρά σου, όσο δύσκολες και αν είναι οι συνθήκες.

Περίληψη

Η παρούσα μεταπτυχιακή διατριβή επικεντρώνεται στην εφαρμογή της ανάλυσης πελατών σε πρακτικό πλαίσιο, μελετώντας την περίπτωση μιας διαδικτυακής πλατφόρμας παράδοσης φαγητού. Στοχεύοντας στην κατανόηση της συμπεριφοράς των πελατών και την ανάπτυξη βελτιωμένων εμπειριών πελατών, η εργασία αναλύει δεδομένα πραγματικών συναλλαγών για να εξάγει πολύτιμες πληροφορίες που μπορούν να ωφελήσουν τόσο τους καταναλωτές όσο και την επιχείρηση. Μέσω της χρήσης τεχνικών όπως το Market Basket Analysis και το Customer Segmentation, η εργασία αποκαλύπτει συσχετίσεις και τάσεις που επηρεάζουν τις αποφάσεις αγοράς των πελατών, προσφέροντας παράλληλα στρατηγικές για την προσαρμογή των προσφορών και την βελτίωση της ικανοποίησης του πελάτη. Επιπλέον, η εργασία συζητά τις προκλήσεις και τις ηθικές διαστάσεις που συνδέονται με την συλλογή και ανάλυση δεδομένων πελατών στον κυβερνοχώρο. Τα ευρήματα της εργασίας υπογραμμίζουν την σημασία της ανάλυσης πελατών στην στρατηγική λήψη αποφάσεων για τις επιχειρήσεις εστίασης, παρέχοντας ένα πλαίσιο για την εφαρμογή αυτών των αναλύσεων στην πράξη με στόχο την βελτίωση της απόδοσης και την αύξηση της ανταγωνιστικότητας.

Abstract

This thesis focuses on the application of customer analytics in a practical context, studying the case of an online food delivery platform. Aiming to understand customer behavior and develop enhanced customer experiences, the research analyzes real transaction data to extract valuable insights beneficial for both consumers and the business. Through the utilization of techniques such as Market Basket Analysis and Customer Segmentation, the thesis reveals correlations and trends influencing customers' purchasing decisions, while also offering strategies to tailor offers and improve customer satisfaction. Furthermore, the thesis discusses the challenges and ethical dimensions associated with the collection and analysis of customer data in the cyber domain. The findings underscore the significance of customer analytics in strategic decision-making for focused businesses, providing a framework for the practical implementation of these analyses to enhance performance and increase competitiveness.

ΠΕΡΙΕΧΟΜΕΝΑ

Περίληψη.....	4
Abstract.....	4
Κεφάλαιο 1	8
Εισαγωγή	8
1.1 Ο όγκος και η αξία των δεδομένων στις επιχειρήσεις	8
1.2 Στόχος της μεταπτυχιακής διατριβής.....	11
Κεφάλαιο 2	11
2.1 Literature review	11
2.2 Big data και εφαρμογές στη διαχείριση ανθρώπινου δυναμικού (human resources - hr).....	13
2.3 Big data και εφαρμογές στο marketing.....	14
Κεφάλαιο 3.....	16
Μελέτη περίπτωσης μιας ηλεκτρονικής πλατφόρμας διανομής φαγητού	16
3.1 Customer Segmentation.....	16
3.1.1 RFM Analysis	17
3.1.2 Clustering Algorithms	18
3.2 Data management.....	22
3.2.1 Data cleansing and preparation	22
3.2.2 Data exploration and visualization	24
3.2.3 Strategic examination of customer behavior based on Recency, Frequency, and Monetary metrics	25
3.2.4 Customer categorization analysis using restaurant data	31
3.2.5 Customer demographic and behavioral analysis	33
3.3 Market basket analysis.....	43
3.3.1 Αλγόριθμος Apriori	44
3.3.2 Αλγόριθμος FP – Growth	44
3.3.3 Αλγόριθμος Eclat.....	45
3.3.4 Αποτελέσματα market basket analysis	45
Κεφάλαιο 4	50
4.1 Συμπεράσματα – Προτάσεις	50
ΒΙΒΛΙΟΓΡΑΦΙΑ	53

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1: Αναλογία Διακίνησης Δεδομένων	8
Πίνακας 2: Μεταβλητές Αρχείου Δεδομένων	22
Πίνακας 3: Food Score, Delivery Score and Service Score BoxPlot	24
Πίνακας 4: Μικρό πλήθος παραγγελιών	25
Πίνακας 5: Κορυφαίες επιλογές σε καταλόγους των εστιατορίων	25
Πίνακας 6: Μεταβλητές Recency, Frequency και Monetary	26
Πίνακας 7: Μεταβλητές Recency, Frequency και Monetary	27
Πίνακας 8: Μεταβλητή frequency	28
Πίνακας 9: Μεταβλητή recency	29
Πίνακας 10: Μεταβλητή monetary	30
Πίνακας 11: RFM Αρχείο Δεδομένων	30
Πίνακας 12: Elbow Method	32
Πίνακας 13: Ροή εργασιών	32
Πίνακας 14: Αποτέλεσμα της εκτέλεσης της ροής εργασιών	33
Πίνακας 15: Τελικό Αρχείο Δεδομένων	34
Πίνακας 16: non missing παρατηρήσεις, μέση τιμή, την τυπική απόκλιση, την ελάχιστη τιμή και τη μέγιστη τιμή	35
Πίνακας 17: Casual Shoppers Bar Chart – Orders Per Most Frequent Region	35
Πίνακας 18: Casual Shoppers Bar Chart – Orders Per Most Frequent Cuisine	36
Πίνακας 19: Casual Shoppers Histogram – Frequency of Pinata Orders	36
Πίνακας 20: Casual Shoppers Histograms – Frequency of Distinct Items and Frequency of Shops	37
Πίνακας 21: Elite Customers Bar Chart – Orders Per Most Frequent Region	37
Πίνακας 22: Elite Customers Bar Chart – Orders Per Most Frequent Cuisine	38
Πίνακας 23: Elite Customers Histogram – Frequency of Pinata Orders	38
Πίνακας 24: Elite Customers Histograms – Frequency of Distinct Items and Frequency of Shops	39
Πίνακας 25: Clusters' Pie Chart	41
Πίνακας 26: Ποσοστό των εσόδων της πλατφόρμας	41
Πίνακας 27: Πελάτες	42
Πίνακας 28: Πελάτες	42
Πίνακας 29: Δύναμη των συσχετίσεων μεταξύ προϊόντων	44
Πίνακας 30: Κανόνες με το μεγαλύτερο Lift	46
Πίνακας 31: Κανόνες με το υψηλότερο lift	48

Κεφάλαιο 1

Εισαγωγή

1.1 Ο όγκος και η αξία των δεδομένων στις επιχειρήσεις

Ο όγκος των πληροφοριών στη σημερινή κοινωνία αυξάνεται ραγδαία, ειδικά με το Διαδίκτυο να παίζει καθοριστικό ρόλο στον τρόπο με τον οποίο οι άνθρωποι έχουν πρόσβαση και χρησιμοποιούν δεδομένα. Σύμφωνα με στατιστικά στοιχεία της Statista, υπολογίζεται ότι 97 zettabytes δεδομένων θα δημιουργηθούν και θα χρησιμοποιηθούν παγκοσμίως το 2022. Αυτά τα δεδομένα δεν κοινοποιούνται μόνο μέσω του διαδικτύου αλλά αποθηκεύονται και σε διάφορες συσκευές για διαφορετικούς σκοπούς. Για παράδειγμα, ακόμη και smartphone με δευτερεύουσες δυνατότητες έχουν χωρητικότητα αποθήκευσης άνω των 32 gigabyte και το iPhone με τη μεγαλύτερη χωρητικότητα έχει 512 gigabyte. Υπάρχουν περίπου 600 κέντρα δεδομένων υπερκλίμακας παγκοσμίως, με περισσότερους από 5.000 διακομιστές. Η πλειοψηφία (39%) βρίσκεται στις Ηνωμένες Πολιτείες, ενώ η Κίνα, η Ιαπωνία, το Ηνωμένο Βασίλειο, η Γερμανία και η Αυστραλία αντιπροσωπεύουν περίπου το 30% του συνόλου (Robinson, 2021).

Το 2018, ο κόσμος δημιούργησε, κατέγραψε, αντέγραψε και κατανάλωσε συνολικά 33 zettabyte δεδομένων, τα οποία αυξήθηκαν σε 59 zettabyte το 2020. Προβλέπεται ότι θα φτάσει τα εκπληκτικά 175 zettabytes μέχρι το 2025. Αν κάθε bit δεδομένων ήταν ένα νόμισμα £ 1, μια στοίβα ενός zettabyte θα εκτεινόταν 2.550 έτη φωτός, αρκετά για να φτάσει στο πλησιέστερο αστρικό σύστημα, το Alpha Centauri, 600 φορές. Επί του παρόντος, παράγεται 59 φορές μεγαλύτερος όγκος δεδομένων κάθε χρόνο και ο ρυθμός ανάπτυξης των σύνθετων εκτιμάται ότι είναι περίπου 61% (Desjardins, 2019).

Περισσότερο από το μισό (53,72%) της παγκόσμιας διακίνησης δεδομένων αποδίδεται σε περιεχόμενο βίντεο. Οι πλατφόρμες μέσω κοινωνικής δικτύωσης είναι γεμάτες με βίντεο, με το TikTok να είναι εξ ολοκλήρου αφιερωμένο στο περιεχόμενο βίντεο και να γνωρίζει σταθερή ανάπτυξη στη βάση χρηστών του. Ακόμη και το Facebook έχει εξελιχθεί σε σημείο που το 51% του περιεχομένου που μοιράζεται στην πλατφόρμα αποτελείται από βίντεο. Ενώ υπάρχουν περιορισμένα διαθέσιμα δημόσια δεδομένα για το Snapchat, εκτιμάται ότι κάθε στιγμιότυπο που αποστέλλεται, πολλά από τα οποία είναι βίντεο, απαιτεί 1 MB δεδομένων. Σε συνδυασμό με τα μέσα κοινωνικής δικτύωσης (12,69%) και τα παιχνίδια (9,86%), αυτές οι τρεις κατηγορίες αντιπροσωπεύουν περισσότερα από τα τρία τέταρτα (76,27%) της συνολικής κίνησης δεδομένων στο διαδίκτυο. Από τον Απρίλιο του 2022, σχεδόν 250 εκατομμύρια email αποστέλλονται κάθε λεπτό, με αποτέλεσμα ένα εκπληκτικό σύνολο 333,22 δισεκατομμυρίων email σε ένα 24ωρο (Johnson, 2021).

Πίνακας 1: Αναλογία Διακίνησης Δεδομένων

Κατηγορία	Αναλογία Διακίνησης Δεδομένων Διαδικτύου
Βίντεο	53,72%
Social Media	12,69%
Παιχνίδι	9,86%
Σερφάρισμα στο διαδίκτυο	5,67%
Μηνύματα	5,35%
Αγορά	4,54%

Κοινή χρήση αρχείων	3,74%
Cloud	2,73%
VPN	1,39%
Ήχος	0,31%

Πηγή: (Johnson, 2021).

Τα μεγάλα δεδομένα διαθέτουν σημαντικές αναξιοποίητες δυνατότητες για τη δημιουργία αξίας. Τα 5V των Big data, τα οποία είναι η ταχύτητα, ο όγκος, η τιμή, η ποικιλία και η ακρίβεια, είναι τα θεμελιώδη χαρακτηριστικά που τα καθορίζουν. Η κατανόηση αυτών των επιτρέπει στους επιστήμονες να αντλήσουν περισσότερη αξία από τα δεδομένα τους και επιτρέπει στους οργανισμούς τους να γίνουν πιο πελατοκεντρικοί (Shafer, 2017). Η χρήση τους σε επιχειρήσεις και βιομηχανίες έχει τη δυνατότητα να βελτιώσει την κατανομή και τον συντονισμό των πόρων, να αυξήσει τη διαφάνεια και τη λογοδοσία, να ελαχιστοποιήσει τη σπατάλη και να προωθήσει νέες ιδέες και προοπτικές. Οι επιχειρήσεις σε όλους τους κλάδους κατανοούν τη σημασία της αξίας από τον συνεχώς αυξανόμενο όγκο δεδομένων. Η συλλογή, η ανάλυση και η χρήση δεδομένων από πηγές όπως συσκευές IoT αναγνωρίζεται ευρέως ότι αποφέρει πολλά οφέλη. Αυτές οι πηγές δεδομένων μπορούν να βοηθήσουν τους οργανισμούς να βελτιώσουν την παραγωγικότητα, να αυξήσουν την αποδοτικότητα, να μειώσουν το κόστος και να αποκτήσουν καλύτερη κατανόηση της αγοράς-στόχου τους, με σημαντικές επιπτώσεις στην εξυπηρέτηση πελατών. Τα δεδομένα παρέχουν χρήσιμες πληροφορίες που επιτρέπουν στις επιχειρήσεις να παραμείνουν ανταγωνιστικές και να αναπτύξουν νέα επιχειρηματικά μοντέλα με βάση τη χρήση δεδομένων (Burton-Jones and Grange, 2012).

Τα δεδομένα διαδραματίζουν κρίσιμο ρόλο στην προώθηση της καινοτομίας σε διαταραγμένα επιχειρηματικά περιβάλλοντα. Για παράδειγμα, στην κατασκευή, τα δεδομένα από τις επιτόπιες συσκευές μπορούν να χρησιμοποιηθούν για τη δημιουργία εμπειριών επαυξημένης πραγματικότητας που βελτιώνουν την ποιότητα και την ταχύτητα των υπηρεσιών. Η δυνατότητα απόκτησης ουσιαστικής επιχειρηματικής και κοινωνικής αξίας από δεδομένα μέσω αναλυτικών στοιχείων και επιχειρηματικής ευφύιας είναι σημαντική. Η εύκολη και έγκαιρη πρόσβαση σε μεγάλα δεδομένα δίνει τη δυνατότητα στους ενδιαφερόμενους να δημιουργήσουν σημαντική αξία. Αυτό είναι ιδιαίτερα εμφανές σε περιπτώσεις όπου δεν υπάρχουν κίνητρα για την ενίσχυση της διαφάνειας των δεδομένων. Για παράδειγμα, στον δημόσιο τομέα, τα μέλη του προσωπικού συχνά σπαταλούν χρόνο αναζητώντας πληροφορίες σε διάφορες κρατικές υπηρεσίες χρησιμοποιώντας μη ψηφιακά μέσα. Αξιοποιώντας μεγάλα δεδομένα, οι οργανισμοί έχουν μειώσει με επιτυχία αυτή την αναποτελεσματικότητα (Baecker et al., 2021).

Η εφαρμογή δεδομένων επιτρέπει επίσης την τμηματοποίηση του πληθυσμού για προσαρμοσμένες ενέργειες. Οι εταιρείες που επικεντρώνονται στην ιδιωτική κατανάλωση εφαρμόζουν εδώ και καιρό στοχευμένες πολιτικές υπηρεσιών ή στρατηγικές μάρκετινγκ για να καλύψουν μεμονωμένες ανάγκες. Η τμηματοποίηση της αγοράς και η ανάλυση με χρήση δημογραφικών πληροφοριών, μετρήσεων αγοράς και αγοραστικής συμπεριφοράς είναι ήδη κοινές. Οι ασφαλιστικές εταιρείες, για παράδειγμα, χρησιμοποιούν μεγάλα δεδομένα για τμηματοποίηση εδώ και αρκετό καιρό. Με τις τεχνολογικές εξελίξεις, πολλές εταιρείες έχουν πλέον δυνατότητες τμηματοποίησης και ανάλυσης σε πραγματικό χρόνο. Ο δημόσιος τομέας έχει επίσης αρχίσει να χρησιμοποιεί μεγάλα δεδομένα για τμηματοποίηση, όπως στις φορολογικές υπηρεσίες όπου οι φορολογούμενοι κατακερματίζονται με βάση παράγοντες όπως το εισόδημα και το πιστωτικό ιστορικό (Burton-Jones and Grange, 2012).

Οι αναδυόμενες τεχνολογίες όπως η τεχνητή νοημοσύνη και στρατηγικές όπως η ψηφιοποίηση παρουσιάζουν σημαντικές ευκαιρίες. Οι εταιρείες στοχεύουν στη δημιουργία αξίας ενσωματώνοντας τεχνητή νοημοσύνη σε όλες τις δραστηριότητές τους, τόσο για πελάτες όσο και για εσωτερικές διαδικασίες (Baecker et al., 2021). Ωστόσο, ο αντίκτυπός τους στους οργανισμούς μπορεί να ενισχυθεί σημαντικά με την αξιοποίηση του τεράστιου όγκου αναξιοποίητων πηγών δεδομένων,

συμπεριλαμβανομένων των εσωτερικών δεδομένων, των ανοιχτών (κυβερνητικών) δεδομένων και των δεδομένων που δημιουργούνται από το IoT (Llave, 2017).

Η συνειδητοποίηση της αξίας των δεδομένων απαιτεί μια πολιτιστική αλλαγή εντός των οργανισμών. Οι επιχειρήσεις πρέπει να επανεξετάσουν τις βασικές τους στρατηγικές και να αγκαλιάσουν τις ψηφιακές διαδικασίες σε διάφορους τομείς, από τα χρηματοοικονομικά μέχρι την υγειονομική περίθαλψη. Οι πλατφόρμες που τυποποιούν και συγκεντρώνουν δεδομένα από διαφορετικές πηγές, όπως μέσα κοινωνικής δικτύωσης, δεδομένα που δημιουργούνται από μηχανή και υπηρεσίες, προσφέρουν πολύτιμο πλαίσιο και γνώσεις για άτομα και πληθυσμούς (Burton-Jones and Grange, 2012).

Ο εντοπισμός των αναγκών, η μεταβλητότητα και η βελτίωση της απόδοσης είναι ένα άλλο πλεονέκτημα της χρήσης μεγάλων δεδομένων. Καθώς οι εταιρείες ψηφιοποιούνται, συγκεντρώνουν τεράστιες ποσότητες εξαιρετικά λεπτομερών δεδομένων συναλλαγών. Αισθητήρες ενσωματωμένοι σε διάφορες φυσικές συσκευές, από εξοπλισμό γραμμής παραγωγής μέχρι αυτοκίνητα και κινητά τηλέφωνα, μετρούν τις διαδικασίες, τη χρήση του προϊόντος και τις ανθρώπινες συμπεριφορές. Οι καταναλωτές δημιουργούν επίσης σημαντικό όγκο δεδομένων μέσω ιστολογίων, φωτογραφιών και βίντεο. Η πρόσβαση και η διαχείριση αυτών των δεδομένων έχει φέρει επανάσταση στη λήψη αποφάσεων. Η ελεγχόμενη έρευνα χρησιμοποιείται ήδη σε πολλούς τομείς της επιχείρησης για τη λήψη καλύτερων αποφάσεων (Baecker et al., 2021).

Τα μεγάλα δεδομένα χρησιμεύουν ως πρώτη ύλη για την ανάπτυξη και τη λειτουργία αυτών των αλγορίθμων. Για παράδειγμα, οι φορολογικές υπηρεσίες χρησιμοποιούν αυτοματοποιημένες μηχανές ανίχνευσης απάτης που χρησιμοποιούν μεγάλα δεδομένα για τον εντοπισμό υποψηφίων για περαιτέρω έρευνα (Burton-Jones and Grange, 2012).

Οι έμποροι λιανικής μπορούν να βελτιστοποιήσουν τις διαδικασίες λήψης αποφάσεων μέσω αλγορίθμων μεγάλων δεδομένων, επιτρέποντας αυτόματες προσαρμογές αποθέματος και τιμολόγηση σε πραγματικό χρόνο τόσο για πωλήσεις εντός καταστήματος όσο και για διαδικτυακές πωλήσεις. Οι κατασκευαστικές εταιρείες μπορούν να προσαρμόσουν αυτόματα τις γραμμές παραγωγής για να βελτιστοποιήσουν την απόδοση, να ελαχιστοποιήσουν τα απόβλητα και να μετριάσουν τις επικίνδυνες συνθήκες. Σε ορισμένες περιπτώσεις, η ανάλυση μεγάλων δεδομένων δεν αυτοματοποιεί πλήρως τις αποφάσεις, αλλά τις διευκολύνει με το χειρισμό ενός τεράστιου όγκου δεδομένων που υπερβαίνει τη χωρητικότητα της μη αυτόματης ανάλυσης υπολογιστικών φύλλων. Οι οργανισμοί έχουν ήδη λάβει πιο αποτελεσματικές αποφάσεις αναλύοντας ολοκληρωμένα σύνολα δεδομένων από πελάτες, υπαλλήλους και αισθητήρες ενσωματωμένους σε προϊόντα.

Τα δεδομένα αναφέρονται συχνά ως το «νέο πετρέλαιο» λόγω της υψηλής αξίας του ως πόρου. Η κατανόηση των δεδομένων μπορεί να οδηγήσει σε καινοτόμο σκέψη και να καθοδηγήσει επιτυχημένες επιχειρηματικές στρατηγικές. Οι κολοσσοί του κλάδου όπως η GE και η Siemens τοποθετούνται πλέον ως εταιρείες δεδομένων, αναγνωρίζοντας τον κρίσιμο ρόλο των δεδομένων στον επαναπροσδιορισμό των λειτουργιών τους (Baecker et al., 2021).

Ωστόσο προκύπτουν και κάποιες προκλήσεις που μπορεί να περιλαμβάνουν τεχνικά ζητήματα, όπως η ασφάλεια των δεδομένων και η προστασία τους από ανεπιθύμητες παρεμβάσεις ή κυβερνοεπιθέσεις, καθώς και δυσκολία στη διαχείριση και ανάλυση τεράστιων όγκων δεδομένων (big data) με ακρίβεια και αποτελεσματικότητα.

Επίσης, οι ηθικές διαστάσεις αφορούν στην προστασία της ιδιωτικότητας των πελατών και στον σεβασμό των δικαιωμάτων τους στα προσωπικά τους δεδομένα. Περιλαμβάνει ζητήματα όπως η διαφάνεια σχετικά με το πώς και για ποιον σκοπό χρησιμοποιούνται τα δεδομένα αυτά, η συναίνεση των πελατών για τη συλλογή και χρήση των δεδομένων τους, και η διασφάλιση ότι η χρήση των δεδομένων δεν θα οδηγήσει σε αδικαιολόγητη διακρίση ή άλλες αρνητικές συνέπειες για τους πελάτες.

Σε γενικές γραμμές, οι επιχειρήσεις χρειάζεται να μπορούν να ισορροπήσουν ανάμεσα στην αξιοποίηση των δεδομένων για τη βελτίωση των υπηρεσιών και προϊόντων τους και την ανάγκη να προστατεύσουν και να σέβονται τα δικαιώματα και τις προσδοκίες των πελατών τους ως προς την ιδιωτικότητα και την ηθική χρήση των προσωπικών τους δεδομένων.

1.2 Στόχος της μεταπτυχιακής διατριβής

Στην παρούσα μεταπτυχιακή διατριβή θα πραγματοποιηθεί η μελέτη περίπτωσης δεδομένων εστιατορίων προερχόμενων από μία ψηφιακή πλατφόρμα παράδοσης φαγητού. Ορισμένες από τις δημοφιλείς πλατφόρμες περιλαμβάνουν τις παρακάτω:

- Efood, η οποία είναι μία από τις μεγαλύτερες πλατφόρμες παράδοσης φαγητού στην Ελλάδα η οποία συνεργάζεται με πολλά εστιατόρια, καταστήματα και super market.
- Wolt , η οποία προσφέρει υπηρεσίες παράδοσης φαγητού από διάφορα εστιατόρια και καφετέριες σε πολλές πόλεις της Ελλάδας.
- Delivery.gr, άλλη μία δημοφιλής πλατφόρμα παράδοσης φαγητού, που συνεργάζεται με πολλά καταστήματα σε διάφορες περιοχές.
- Δες Φαγητό, η οποία παρέχει υπηρεσίες παράδοσης φαγητού και προϊόντων σε διάφορες πόλεις της Ελλάδας.
- Vrisko.gr, η οποία εκτός από την παράδοση φαγητού, επιτρέπει στους χρήστες να βρίσκουν και να αγοράζουν τρόφιμα από καταστήματα τοπικά στην περιοχή τους.
- Box.gr, ανερχόμενη δημοφιλής πλατφόρμα διανομής φαγητού σε συνεργασία με μεγάλη εταιρία τηλεπικοινωνιών.

Η υλοποίηση της ανάλυσης αφορά στην τμηματοποίηση των πελατών με βάση την αλληλεπίδραση και συμπεριφορά των πελατών με την επιχείρηση. Η τμηματοποίηση των πελατών σε ομοιογενή τμήματα, η ανάλυση του προφίλ και των χαρακτηριστικών κάθε τμήματος, στη συνέχεια η στόχευση σε ένα ή περισσότερα τμήματα και τέλος η τοποθέτηση των προϊόντων και των υπηρεσιών ανάλογα με τους καταναλωτές κάθε τμήματος, θα οδηγήσει στην καλύτερη κατανόηση των καταναλωτών και των αναγκών τους. Επίσης θα συμβάλει στην δημιουργία μίας πιο προσωπικής προσέγγισης των πελατών, στην αποφυγή των τμημάτων των καταναλωτών που θεωρούνται μη κερδοφόρα και τέλος στην διατήρηση και καλύτερη εξυπηρέτηση των πελατών. Επιπλέον, θα πραγματοποιηθεί ανάλυση καλαθιού αγορών, η οποία θα βοηθήσει στην κατανόηση των συνδυασμών των προϊόντων που αγοράζονται και συνδυάζονται μαζί, επιτρέποντας την πιο στοχευμένη και αποτελεσματική διαμόρφωση των προσφορών και των προωθητικών ενεργειών προς τους πελάτες.

Κεφάλαιο 2

2.1 Literature review

Η απόκτηση πελατών είναι ζωτικής σημασίας, αλλά η διατήρηση των πελατών είναι ακόμη πιο σημαντική, καθώς η απώλεια ενός πελάτη έχει ως αποτέλεσμα την απώλεια πιθανών αγορών καθ' όλη τη διάρκεια ζωής (Kotler & Armstrong, 2006). Οι Yeh et al. (2009) τονίζουν ότι η Διαχείριση Πελατειακών Σχέσεων (CRM) στοχεύει στην απόκτηση και διατήρηση των πιο κερδοφόρων πελατών κατανοώντας τις αξίες τους. Σε ανταγωνιστικές βιομηχανίες, ο εντοπισμός και η διατήρηση πελατών υψηλής αξίας είναι απαραίτητος (Chang et al., 2007; Chhliya κ.ά., 2009; Mutandwa κ.ά., 2009). Η αποτελεσματική διατήρηση πελατών και κερδοφορία απαιτεί προσαρμοσμένες στρατηγικές μάρκετινγκ που αντιμετωπίζουν διαφορετικές ανάγκες των πελατών, κατανέμοντας αποτελεσματικά τους πόρους (Huang et al., 2009; Chang κ.ά., 2010). Οι Sohrabi και Khanlari (2007) υποστηρίζουν ότι είναι ζωτικής σημασίας να αξιολογηθεί η κερδοφορία των πελατών και να κατανοηθούν ανάλογα οι πόροι.

Καθώς οι εταιρείες συσσωρεύουν μεγαλύτερα αρχεία συναλλαγών, η κατάτμηση των πελατών σε ομοιογενείς ομάδες με βάση τις ομοιότητες καθίσταται απαραίτητη για την αποτελεσματική κατάτμηση της αγοράς (Hung & Tsai, 2008; Chang κ.ά., 2010). Η αξιολόγηση της αξίας διαφορετικών ομάδων πελατών μπορεί να προσφέρει πολύτιμες διαχειριστικές πληροφορίες, επιτρέποντας προσαρμοσμένες στρατηγικές μάρκετινγκ για την κάλυψη ποικίλων αναγκών πελατών. Οι Allenby et al. (1998) σημειώνουν ότι δεν υπάρχει ενιαίο σύνολο μεταβλητών για πλήρη κατάτμηση

της αγοράς. Ο Kotler (2003) προτείνει την ταξινόμηση των πελατών με βάση τα χαρακτηριστικά (γεωγραφικά, δημογραφικά, ψυχογραφικά) και τις μεταβλητές συμπεριφοράς (στάσεις προϊόντων, απόκριση οφέλους, αφοσίωση στο εμπορικό σήμα) (Wu & Pan, 2009).

Το μοντέλο RFM (Recency, Frequency, Monetary), που εισήχθη από τους Bult και Wansbeek (1995), έχει αποδειχθεί αποτελεσματικό σε βάσεις δεδομένων μάρκετινγκ (Blattberg et al., 2008). Η ανάλυση RFM αξιολογεί τρεις βασικές μεταβλητές που σχετίζονται με την αγορά και επηρεάζουν τις μελλοντικές αγορές των πελατών. Η επικαιρότητα υποδεικνύει το χρόνο από την τελευταία αγορά, η συχνότητα αντιπροσωπεύει τον αριθμό των αγορών εντός μιας συγκεκριμένης περιόδου και η νομισματική αναφέρεται στο συνολικό ποσό που δαπανήθηκε κατά τη διάρκεια αυτής της περιόδου (Wang, 2010). Αυτές οι μεταβλητές συμπεριφοράς μπορούν να τμηματοποιήσουν τους πελάτες με βάση το προϊόν, τη στάση της επωνυμίας και την αφοσίωση.

Τα τελευταία χρόνια παρατηρείται σημαντική αύξηση των διαδικτυακών λιανικών πωλήσεων. Σύμφωνα με το Interactive Media in Retail Group (IMRG), οι online αγοραστές στο Ηνωμένο Βασίλειο ξόδεψαν περίπου 50 δισεκατομμύρια λίρες το 2011, αύξηση άνω του 5000% από το 2000. Αυτή η αύξηση των διαδικτυακών πωλήσεων σηματοδοτεί μια θεμελιώδη αλλαγή στη συμπεριφορά των καταναλωτών, με μοναδικά χαρακτηριστικά που διακρίνουν τις διαδικτυακές αγορές από το παραδοσιακό λιανικό εμπόριο. Οι ηλεκτρονικές αγορές επιτρέπουν τη λεπτομερή παρακολούθηση των δραστηριοτήτων των πελατών, συμπεριλαμβανομένου του ιστορικού περιήγησης, των λεπτομερειών παραγγελιών και των πληροφοριών λογαριασμού, επιτρέποντας στους λιανοπωλητές να εξατομικεύουν τις αλληλεπιδράσεις και να δημιουργούν πελατοκεντρική επιχειρηματική ευφυΐα.

Οι βασικές επιχειρηματικές ανησυχίες για τους διαδικτυακούς λιανοπωλητές περιλαμβάνουν την κατανόηση της συμπεριφοράς των πελατών, τον εντοπισμό πολύτιμων και πιστών πελατών, την ανάλυση μοτίβων αγοράς και τη βελτιστοποίηση στρατηγικών προώθησης. Οι τεχνικές εξόρυξης δεδομένων, σε συνδυασμό με επιχειρηματικές μετρήσεις όπως το μοντέλο RFM και η αξία ζωής του πελάτη, χρησιμοποιούνται ευρέως για την αντιμετώπιση αυτών των ανησυχιών. Κορυφαίοι διαδικτυακοί λιανοπωλητές όπως η Amazon, η Walmart και η Tesco έχουν ενσωματώσει την εξόρυξη δεδομένων στις επιχειρηματικές τους διαδικασίες για να αποκτήσουν ανταγωνιστικό πλεονέκτημα μέσω πελατοκεντρικού μάρκετινγκ. Ωστόσο, οι μικρότερες επιχειρήσεις και οι νεοεισερχόμενοι συχνά δεν διαθέτουν την τεχνική εμπειρογνώμοσύνη και τους πόρους για την αποτελεσματική αξιοποίηση αυτών των τεχνικών.

Τα τελευταία χρόνια, αρκετοί ερευνητές έχουν ενσωματώσει μεταβλητές RFM σε μοντέλα ομαδοποίησης για την ενίσχυση του CRM. Οι Hosseini et al. (2010) συνδύασαν ένα σταθμισμένο μοντέλο RFM με τον αλγόριθμο K-Means για τη βελτίωση του CRM στις επιχειρήσεις, επιτρέποντας πιο λεπτή τμηματοποίηση πελατών και αποτελεσματικές στρατηγικές μάρκετινγκ. Οι Wu et al. (2009) εφάρμοσαν το μοντέλο RFM και τη μέθοδο K-Means για να αναλύσουν τη βάση δεδομένων πελατών ενός ταϊβανέζικου εξοπλισμού, με στόχο τη δημιουργία ισχυρών σχέσεων με τους πελάτες και την εδραίωση της αφοσίωσης μεταξύ εξαιρετικά κερδοφόρων, μακροπρόθεσμων πελατών. Οι Chuang και Shen (2008) χρησιμοποίησαν μια διαδικασία πολλαπλών σταδίων που περιλαμβάνει στάθμιση μεταβλητών RFM μέσω διαδικασίας αναλυτικής ιεραρχίας, ανάλυση ομαδοποίησης για αξίες διάρκειας ζωής πελατών και μια μέθοδο αυτο-οργάνωσης χάρτη για τον εντοπισμό ομάδων πελατών υψηλής αξίας, επιτρέποντας τη στρατηγική κατανομή πόρων προς τους πιο πολύτιμους πελάτες.

Ακόμη, το έργο της χρήσης της ανάλυσης καλάθιού αγοράς στην έρευνα διαχείρισης έχει πραγματοποιηθεί από τους Aguinis et al. Η ανάλυση καλάθιού αγοράς είναι επίσης γνωστή ως association rule mining. Βοηθά τον αναλυτή να κατανοήσει τη συμπεριφορά των πελατών, π.χ. ποια προϊόντα αγοράζονται μαζί.

Με μια ευρύτερη έννοια, η ανάλυση καλάθιού αγοράς στοχεύει στα καλάθια πελατών προκειμένου να παρακολουθεί τα αγοραστικά πρότυπα και να βελτιώνει την ικανοποίηση των πελατών (Microstrategy, 2003). Μπορούν να χρησιμοποιηθούν τα ακόλουθα αναλυτικά στοιχεία: ποσοστά προσκόλλησης, δημογραφικά καλάθια, εναλλαγή επωνυμίας, αφοσίωση πελατών, βασικά στοιχεία, είδη ανά καλάθι, τιμή στο καλάθι, συνεισφορά εσόδων, διείσδυση αγοραστών και άλλα. Με μια στενότερη έννοια, η ανάλυση του καλάθιού αγοράς μας δίνει την απάντηση στο ακόλουθο ερώτημα που έθεσαν οι Svetina, M., & Zupančič, J. (2005): ποια αγαθά πωλούνται μαζί στο πλαίσιο

της ίδιας συναλλαγής ή στον ίδιο πελάτη; Αναλύοντας αυτές τις πληροφορίες, προσπαθούμε να εντοπίσουμε επαναλαμβανόμενα μοτίβα προκειμένου να προσφέρουμε σχετικά αγαθά μαζί και, ως εκ τούτου, να αυξήσουμε τις πωλήσεις. Μπορούμε να παρακολουθούμε τις σχετικές πωλήσεις σε διαφορετικά επίπεδα ταξινόμησης προϊόντων ή σε διαφορετικές ομάδες πελατών.

2.2 Big data και εφαρμογές στη διαχείριση ανθρώπινου δυναμικού (human resources - hr)

Τα Big Data μπορούν να χρησιμοποιηθούν από εταιρείες για τη βελτίωση της διατήρησης των εργαζομένων, τη βελτίωση της μάθησης και της ανάπτυξης, τον εξορθολογισμό των διαδικασιών πρόσληψης και την αύξηση της δέσμευσης των εργαζομένων. Αναλύοντας δεδομένα για την ικανοποίηση των εργαζομένων, οι ομάδες ανθρώπινου δυναμικού μπορούν να εντοπίσουν τάσεις και πρότυπα που τις βοηθούν να κατανοήσουν τις ανάγκες και τις επιθυμίες των εργαζομένων τους. Οι συνεντεύξεις αποχώρησης εργαζομένων δεν βοηθούν πολύ στην κατανόηση του κινήτρου των εργαζομένων να εγκαταλείψουν τις εταιρείες (Barbosa et al., 2018).

Με την ανάλυση μεγάλων δεδομένων, οι ομάδες ανθρώπινου δυναμικού σήμερα μπορούν να εφαρμόσουν προγράμματα που βασίζονται σε δεδομένα, να διεξάγουν τακτικές έρευνες για την ικανοποίηση του προσωπικού και να εντοπίζουν τάσεις και μοτίβα που παρέχουν ποσοτικά δεδομένα. Οι επαγγελματίες ανθρώπινου δυναμικού μπορούν να χρησιμοποιήσουν αυτά τα δεδομένα για να αναλύσουν τους υπαλλήλους τους με νέους τρόπους για να ανακαλύψουν τις ανάγκες και τις επιθυμίες τους. Αυτό τους επιτρέπει να παρέχουν το σωστό επίπεδο ευελιξίας, οφέλη και προνόμια για να διατηρήσουν κορυφαία ταλέντα (Chen et al., 2021).

Όσον αφορά τη μάθηση και την ανάπτυξη, τα μεγάλα δεδομένα και τα αναλυτικά στοιχεία επιτρέπουν την εξατομικευμένη και αυτοδύναμη διαδικτυακή μάθηση, η οποία είναι πιο οικονομική και μπορεί να ενσωματωθεί απρόσκοπτα στην καθημερινή ρουτίνα των εργαζομένων. Η Danone, για παράδειγμα, έχει δημιουργήσει μια διαδικτυακή πλατφόρμα που ονομάζεται Danone Campus 2.0, όπου οι εργαζόμενοι μπορούν να έχουν πρόσβαση σε ένα ευρύ φάσμα μαθησιακών πόρων και να συνεργάζονται με συναδέλφους (Vassakis et al., 2018).

Τα Big Data διαδραματίζουν επίσης κρίσιμο ρόλο στον εξορθολογισμό της διαδικασίας πρόσληψης. Συλλέγοντας και αναλύοντας δεδομένα υψηλής ποιότητας, οι εταιρείες μπορούν να εντοπίσουν γρήγορα τα σωστά ταλέντα και να ενισχύσουν την ποικιλομορφία στο εργατικό τους δυναμικό. Η Unity Technologies, για παράδειγμα, τοποθετεί δεδομένα σε μια διοχέτευση προσλήψεων, τα αξιολογεί με βάση τα κριτήρια αναφοράς και μοιράζεται τα ευρήματα με διευθυντές προσλήψεων και ηγέτες για τη βελτίωση της λήψης αποφάσεων. Αυτό όχι μόνο μειώνει τον χρόνο πρόσληψης, αλλά διασφαλίζει επίσης ότι οι θέσεις εργασίας γεμίζουν με τα κατάλληλα ταλέντα (Darvazeh et al., 2020).

Επιπλέον, τα Big Data μπορούν να βοηθήσουν τις εταιρείες να ερμηνεύσουν και να ενεργήσουν βάσει των αποτελεσμάτων των ερευνών αφοσίωσης των εργαζομένων. Συλλέγοντας σχετικές μετρήσεις ανθρώπινου δυναμικού, οι εργοδότες μπορούν να αποκτήσουν γνώσεις για το τι παρακινεί τους εργαζόμενους, τι τους αποτρέπει και πώς να βελτιώσουν την απόδοσή τους. Η Humanyze, για παράδειγμα, χρησιμοποιεί αναλυτικά στοιχεία για την παρακολούθηση των προτύπων επικοινωνίας των εργαζομένων και τη βελτίωση της απόδοσης. Η Bank of America υλοποίησε συλλογικά διαλείμματα με βάση γνώσεις από την τεχνολογία της Humanyze, με αποτέλεσμα βελτιωμένη απόδοση και μειωμένα επίπεδα άγχους (Vassakis et al., 2018).

Οι εταιρείες που διαχειρίζονται αποτελεσματικά και αξιοποιούν δεδομένα μπορούν να αποκτήσουν ανταγωνιστικό πλεονέκτημα. Παραμένοντας ενημερωμένοι για τις τελευταίες τάσεις και επενδύοντας στην επιστήμη δεδομένων, οι οργανισμοί ανθρώπινου δυναμικού μπορούν να αντλήσουν πολύτιμες γνώσεις και να λάβουν τεκμηριωμένες αποφάσεις για την επιτυχία (Chen et al., 2021).

2.3 Big data και εφαρμογές στο marketing

Στο παρελθόν, οι επιχειρήσεις δεν διέθεταν την τεχνολογία για να αναλύσουν αποτελεσματικά μεγάλο όγκο δεδομένων πελατών και να αντλήσουν χρήσιμες πληροφορίες. Ωστόσο, καθώς αυτή η τεχνολογία έγινε διαθέσιμη, οι εταιρείες που τη χρησιμοποίησαν απέκτησαν σημαντικό ανταγωνιστικό πλεονέκτημα μέσω της ανάλυσης δεδομένων πελατών (Le et al., 2017).

Σήμερα, οι περισσότερες εταιρείες έχουν υιοθετήσει κάποια μορφή ανάλυσης πελατών, η οποία έχει μειώσει το ανταγωνιστικό πλεονέκτημα που παρείχε κάποτε. Ως αποτέλεσμα, οι επιχειρήσεις προσπαθούν να εμβαθύνουν σε βάθος από τις παραδοσιακές αναλύσεις δεδομένων πελατών για να βελτιώσουν την αποτελεσματικότητα του μάρκετινγκ και να ανακτήσουν ένα σημαντικό ανταγωνιστικό πλεονέκτημα, παρόμοιο με αυτό που γνώρισαν οι πρώτοι που υιοθετούσαν τα αναλυτικά στοιχεία πελατών (Le et al., 2017).

Τα δεδομένα διαδραματίζουν κρίσιμο ρόλο στην παροχή εμπειρίας στους πελάτες και αποτελούν τη βάση των επιχειρηματικών λειτουργιών. Η ανάλυση πελατών περιλαμβάνει τη συλλογή και την ανάλυση δεδομένων πελατών για την κατανόηση των αναγκών τους, της ευαισθησίας των τιμών και των αμφιβολιών για την επιλογή προϊόντων ή υπηρεσιών. Αυτές οι πληροφορίες χρησιμοποιούνται για την ανάπτυξη στοχευμένων δραστηριοτήτων μάρκετινγκ και πωλήσεων. Τα μεγάλα δεδομένα προσφέρουν μια ευρύτερη κατανόηση του κοινού-στόχου και του τρόπου με τον οποίο τα προϊόντα ή οι υπηρεσίες μπορούν να παρουσιαστούν αποτελεσματικά σε αυτά.

Τα μεγάλα δεδομένα συμβάλλουν με δύο τρόπους. Πρώτον, επιτρέπουν την πλήρη τμηματοποίηση του κοινού-στόχου. Δεύτερον, επιτρέπουν στην ανάλυση συναισθήματος να κατανοήσει πώς τα διαφορετικά τμήματα αντιλαμβάνονται και αλληλεπιδρούν με το προϊόν. Αυτή η ποικιλία προοπτικών ανοίγει δυνατότητες για την παρουσίαση του προϊόντος με τον πιο αποτελεσματικό τρόπο με βάση τις συγκεκριμένες τάσεις του τμήματος (Le et al., 2017).

Η διαδικασία ξεκινά με τη συλλογή πληροφοριών από σχετικές πηγές, όπως ιστότοπους, εφαρμογές για κινητά και web, καμπάνιες μάρκετινγκ και πηγές συνεργατών. Στη συνέχεια, αυτά τα δεδομένα υποβάλλονται σε επεξεργασία και παρέχουν πληροφορίες σχετικά με τον τρόπο με τον οποίο οι άνθρωποι χρησιμοποιούν το προϊόν ή την υπηρεσία και επισημαίνουν τομείς για βελτίωση. Με βάση αυτές τις πληροφορίες, οι ιδιοκτήτες επιχειρήσεων μπορούν να προσαρμόσουν τα μηνύματά τους και να απευθυνθούν σε τομείς που χρειάζονται βελτίωση (Kshetri, 2014).

Η Amazon είναι ένα χαρακτηριστικό παράδειγμα χρήσης αναλυτικών στοιχείων πελατών για τη διαχείριση του αποθέματος των προϊόντων τους με βάση τα δεδομένα των πελατών και την προσαρμογή των συστάσεων ώστε να ταιριάζουν στις μεμονωμένες προτιμήσεις. Η ανάλυση συναισθήματος χρησιμοποιείται επίσης για την κατανόηση του τρόπου με τον οποίο το προϊόν γίνεται αντιληπτό από διαφορετικούς χρήστες μέσω αλγορίθμων αναγνώρισης προτύπων. Αυτή η κατανόηση επιτρέπει προσαρμογές στην παρουσίαση και πιο ακριβή στόχευση των προσπαθειών μάρκετινγκ (Le et al., 2017).

Για να προσφέρουμε μια εξατομικευμένη εμπειρία και να παραμείνουμε μπροστά, το μέλλον του μάρκετινγκ πιστεύεται ότι βρίσκεται στο γενετικό εξατομικευμένο μάρκετινγκ AI, σύμφωνα με 9 στους 10 επαγγελματίες. Υπάρχουν πολλές πολυπόθητες τεχνολογίες διαθέσιμες για να επιτευχθεί αυτό.

Τα μεγάλα δεδομένα, οι αλγόριθμοι και τα μοντέλα πρόβλεψης αξιοποιούνται για άμεση στόχευση. Η υπερ-εξατομίκευση απαιτεί σημαντικά μεγαλύτερο όγκο δεδομένων υψηλής ποιότητας, λεπτομερών και ποικίλων δεδομένων, που γίνονται εφικτά από τα Big Data. Το λογισμικό πρόβλεψης εξατομίκευσης (PPS) είναι μεταξύ αυτών των τεχνολογιών, το οποίο προσφέρει υψηλή απόδοση επένδυσης (ROI) εξατομικεύοντας την επιλογή προϊόντων για κάθε καταναλωτή (Le et al., 2017).

Η ποικιλία και η ποικιλομορφία των δεδομένων που επεξεργάζονται επιτρέπουν στις εταιρείες να ενσωματώνουν νέα δεδομένα στην ανάλυσή τους, κάτι που προηγουμένως ήταν πρόκληση από την άποψη του μάρκετινγκ. Η ανάλυση μη δομημένων δεδομένων, όπως το ελεύθερο κείμενο ή τα βίντεο, επιτρέπει επίσης την εξέταση των ποιοτικών δεδομένων που απαιτούνται για τη συμπεριφορά και την ψυχογραφική ανάλυση σε ατομικό επίπεδο (Kuo et al., 2009).

Η ταχύτητα συλλογής και επεξεργασίας δεδομένων επιτρέπει την ανάλυση σε πραγματικό χρόνο και τη συνεχή άμεση στόχευση. Αυτή η στόχευση μπορεί να ενημερωθεί ή να προσαρμοστεί με βάση το περιβάλλον και το περιβάλλον του πελάτη, ευθυγραμμισμένη με τις αρχές της υπερπλαισίωσης. Η ταχύτητα ανάλυσης δεδομένων επιτρέπει επίσης στις ομάδες μάρκετινγκ να μετρούν γρήγορα το ROI των στοχευμένων ενεργειών μάρκετινγκ και επικοινωνίας (Kshetri, 2014). Το προτεινόμενο μοντέλο εξαρτάται από την πρόταση "ο καλύτερος προγνωστικός παράγοντας της μελλοντικής συμπεριφοράς των πελατών είναι η συμπεριφορά των πελατών του παρελθόντος". (Swearingen, 2009)

Στον σημερινό ψηφιακό κόσμο, τα ηλεκτρονικά μέσα χρησιμοποιούνται ευρέως για διάφορες εργασίες, συμπεριλαμβανομένου του ηλεκτρονικού εμπορίου που διευκολύνεται από το Διαδίκτυο. Τα σχόλια των πελατών θεωρούνται αξιόπιστη πηγή πληροφοριών και οι άνθρωποι συχνά βασίζονται σε αυτά όταν λαμβάνουν αποφάσεις για αγορές. Αυτές οι κριτικές παίζουν σημαντικό ρόλο στη διαδικασία λήψης αποφάσεων του πελάτη και επηρεάζουν το επιχειρηματικό μοντέλο των παρόχων. Η ανάλυση και η κατανόηση των σχολίων συμβάλλει στην οικοδόμηση εμπιστοσύνης με τους αγοραστές και υποστηρίζει την ανάπτυξη της επιχείρησης. Οι πλατφόρμες μέσω κοινωνικής δικτύωσης δημιουργούν τεράστιο όγκο δεδομένων, οδηγώντας σε αυξημένο ενδιαφέρον για την εξόρυξη αυτών των δεδομένων για ανάλυση συναισθήματος. Η ανάλυση των μέσω κοινωνικής δικτύωσης παρέχει πολύτιμες γνώσεις και καινοτόμες ιδέες για την ανάπτυξη προϊόντων. Αυτές οι συνομιλίες προσφέρουν επίσης στις επιχειρήσεις πληροφορίες για το πώς αντιλαμβάνονται οι πελάτες την επωνυμία τους και τις βοηθούν να λαμβάνουν τεκμηριωμένες αποφάσεις για να διατηρήσουν τη φήμη τους (Kuo et al., 2009).

Η πρόβλεψη της αφοσίωσης των πελατών είναι ζωτικής σημασίας για τις εταιρείες να εκτιμήσουν το πιθανό εισόδημά τους από ενεργούς πελάτες. Το κόστος απόκτησης νέων πελατών είναι συνήθως υψηλότερο από το κόστος διατήρησης των υπαρχόντων. Η δυνατότητα πρόβλεψης της αφοσίωσης των πελατών βοηθά στον εντοπισμό πελατών με ισχυρή πίστη και εκείνων που ενδέχεται να στραφούν σε ανταγωνιστές, κάτι που είναι ιδιαίτερα σημαντικό στη σημερινή ανταγωνιστική βιομηχανία τηλεπικοινωνιών (Kuo et al., 2009).

Για να βελτιώσουν την ικανοποίηση και τη διατήρηση των πελατών, οι εταιρείες πρέπει να αναπτύξουν καινοτόμες δραστηριότητες. Η μεγιστοποίηση της κερδοφορίας είναι ένας κοινός στόχος για τις περισσότερες εταιρείες και για να επιτευχθεί αυτό, πρέπει να αναλύσουν τη διαχείριση σχέσεων με τους πελάτες (CRM) και να εφαρμόσουν αποτελεσματικές στρατηγικές μάρκετινγκ. Μελέτες έχουν δείξει ότι η τιμή δεν είναι ο μόνος παράγοντας που επηρεάζει τις αποφάσεις αγοράς των πελατών. τόσο η εταιρεία όσο και ο πελάτης πρέπει να συμφωνήσουν για την αξία του προϊόντος και την καλή εξυπηρέτηση πελατών. Επομένως, η τμηματοποίηση των πελατών με βάση την αγοραστική τους συμπεριφορά είναι απαραίτητη για την ανάπτυξη επιτυχημένων στρατηγικών μάρκετινγκ και την απόκτηση ανταγωνιστικού πλεονεκτήματος.

Πολυάριθμες μελέτες έχουν διερευνήσει τον τρόπο με τον οποίο οι εταιρείες χρησιμοποιούν την ανάλυση μεγάλων δεδομένων για να βελτιώσουν τις δραστηριότητές τους. Για παράδειγμα, η σωστή διαχείριση των δυνατοτήτων μεγάλων δεδομένων μπορεί να ενθαρρύνει την ανάπτυξη πράσινων καινοτόμων προϊόντων και βιώσιμων αλυσίδων εφοδιασμού (Kuo et al., 2009).

Η ανάλυση μεγάλων δεδομένων (BDA) μπορεί να ωφελήσει τους ηλεκτρονικούς πωλητές βελτιώνοντας την αποδοτικότητα του κόστους συναλλαγών στην αγορά, την αποδοτικότητα κόστους διαχείρισης συναλλαγών και την αποδοτικότητα κόστους χρόνου. Ειδικά στο πλαίσιο του ηλεκτρονικού εμπορίου, τα δεδομένα επιτρέπουν στους εμπόρους να παρακολουθούν τη συμπεριφορά μεμονωμένων χρηστών και να βελτιστοποιούν στρατηγικές για τη μετατροπή των εφάπαξ πελατών σε τακτικούς πελάτες. Οι εταιρείες που ενσωματώνουν αναλυτικά στοιχεία μεγάλων δεδομένων στην αλυσίδα αξίας τους μπορούν να επιτύχουν υψηλότερη παραγωγικότητα σε σύγκριση με τους ανταγωνιστές τους. Ενώ έχει γίνει έρευνα για τα θετικά αποτελέσματα των αναλυτικών δεδομένων μεγάλων δεδομένων, έχει δοθεί λιγότερη προσοχή στις αρνητικές επιπτώσεις, όπως ανησυχίες για το απόρρητο και την ασφάλεια, τον εθισμό στα ψώνια και τις ομαδικές επιρροές. Ο αντίκτυπος της ανάλυσης μεγάλων δεδομένων στις απαντήσεις των πελατών, τόσο θετικές όσο και αρνητικές, δεν έχει μελετηθεί εκτενώς (Kuo et al., 2009).

Πριν από την παγκόσμια οικονομική κρίση το 2008, οι πελάτες έτειναν να καταναλώνουν περισσότερα προϊόντα. Ωστόσο, η κρίση οδήγησε τους πελάτες να είναι πιο προσεκτικοί στις αγοραστικές τους αποφάσεις, με αποτέλεσμα τη μείωση των αγορών και την αμυντική συμπεριφορά τους. Σήμερα, οι πελάτες αντιμετωπίζουν έναν τεράστιο όγκο διαφορετικών πληροφοριών, καθιστώντας τη διαδικασία λήψης αποφάσεων πιο περίπλοκη και τη συμπεριφορά τους απρόβλεπτη. Η κατανόηση της συμπεριφοράς των πελατών απαιτεί νέες μεθόδους και η ανάλυση μεγάλων δεδομένων έχει δυνατότητες από αυτή την άποψη (Le et al., 2017).

Προηγούμενες μελέτες είχαν επικεντρωθεί στον αντίκτυπο των αναλυτικών στοιχείων μεγάλων δεδομένων στις επιχειρηματικές αξίες και προκλήσεις, αλλά υπάρχει έλλειψη έρευνας σχετικά με τις απόψεις των πελατών σχετικά με την εφαρμογή των αναλυτικών στοιχείων μεγάλων δεδομένων στις ηλεκτρονικές αγορές. Ως εκ τούτου, η μελέτη των απαντήσεων των πελατών στα πλεονεκτήματα και τα μειονεκτήματα της εφαρμογής ανάλυσης μεγάλων δεδομένων έχει γίνει μια προηγμένη τάση στη στρατηγική μάρκετινγκ (Kuo et al., 2009).

Από την άποψη του μάρκετινγκ, το μοντέλο AIDA (Προσοχή, Ενδιαφέρον, Επιθυμία και Δράση) χρησιμοποιείται συχνά για τη μέτρηση των απαντήσεων των πελατών. Αντιπροσωπεύει τα στάδια από τα οποία οι ηλεκτρονικοί πωλητές οδηγούν τους πελάτες τους στη διαδικασία πώλησης, από την προσέλκυση της προσοχής έως τη δημιουργία ενδιαφέροντος, τη δημιουργία επιθυμίας και, τελικά, την ώθηση της δράσης (Li et al., 2013).

Κεφάλαιο 3

Μελέτη περίπτωσης μιας ηλεκτρονικής πλατφόρμας διανομής φαγητού

3.1 Customer Segmentation

Η τμηματοποίηση πελατών είναι η μέθοδος ομαδοποίησης των πελατών σε υποομάδες με βάση τις ομοιότητες (Choi, Choi, Yoon, & Joung, 2020; Deng & Gao, 2020; Singh & Mittal, 2021; Zeybek, 2018). Μια απλή προσέγγιση στην τμηματοποίηση των πελατών περιλαμβάνει γεωγραφική, δημογραφική, ψυχογραφική και συμπεριφορική τμηματοποίηση. Η συμπεριφορική τμηματοποίηση, η οποία είναι η πιο συχνά χρησιμοποιούμενη μέθοδος, περιλαμβάνει τα ακόλουθα κύρια βήματα:

- (1) Επιχειρηματική κατανόηση και σχεδιασμός διαδικασίας τμηματοποίησης
- (2) Κατανόηση, προετοιμασία και εμπλουτισμός δεδομένων
- (3) Ταυτοποίηση τμημάτων μέσω μοντελοποίησης ομαδοποίησης
- (4) Αξιολόγηση και κατάρτιση προφίλ των αποκαλυφθέντων τμημάτων
- (5) Ανάπτυξη λύσης τμηματοποίησης, σχεδιασμός και παράδοση διαφοροποιημένων στρατηγικών (Chorianopoulos, 2016; Tsirtsis & Chorianopoulos, 2011).

Η τμηματοποίηση μπορεί να χρησιμοποιήσει διάφορους εναλλακτικούς αλγόριθμους τμηματοποίησης, συμπεριλαμβανομένου του αλγορίθμου συσχέτισης, του αλγορίθμου ομαδοποίησης, του αλγορίθμου ταξινόμησης ή του αλγορίθμου παλινδρόμησης (Tsirtsis & Chorianopoulos, 2011). Μεταξύ αυτών, η ομαδοποίηση είναι η πιο ακριβής και αποτελεσματική μέθοδος τμηματοποίησης πελατών.

Η τμηματοποίηση, η στόχευση και η τοποθέτηση μάρκετινγκ είναι μια θεμελιώδης προσέγγιση μάρκετινγκ για τη δημιουργία ανώτερης αξίας πελατών και την υποστήριξη της ανάπτυξης προϊόντων και υπηρεσιών (Gupta, Justy, Kamboj, Kumar & Kristoffersen, 2021; Munusamy & Murugesan, 2020). Το πρώτο πράγμα που έχει μεγαλύτερη σημασία είναι ο εντοπισμός βασικών χαρακτηριστικών για τη διαφοροποίηση κάθε τμήματος της αγοράς που καταλήγει σε στόχευση και τοποθέτηση προϊόντων. Στις ανταγωνιστικές αγορές, οι επιχειρήσεις πρέπει να κατανοήσουν τις προσαρμοσμένες στρατηγικές τους. Ωστόσο, είναι δύσκολο να κατανοήσουμε

σαφώς έναν τεράστιο αριθμό πελατών (Chen, Zhang, Chu, & Yan, 2019; Deng & Gao, 2020). Ως εκ τούτου, οι εταιρείες χρησιμοποιούν συχνά τμηματοποίηση πελατών (που ονομάζεται επίσης τμηματοποίηση της αγοράς) προτού μπορέσουν να εφαρμόσουν στρατηγικές μάρκετινγκ στους πελάτες τους.

Η τμηματοποίηση πελατών χωρίζει τους πελάτες σε ομάδες με βάση τις ομοιότητες τους στις ανάγκες, τα χαρακτηριστικά ή τις συμπεριφορές τους, προκειμένου να διατηρηθούν οι σχέσεις με τους πελάτες και να αυξηθούν τα κέρδη. Επιπλέον, η τμηματοποίηση πελατών είναι εργαλείο για τον εντοπισμό πελατών που αποτελούν την πρωταρχική διαδικασία της κατάλληλης διαχείρισης δημιουργίας σχέσεων μάρκετινγκ των πελατών (CRM) (Deng & Gao, 2020; Dolnicar, Grün, & Leisch, 2018a). Συγκεκριμένα, η τμηματοποίηση πελατών υποστηρίζει τις επιχειρήσεις να ενισχύσουν τη διατήρηση και την αφοσίωση των πελατών και βοηθά επίσης στον προσδιορισμό της αξίας ενός πελάτη (Ballestar, Grau-Carles, & Sainz, 2018; Nandapala & Jayasena, 2020; Wu & Liu, 2020).

3.1.1 RFM Analysis

Η ανάλυση RFM αποτελεί μία κρίσιμη μεθοδολογία στον τομέα της ανάλυσης δεδομένων πελατών, η οποία επικεντρώνεται στην κατανόηση και την κατάταξη των πελατών βάσει τριών κύριων δεικτών: της Προσφατότητας (Recency) - πόσο πρόσφατα έκανε ένας πελάτης μία αγορά, της Συχνότητας (Frequency) - πόσο συχνά ένας πελάτης κάνει αγορές μέσα σε έναν ορισμένο χρονικό διάστημα, και της Οικονομικής Αξίας (Monetary value) - το συνολικό χρηματικό ποσό που έχει ξοδέψει ένας πελάτης. Αυτή η ανάλυση επιτρέπει στις επιχειρήσεις να προσδιορίσουν και να κατανοήσουν τις πρότυπες συμπεριφορές των πελατών τους, να αναγνωρίσουν τις πιο πολύτιμες ομάδες πελατών και να προσαρμόσουν τις στρατηγικές μάρκετινγκ και εξυπηρέτησης πελατών ανάλογα, με σκοπό την ενίσχυση της πελατειακής αφοσίωσης και την αύξηση της συνολικής επιχειρηματικής αποδοτικότητας.

Υπάρχουν διάφορες μέθοδοι τμηματοποίησης πελατών, αλλά οι περισσότερες από αυτές βασίζονται σε συμπεριφορικές, ψυχογραφικές, γεωγραφικές και δημογραφικές πληροφορίες των πελατών. Ωστόσο, οι πληροφορίες συμπεριφοράς των πελατών που βασίζονται στην ανάλυση RFM τονίζονται λόγω της χρήσης ενός μικρού συνόλου χαρακτηριστικών για την τμηματοποίηση των πελατών που ερωτήθηκαν στο Alves Gomes and Meisen (2023). Αυτοί οι παράγοντες θα χρησιμοποιηθούν για την τμηματοποίηση πελατών με πολλούς αλγόριθμους.

Η διαχείριση πελατειακών σχέσεων (CRM) αναφέρεται στις διαδικασίες που χρησιμοποιούν οι εταιρείες για να κερδίσουν την αφοσίωση των πελατών καθ' όλη τη διάρκεια ζωής και έτσι να αυξήσουν το ανταγωνιστικό πλεονέκτημα και τα κέρδη. Λέγεται για το CRM: "η σωστή στιγμή, το σωστό κανάλι, η σωστή τιμή και οι σωστοί πελάτες". Λόγω της ωριμότητας των αγορών ηλεκτρονικών αγορών και των εξελίξεων στην τεχνολογία του Διαδικτύου, οι εταιρείες υιοθετούν όλο και περισσότερο CRM για τη διαχείριση των σχέσεων με τους πελάτες τους (Cheng and Chen 2009). Ένα επιτυχημένο διαδικτυακό CRM όχι μόνο μπορεί να ενισχύσει τις εικονικές αλληλεπιδράσεις μιας εταιρείας με τους πελάτες της, αλλά παρέχει επίσης έναν τρόπο δημιουργίας περισσότερων εσόδων αναπτύσσοντας μια στρατηγική ηλεκτρονικού εμπορίου και επανασχεδιάζοντας τις ιστοσελίδες της. Σύμφωνα με τον κανόνα 80/20 του Pareto (Bult and Wansbeek 1995, Stone and Jacobs 2001), ένα μικρό ποσοστό πελατών συμβάλλει σημαντικά στα έσοδα μιας εταιρείας. Έτσι, είναι καλύτερο να διατηρήσετε εκείνους τους πελάτες που ξοδεύουν τα περισσότερα ή έχουν μείνει στην εταιρεία το μεγαλύτερο χρονικό διάστημα από το να αποκτήσετε νέους πελάτες. Αλλά πώς μπορεί κανείς να διατηρήσει πολύτιμους πελάτες; Οι εταιρείες πρέπει να μάθουν για την αγοραστική συμπεριφορά των πελατών και στη συνέχεια να υιοθετήσουν την πιο αποτελεσματική στρατηγική μάρκετινγκ για κάθε ομάδα πελατών. Οι πιο συχνά χρησιμοποιούμενες στρατηγικές είναι το μάρκετινγκ one-to-one και το σύστημα συστάσεων (Jiang and Tuzhilin 2006). Και οι δύο στρατηγικές εξατομικεύουν προϊόντα και υπηρεσίες. Με το σύστημα συστάσεων, η ιδέα είναι να επιτρέπεται στους πελάτες να αναζητούν λεπτομερείς πληροφορίες που τους ενδιαφέρουν σε εξατομικευμένες ιστοσελίδες που προσφέρουν περιεχόμενο προσαρμοσμένο στο συγκεκριμένο τμήμα της αγοράς τους

3.1.2 Clustering Algorithms

Οι αλγόριθμοι τμηματοποίησης μπορούν να χωριστούν σε τέσσερις κύριες ομάδες. Περιλαμβάνουν αλγόριθμους συσχέτισης (e.g., Apriori, FP growth, ECLAT partition), συσταδοποίησης (e.g., hierarchical clustering, K-means clustering, fuzzy C Means clustering, density-based clustering, affinity propagation clustering), αλγόριθμους ταξινόμησης (e.g., KNN, Naïve Bayes, SVM, C4.5, Decision Tree) και αλγόριθμους παλινδρόμησης (e.g., Logistic regression) (Tsipstsis & Chorianopoulos, 2011).

Η συσταδοποίηση είναι μια προσέγγιση μάθησης χωρίς επίβλεψη και μπορεί να χωριστεί σε δύο κατηγορίες: σκληρή συσταδοποίηση και μαλακή συσταδοποίηση (Singh & Srivastava, 2020). Στο σκληρό σύμπλεγμα, κάθε αντικείμενο εκχωρείται σε ένα μόνο σύμπλεγμα, ενώ το μαλακό σύμπλεγμα μπορεί επίσης να γίνει με επικαλυπτόμενο τρόπο με τον οποίο ένα αντικείμενο μπορεί να είναι μέρος περισσότερων από ένα συμπλεγμάτων. Αφού δημιουργήσουμε μια κατάτμηση χρησιμοποιώντας έναν αλγόριθμο ομαδοποίησης, πρέπει να αξιολογήσουμε την εγκυρότητα ή την ποιότητα αυτής της κατάτμησης. Εάν δημιουργηθούν πολλές κατατμήσεις (π.χ. με διαφορετικό αριθμό συμπλεγμάτων), χρειαζόμαστε τρόπους σύγκρισης πριν ολοκληρώσουμε το αποτέλεσμα ομαδοποίησης. Υπάρχουν τρεις ευρείες κατηγορίες για την εγκυρότητα του συμπλέγματος: εσωτερικοί, εξωτερικοί και σχετικοί δείκτες (Tsipstsis & Chorianopoulos, 2011).

3.1.2.1 K-Means Algorithm

Ο K-means αλγόριθμος είναι η πιο δημοφιλής μέθοδος διαμέρισης ομαδοποίησης. Προτάθηκε για πρώτη φορά από τον MacQueen το 1967. Ο K-means είναι μια μη εποπτευόμενη, μη ντετερμινιστική, αριθμητική, επαναληπτική μέθοδος ομαδοποίησης. Στον K-means αλγόριθμο κάθε σύμπλεγμα αντιπροσωπεύεται από τη μέση τιμή των αντικειμένων στο σύμπλεγμα. Εδώ χωρίζουμε ένα σύνολο n αντικειμένων σε k clusters έτσι ώστε η ομοιότητα μεταξύ συμπλεγμάτων να είναι χαμηλή και η ομοιότητα εντός συμπλέγματος να είναι υψηλή. Η ομοιότητα μετράται με βάση τη μέση τιμή των αντικειμένων σε ένα σύμπλεγμα .

Ο αλγόριθμος αποτελείται από δύο ξεχωριστές φάσεις.

1η φάση

Επιλέξτε k κεντροειδές τυχαία, όπου η τιμή k είναι σταθερή εκ των προτέρων.

2η Φάση

Κάθε αντικείμενο στο σύνολο δεδομένων συσχετίζεται με το πλησιέστερο κεντροειδές. Η Ευκλείδεια απόσταση χρησιμοποιείται για τη μέτρηση της απόστασης μεταξύ κάθε αντικειμένου δεδομένων και κεντροειδούς συμπλέγματος.

Είσοδος

K : αριθμός επιθυμητών συστάδων

D : $\{d_1, d_2, \dots, d_n\}$ ένα σύνολο δεδομένων που περιέχει n αντικείμενα.

Έξοδος

Ένα σύνολο k συστάδων όπως καθορίζεται στην είσοδο

Μέθοδος

1) Επιλέξτε αυθαίρετα το στοιχείο δεδομένων k από το σύνολο δεδομένων D ως αρχικό κεντροειδές σύμπλεγμα.

2) Επαναλάβετε

3) Αντιστοιχίστε κάθε στοιχείο δεδομένων d_i στο σύμπλεγμα με το οποίο το αντικείμενο είναι πιο παρόμοιο με βάση τη μέση τιμή του αντικειμένου στο σύμπλεγμα.

4) Υπολογίστε τη νέα μέση τιμή των στοιχείων δεδομένων για κάθε σύμπλεγμα και ενημερώστε τη μέση τιμή.

5) Μέχρι να μην υπάρχει καμία αλλαγή.

3.1.2.2 DBSCAN Algorithm (Density-Based Spatial Clustering of Applications with Noise)

Ο DBSCAN είναι ο πρώτος αλγόριθμος ομαδοποίησης που βασίζεται στην πυκνότητα. Αυτό προτάθηκε από τους Ester et al. το 1996 και σχεδιάστηκε για να συσταδοποιήσει δεδομένα αυθαίρετων σχημάτων παρουσία θορύβου σε χωρικές και μη χωρικές βάσεις δεδομένων υψηλών διαστάσεων.

Σε σύγκριση με τις μεθόδους ομαδοποίησης που δεν βασίζονται στην πυκνότητα, ο DBSCAN αλγόριθμος διαθέτει μοναδικές και προηγμένες δυνατότητες που είναι χρήσιμες στην ανίχνευση αντικειμένων/κλάσεων/μοτίβων/δομών διαφορετικών σχημάτων και μεγεθών. Ο DBSCAN είναι πολύ καλός αλγόριθμος για την εύρεση «φυσικών» συστάδων και της διάταξής τους εντός του χώρου δεδομένων όταν έχουν συγκρίσιμη πυκνότητα χωρίς προκαταρκτικές πληροφορίες σχετικά με τις ομάδες που υπάρχουν σε ένα σύνολο δεδομένων.

Αρχικά, ο αλγόριθμος DBSCAN προτάθηκε για ομαδοποίηση χωρικών δεδομένων. Λόγω των μοναδικών χαρακτηριστικών του, ο αλγόριθμος έγινε γρήγορα δημοφιλής και εφαρμόστηκε ως προσέγγιση ομαδοποίησης για άλλους τύπους δεδομένων. Μερικά παραδείγματα εφαρμογών του, περιλαμβάνουν εφαρμογές σε τομείς της επιστήμης όπως σε έργα πολιτικών μηχανικών (ομαδοποίηση χωροταξικών δικτύων υποδομής), στη χημεία, στη φασματοσκοπία (ομαδοποίηση φασμάτων μάζας ενός σωματιδίου) και φασματομετρία μάζας χρόνου πτήσης αερολύματος, στις κοινωνικές επιστήμες (ομαδοποίηση εντόμων βάσει δεδομένων χημικής φερομόνης) και σε ιατρική διάγνωση βασισμένη σε ιατρικές εικόνες (για την ανίχνευση μοτίβων ατροφίας του εγκεφάλου και για την ανίχνευση δερματικών βλαβών). Επίσης μπορεί να εφαρμοστεί στον τομέα της τηλεπισκόπησης για την τμηματοποίηση τρισδιάστατων εικόνων (υπερφασματικές εικόνες).

Η βασική ιδέα του DBSCAN είναι ότι για κάθε αντικείμενο ενός συμπλέγματος η γειτονιά δεδομένης ακτίνας (Eps) πρέπει να περιέχει τουλάχιστον έναν ελάχιστο αριθμό αντικειμένων ($MinPts$), που σημαίνει ότι η πληθικότητα της γειτονιάς πρέπει να υπερβεί κάποιο όριο. Η ϵ -γειτονιά ενός αυθαίρετου σημείου 'p' ορίζεται ως,

$$N_{Eps} = \{q \in D / dist(p, q) < Eps\} \quad (1)$$

Εδώ, το D είναι η βάση δεδομένων των αντικειμένων. Αν οι ϵ -γειτονιές ενός σημείου P περιέχουν τουλάχιστον έναν ελάχιστο αριθμό σημείων, και στη συνέχεια, αυτό το σημείο ονομάζεται βασικό σημείο. Το βασικό σημείο ορίζεται ως:

$$N_{Eps}(P) > MinPts \quad (2)$$

Εδώ τα Eps και $MinPts$ είναι οι καθορισμένες παράμετροι του χρήστη που ορίζουν την ακτίνα της γειτονιάς και τον ελάχιστο αριθμό σημείων στη ϵ -γειτονιά ενός βασικού σημείου αντίστοιχα. Εάν δεν πληρείται αυτή η προϋπόθεση, τότε αυτό το σημείο θεωρείται ως μη βασικό σημείο.

3.1.2.3 Mean Shift Algorithm

Οι μέθοδοι ανάλυσης χώρου χαρακτηριστικών που βασίζονται σε gradients χρησιμοποιούν gradients της συνάρτησης πυκνότητας πιθανότητας για να βρουν τα μέγιστα. Οι μέθοδοι αυτές είναι πολύπλοκες λόγω, μεταξύ άλλων, της ανάγκης για εκτίμηση της πιθανότητας πυκνότητας.

Οι μέθοδοι ανάλυσης χώρου χαρακτηριστικών που βασίζονται σε gradients πρώτα το gradient και στη συνέχεια, ο πυρήνας μετατοπίζεται από ένα συγκεκριμένο διάνυσμα μήκους προς την κατεύθυνση μιας μέγιστης αύξησης της πυκνότητας. Η σπουδαιότητα βρίσκεται στο μέγεθος βήματος που πρέπει να επιλεγεί κατάλληλα, δηλαδή πώς θα επιλεγεί ένα κατάλληλο μέγεθος βήματος, επειδή ένα μικρό μέγεθος βήματος θα επιβραδύνει τη σύγκλιση.

Ο αλγόριθμος μέσης μετατόπισης λύνει το κύριο πρόβλημα των μεθόδων κλίσης. Η κύρια ιδέα του μέση μετατόπιση είναι η αντιμετώπιση των σημείων στο D-διαστάσεων χαρακτηριστικό χώρο ως εμπειρική πυκνότητα πιθανότητας συνάρτηση όπου οι πυκνές περιοχές αντιστοιχούν στα τοπικά μέγιστα της υποκείμενης κατανομής.

Η ανάβαση κλίσης εκτελείται στο χώρο χαρακτηριστικών στην εκτίμηση τοπικής πυκνότητας μέχρι τη σύγκλιση. Μετά την διαδικασία, τα σταθερά σημεία αντιστοιχούν στους τρόπους διανομής και τα ίδια στάσιμα σημεία θεωρούνται μέλη του ίδιου συμπλέγματος.

Το μέγεθος βήματος της μέσης μετατόπισης είναι προσαρμοστικό και εξαρτάται από την κλίση της πυκνότητας της πιθανότητας. Η κλίση δεν υπολογίζεται, αντ' αυτού, υπολογίζεται ένα πιο αποτελεσματικό διάνυσμα μέσης μετατόπισης. Το διάνυσμα μέσης μετατόπισης δείχνει προς την ίδια κατεύθυνση με την κλίση στις μεθόδους που βασίζονται σε κλίση. Σε αντίθεση με τη γνωστή προσέγγιση ομαδοποίησης μέσω K , η μέση μετατόπιση δεν χρειάζεται υποθέσεις σχετικά με τον αριθμό των συμπλεγμάτων και το σχήμα της κατανομής, αλλά η απόδοσή της εξαρτάται από την επιλογή παραμέτρων κλίμακας. Το εύρος ζώνης είναι η μόνη παράμετρος για συντονισμό, οπότε για τη μονοδιάστατη περίπτωση αυτό είναι μια σχετικά απλή διαδικασία, αλλά σε μια πολυδιάστατη περίπτωση, μπορεί να είναι δύσκολη. Η μέση μετατόπιση μπορεί να μη λειτουργεί καλά σε υψηλότερες διαστάσεις.

Η διαδικασία μέσης μετατόπισης αποτελείται από δύο στάδια:

1. Κατασκευή πυκνότητας πιθανότητας σε κάποιο χώρο χαρακτηριστικών,
 2. Τη χαρτογράφηση κάθε σημείου στο μέγιστο της πυκνότητας που βρίσκεται πλησιέστερα σε αυτό.
- Κάθε σημείο δεδομένων μετατοπίζεται στον σταθμισμένο μέσο όρο του συνόλου δεδομένων. Ο αλγόριθμος μέσης μετατόπισης προσπαθεί να βρει σταθερά σημεία μιας εκτιμώμενης συνάρτησης πυκνότητας πιθανότητας (PDF).

3.1.2.4 Gaussian Mixture Model Algorithm (GMM)

Η μοντελοποίηση Gaussian Mixture είναι ένα σημαντικό έργο στον τομέα της ομαδοποίησης και της αναγνώρισης προτύπων. Υπάρχουν διάφορες προσεγγίσεις για την εκτίμηση των παραμέτρων ενός μείγματος, όπως γραφικές μέθοδοι, αντιστοίχιση ροπών, μπεϋζιανές προσεγγίσεις και η μέθοδος εκτίμησης μέγιστης πιθανότητας (MLE). Μια ευρέως χρησιμοποιούμενη προσέγγιση για το τελευταίο πρόβλημα είναι ο αλγόριθμος Expectation Maximization (EM). Ξεκινώντας με κάποιο αρχικό μοντέλο μείγματος, ο αλγόριθμος EM εναλλάσσεται μεταξύ του υπολογισμού ενός κατώτερου ορίου της λογαριθμικής πιθανότητας και της βελτίωσης του τρέχοντος μοντέλου σε σχέση με αυτό το κατώτερο όριο. Ο αλγόριθμος συγκλίνει σε ένα συγκεκριμένο σταθερό σημείο της συνάρτησης πιθανότητας. Δυστυχώς, η συνάρτηση πιθανότητας είναι γενικά μη κυρτή, διαθέτοντας πολλά σταθερά σημεία, συμπεριλαμβανομένων μικρών τοπικών μεγίστων, και ακόμη χειρότερα, τοπικών ελάχιστων και ελάχιστων σημείων (saddle points). Επιπλέον, η σύγκλιση του αλγορίθμου EM σε οποιονδήποτε τύπο σημείου εξαρτάται σε μεγάλο βαθμό από το επιλεγμένο αρχικό μοντέλο.

3.1.2.5 BIRCH Algorithm (Balanced Iterative Reducing and Clustering using Hierarchies)

Ο αλγόριθμος BIRCH είναι μια μέθοδος ομαδοποίησης και είναι ιδιαίτερα κατάλληλη για πολύ μεγάλες βάσεις δεδομένων. Το κόστος εισόδου/εξόδου είναι γραμμικό στο μέγεθος του συνόλου δεδομένων: μία μόνο σάρωση του συνόλου δεδομένων δίνει μια καλή ομαδοποίηση και ένα ή περισσότερα πρόσθετα περάσματα μπορούν (προαιρετικά) να χρησιμοποιηθούν για την περαιτέρω βελτίωση της ποιότητας. Αξιολογώντας την αποδοτικότητα χρόνου / χώρου του BIRCH, την ευαισθησία σειράς εισαγωγής δεδομένων και την ποιότητα ομαδοποίησης και συγκρίνοντας με άλλους υπάρχοντες αλγόριθμους μέσω πειραμάτων, υποστηρίζουμε ότι ο αλγόριθμος BIRCH είναι η καλύτερη διαθέσιμη μέθοδος ομαδοποίησης για πολύ μεγάλες βάσεις δεδομένων.

Η αρχιτεκτονική του BIRCH προσφέρει επίσης ευκαιρίες για παραλληλισμό και για διαδραστικό ή δυναμικό συντονισμό απόδοσης με βάση τη γνώση σχετικά με το σύνολο δεδομένων, που αποκτήθηκε κατά τη διάρκεια της εκτέλεσης.

Τέλος, ο αλγόριθμος Birch είναι η πρώτη μέθοδος ομαδοποίησης που προτείνεται στην περιοχή της βάσης δεδομένων που αντιμετωπίζει ακραίες τιμές (δισαιθητικά, σημεία δεδομένων που, θα πρέπει να θεωρούνται ως "θόρυβος") και προτείνει μια εύλογη λύση.

3.1.2.6 Affinity Propagation Algorithm

Η ομαδοποίηση διάδοσης συνάφειας (AP) είναι ένας αλγόριθμος γρήγορης ομαδοποίησης, ειδικά στην περίπτωση μεγάλου αριθμού συμπλεγμάτων, και έχει ορισμένα πλεονεκτήματα: ταχύτητα, γενική εφαρμοσιμότητα και καλή απόδοση. Το AP λειτουργεί με βάση τις ομοιότητες μεταξύ ζευγών σημείων δεδομένων (ή $n \times n$ πίνακα ομοιότητας S για n σημεία δεδομένων) και ταυτόχρονα θεωρεί όλα τα σημεία δεδομένων ως πιθανά κέντρα συμπλέγματος (που ονομάζονται παραδείγματα). Για να βρει τα κατάλληλα παραδείγματα, η AP συσσωρεύει αποδεικτικά στοιχεία "ευθύνης" $R(i,k)$ από το σημείο δεδομένων i για το πόσο κατάλληλο είναι το σημείο k για να χρησιμεύσει ως παράδειγμα για το σημείο i , και συσσωρεύει αποδεικτικά στοιχεία "διαθεσιμότητας" $A(i,k)$ από το υποψήφιο παράδειγμα σημείο k για το πόσο κατάλληλο θα ήταν για το σημείο i να επιλέξει το σημείο k ως υπόδειγμά του. Από την άποψη των αποδεικτικών στοιχείων, όσο μεγαλύτερο το $R(:,k)+A(:,k)$, τόσο μεγαλύτερη πιθανότητα το σημείο k να είναι το τελικό κέντρο συμπλέγματος. Με βάση τη συσσώρευση αποδεικτικών στοιχείων, το AP αναζητά συστάδες μέσω μιας επαναληπτικής διαδικασίας μέχρι να προκύψει ένα υψηλής ποιότητας σύνολο υποδειγμάτων και αντίστοιχων συστάδων. Στην επαναληπτική διαδικασία, τα αναγνωρισμένα παραδείγματα ξεκινούν από τα μέγιστα n παραδείγματα σε λιγότερα παραδείγματα έως ότου εμφανιστούν m παραδείγματα και δεν αλλάζουν πλέον (ή συγκλίνει ο αλγόριθμος AP). Οι m συστάδες που βρέθηκαν με βάση τα m παραδείγματα είναι η λύση ομαδοποίησης του AP.

3.1.2.7 OPTICS Algorithm (Ordering Points To Identify the Clustering Structure)

Ο αλγόριθμος OPTICS είναι αλγόριθμος ομαδοποίησης βάσει πυκνότητας, δημιουργώντας μια διάταξη των σημείων που επιτρέπει την εξαγωγή συστάδων με αυθαίρετες τιμές για ϵ . Η παράμετρος ϵ είναι μια απόσταση, είναι η ακτίνα γειτονιάς. Επομένως, προκειμένου να δημιουργήσουμε ένα σύνολο ή μια σειρά συμπλεγμάτων βάσει πυκνότητας, παρέχουμε ένα σύνολο τιμών παραμέτρων απόστασης. Για την ταυτόχρονη κατασκευή των διαφορετικών συμπλεγμάτων, τα αντικείμενα πρέπει να υποβληθούν σε επεξεργασία με συγκεκριμένη σειρά. Αυτή η σειρά επιλέγει ένα αντικείμενο που είναι προσβάσιμο από την πυκνότητα σε σχέση με τη χαμηλότερη τιμή, έτσι ώστε τα συμπλέγματα με υψηλότερη πυκνότητα (χαμηλότερο ϵ) να ολοκληρωθούν πρώτα. Η απόσταση παραγωγής ϵ είναι η μεγαλύτερη απόσταση που λαμβάνεται υπόψη για συστάδες. Τα συμπλέγματα μπορούν να εξαχθούν για όλα τα ϵ_i τέτοια ώστε $0 \leq \epsilon_i \leq \epsilon$. Με βάση αυτή την ιδέα, πρέπει να αποθηκευτούν δύο τιμές για κάθε απόσταση πυρήνα αντικειμένου και την απόσταση προσβασιμότητας. Η απόσταση πυρήνα ενός αντικειμένου p είναι η μικρότερη τιμή ϵ' που καθιστά το p ως βασικό αντικείμενο. Εάν το p δεν είναι βασικό αντικείμενο, η απόσταση πυρήνα του p είναι απροσδιόριστη. Η απόσταση προσβασιμότητας ενός αντικειμένου p και ενός άλλου αντικειμένου q είναι η μεγαλύτερη τιμή της απόστασης πυρήνα του p και της Ευκλείδειας απόστασης μεταξύ p και q . Εάν το p δεν είναι βασικό αντικείμενο, η απόσταση προσβασιμότητας μεταξύ p και q είναι απροσδιόριστη. Ένα διάγραμμα προσβασιμότητας για ένα απλό σύνολο δεδομένων 2 διαστάσεων, το οποίο δείχνει ότι τα δεδομένα είναι ομαδοποιημένα.

3.1.2.8 Agglomerative Hierarchical Clustering Algorithm

Η συσσωρευτική ιεραρχική ομαδοποίηση είναι μια μέθοδος ομαδοποίησης από κάτω προς τα πάνω, όπου οι συστάδες έχουν υποσυστάδες, οι οποίες με τη σειρά τους έχουν υποσυστάδες κ.λπ. Το

κλασικό παράδειγμα αυτού είναι η ταξινόμηση των ειδών. Τα δεδομένα γονιδιακής έκφρασης μπορεί επίσης να παρουσιάζουν αυτή την ιεραρχική ποιότητα (π.χ. οικογένειες γονιδίων νευροδιαβιβαστών). Η συσσωρευτική ιεραρχική ομαδοποίηση ξεκινά με κάθε αντικείμενο (γονίδιο ή δείγμα) σε ένα μόνο σύμπλεγμα. Στη συνέχεια, σε κάθε διαδοχική επανάληψη, συγκεντρώνει (συγχωνεύει) το πλησιέστερο ζεύγος συστάδων ικανοποιώντας ορισμένα κριτήρια ομοιότητας, έως ότου όλα τα δεδομένα είναι σε ένα σύμπλεγμα. Στα πλεονεκτήματα αυτού του αλγορίθμου είναι ότι μπορεί να παράγει μια διάταξη των αντικειμένων, η οποία μπορεί να είναι ενημερωτική για την εμφάνιση δεδομένων. Δημιουργούνται μικρότερα σμήνη, τα οποία μπορεί να είναι χρήσιμα για την περαιτέρω ανακάλυψη.

Υπάρχουν μερικοί πιθανοί τρόποι για να εκτελέσετε συγκεντρωτική ιεραρχική ομαδοποίηση. Ωστόσο, ακολουθούν γενικά τα ακόλουθα σημαντικά βήματα:

1. Υπολογισμός του πίνακα εγγύτητας για τις αρχικές συστάδες που είναι η έξοδος διαδικασία K-means.
2. Αναζήτηση της ελάχιστης απόστασης στη μήτρα.
3. Συνδυασμός των δύο συστάδων με την ελάχιστη απόσταση.
4. Ενημέρωση του πίνακα εγγύτητας υπολογίζοντας τις αποστάσεις μεταξύ του νέου συμπλέγματος με τα άλλα σμήνη
5. Επανάληψη των προηγούμενων τριών βημάτων εάν παραμένουν περισσότερα από ένα συμπλέγματα.

Η πολυπλοκότητα μιας τέτοιας διαδικασίας είναι τουλάχιστον $O(n^2)$ και μπορεί να φτάσει μέχρι το $O(n^2 \log n)$.

3.2 Data management

3.2.1 Data cleansing and preparation

Το βήμα προεπεξεργασίας δεδομένων χειρίζεται ακραίες τιμές, συμπληρώνει τιμές που λείπουν και πραγματοποιεί μείωση στις διαστάσεις, μετασχηματισμό, δημιουργία ιεραρχίας εννοιών, κανονικοποίηση και διακριτοποίηση.

Το αρχείο δεδομένων των εστιατορίων της ψηφιακής πλατφόρμας διανομής φαγητού περιέχει 24 μεταβλητές όπως φαίνεται και στον πίνακα 1, και περιέχει όλες τις παραγγελίες πελατών για τους μήνες Απρίλιο και Μάιο του 2021. Ο όγκος του αρχικού αρχείου δεδομένων είναι 1448011 εγγραφές. Οι διακριτοί πελάτες είναι 24998 και οι διακριτές κατηγορίες προϊόντων είναι 17054.

Πίνακας 2: Μεταβλητές Αρχείου Δεδομένων

Variable Name	Data Type	Description
Var1	Numeric	Primary Key
Customer_id	Nominal	A unique code on an invoice that is used to reference a customer's account.
Order_id	Nominal	A unique identifier that is assigned to an order when it is placed.
Item_name	Nominal	The name of the item ordered.
Category_name	Nominal	The name of the category of the item ordered.
Item_price	Numeric	The cost of the item.
Item_quantity	Numeric	The quantity of the item in the order.
Shop_id	Nominal	A unique identifier that each shop in the platform has.

Shop_is_chain	Nominal	If a shop is part of a chain of similar stores then the variable get the value 'True', else 'False'
Vertical	Nominal	Vertical markets are business niches where vendors serve a specific audience and their set of needs. In this dataset vertical is always restaurants.
Delivered_by_efood	Nominal	If an order is delivered by efood, then the variable get the value 'True', else 'False'.
Shop_Cuisine	Nominal	The cuisine that the shop serves.
Is_online_payment	Nominal	If an order is payed online, then the variable get the value 'True', else 'False'.
Order_has_discount_coupon	Nominal	If an order has discount coupon, then the variable get the value 'True', else 'False'.
Order_from_pinata	Nominal	If an order is made from pinata choice, then the variable get the value 'True', else 'False'.
Geographical_region_of_user	Nominal	The geographical region of the user.
Shop_city	Nominal	The city where the shop is located.
Geographical_region_of_shop	Nominal	The geographical region of the shop.
Food_score	Numeric	The score that the customer gave for the food quality.
Service_Score	Numeric	The score that the customer gave for the service quality.
Delivery_score	Numeric	The score that the customer gave for the delivery quality.
Order_timestamp	Date	The date that the order has been made.
First_order_at	Date	The date that this customer made his first order.
Last_order_at	Date	The date that this customer made his last order.

Λόγω του μεγάλου όγκου των δεδομένων (1448011 εγγραφές) και το γεγονός ότι η υλοποίηση της ανάλυσης γίνεται στο εκπαιδευτικό SAS Viya, υπήρχε η ανάγκη διαχωρισμού του αρχείου δεδομένων σε πέντε αρχεία τύπου csv. Αυτό έγινε με τη βοήθεια της Python και της βιβλιοθήκης Pandas στην πλατφόρμα Jupyter Notebook. Στη συνέχεια έγινε η μεταφόρτωση των 5 αρχείων csv στο SAS Viya.

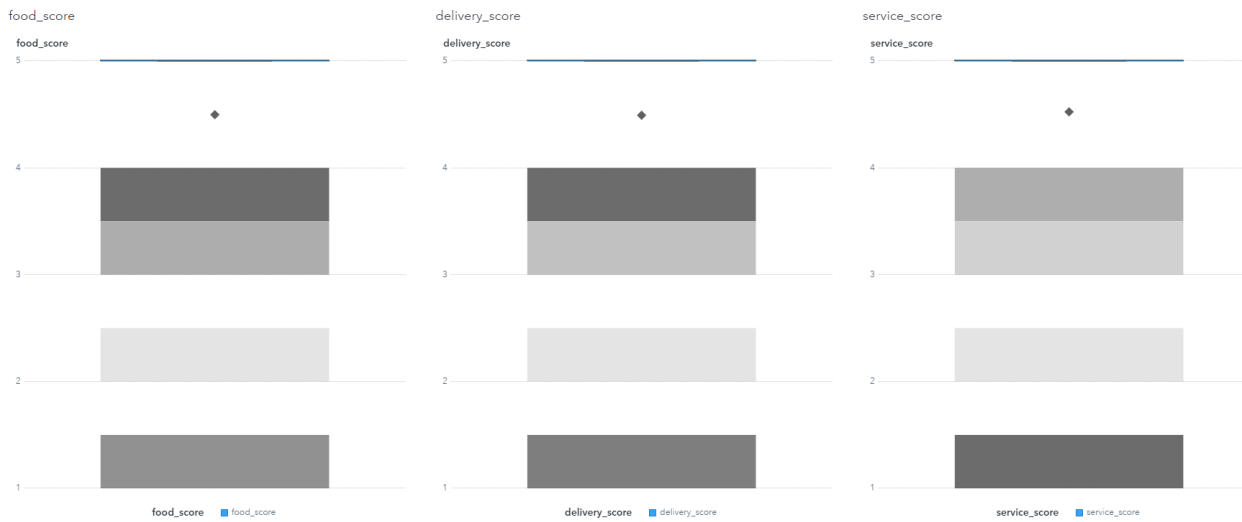
Ανοίγοντας κάθε ένα από τα πέντε επιμέρους αρχεία, παρατηρήθηκαν κάποιες ανωμαλίες στα δεδομένα και κρίθηκε σκόπιμο να διορθωθούν. Η στήλη order_timestamp ήταν της μορφής 2021-04-29 04:27:45 UTC οπότε έπρεπε να αφαιρεθεί το λεκτικό 'UTC', δηλαδή να κρατηθούν οι 19 πρώτοι χαρακτήρες της στήλης και στη συνέχεια να μετατραπεί από στήλη χαρακτήρα σε αριθμητικό με το κατάλληλο format για ημερομηνίες. Αντίστοιχη διαδικασία υλοποιήθηκε για τις στήλες First_order_at και Last_order_at. Επίσης χρειάστηκε να μετατραπούν οι στήλες food_score, delivery_score και

service_score από στήλες χαρακτήρα σε αριθμητικό, διαγράφηκαν οι αρχικές στήλες και μετονομάστηκαν στις αρχικές τους ονομασίες. Τέλος παρατηρήθηκε επίσης η ύπαρξη πολλαπλών ίδιων εγγραφών οπότε και κρίθηκε απαραίτητη η απομάκρυνση τους με τη βοήθεια του proc sql.

3.2.2 Data exploration and visualization

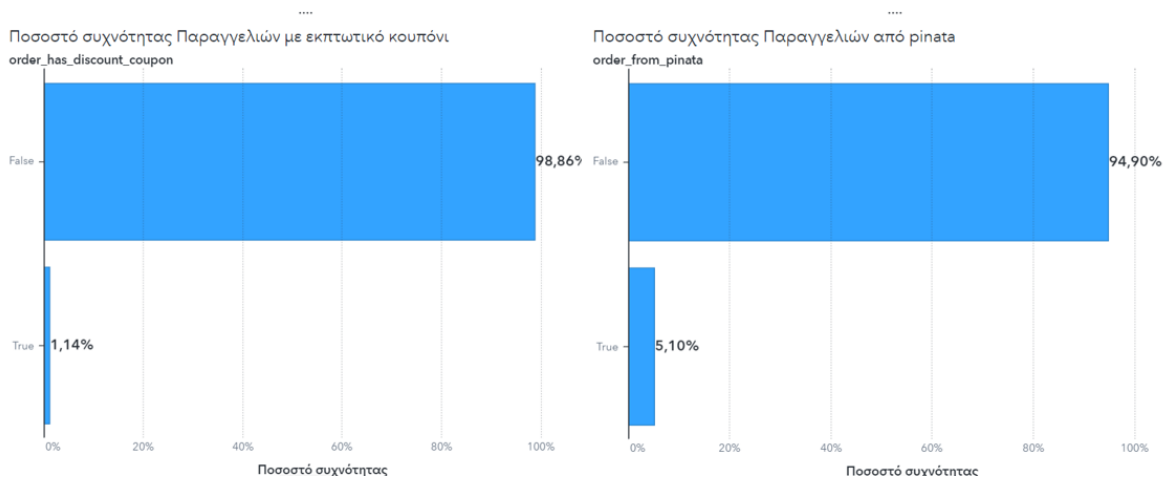
Στη συνέχεια μέσα από τη πλατφόρμα SAS Viya, αξιοποιούμε την επιλογή της 'Εξερεύνησης και Απεικόνισης' για να δούμε πιο πολλά στοιχεία οπτικοποιημένα, για τις επιμέρους μεταβλητές του dataset, δηλαδή αν υπάρχουν ακραίες τιμές, κενές τιμές, κτλ. Δημιουργήθηκε λοιπόν ένα θηκόγραμμα (boxplot) για τις μεταβλητές food_score, delivery_score και service_score. Σε αυτό παρατηρούμε ότι το μεγαλύτερο πλήθος των τιμών βρίσκεται κοντά στο πάνω άκρο, που σημαίνει ότι τα περισσότερα εστιατόρια σύμφωνα με το πλήθος των παραγγελιών, έχουν αξιολογηθεί με πολύ υψηλές βαθμολογίες σε ποιότητα φαγητού, εξυπηρέτησης και διανομής.

Πίνακας 3: Food Score, Delivery Score and Service Score BoxPlot



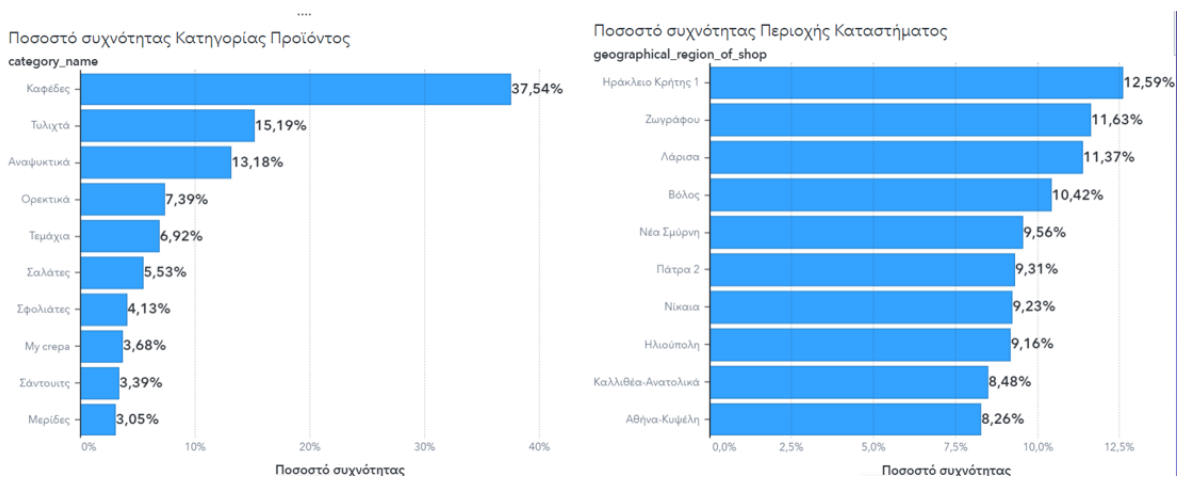
Τα παρακάτω διάγραμμα παρουσιάζουν το μικρό πλήθος παραγγελιών που έχουν γίνει με προσφορά εκπτωτικό κουπόνι και pinata. Αυτό ίσως οφείλεται στις χαμηλές αξιολογήσεις που ως επί το πλείστον έχουν τα εστιατόρια που προσφέρουν pinata προσφορές, στην έλλειψη επαρκούς ενημέρωσης των πελατών για τα προσφερόμενα εκπτωτικά κουπόνια ή στο γεγονός ότι υπάρχει ελάχιστο ποσό χρέωσης για την εφαρμογή οποιουδήποτε εκπτωτικού κουπονιού.

Πίνακας 4: Μικρό πλήθος παραγγελιών



Τα επόμενα διαγράμματα εμφανίζουν το ποσοστό συχνότητας των 10 κορυφαίων κατηγοριών προϊόντων και το ποσοστό συχνότητας των 10 κορυφαίων περιοχών καταστημάτων.

Πίνακας 5: Κορυφαίες επιλογές σε καταλόγους των εστιατορίων



3.2.3 Strategic examination of customer behavior based on Recency, Frequency, and Monetary metrics

Στο ανταγωνιστικό τοπίο της βιομηχανίας των διαδικτυακών πλατφορμών διανομής φαγητού, η κατανόηση και η αποτελεσματική στόχευση των πελατειακών ομάδων είναι ζωτικής σημασίας για τη βιώσιμη επιτυχία. Η ανάλυση RFM (Recency, Frequency, Monetary) είναι ένα πολύτιμο εργαλείο στην τμηματοποίηση πελατών, επιτρέποντας στις επιχειρήσεις να εντοπίζουν και να καλύπτουν τις μοναδικές ανάγκες διαφορετικών ομάδων πελατών. Αυτή η ανάλυση είναι ιδιαίτερα χρήσιμη στο λογισμικό SAS, ένα ισχυρό αναλυτικό εργαλείο, για την απόκτηση αξιοποιήσιμων πληροφοριών από δεδομένα.

Τι είναι η ανάλυση RFM;

1. Recency (R)

Το recency μετρά πόσο πρόσφατα ένας πελάτης έχει αλληλεπιδράσει με ένα εστιατόριο. Στο πλαίσιο της SAS, η μέτρηση του πρόσφατου recency μπορεί να υπολογιστεί προσδιορίζοντας τη χρονική διαφορά μεταξύ της ημερομηνίας τελευταίας παραγγελίας ή συναλλαγής και της τρέχουσας ημερομηνίας. Αυτό βοηθά στον εντοπισμό των πελατών που είναι ενεργοί αυτήν τη στιγμή και εκείνων που μπορεί να χρειάζονται εκ νέου προσέγγιση.

2. Frequency (F)

Το frequency αξιολογεί πόσο συχνά ένας πελάτης αλληλεπιδρά με ένα εστιατόριο. Χρησιμοποιώντας τη SAS, το frequency καθορίζεται μετρώντας τον αριθμό των συναλλαγών ή των παραγγελιών που πραγματοποιούνται από κάθε πελάτη. Αυτή η μέτρηση προσδιορίζει τους πιο πιστούς και συχνούς πελάτες του εστιατορίου.

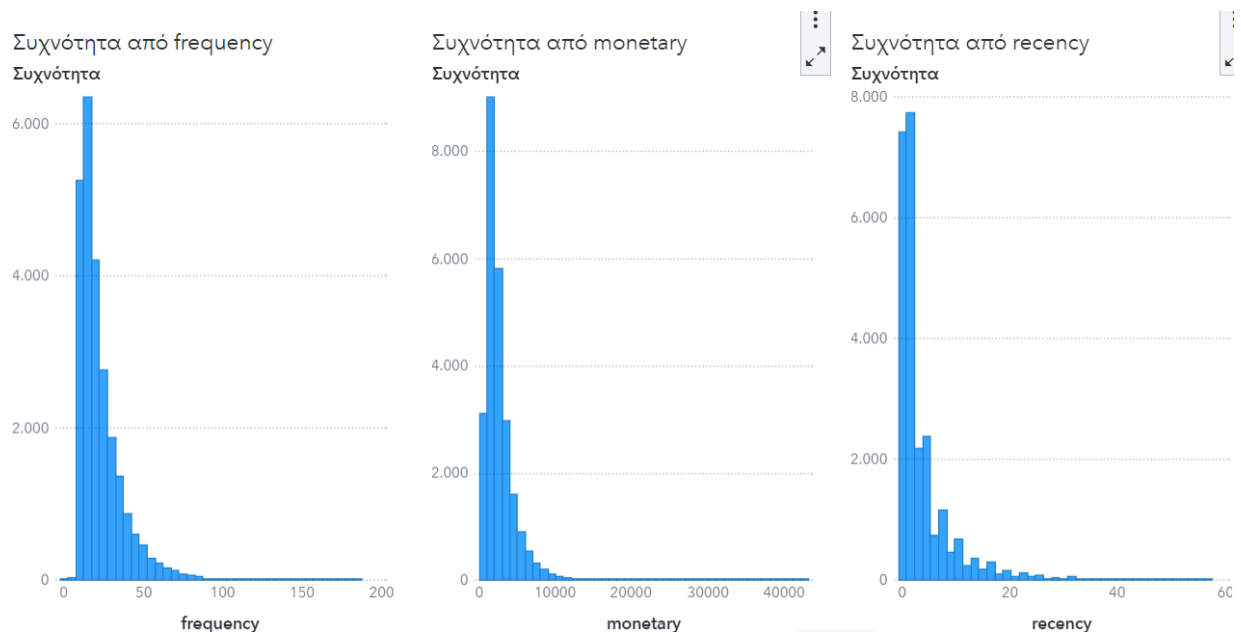
3. Monetary (M)

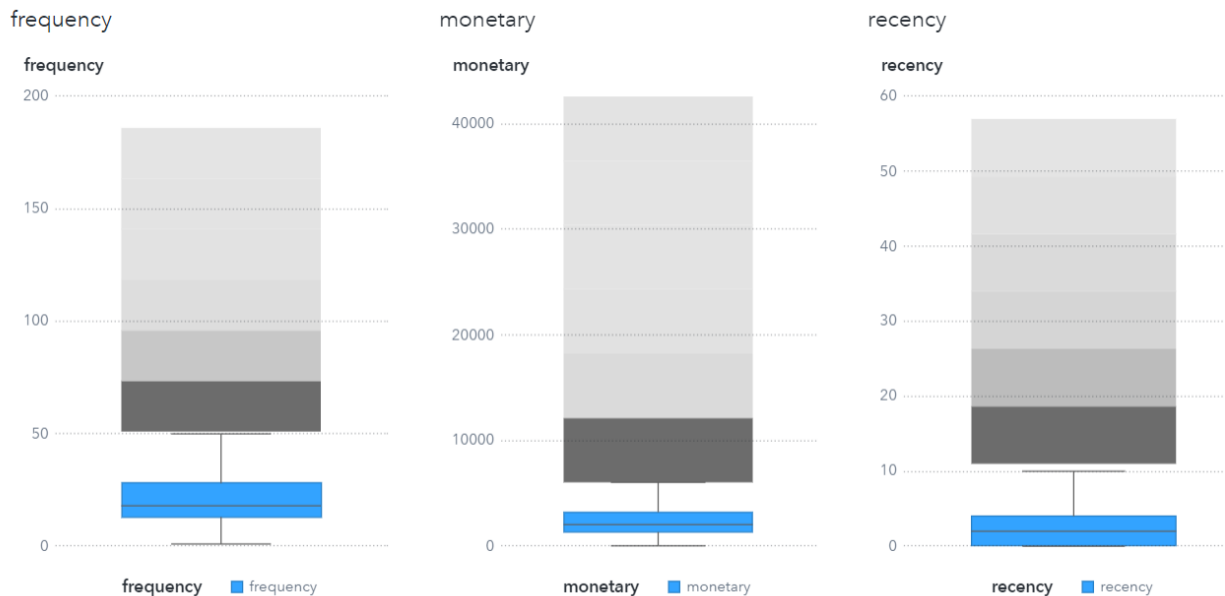
Το monetary αντιπροσωπεύει το συνολικό χρηματικό ποσό που δαπανάται από έναν πελάτη. Στη SAS, αυτό υπολογίζεται αθροίζοντας τη χρηματική αξία κάθε συναλλαγής. Βοηθά στον εντοπισμό πελατών υψηλής αξίας που συμβάλλουν σημαντικά στα έσοδα του εστιατορίου.

Πιο συγκεκριμένα, η μεταβλητή Recency κατασκευάστηκε υπολογίζοντας το χρονικό διάστημα μεταξύ της τελευταίας ημερομηνίας συναλλαγής και της ώρας παραγγελίας για κάθε πελάτη. Η μεταβλητή Frequency δημιουργήθηκε βρίσκοντας τον αριθμό των διακριτών συναλλαγών που έχει πραγματοποιήσει κάθε πελάτης εντός της συγκεκριμένης περιόδου. Η μεταβλητή Monetary κατασκευάστηκε υπολογίζοντας το σωρευτικό σύνολο των χρημάτων που δαπανήθηκαν από κάθε πελάτη με βάση την τιμή και την ποσότητα των προϊόντων.

Στα παρακάτω ιστογράμματα των μεταβλητών Recency, Frequency και Monetary παρατηρείται ότι τα ιστογράμματα αυτά είναι skewed προς τα δεξιά και ίσως στη συνέχεια χρειαστεί να χρησιμοποιήσουμε τον λογάριθμο για διευκόλυνση της ανάλυσης.

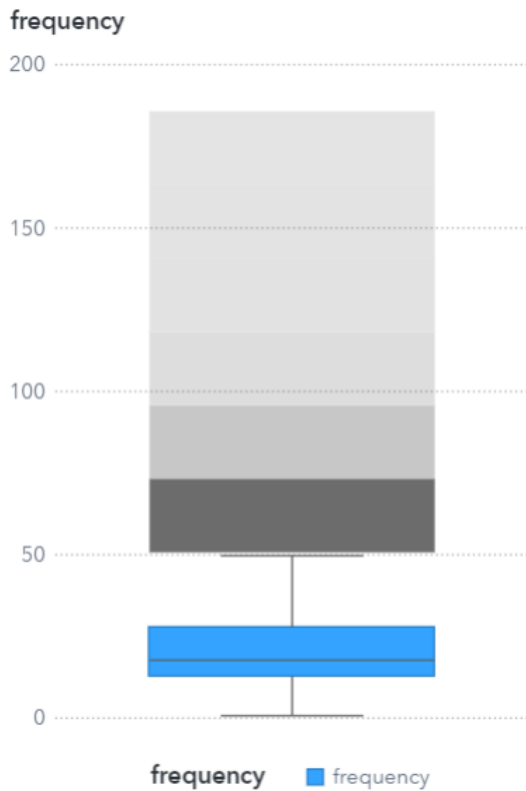
Πίνακας 6: Μεταβλητές Recency, Frequency και Monetary



Πίνακας 7: Μεταβλητές Recency, Frequency και Monetary

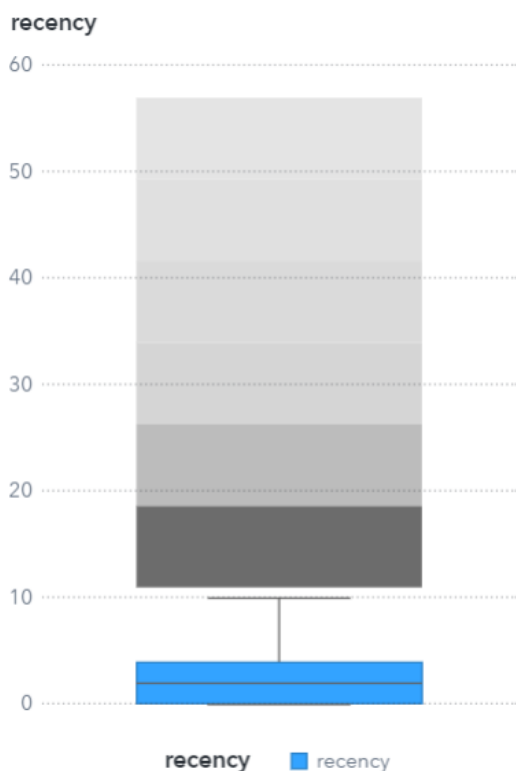
Το boxplot για το Frequency αντιπροσωπεύει την κατανομή της συχνότητας των συναλλαγών ή των αγορών των πελατών. Εμφανίζει το ενδοτεταρτημοριακό εύρος (IQR) της συχνότητας συναλλαγών και τα upper και lower whiskers δείχνουν το εύρος της συχνότητας συναλλαγών. Τέλος οι ακραίες τιμές αντιπροσωπεύουν πελάτες των οποίων η συχνότητα συναλλαγών είναι σημαντικά υψηλότερη ή χαμηλότερη από την πλειοψηφία.

Πιο συγκεκριμένα, για την μεταβλητή frequency το 50% των παρατηρήσεων είναι κάτω από 18 που είναι η διάμεσος, που σημαίνει ότι οι μισοί από τους πελάτες έχουν συχνότητα συναλλαγών κάτω από 18 φορές. Επίσης το 75% των παρατηρήσεων είναι κάτω από 28, που σημαίνει ότι τα $\frac{3}{4}$ των πελατών έχουν συχνότητα συναλλαγών κάτω από 28 φορές. Και τέλος υπάρχουν και τιμές εκτός upper whisker, δηλαδή πελάτες με συχνότητα συναλλαγών άνω των 50.

Πίνακας 8: Μεταβλητή frequency

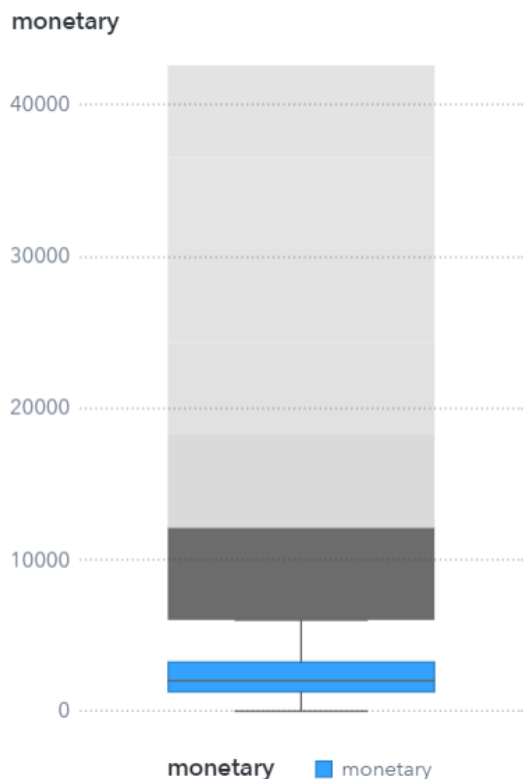
Το boxplot για το Recency απεικονίζει την κατανομή του διαστήματος μεταξύ της ημερομηνίας της τελευταίας συναλλαγής και της πιο πρόσφατης ημερομηνίας για την οποία μας ενδιαφέρει να γίνει η ανάλυση, δηλαδή του χρόνου που έχει παρέλθει από την τελευταία αγορά ενός πελάτη. Εμφανίζει το ενδοτεταρτημοριακό εύρος (IQR) του recency συναλλαγών και τα upper και lower whiskers δείχνουν το εύρος των τιμών του. Τέλος οι ακραίες τιμές επισημαίνουν πελάτες με ασυνήθιστα μεγάλο ή σύντομο χρονικό διάστημα από την τελευταία αγορά τους.

Πιο συγκεκριμένα, για την μεταβλητή recency το 50% των παρατηρήσεων είναι κάτω από 2 που είναι η διάμεσος, που σημαίνει ότι οι μισοί από τους πελάτες έχουν πραγματοποιήσει την πιο πρόσφατη συναλλαγή τους 2 μέρες πριν. Επίσης το 75% των παρατηρήσεων είναι κάτω από 4, που σημαίνει ότι τα $\frac{3}{4}$ των πελατών έχουν κάνει την πιο πρόσφατη συναλλαγή τους 4 μέρες πριν. Και τέλος υπάρχουν και τιμές εκτός upper whisker, δηλαδή πελάτες με την πιο πρόσφατη επιλογή τους να είναι άνω των 10 ημερών πριν.

Πίνακας 9: Μεταβλητή recency

Το boxplot για το Monetary εμφανίζει την κατανομή της νομισματικής αξίας των συναλλαγών των πελατών, αντιπροσωπεύει το ενδοτεταρτημοριακό εύρος (IQR) των νομισματικών αξιών και τα upper και lower whiskers δείχνουν το εύρος των χρημάτων. Τέλος οι ακραίες τιμές προσδιορίζουν πελάτες με εξαιρετικά υψηλές ή χαμηλές δαπάνες.

Πιο συγκεκριμένα, για την μεταβλητή monetary το 50% των παρατηρήσεων είναι κάτω από 2051 που είναι η διάμεσος, που σημαίνει ότι οι μισοί από τους πελάτες έχουν πραγματοποιήσει συναλλαγές αξίας έως 2051 ευρώ σύνολο. Επίσης το 75% των παρατηρήσεων είναι κάτω από 3221, που σημαίνει ότι τα $\frac{3}{4}$ των πελατών έχουν πραγματοποιήσει συναλλαγές αξίας έως 3221 ευρώ σύνολο. Και τέλος υπάρχουν ακραίες τιμές, που αντιπροσωπεύουν πελάτες που έχουν πραγματοποιήσει συναλλαγές άνω των 6046 ευρώ σύνολο.

Πίνακας 10: Μεταβλητή monetary

Στο RFM_dataset εκτός των τριών μεταβλητών, Recency, Frequency, Monetary, έχουν δημιουργηθεί κάποιες επιπλέον υπολογιζόμενες μεταβλητές. Το RFM_dataset έχει 24998 εγγραφές που η κάθε εγγραφή αντιστοιχεί σε έναν διακριτό πελάτη και οι διακριτές κατηγορίες προϊόντων είναι 17054.

Παρακάτω, ο πίνακας με τις επιμέρους μεταβλητές του RFM dataset.

Πίνακας 11: RFM Αρχείο Δεδομένων

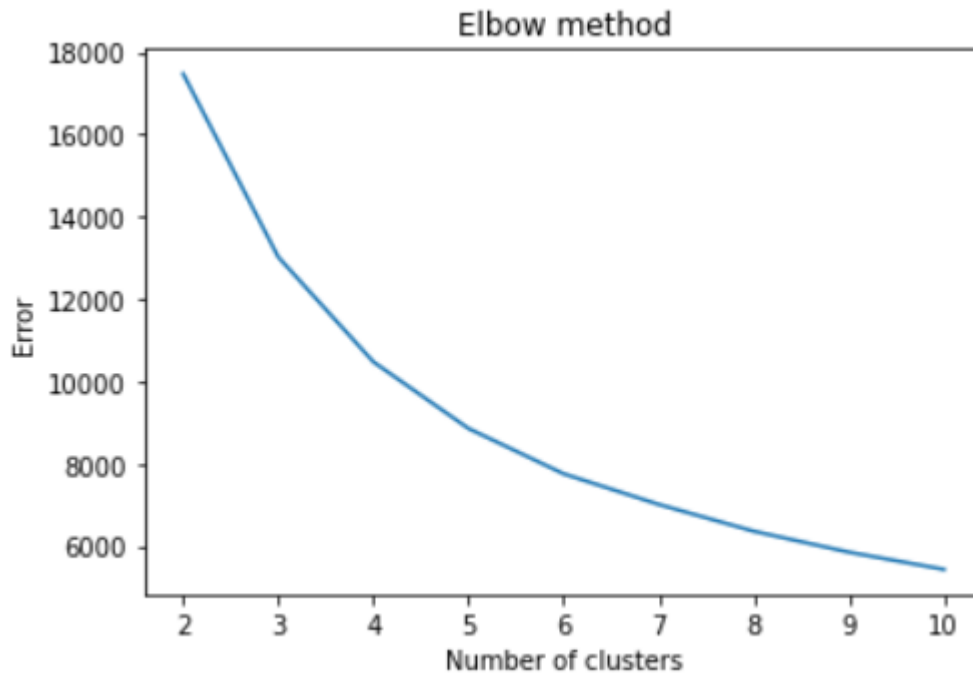
Variable Name	Data Type	Description
Customer	Nominal	A unique code on an invoice that is used to reference a customer's account.
Num_distinct_items	Numeric	The number of distinct items purchased by a customer.
Avg_food_score	Numeric	Represents the average score given by customers for the quality of food.
Avg_delivery_score	Numeric	The average score given by customers for the delivery service.
Avg_service_score	Numeric	The average score given by customers for the overall service.

Total_amount_spent	Numeric	Represents the total monetary spending by customers.
Num_orders	Numeric	The total number of orders placed by each customer.
Avg_order_value_per_customer	Numeric	The average value of orders placed by each customer.
Frequency	Numeric	Represents how often a customer makes purchases.
Recency	Numeric	Indicates the time elapsed since the customer's last purchase.
Monetary	Numeric	Represents the total amount spent by each customer.
Target	Numeric	A variable with only value the number 1
Num_shops_visited	Numeric	The number of different shops visited by a customer.

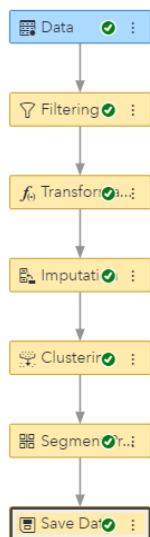
3.2.4 Customer categorization analysis using restaurant data

Στη συνέχεια γίνεται η υλοποίηση και εκτέλεση της παρακάτω ροής εργασιών. Η όλη διαδικασία στοχεύει να μετατρέψει τα ανεπεξέργαστα δεδομένα σε ουσιαστικές πληροφορίες ώστε να γίνει ο διαχωρισμός των πελατών σε διαφορετικά segments. Το φιλτράρισμα και ο μετασχηματισμός βοηθούν στον καθαρισμό και την προετοιμασία των δεδομένων, ο καταλογισμός διασφαλίζει την πληρότητα των δεδομένων, η ομαδοποίηση αποκαλύπτει μοτίβα και το προφίλ τμήματος παρέχει λεπτομερείς πληροφορίες για κάθε ομάδα. Αξίζει να αναφερθεί ότι το βήμα της ομαδοποίησης έχει γίνει με την μέθοδο $k - means$ και η επιλογή του πλήθους των clusters γίνεται με τη βοήθεια του Aligned Box Criterion όπου ο αριθμός των clusters υπολογίζεται αυτόματα. Με την μέθοδο αυτή δημιουργούνται 5 clusters. Επίσης υλοποιήθηκε η μέθοδος $k - means$ και με τη βοήθεια του Elbow method μέσω της Python με αντίστοιχο πλήθος clusters, δηλαδή 5 clusters. Στην περίπτωση αυτή, το γράφημα σχετικά με την ποσότητα των τετραγώνων εκτιμάται σε κάθε επιλογή συστάδων. Εάν η κλίση του γραφήματος αλλάξει από υψηλή σε ρηχή, η τέλεια ποσότητα συστάδων, k , θα οδηγηθεί, σε αυτό το σημείο "αγκώνα". Παρακάτω βρίσκεται το γράφημα που περιγράφει τη μέθοδο elbow για τα δεδομένα μας και η ροή εργασιών.

Πίνακας 12: Elbow Method



Πίνακας 13: Ροή εργασιών



Παρακάτω παρουσιάζεται το αποτέλεσμα της εκτέλεσης της προηγούμενης ροής οπτικοποιημένο σε πίνακα διασταύρωσης.

Πίνακας 14: Αποτέλεσμα της εκτέλεσης της ροής εργασιών

Segment Name	frequency	recency	monetary
VIP Customers	20,23	1,31	2511,08
Premium Customers	37,83	0,48	3018,55
Inactive Customers	14,93	8,3	1786,14
Elite Customers	36,56	2,07	5629,69
Casual Shoppers	13,58	1,22	1104,02
Άθροισμα	21,65	3,22	2422,18

Το πρώτο cluster μπορεί να περιγραφεί ως cluster με πελάτες που πραγματοποιούν συχνές αγορές, έχουν πραγματοποιήσει πρόσφατες αγορές υποδηλώνοντας ενεργό ενασχόληση και συνεισφέρουν σημαντικά στα έσοδα με υψηλή δαπάνη. Οι VIP πελάτες είναι από τους πιο πολύτιμους πελάτες που κάνουν συχνές αγορές, παραμένουν ενεργοί και συνεισφέρουν σημαντικά στα έσοδα. Προσφορές αποκλειστικών προνομίων και εξατομικευμένες εμπειρίες μπορεί να βοηθήσουν στη διατήρηση και ενίσχυση της πιστότητάς τους.

Το δεύτερο cluster μπορεί να περιγραφεί ως cluster με πελάτες που πραγματοποιούν ακόμα πιο συχνές αγορές, έχουν πραγματοποιήσει πολύ πρόσφατες αγορές, υποδηλώνοντας μία ισχυρή και ενεργή σχέση και τέλος αυτοί οι πελάτες είναι υψηλοί δαπανητές, που συνεισφέρουν σημαντικά στα έσοδα. Οι Premium πελάτες αντιπροσωπεύουν μια ομάδα υψηλά ενεργών και υψηλών δαπανητών. Η προσαρμογή των στρατηγικών μάρκετινγκ για την ενίσχυση της εμπειρίας τους και η προσφορά προηγμένων υπηρεσιών μπορεί να συμβάλει στη διατήρηση της πιστότητάς τους.

Το τρίτο cluster μπορεί να περιγραφεί ως cluster με πελάτες που πραγματοποιούν λιγότερες αγορές, έχουν πραγματοποιήσει πρόσφατες αγορές, εμφανίζοντας κάποιο επίπεδο ενασχόλησης και τέλος αυτοί οι πελάτες συνεισφέρουν λιγότερο στα έσοδα σε σύγκριση με άλλες ομάδες πελατών. Οι Casual αγοραστές είναι πελάτες που κάνουν περιστασιακές αγορές. Στρατηγικές για να προωθηθούν περισσότερες συναλλαγές και για να προβληθεί η αξία του προϊόντος μπορεί να αυξήσουν τη συμμετοχή τους.

Το τέταρτο cluster μπορεί να περιγραφεί ως cluster με πελάτες που πραγματοποιούν λιγότερες αγορές, δεν έχουν πραγματοποιήσει πρόσφατες αγορές, υποδηλώνοντας μια πώση στη συμμετοχή και παρόλο που δεν είναι υψηλοί δαπανητές, συνεισφέρουν σημαντικά στα έσοδα. Οι Inactive πελάτες είναι πελάτες που παλιότερα είχαν ενεργό συμμετοχή, αλλά έχουν γίνει λιγότερο συχνοί. Υλοποίηση καμπανιών ανανέωσης και στοχευμένων προωθητικών ενεργειών μπορεί να βοηθήσει στον επαναφορά τους.

Το πέμπτο cluster μπορεί να περιγραφεί ως cluster με πελάτες που πραγματοποιούν πολύ συχνές αγορές, έχουν πραγματοποιήσει αγορές πρόσφατα, διατηρώντας ένα δυνατό επίπεδο συμμετοχής και είναι υψηλοί δαπανητές, συνεισφέροντας σημαντικά στα έσοδα. Οι Elite πελάτες αντιπροσωπεύουν τους πελάτες κορυφαίας κατηγορίας που και πραγματοποιούν συχνές αγορές και δαπανούν γενναιόδωρα. Παροχή αποκλειστικών ανταμοιβών, εξατομικευμένων εμπειριών και πρόσβασης σε νέα προϊόντα εγκαίρως μπορεί να ενισχύσει την πιστότητά τους και την ικανοποίησή τους.

3.2.5 Customer demographic and behavioral analysis

Για τον σκοπό δημιουργίας συγκεκριμένου προφίλ της εκάστοτε ομάδας πελατών, έχει δημιουργηθεί ένα dataset με ονομασία thesis.final_results με τις εξής μεταβλητές :

Πίνακας 15: Τελικό Αρχείο Δεδομένων

Variable Name	Data Type	Description
Cluster_Name	Nominal	The name of the cluster that each customer belongs to
Customer_id	Nominal	A unique code on an invoice that is used to reference a customer's account.
Num_distinct_items	Numeric	The number of distinct items purchased by a customer.
Avg_food_score	Numeric	Represents the average score given by customers for the quality of food.
Avg_delivery_score	Numeric	The average score given by customers for the delivery service.
Avg_service_score	Numeric	The average score given by customers for the overall service.
Total_amount_spent	Numeric	Represents the total monetary spending by customers.
Num_orders	Numeric	The total number of orders placed by each customer.
Avg_order_value_per_customer	Numeric	The average value of orders placed by each customer.
Frequency	Numeric	Represents how often a customer makes purchases.
Recency	Numeric	Indicates the time elapsed since the customer's last purchase.
Monetary	Numeric	Represents the total amount spent by each customer.
Num_shops_visited	Numeric	The number of different shops visited by a customer.
Pinata_order_count	Numeric	The number of orders made with pinata discount
Avg_order_value	Numeric	The average order value per customer
Most_frequent_region	Nominal	The most frequent region of orders made by a customer
Order_count_per_region	Numeric	The number of orders made from the most frequent region
Shop_cuisine	Nominal	The most preferable cuisine
Order_count_per_cuisine	Numeric	The number of orders made with the most preferable cuisine
Cluster_id	Numeric	A unique code indicating the cluster a customer belongs to

Λόγω της αφαίρεσης των ακραίων τιμών στο στάδιο της RFM ανάλυσης, παρατήρηθηκε ένα cluster πελατών με την ονομασία «Unknown» που δεν ανήκει σε κανένα από τα 5 clusters. Οπότε για

την πραγματοποίηση της συγκεκριμένης εργασίας κρίθηκε σκόπιμο να αφαιρεθεί το «Unknown» από τα δεδομένα.

Για κάθε μεταβλητή που παρατίθεται (Num_distinct_items, Total_amount_spent, Num_orders, Avg_order_value_per_customer, Frequency, Recency, Monetary), παρακάτω βλέπουμε έναν πίνακα που περιλαμβάνει τον αριθμό των non missing παρατηρήσεων, τη μέση τιμή, την τυπική απόκλιση, την ελάχιστη τιμή και τη μέγιστη τιμή. Αυτό μπορεί να βοηθήσει στην κατανόηση της κατανομής κάθε μεταβλητής, τον προσδιορισμό ακραίων τιμών και στην απόκτηση μιας αίσθησης της κλίμακας και της μεταβλητότητας των δεδομένων.

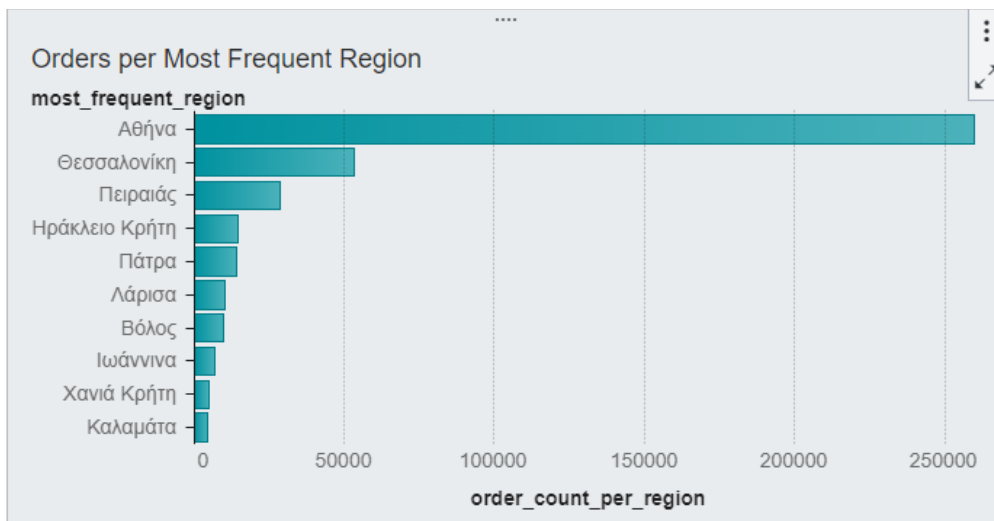
Πίνακας 16: non missing παρατηρήσεις, μέση τιμή, την τυπική απόκλιση, την ελάχιστη τιμή και τη μέγιστη τιμή

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
num_distinct_items	24998	29.5595248	18.1711656	1.0000000	294.0000000
total_amount_spent	24998	2373.35	1907.39	42.0000000	37983.00
num_orders	24998	23.0023602	14.6198469	1.0000000	186.0000000
avg_order_value_per_customer	24998	46.6437823	23.4189212	4.8100000	485.0400000
frequency	24998	23.0023602	14.6198469	1.0000000	186.0000000
recency	24998	3.8366669	6.1987398	0	57.0000000
monetary	24998	2626.23	2114.79	42.0000000	42597.00

Στη συνέχεια έχουν δημιουργηθεί τα παρακάτω γραφήματα που δείχνουν τις παραγγελίες ανά πιο συχνή περιοχή, τις παραγγελίες ανά πιο συχνή κουζίνα, την συχνότητα των παραγγελιών με προσφορά ριπάτα και την συχνότητα των διαφορετικών προϊόντων και καταστημάτων. Όλα τα γραφήματα έχουν φιλτραριστεί ανά cluster.

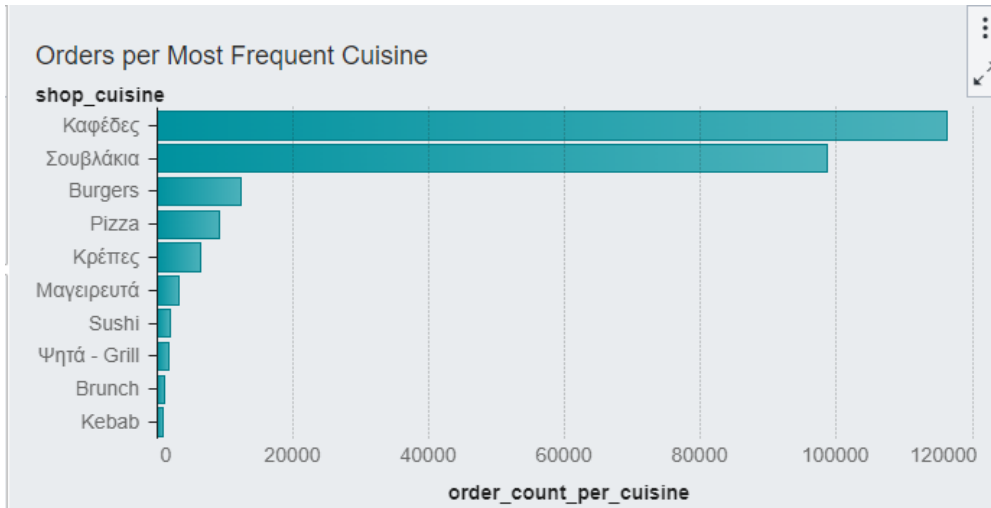
Πίνακας 17: Casual Shoppers Bar Chart – Orders Per Most Frequent Region



Το γράφημα "Casual Shoppers Bar Chart – Orders Per Most Frequent Region" παρουσιάζει τον αριθμό των παραγγελιών που πραγματοποιήθηκαν από τους Casual Shoppers σε κάθε περιοχή που ήταν η πιο συχνή γι' αυτούς. Κάθε μπάρα στο γράφημα αντιστοιχεί σε μια περιοχή, ενώ το ύψος της μπάρας αντιπροσωπεύει τον αριθμό των παραγγελιών. Μέσω αυτού του γραφήματος μπορούμε

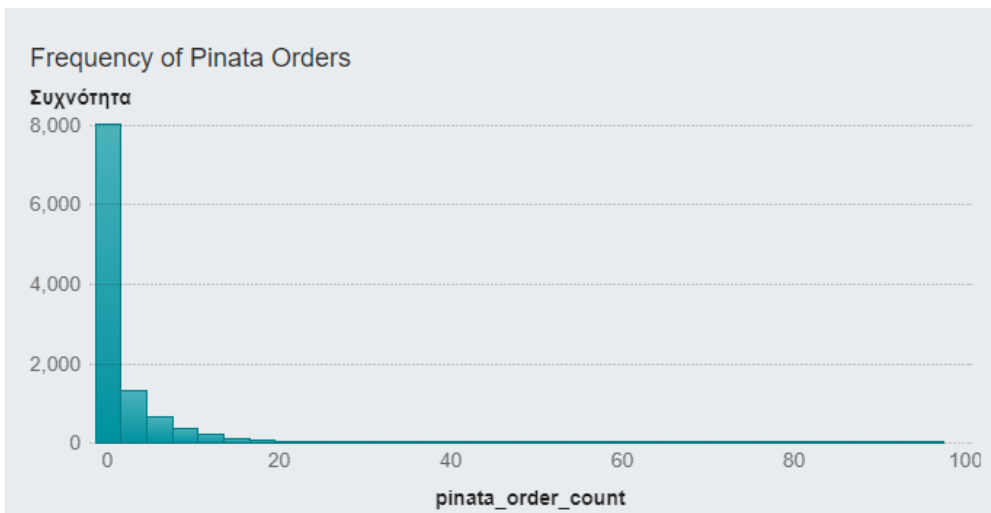
να κατανοήσουμε ποιες περιοχές είναι πιο δημοφιλείς μεταξύ των Casual Shoppers και πόσες παραγγελίες καταγράφηκαν σε κάθε μια από αυτές.

Πίνακας 18: Casual Shoppers Bar Chart – Orders Per Most Frequent Cuisine

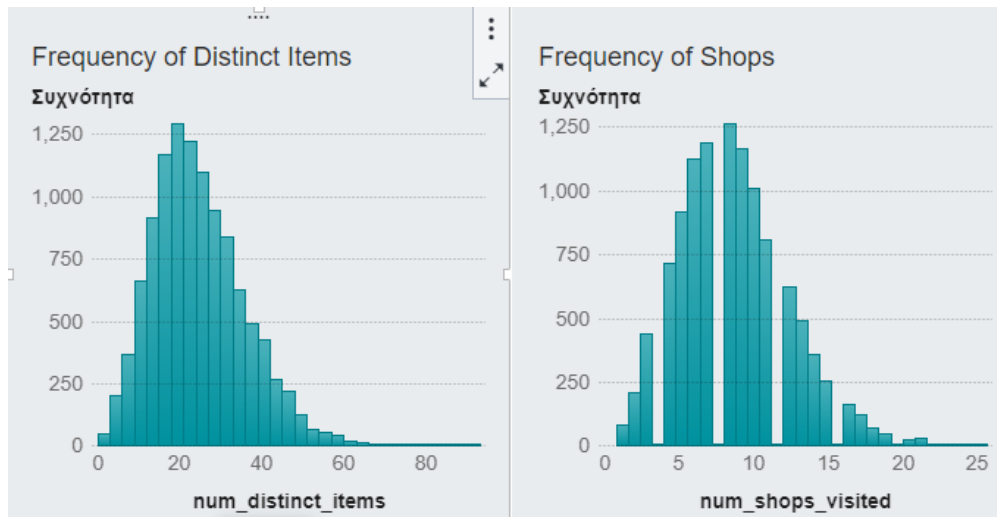


Το γράφημα "Casual Shoppers Bar Chart – Orders Per Most Frequent Cuisine" παρουσιάζει τον αριθμό των παραγγελιών ανά πιο συχνή κουζίνα για την ομάδα των "Casual Shoppers". Αυτό το γράφημα προσφέρει μια επισκόπηση των προτιμήσεων διατροφής των Casual Shoppers και μπορεί να βοηθήσει στην κατανόηση των κυρίαρχων κουζινών που ενδιαφέρουν αυτήν την ομάδα πελατών.

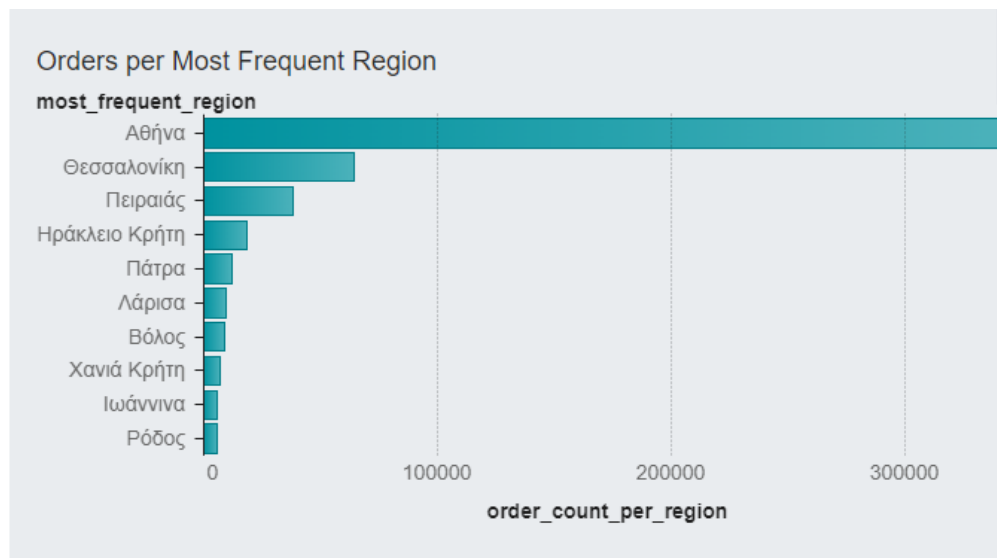
Πίνακας 19: Casual Shoppers Histogram – Frequency of Pinata Orders



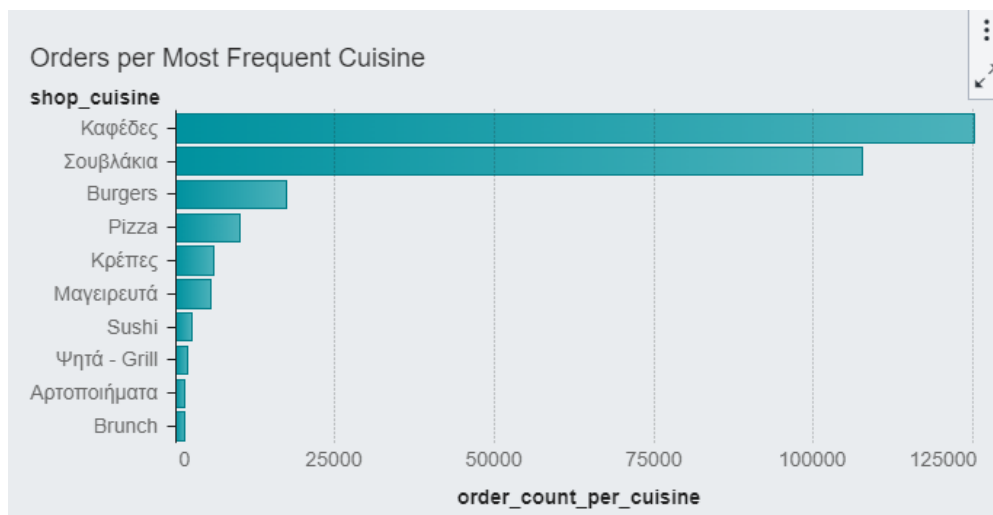
Το γράφημα "Casual Shoppers Histogram – Frequency of Pinata Orders" δείχνει τη συχνότητα των παραγγελιών από τους "Casual Shoppers" σε σχέση με τη συμμετοχή τους σε παραγγελίες από το Pinata, μια διαδικτυακή υπηρεσία που παρέχει προσφορές και εκπτώσεις. Ουσιαστικά, απεικονίζει πόσες παραγγελίες έχουν πραγματοποιήσει οι "Casual Shoppers" σε διαφορετικά επίπεδα συμμετοχής στο Pinata.

Πίνακας 20: Casual Shoppers Histograms – Frequency of Distinct Items and Frequency of Shops

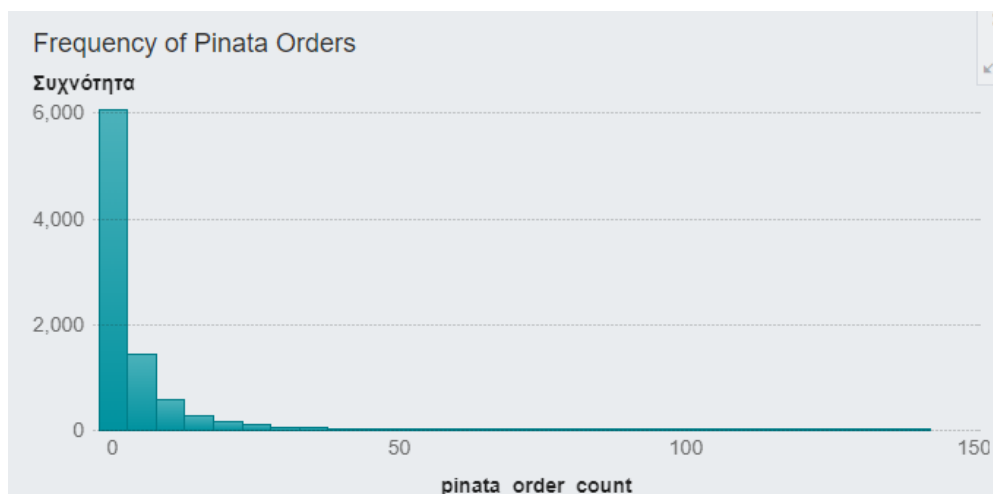
Το γράφημα "Casual Shoppers Histogram – Frequency of Distinct Items" δείχνει τη συχνότητα εμφάνισης διαφορετικών αντικειμένων που αγοράζουν οι "Casual Shoppers". Με άλλα λόγια, δείχνει πόσο συχνά οι ανεπίσημοι αγοραστές αγοράζουν διαφορετικά προϊόντα από την πλατφόρμα. Το γράφημα "Casual Shoppers Histogram – Frequency of Shops" δείχνει τη συχνότητα των casual shoppers σε σχέση με τον αριθμό των καταστημάτων που επισκέπτονται. Συγκεκριμένα, δείχνει πόσοι casual shoppers επισκέπτονται διαφορετικό αριθμό καταστημάτων, δημιουργώντας ένα ιστόγραμμα που απεικονίζει τη διανομή της συχνότητας των επισκέψεών τους.

Πίνακας 21: Elite Customers Bar Chart – Orders Per Most Frequent Region

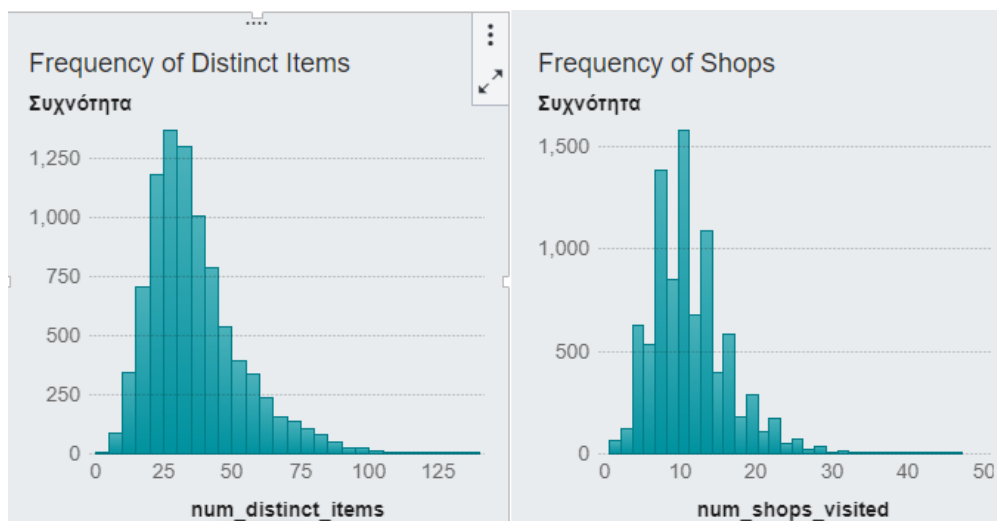
Το γράφημα "Elite Customers Bar Chart – Orders Per Most Frequent Region" δείχνει τον αριθμό των παραγγελιών που έχουν τοποθετήσει οι "Elite Customers" σε κάθε περιοχή παράδοσης που είναι η πιο συχνή γι' αυτούς. Αυτό μπορεί να βοηθήσει στην κατανόηση των προτιμήσεων των πιο αξιόλογων πελατών σε σχέση με τη γεωγραφική τους τοποθεσία.

Πίνακας 22: Elite Customers Bar Chart – Orders Per Most Frequent Cuisine

Το γράφημα "Elite Customers Bar Chart – Orders Per Most Frequent Cuisine" δείχνει τον αριθμό των παραγγελιών από τους Elite πελάτες για κάθε πιο συχνή κουζίνα. Αυτό μας δίνει μια εικόνα των προτιμήσεων διατροφής των πελατών με υψηλή επίπεδο εξυπηρέτησης και αγοραστικής δύναμης, και μπορεί να παρέχει στρατηγικές εισηγήσεις για την ανάπτυξη μενού και την προώθηση προϊόντων σε αυτήν την κατηγορία πελατών.

Πίνακας 23: Elite Customers Histogram – Frequency of Pinata Orders

Το γράφημα "Elite Customers Histogram – Frequency of Pinata Orders" δείχνει τη συχνότητα των παραγγελιών από τους πελάτες της κατηγορίας "Elite" που χρησιμοποιούν την υπηρεσία Pinata σε ένα δεδομένο χρονικό διάστημα. Ο άξονας των x δείχνει τον αριθμό των παραγγελιών, ενώ ο άξονας των y δείχνει τη συχνότητα ή το ποσοστό των πελατών που έχουν πραγματοποιήσει αυτό τον αριθμό παραγγελιών. Από το γράφημα μπορούμε να κατανοήσουμε πόσο συχνά οι πελάτες "Elite" χρησιμοποιούν την υπηρεσία Pinata και πώς κατανέμεται η συχνότητα των παραγγελιών μεταξύ τους.

Πίνακας 24: Elite Customers Histograms – Frequency of Distinct Items and Frequency of Shops

Το γράφημα "Elite Customers Histogram – Frequency of Distinct Items" δείχνει τη συχνότητα του αριθμού των διακριτών αντικειμένων που αγοράζουν οι Elite πελάτες. Συγκεκριμένα, παρουσιάζει πόσο συχνά οι Elite πελάτες αγοράζουν διαφορετικά αντικείμενα, δηλαδή πόσο συχνά επιλέγουν να αγοράσουν προϊόντα που δεν έχουν αγοράσει ξανά.

Το γράφημα "Elite Customers Histogram – Frequency of Shops" δείχνει τη συχνότητα με την οποία οι Elite πελάτες της πλατφόρμας διανομής φαγητού επισκέπτονται διαφορετικά καταστήματα. Κάθε μπάρα στο ιστόγραμμα αντιστοιχεί σε ένα εύρος του αριθμού των καταστημάτων που επισκέπτονται οι Elite πελάτες, ενώ η ύψωση της μπάρας αντιπροσωπεύει τον αριθμό των Elite πελατών που εμφανίζονται σε αυτό το εύρος. Με άλλα λόγια, το γράφημα αναδεικνύει το πρότυπο του πόσο συχνά οι Elite πελάτες επισκέπτονται διαφορετικά καταστήματα στο πλαίσιο της πλατφόρμας διανομής φαγητού.

Σύμφωνα με την ομαδοποίηση των πελατών που έχει προηγηθεί, οφείλουν να γίνουν μεμονωμένες και εξατομικευμένες ενέργειες εκ μέρους της επιχείρησης για κάθε ομάδα πελατών που περιγράφεται παραπάνω.

Cluster 1 – VIP Customers

Αυτή η ομάδα επιδεικνύει συνεπές πρότυπο συχνών αγορών, υποδηλώνοντας ενεργό ενδιαφέρον για τη διαδικτυακή πλατφόρμα. Παρά τις συχνές αγορές, η επιλογή για πρόσφατες αγορές υποδηλώνει τρέχον και συνεχές ενδιαφέρον για προϊόντα ή υπηρεσίες. Η μέτρια νομισματική συνεισφορά αντανακλά σταθερό και αξιόπιστο ρεύμα εσόδων. Οι πελάτες αυτοί είναι διακριτικοί και ενδιαφέρονται για μια ποικιλία προϊόντων, καθιστώντας τους ανταποκριτικούς σε προσφορές που προωθούν συχνές αγορές.

Για τους VIP πελάτες προτείνονται εξατομικευμένες προσφορές, δηλαδή δημιουργία προσφορών και εκπτώσεων βασισμένων στις προτιμήσεις αγοράς τους. Επιπλέον, υπάρχει η επιλογή αποκλειστικών προνομίων, όπως δωρεάν πρόσβαση σε νέα εστιατόρια.

Cluster 2 - Premium Customers

Αυτή η ομάδα αντιπροσωπεύει πελάτες που κάνουν εξαιρετικά συχνές αγορές, επιδεικνύοντας συνεχή αλληλεπίδραση με την πλατφόρμα. Κάνουν αγορές πολύ πρόσφατα, δείχνοντας υψηλό επίπεδο τρέχουσας δραστηριότητας και ενδιαφέροντος. Συνεισφέρουν σημαντικά

στα συνολικά έσοδα, υποδηλώνοντας κρίσιμη και υψηλή αξία. Οι πελάτες αυτοί εκδηλώνουν προτίμηση για προϊόντα και υπηρεσίες υψηλής ποιότητας, περιμένοντας ένα επίπεδο αποκλειστικότητας. Εξατομικευμένες και αποκλειστικές προσφορές συνάδουν με τις προτιμήσεις τους.

Για τους Premium πελάτες προτείνεται η παροχή εξατομικευμένων συμβουλών για νέα εστιατόρια βασισμένες στις αγορές τους, η προσφορά πακέτων γευστικών συνδυασμών από δημοφιλή εστιατόρια και εξατομικευμένες προτάσεις μενού βασισμένες στις προτιμήσεις τους.

Cluster 3 - Casual Shoppers

Αυτοί οι πελάτες πραγματοποιούν αγορές με χαμηλή συχνότητα, υποδηλώνοντας μικρότερο επίπεδο δραστηριότητας σε σύγκριση με άλλες ομάδες. Παρόλα αυτά, τείνουν να κάνουν πρόσφατες αγορές, υποδηλώνοντας τρέχον και ασταθές ενδιαφέρον. Η συνεισφορά στα έσοδα είναι χαμηλή έως μέτρια, κάτι που αποτελεί ευκαιρία για αύξηση. Ανταποκρίνονται θετικά σε έκπτωση και προσφορές, απαιτώντας εξατομικευμένες κινητοποιήσεις για αυξημένη συχνότητα και δαπάνη.

Για τους Casual πελάτες προτείνονται πακέτα φαγητού με ελκυστικές εκπτώσεις με στόχο την αύξηση της συχνότητας αγορών τους και ειδικές προσφορές για την επόμενη αγορά μετά από μεγάλο διάστημα αδράνειας.

Cluster 4 - Inactive Customers

Επιδεικνύουν χαμηλή συχνότητα αγορών, υποδηλώνοντας μειωμένο επίπεδο αλληλεπίδρασης με την πλατφόρμα. Δεν έχουν εμπλακεί πρόσφατα, υποδηλώνοντας έλλειψη αλληλεπίδρασης. Η συνεισφορά στα έσοδα είναι μέτρια, υποδηλώνοντας ευκαιρία για προσπάθειες αναστάτωσης. Οι Ανενεργοί Πελάτες μπορεί να έχουν αντιμετωπίσει προβλήματα ή να έχουν χάσει το ενδιαφέρον τους, απαιτώντας προσπάθειες επαναφοράς για την κατανόηση των ανησυχιών τους.

Για τους Inactive πελάτες προτείνονται εκστρατείες με email και ειδοποιήσεις για επαναφορά στην πλατφόρμα καθώς και αυξημένες εκπτώσεις για την επόμενη παραγγελία.

Cluster 5 - Elite Customers

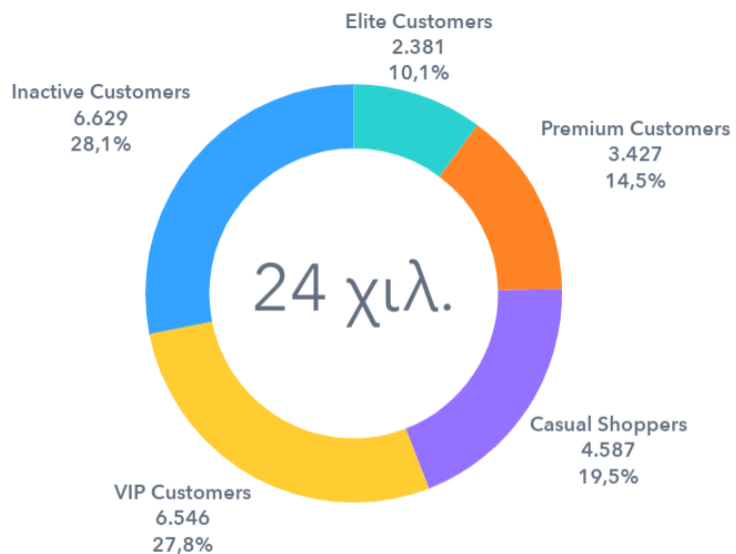
Αυτοί οι αγοραστές αντιπροσωπεύουν την υψηλότερη συχνότητα αγορών, επιδεικνύοντας ένα εξαιρετικό επίπεδο συμμετοχής. Κάνουν πρόσφατες αγορές, υποδηλώνοντας συνεχές ενδιαφέρον. Περιμένουν προϊόντα και υπηρεσίες υψηλής ποιότητας, επιδιώκοντας ένα επίπεδο πολυτέλειας.

Για τους Elite πελάτες προτείνεται η οργάνωση εκδηλώσεων ή γευσιγνωσίας και η αποστολή ειδικών προσφορών για υπηρεσίες catering.

Στη συνέχεια παρουσιάζονται διαγράμματα τα οποία συνεισφέρουν στην υλοποίηση της δημιουργίας προφίλ για κάθε ομάδα πελατών.

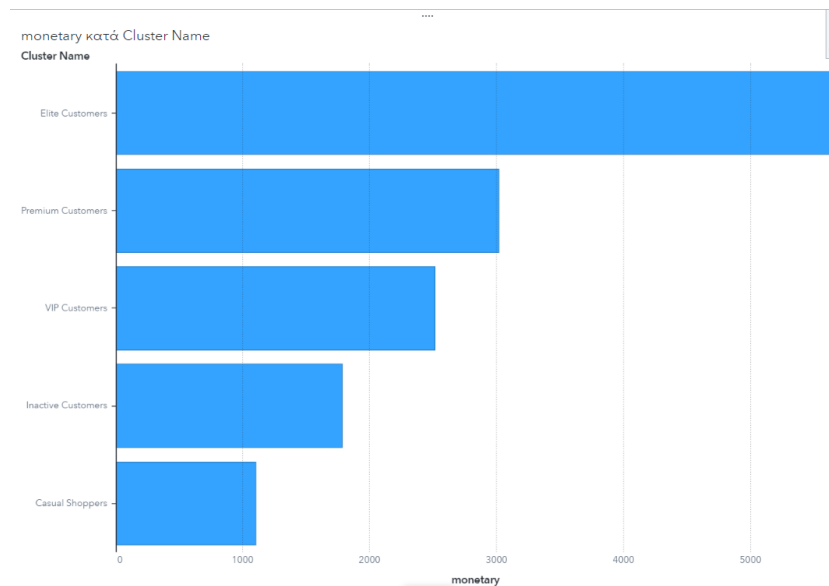
Στο παρακάτω διάγραμμα παρουσιάζεται η κατανομή και το πλήθος των πελατών ανά ομάδα.

Πίνακας 25: Clusters' Pie Chart



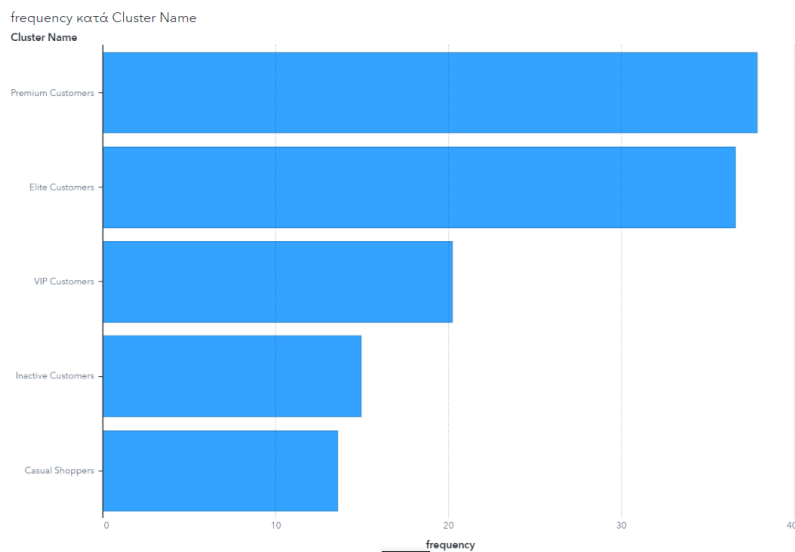
Στο επόμενο ραβδόγραμμα παρατηρείται ότι το μεγαλύτερο ποσοστό των εσόδων της πλατφόρμας προέρχεται από τους Elite πελάτες.

Πίνακας 26: Ποσοστό των εσόδων της πλατφόρμας



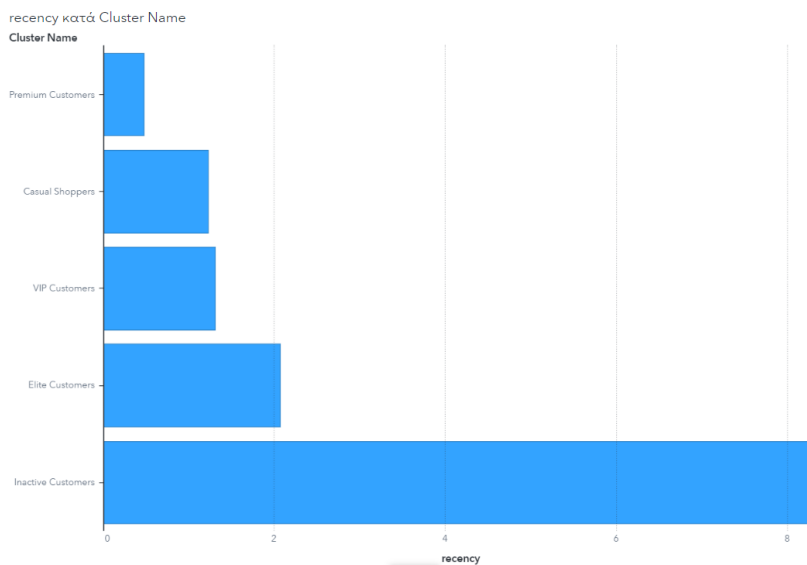
Στο επόμενο ραβδόγραμμα παρατηρείται ότι οι πιο συχνοί πελάτες είναι οι αυτοί που ανήκουν στην ομάδα των Premium πελατών.

Πίνακας 27: Πελάτες



Στο επόμενο ραβδόγραμμα παρατηρείται ότι οι πελάτες που έχουν απέχουν πολύ μεγάλο χρονικό διάστημα από την πραγματοποίηση αγοράς είναι οι αυτοί που ανήκουν στην ομάδα των Inactive πελατών.

Πίνακας 28: Πελάτες



3.3 Market basket analysis

Η ανάλυση καλαθιού αγοράς (MBA) επίσης γνωστή ως μάθηση κανόνων συσχέτισης ή ανάλυση συνάφειας, είναι μια τεχνική εξόρυξης δεδομένων που μπορεί να χρησιμοποιηθεί σε διάφορους τομείς, όπως το μάρκετινγκ, τη βιοπληροφορική, την εκπαίδευση, τη πυρηνική επιστήμη κ.λπ. Ο κύριος στόχος του MBA στο μάρκετινγκ είναι να παρέχει τις πληροφορίες στον λιανοπωλητή για να κατανοήσει την αγοραστική συμπεριφορά του αγοραστή, η οποία μπορεί να βοηθήσει τον λιανοπωλητή στη σωστή λήψη αποφάσεων.

Χρησιμοποιώντας τη λειτουργία ανάλυσης καλαθιού αγοράς, μπορεί να αποκτηθεί γνώση των αντικειμένων που αγοράζονται συχνά από τους πελάτες ταυτόχρονα και έχουν την ευκαιρία να προωθηθούν. Όσον αφορά τον στόχο της λειτουργίας ανάλυσης καλαθιού αγοράς για να αποφασίσετε ποια προϊόντα αγοράζουν ταυτόχρονα οι πελάτες, όπου το όνομα αυτής της λειτουργίας λαμβάνεται από τη συμπεριφορά των πελατών στην τοποθέτηση προϊόντων αγορών στα καλάθια αγορών ή στη λίστα αγορών τους. Πέρα από τον προσδιορισμό του μοτίβου καλαθιού αγορών ενός πελάτη θα είναι σε θέση να βοηθήσει σημαντικά μια εταιρεία στη χρήση αυτών των πληροφοριών όσον αφορά τις ανάγκες της επιχειρηματικής στρατηγικής, ένας από αυτούς είναι η τοποθέτηση των προϊόντων που αγοράζονται συχνότερα ταυτόχρονα σε έναν συγκεκριμένο τομέα.

Ο κανόνας συσχέτισης σχετίζεται με τη δήλωση "τι' με τι". Αυτό το θέμα μπορεί να είναι σε μορφή δήλωσης σχετικά με τη δραστηριότητα συναλλαγών που πραγματοποιείται από τους πελάτες σε ένα σούπερ μάρκετ ή μία επιχείρηση γενικότερα. Από αυτή τη δήλωση, υπάρχει μια ισχυρή σχέση με τη μελέτη της βάσης δεδομένων συναλλαγών πελατών για τον προσδιορισμό της συνήθειας ενός αγορασμένου προϊόντος με ποιο προϊόν, επομένως, ο κανόνας συσχέτισης αναφέρεται συχνά ως ανάλυση καλαθιού αγοράς.

Η σημασία ενός association rule μπορεί να υπολογιστεί με την παρουσία δύο παραμέτρων, δηλαδή της υποστήριξης και της εμπιστοσύνης. Υποστήριξη (υποστηρικτική αξία, support) είναι το ποσοστό των συνδυασμών στοιχείων προϊόντος στη βάση δεδομένων. Ενώ η εμπιστοσύνη (τιμή βεβαιότητας, confidence) είναι μια τιμή για τον προσδιορισμό της ισχύος των σχέσεων μεταξύ των στοιχείων στους κανόνες σύνδεσης. Η υποστήριξη είναι η πιθανότητα εμφάνισης ενός συνόλου στοιχείων στα δεδομένα και η εμπιστοσύνη είναι η δεσμευμένη πιθανότητα ότι μια συναλλαγή που περιέχει ορισμένα στοιχεία περιέχει επίσης ένα άλλο στοιχείο. Ο λόγος ανύψωσης δείχνει τη δύναμη που διαθέτει το αποτέλεσμα των κανόνων σύνδεσης. Συνολικά, ο λόγος ανύψωσης συνοψίζει τη δύναμη των συσχετίσεων μεταξύ προϊόντων.

Πίνακας 29: Δύναμη των συσχετίσεων μεταξύ προϊόντων

$$S = \frac{\sum(Ta+Tc)}{\sum(T)} \quad (1)$$

$$Support(A) = \frac{\text{The Number of Transaction that Contains } A}{\text{Total Transaction}} \quad (2)$$

where:

S = Support

$\sum(Ta + Tc)$ = the number of transaction that contains *antecedent* and *consequent*.

$\sum(T)$ = the number of transaction.

$$C = \frac{\sum(Ta+Tc)}{\sum(Ta)} \quad (3)$$

$$Confidence = P(A | B) = \frac{\text{The Number of Transaction that Contains } A \text{ and } B}{\text{Total Transaction that Contains } A} \quad (4)$$

where:

C = Confidence

$\sum(Ta + Tc)$ = the number of transaction that contains *antecedent* and *consequent*.

$\sum(Ta)$ = the number of transaction that contains *antecedent*.

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

3.3.1 Αλγόριθμος Apriori

Το Apriori είναι ο πρώτος και βασικός αλγόριθμος για την εύρεση συχνών συνόλων αντικειμένων που προτάθηκε από τους R. Agrawal και R. Srikant το 1994. Το Apriori περιλαμβάνει μια προσέγγιση γνωστή ως αναζήτηση επιπέδου, όπου τα k-itemsets χρησιμοποιούνται για την εξερεύνηση (k + 1)-itemsets. Εδώ, αρχικά εντοπίζονται συχνά σύνολα στοιχείων 1 με σάρωση της βάσης δεδομένων που ικανοποιούν την ελάχιστη υποστήριξη. Και πάλι, τα συχνά σύνολα στοιχείων 2 εντοπίζονται χρησιμοποιώντας συχνά σύνολα στοιχείων 1. Έτσι, αυτή η διαδικασία συνεχίζεται μέχρι να βρεθούν συχνά kitemsets [3]. Στην πραγματικότητα το Apriori ακολουθεί μια αντιστονοτική ιδιότητα που δηλώνει ότι κάθε υποσύνολο ενός συχνού itemset πρέπει επίσης να είναι συχνό και χρησιμοποιεί μια αναζήτηση εύρους για να μετράει συχνά τα υποψήφια στοιχεία.

Αυτός ο αλγόριθμος έχει δύο κύρια βήματα, τα εξής:

- Joining Step: Για να βρείτε το L_k , δημιουργείται ένα σύνολο υποψήφιων k-itemsets ενώνοντας (L_{k-1}) με τον εαυτό του.
- Pruning Step: Οποιοδήποτε (k - 1)-itemset που δεν είναι συχνό δεν μπορεί να είναι υποσύνολο του συχνού k-itemset.

3.3.2 Αλγόριθμος FP – Growth

Ο αλγόριθμος FP-Growth είναι ένας αποτελεσματικός τρόπος για να εντοπίσουμε τα πιο συχνά σύνολα στοιχείων σε ένα σύνολο δεδομένων. Αυτός ο αλγόριθμος είναι μια εξέλιξη του αλγορίθμου Apriori, αντιμετωπίζοντας τις αδυναμίες του Apriori και προσφέροντας αποτελεσματικότερη λύση για τον εντοπισμό συχνών συνόλων.

Ο αλγόριθμος αυτός ξεκινά συμπιέζοντας τη βάση δεδομένων εισόδου, αναπτύσσοντας έτσι μια παρουσία ενός συχνού δέντρου μοτίβου. Στη συνέχεια, η συμπιεσμένη βάση δεδομένων διαιρείται σε μερικές υπό όρους βάσεις δεδομένων, όπου κάθε βάση δεδομένων αντιπροσωπεύει ένα μοναδικό συχνό μοτίβο. Τέλος, πραγματοποιείται εξόρυξη κάθε βάσης δεδομένων έξω ξεχωριστά. Ως εκ τούτου, το κόστος αναζήτησης είναι σημαντικά μειωμένο, προσφέροντας καλή εκλεκτικότητα.

Τα βήματα του αλγορίθμου FP-Growth είναι τα εξής:

1. Δημιουργία FP-Tree: Διαβάζουμε τα δεδομένα και χτίζουμε το FP-Tree. Κάθε αντικείμενο προστίθεται στο δέντρο με βάση τη συχνότητά του στα δεδομένα. Τα αντικείμενα ταξινομούνται βάσει της συχνότητας εμφάνισής τους.
2. Κατασκευή Conditional FP-Tree: Δημιουργούμε έναν Conditional FP-Tree για κάθε συνδυασμό στοιχείων. Αφαιρούμε τα λιγότερο συχνά στοιχεία από τα δεδομένα και επαναλαμβάνουμε τη διαδικασία για κάθε νέο Conditional FP-Tree.
3. Αναδρομική εφαρμογή του αλγορίθμου στα Conditional FP-Trees: Επαναλαμβάνουμε τη διαδικασία κατασκευής FP-Tree και Conditional FP-Tree για κάθε νέο σύνολο δεδομένων. Επαναλαμβάνουμε αυτό το βήμα μέχρι να μην υπάρχουν πλέον στοιχεία για εξέταση.
4. Εξαγωγή συχνών συνόλων στοιχείων: Εξάγουμε τα συχνά σύνολα στοιχείων από το FP-Tree, λαμβάνοντας υπόψη τα Conditional FP-Trees που έχουν δημιουργηθεί. Κάθε σύνολο στοιχείων που εξάγεται αποτελεί ένα συχνό σύνολο.

3.3.3 Αλγόριθμος Eclat

Οι αλγόριθμοι Apriori και FP-Growth είναι παραδείγματα μεθόδων κανόνων συσχέτισης με οριζόντιες μορφές δεδομένων. Μια μέθοδος κανόνων συσχέτισης που χρησιμοποιεί κατακόρυφη μορφή δεδομένων είναι η ECLAT ή ο μετασχηματισμός κλάσης ισοδυναμίας. Το κύριο καθήκον του αλγορίθμου μετασχηματισμού κλάσης ισοδυναμίας είναι να βρίσκει συχνά στοιχεία σε μια ομάδα συναλλαγών και λειτουργεί μόνο σε μια βάση δεδομένων κάθετης διάταξης. Ο αλγόριθμος ECLAT σαρώνει τη βάση δεδομένων για να βρει το συχνό σύνολο στοιχείων μόνο μία φορά, διαφορετικά από τον αλγόριθμο Apriori, χρειάζεται περισσότερο χρόνος για την εύρεση συχνών συνόλων στοιχείων επειδή πρέπει να σαρώσει επανειλημμένα τη βάση δεδομένων, έτσι ώστε η διαδικασία να απαιτεί περισσότερο χρόνο. Ο αλγόριθμος ECLAT βρίσκει στοιχεία από κάτω, όπως η πρώτη αναζήτηση βάθους. Αυτός ο αλγόριθμος υπολογίζει μόνο την τιμή υποστήριξης, ενώ η τιμή εμπιστοσύνης δεν υπολογίζεται σε αυτόν τον αλγόριθμο,

Τα βήματα του αλγορίθμου Eclat είναι τα εξής:

1. Δημιουργία Ισοδύναμων Κλάσεων (Equivalence Classes): Αρχικά, τα δεδομένα συναλλαγών ομαδοποιούνται σε ισοδύναμες κλάσεις βάσει των στοιχείων που περιέχουν. Κάθε στοιχείο αποτελεί την ισοδύναμη κλάση του ίδιου του.
2. Εύρεση Συχνών Συνόλων Στοιχείων μέσα στις Κλάσεις: Στη συνέχεια, αναζητούμε συχνά σύνολα στοιχείων εντός κάθε ισοδύναμης κλάσης, δηλαδή τις συνδυασμένες συχνότητες των στοιχείων σε κάθε κλάση.
3. Συνένωση Συχνών Συνόλων Στοιχείων: Τέλος, συνδυάζουμε τα συχνά σύνολα στοιχείων που έχουν βρεθεί σε κάθε ισοδύναμη κλάση για να δημιουργήσουμε τα συχνά σύνολα στοιχείων για ολόκληρο το σύνολο των δεδομένων.

3.3.4 Αποτελέσματα market basket analysis

Παρακάτω αναλύονται τα καλάθια αγοράς για τα 2 σημαντικότερα clusters.

Το πρώτο cluster που έχει επιλεγεί είναι το cluster των «Casual Shoppers» καθώς αυτοί οι πελάτες πραγματοποιώντας αγορές με χαμηλή συχνότητα, κάνουν όμως πρόσφατες αγορές, και

με χαμηλή έως μέτρια συνεισφορά στα έσοδα, είναι πελάτες που ανταποκρίνονται θετικά σε ενέργειες εκ μέρους της επιχείρησης και άρα θα ήταν θετικό να αναλυθεί η συμπεριφορά τους και με τη βοήθεια του καλαθιού αγοράς. Αντιλαμβανόμαστε ότι οι συγκεκριμένοι πελάτες έχουν προοπτική να εξελιχθούν σε πιο ενεργούς αγοραστές και επομένως πρέπει να τους κρατήσουμε.

Παρακάτω βλέπουμε τους 20 πρώτους κανόνες με το μεγαλύτερο Lift.

Πίνακας 30: Κανόνες με το μεγαλύτερο Lift

Obs	RULEID	LHS	RHS	COUNT	SUPPORT	CONF	LIFT	ITEM1	ITEM2	ITEM3	RULE
1	219	1	2	73	0.11716	51.7730	282.981	chicken nuggets	cheeseburger	goody's club	chicken nuggets ==> cheeseburger & goody's club
2	255	2	1	73	0.11716	64.0351	282.981	cheeseburger	goody's club	chicken nuggets	cheeseburger & goody's club ==> chicken nuggets
3	5	1	1	34	0.05457	49.2754	230.853	coca cola 1.5lt	pizza κλασική special		coca cola 1.5lt ==> pizza κλασική special
4	41	1	1	34	0.05457	25.5639	230.853	pizza κλασική special	coca cola 1.5lt		pizza κλασική special ==> coca cola 1.5lt
5	147	2	1	73	0.11716	93.5897	230.497	cheeseburger	chicken nuggets	goody's club	cheeseburger & chicken nuggets ==> goody's club
6	291	1	2	73	0.11716	28.8538	230.497	goody's club	cheeseburger	chicken nuggets	goody's club ==> cheeseburger & chicken nuggets
7	28	1	1	33	0.05296	17.9348	177.384	μπουγάτσα με κρέμα	μπουγάτσα με τυρί		μπουγάτσα με κρέμα ==> μπουγάτσα με τυρί
8	64	1	1	33	0.05296	62.3810	177.384	μπουγάτσα με κρέμα	μπουγάτσα με τυρί		μπουγάτσα με τυρί ==> μπουγάτσα με κρέμα
9	795	1	1	85	0.13641	60.2837	148.469	chicken nuggets	goody's club		chicken nuggets ==> goody's club
10	831	1	1	85	0.13641	33.5968	148.469	goody's club	chicken nuggets		goody's club ==> chicken nuggets
11	264	1	1	42	0.06740	16.6008	105.551	goody's club	golden burger		goody's club ==> golden burger
12	228	1	1	42	0.06740	42.8571	105.551	golden burger	goody's club		golden burger ==> goody's club
13	116	1	1	40	0.06420	61.5385	93.069	τυρόπιττα του γρηγόρη	espresso freddo		τυρόπιττα του γρηγόρη ==> espresso freddo
14	80	1	1	40	0.06420	9.7087	93.069	espresso freddo	τυρόπιττα του γρηγόρη		espresso freddo ==> τυρόπιττα του γρηγόρη
15	583	1	1	51	0.08185	23.9437	81.974	cold cappuccino	cold esπρέσσο		cold cappuccino ==> cold esπρέσσο
16	619	1	1	51	0.08185	28.0220	81.974	cold esπρέσσο	cold cappuccino		cold esπρέσσο ==> cold cappuccino
17	327	2	1	73	0.11716	85.8824	74.016	chicken nuggets	goody's club	cheeseburger	chicken nuggets & goody's club ==> cheeseburger
18	183	1	2	73	0.11716	10.0968	74.016	cheeseburger	chicken nuggets	goody's club	cheeseburger ==> chicken nuggets & goody's club
19	2	1	1	49	0.07864	43.3628	65.581	cappuccino freddo	espresso freddo		cappuccino freddo ==> espresso freddo
20	38	1	1	49	0.07864	11.8932	65.581	espresso freddo	cappuccino freddo		espresso freddo ==> cappuccino freddo

Ας αναλύσουμε τους 5 πρώτους κανόνες.

Ο πρώτος κανόνας αφορά τη σχέση μεταξύ τριών προϊόντων: "chicken nuggets", "cheeseburger", και "Goody's Club".

Πιο αναλυτικά:

Support = 0.11716: Αυτή η τιμή σημαίνει ότι το 11.716% των συναλλαγών περιλαμβάνει την αγορά των "chicken nuggets", "cheeseburger", και "Goody's Club" μαζί. Είναι ένας δείκτης που μας δείχνει πόσο συχνά εμφανίζεται ο συγκεκριμένος συνδυασμός προϊόντων στο σύνολο των συναλλαγών.

Confidence = 51.773%: Αυτό σημαίνει ότι όταν ένας πελάτης αγοράζει "chicken nuggets" και "cheeseburger", υπάρχει 51.773% πιθανότητα να αγοράσει επίσης "Goody's Club". Η εμπιστοσύνη μας δείχνει πόσο συχνά το συμβάν προς το οποίο σημειώνεται η συνθήκη (π.χ., αγορά "Goody's Club") συμβαίνει, δεδομένου ότι έχει συμβεί ένα άλλο συμβάν (αγορά "chicken nuggets" και "cheeseburger").

Lift = 282.981: Αυτή η εξαιρετικά υψηλή τιμή lift υποδεικνύει μια πολύ ισχυρή θετική συσχέτιση μεταξύ της αγοράς των "chicken nuggets" και "cheeseburger" και της αγοράς του "Goody's Club". Σημαίνει ότι η πιθανότητα αγοράς "Goody's Club" σε συνδυασμό με αυτά τα δύο άλλα προϊόντα είναι 282.981 φορές μεγαλύτερη από την πιθανότητα να αγοραστεί "Goody's Club" αν οι αγορές ήταν ανεξάρτητες (δηλαδή, χωρίς να λαμβάνονται υπόψη οι συνδυασμοί προϊόντων).

Ο δεύτερος κανόνας αφορά τη σχέση μεταξύ δύο προϊόντων: "Coca Cola 1,5lt" και "Pizza Κλασική Special".

Support = 0.05457: Αυτή η τιμή δηλώνει ότι το 5.457% των συνολικών συναλλαγών στο σύνολο δεδομένων περιλαμβάνει την αγορά και των δύο προϊόντων μαζί, δηλαδή "Coca Cola 1,5lt" και "Pizza Κλασική Special". Είναι ένας δείκτης της συχνότητας εμφάνισης αυτού του συγκεκριμένου συνδυασμού προϊόντων.

Confidence = 49.2754%: Αυτό σημαίνει ότι υπάρχει 49.2754% πιθανότητα όταν ένας πελάτης αγοράζει "Coca Cola 1,5lt", να αγοράσει επίσης "Pizza Κλασική Special". Η εμπιστοσύνη μετρά την πιθανότητα ο συγκεκριμένος συνδυασμός προϊόντων να αγοραστεί όταν έχει ήδη αγοραστεί το πρώτο προϊόν, δίνοντας μια ένδειξη της δύναμης της σχέσης μεταξύ των δύο.

Lift = 230.853: Αυτή η τιμή υποδηλώνει μια πολύ ισχυρή θετική συσχέτιση μεταξύ της αγοράς των δύο προϊόντων. Σημαίνει ότι η πιθανότητα αγοράς "Pizza Κλασική Special" όταν έχει ήδη αγοραστεί "Coca Cola 1,5lt" είναι 230.853 φορές μεγαλύτερη από την πιθανότητα να αγοραστεί "Pizza Κλασική Special" αν η αγορά της ήταν ανεξάρτητη (δηλαδή, χωρίς να λαμβάνονται υπόψη οι συνδυασμοί προϊόντων).

Ο τρίτος κανόνας αφορά τη σχέση μεταξύ δύο διαφορετικών γεύσεων μπουγάτσας: "μπουγάτσα με κρέμα" και "μπουγάτσα με τυρί".

Πιο λεπτομερώς:

Support = 0.05296: Αυτό σημαίνει ότι το 5.296% των συνολικών συναλλαγών περιλαμβάνει την αγορά και των δύο ειδών μπουγάτσας μαζί. Αυτός ο αριθμός δείχνει τη συχνότητα με την οποία αυτός ο συγκεκριμένος συνδυασμός προϊόντων εμφανίζεται στις συνολικές αγορές, δίνοντάς μας μια ένδειξη της δημοφιλίας τους ως ζευγάρι.

Confidence = 17.9348%: Αυτός ο δείκτης μας λέει ότι υπάρχει περίπου 17.9348% πιθανότητα όταν κάποιος αγοράζει μπουγάτσα με κρέμα, να αγοράσει επίσης μπουγάτσα με τυρί. Η εμπιστοσύνη μετράει την πιθανότητα αγοράς του δεύτερου προϊόντος δεδομένης της αγοράς του πρώτου, δίνοντας μια ιδέα για το πόσο συχνά τα δύο προϊόντα αγοράζονται μαζί.

Lift = 177.384: Αυτή η τιμή δείχνει ότι οι πελάτες που αγοράζουν μπουγάτσα με κρέμα είναι 177.384 φορές πιο πιθανό να αγοράσουν μπουγάτσα με τυρί από ό,τι αν η αγορά της μπουγάτσας με τυρί ήταν ανεξάρτητη. Ένας τόσο υψηλός δείκτης lift υποδηλώνει μια πολύ ισχυρή θετική συσχέτιση μεταξύ των δύο προϊόντων, δηλαδή την ύπαρξη μιας ισχυρής τάσης από τους πελάτες να τα αγοράζουν μαζί.

Ο τέταρτος κανόνας αφορά τη σχέση μεταξύ δύο προϊόντων: το "Goody's Club" και το "Golden Burger".

Πιο συγκεκριμένα:

Support = 0,06740: Αυτό σημαίνει ότι το 6,74% των συνολικών συναλλαγών περιλαμβάνει και τα δύο προϊόντα μαζί, δηλαδή το "Goody's Club" και το "Golden Burger". Ο δείκτης support μετρά τη συχνότητα με την οποία αυτά τα δύο προϊόντα εμφανίζονται μαζί στις αγορές.

Confidence = 16.6008%: Αυτή η τιμή υποδηλώνει ότι όταν ένας πελάτης αγοράζει το "Goody's Club", υπάρχει περίπου 16,60% πιθανότητα να αγοράσει επίσης το "Golden Burger". Η confidence μετρά την πιθανότητα αγοράς του δεύτερου προϊόντος δεδομένης της αγοράς του πρώτου, παρέχοντας μια ένδειξη της δύναμης της σχέσης μεταξύ των δύο προϊόντων.

Lift = 105,551: Αυτή η τιμή υποδηλώνει ότι οι πελάτες που αγοράζουν το "Goody's Club" είναι περίπου 105,55 φορές πιο πιθανό να αγοράσουν επίσης το "Golden Burger" σε σχέση με την τυχαία επιλογή του "Golden Burger" χωρίς να λαμβάνονται υπόψη άλλες πληροφορίες. Ένας lift μεγαλύτερος από 1 υποδηλώνει μια θετική συσχέτιση μεταξύ των προϊόντων, δηλαδή ότι τα προϊόντα αγοράζονται μαζί πιο συχνά από ό,τι θα αναμενόταν αν ήταν ανεξάρτητα.

Τέλος, ο πέμπτος κανόνας αφορά τη σχέση μεταξύ δύο συγκεκριμένων προϊόντων: της "τυρόπιτσας του Γρηγόρη" και του "espresso freddo".

Support = 0.06420: Αυτό σημαίνει ότι το 6.42% των συνολικών συναλλαγών περιλαμβάνει και τα δύο προϊόντα μαζί, δηλαδή την "τυρόπιτσα του Γρηγόρη" και το "espresso freddo". Ο support δείχνει τη συχνότητα με την οποία εμφανίζεται ο συνδυασμός αυτών των προϊόντων σε σχέση με το σύνολο των συναλλαγών.

Confidence = 61.5385%: Αυτός ο δείκτης μας λέει ότι όταν κάποιος αγοράζει την "τυρόπιτα του Γρηγόρη", υπάρχει περίπου 61.54% πιθανότητα να αγοράσει επίσης και το "espresso freddo". Η confidence μετράει την πιθανότητα αγοράς του δεύτερου προϊόντος δεδομένης της αγοράς του πρώτου, δηλαδή τη σχετική συχνότητα με την οποία συμβαίνουν αυτές οι δύο αγορές μαζί.

Lift = 93.069: Αυτός ο δείκτης υποδηλώνει ότι οι πελάτες που αγοράζουν την "τυρόπιτα του Γρηγόρη" είναι 93.069 φορές πιο πιθανό να αγοράσουν επίσης το "espresso freddo" σε σχέση με την τυχαία επιλογή του "espresso freddo" χωρίς καμία άλλη πληροφορία. Ένας lift μεγαλύτερος από 1 υποδηλώνει θετική συσχέτιση μεταξύ των προϊόντων, δηλαδή ότι τα προϊόντα αγοράζονται μαζί πιο συχνά απ' ό,τι θα αναμενόταν αν ήταν ανεξάρτητα.

Το δεύτερο cluster που έχει επιλεγεί είναι το cluster των «Inactive Customers» καθώς εμπεριέχει πελάτες που πραγματοποιούν λιγότερες αγορές χωρίς πρόσφατες αγορές. Παρόλο που δεν είναι δαπανούν υψηλά ποσά, συνεισφέρουν σημαντικά στα έσοδα. Οι Inactive πελάτες είναι πελάτες που παλιότερα είχαν ενεργό συμμετοχή αλλά πλέον δεν είναι τόσο ενεργοί.

Παρακάτω βλέπουμε τους 20 πρώτους κανόνες με το υψηλότερο lift.

Πίνακας 31: Κανόνες με το υψηλότερο lift.

Obs	RULEID	LHS	RHS	COUNT	SUPPORT	CONF	LIFT	ITEM1	ITEM2	ITEM3	RULE
1	13	1	1	65	0.06568	58.5586	531.636	half roll volcano	shisan roll california		half roll volcano ==> shisan roll california
2	49	1	1	65	0.06568	58.5586	531.636	shisan roll california	half roll volcano		shisan roll california ==> half roll volcano
3	89	1	1	67	0.06771	39.4118	207.453	γύρος κοτόπουλο σε πίττα του κακού λίκου	γύρος χοιρινός σε πίττα του κακού λίκου		γύρος κοτόπουλο σε πίττα του κακού λίκου ==> γύρος χοιρινός σε πίττα του κακού λίκου
4	125	1	1	67	0.06771	35.6383	207.453	γύρος χοιρινός σε πίττα του κακού λίκου	γύρος κοτόπουλο σε πίττα του κακού λίκου		γύρος χοιρινός σε πίττα του κακού λίκου ==> γύρος κοτόπουλο σε πίττα του κακού λίκου
5	579	1	2	199	0.20110	51.4212	204.359	chicken nuggets	cheeseburger	goody's club	chicken nuggets ==> cheeseburger & goody's club
6	615	2	1	199	0.20110	79.9197	204.359	cheeseburger	goody's club	chicken nuggets	cheeseburger & goody's club ==> chicken nuggets
7	364	1	1	59	0.05962	71.9512	183.037	chicken mcnuggets™ 20μχ	μεγάλες πατάτες ηλιανθιές		chicken mcnuggets™ 20μχ ==> μεγάλες πατάτες ηλιανθιές
8	400	1	1	59	0.05962	15.1671	183.037	μεγάλες πατάτες ηλιανθιές	chicken mcnuggets™ 20μχ		μεγάλες πατάτες ηλιανθιές ==> chicken mcnuggets™ 20μχ
9	1443	1	2	50	0.05053	12.8535	169.594	μεγάλες πατάτες ηλιανθιές	cheeseburger	μαγιονέζα	μεγάλες πατάτες ηλιανθιές ==> cheeseburger & μαγιονέζα
10	1299	2	1	50	0.05053	66.6667	169.594	cheeseburger	μαγιονέζα	μεγάλες πατάτες ηλιανθιές	cheeseburger & μαγιονέζα ==> μεγάλες πατάτες ηλιανθιές
11	444	1	1	67	0.06771	22.9452	169.449	golden burger	green burger		golden burger ==> green burger
12	480	1	1	67	0.06771	50.0000	169.449	green burger	golden burger		green burger ==> golden burger
13	507	2	1	199	0.20110	82.2314	168.477	cheeseburger	chicken nuggets	goody's club	cheeseburger & chicken nuggets ==> goody's club
14	651	1	2	199	0.20110	41.2008	168.477	goody's club	cheeseburger	chicken nuggets	goody's club ==> cheeseburger & chicken nuggets
15	137	1	1	61	0.06164	51.2605	164.696	μπουγάτσα με τυρί	μπουγάτσα με κρέμα		μπουγάτσα με τυρί ==> μπουγάτσα με κρέμα
16	101	1	1	61	0.06164	19.8052	164.696	μπουγάτσα με κρέμα	μπουγάτσα με τυρί		μπουγάτσα με κρέμα ==> μπουγάτσα με τυρί
17	219	2	1	77	0.07781	55.7971	141.943	cheeseburger	chicken burger	μεγάλες πατάτες ηλιανθιές	cheeseburger & chicken burger ==> μεγάλες πατάτες ηλιανθιές
18	363	1	2	77	0.07781	19.7943	141.943	μεγάλες πατάτες ηλιανθιές	cheeseburger	chicken burger	μεγάλες πατάτες ηλιανθιές ==> cheeseburger & chicken burger
19	42	1	1	73	0.07377	21.7910	138.231	pizza κλασική special	coca cola 1.5lit		pizza κλασική special ==> coca cola 1.5lit
20	6	1	1	73	0.07377	46.7949	138.231	coca cola 1.5lit	pizza κλασική special		coca cola 1.5lit ==> pizza κλασική special
21	1407	2	1	50	0.05053	19.7828	120.722	cheeseburger	μεγάλες πατάτες ηλιανθιές	μαγιονέζα	cheeseburger & μεγάλες πατάτες ηλιανθιές ==> μαγιονέζα
22	1371	1	2	50	0.05053	30.8642	120.722	μαγιονέζα	cheeseburger	μεγάλες πατάτες ηλιανθιές	μαγιονέζα ==> cheeseburger & μεγάλες πατάτες ηλιανθιές
23	472	1	1	215	0.21726	44.5135	113.823	goody's club	chicken nuggets		goody's club ==> chicken nuggets
24	436	1	1	215	0.21726	55.5556	113.823	chicken nuggets	soodv's club		chicken nuggets ==> soodv's club

Ο πρώτος κανόνας αφορά τη σχέση μεταξύ δύο προϊόντων: του "half roll volcano" και του "shisan roll california"

Support = 0.06568: Αυτό σημαίνει ότι το 6,568% των συνολικών συναλλαγών περιλαμβάνει και τα δύο προϊόντα μαζί, δηλαδή το "half roll volcano" και το "shisan roll california". Ο δείκτης support μετρά τη συχνότητα με την οποία αυτά τα δύο προϊόντα εμφανίζονται μαζί στις αγορές.

Confidence = 58.5586%: Αυτή η τιμή υποδηλώνει ότι όταν ένας πελάτης αγοράζει το "half roll volcano", υπάρχει περίπου 58,56% πιθανότητα να αγοράσει επίσης το "shisan roll california". Η confidence μετρά την πιθανότητα αγοράς του δεύτερου προϊόντος δεδομένης της αγοράς του πρώτου, παρέχοντας μια ένδειξη της δύναμης της σχέσης μεταξύ των δύο προϊόντων.

Lift = 531.636: Αυτή η τιμή υποδηλώνει ότι οι πελάτες που αγοράζουν το "half roll volcano" είναι περίπου 531,64 φορές πιο πιθανό να αγοράσουν επίσης το "shisan roll california" σε σχέση με την τυχαία επιλογή του "shisan roll california" χωρίς να λαμβάνονται υπόψη άλλες πληροφορίες. Ένας lift μεγαλύτερος από 1 υποδηλώνει μια θετική συσχέτιση μεταξύ των προϊόντων, δηλαδή ότι τα προϊόντα αγοράζονται μαζί πιο συχνά από ό,τι θα αναμενόταν αν ήταν ανεξάρτητα.

Ο δεύτερος κανόνας αφορά τη σχέση μεταξύ δύο προϊόντων: του "γύρου κοτόπουλου σε πίτα του κακού λύκου" και του "γύρου χοιρινού σε πίτα του κακού λύκου"

Support = 0.06771: Αυτό σημαίνει ότι το 6.771% των συνολικών συναλλαγών περιλαμβάνει και τα δύο προϊόντα μαζί, δηλαδή τον "γύρο κοτόπουλου σε πίτα του κακού λύκου" και τον "γύρο χοιρινού σε πίτα του κακού λύκου". Ο δείκτης support μετρά τη συχνότητα με την οποία αυτά τα δύο προϊόντα εμφανίζονται μαζί στις αγορές.

Confidence = 39.4118%: Αυτή η τιμή υποδηλώνει ότι όταν ένας πελάτης αγοράζει τον "γύρο κοτόπουλου σε πίτα του κακού λύκου", υπάρχει περίπου 39.41% πιθανότητα να αγοράσει επίσης τον "γύρο χοιρινού σε πίτα του κακού λύκου". Η confidence μετρά την πιθανότητα αγοράς του δεύτερου προϊόντος δεδομένης της αγοράς του πρώτου, παρέχοντας μια ένδειξη της δύναμης της σχέσης μεταξύ των δύο προϊόντων.

Lift = 207.453: Αυτή η τιμή υποδηλώνει ότι οι πελάτες που αγοράζουν τον "γύρο κοτόπουλου σε πίτα του κακού λύκου" είναι περίπου 207.45 φορές πιο πιθανό να αγοράσουν επίσης τον "γύρο χοιρινού σε πίτα του κακού λύκου" σε σχέση με την τυχαία επιλογή του "γύρου χοιρινού σε πίτα του κακού λύκου" χωρίς να λαμβάνονται υπόψη άλλες πληροφορίες.

Ο τρίτος κανόνας αφορά τη σχέση μεταξύ τριών προϊόντων: των "chicken nuggets", "cheeseburger" και "goody's club".

Support = 0.20110: Αυτό σημαίνει ότι το 20.11% των συνολικών συναλλαγών περιλαμβάνει τα τρία προϊόντα μαζί, δηλαδή τα "chicken nuggets", "cheeseburger" και "goody's club". Ο δείκτης support μετρά τη συχνότητα με την οποία αυτά τα τρία προϊόντα εμφανίζονται μαζί στις αγορές.

Confidence = 51.4212%: Αυτή η τιμή υποδηλώνει ότι όταν ένας πελάτης αγοράζει "chicken nuggets" και "cheeseburger", υπάρχει περίπου 51.42% πιθανότητα να αγοράσει επίσης το "goody's club". Η confidence μετρά την πιθανότητα αγοράς του τρίτου προϊόντος δεδομένης της αγοράς των πρώτων δύο, παρέχοντας μια ένδειξη της δύναμης της σχέσης μεταξύ των προϊόντων.

Lift = 204.359: Αυτή η τιμή υποδηλώνει ότι οι πελάτες που αγοράζουν τα "chicken nuggets" και το "cheeseburger" είναι περίπου 204.36 φορές πιο πιθανό να αγοράσουν επίσης το "goody's club" σε σχέση με την τυχαία επιλογή του "goody's club" χωρίς να λαμβάνονται υπόψη άλλες πληροφορίες.

Ο τέταρτος κανόνας αφορά τη σχέση μεταξύ δύο προϊόντων: των "Chicken McNuggets™ 20 τεμάχια" και των "Μεγάλων Πατάτων Τηγανητών".

Support = 0.05962: Αυτό σημαίνει ότι το 5,962% των συνολικών συναλλαγών περιλαμβάνει και τα δύο προϊόντα μαζί, δηλαδή τα "Chicken McNuggets™ 20 τεμάχια" και τις "Μεγάλες Πατάτες Τηγανητές". Ο δείκτης support μετρά τη συχνότητα με την οποία αυτά τα δύο προϊόντα εμφανίζονται μαζί στις αγορές.

Confidence = 71.9512%: Αυτή η τιμή υποδηλώνει ότι όταν ένας πελάτης αγοράζει τα "Chicken McNuggets™ 20 τεμάχια", υπάρχει περίπου 71,95% πιθανότητα να αγοράσει επίσης τις "Μεγάλες Πατάτες Τηγανητές". Η confidence μετρά την πιθανότητα αγοράς του δεύτερου προϊόντος δεδομένης της αγοράς του πρώτου, παρέχοντας μια ένδειξη της δύναμης της σχέσης μεταξύ των δύο προϊόντων.

Lift = 183.037: Αυτή η τιμή υποδηλώνει ότι οι πελάτες που αγοράζουν τα "Chicken McNuggets™ 20 τεμάχια" είναι περίπου 183,04 φορές πιο πιθανό να αγοράσουν επίσης τις "Μεγάλες Πατάτες Τηγανητές" σε σχέση με την τυχαία επιλογή τους, χωρίς να λαμβάνονται υπόψη άλλες πληροφορίες.

Τέλος, ο πέμπτος κανόνας αφορά τη σχέση μεταξύ τριών προϊόντων: των "Μεγάλων Πατάτων Τηγανητών", του "Cheeseburger" και της "Μαγιονέζας".

Support = 0.05053: Αυτό σημαίνει ότι το 5,053% των συνολικών συναλλαγών περιλαμβάνει και τα τρία προϊόντα μαζί, δηλαδή τις "Μεγάλες Πατάτες Τηγανητές", το "Cheeseburger" και την "Μαγιονέζα". Ο δείκτης support μετρά τη συχνότητα με την οποία αυτά τα τρία προϊόντα εμφανίζονται μαζί στις αγορές.

Confidence = 12.8535%: Αυτή η τιμή υποδηλώνει ότι όταν ένας πελάτης αγοράζει τις "Μεγάλες Πατάτες Τηγανητές", υπάρχει περίπου 12,85% πιθανότητα να αγοράσει επίσης το "Cheeseburger" και την "Μαγιονέζα". Η confidence μετρά την πιθανότητα αγοράς των δύο προϊόντων δεδομένης της αγοράς του πρώτου, παρέχοντας μια ένδειξη της δύναμης της σχέσης μεταξύ τους.

Lift = 169.594: Αυτή η τιμή υποδηλώνει ότι οι πελάτες που αγοράζουν τις "Μεγάλες Πατάτες Τηγανητές" είναι περίπου 169,59 φορές πιο πιθανό να αγοράσουν επίσης το "Cheeseburger" και την "Μαγιονέζα" σε σχέση με την τυχαία επιλογή τους, χωρίς να λαμβάνονται υπόψη άλλες πληροφορίες.

Συμπερασματικά, το Market Basket Analysis αποτελεί ένα ισχυρό εργαλείο για την ανάλυση των προτιμήσεων και συμπεριφορών των καταναλωτών στις αγορές. Με αυτήν την τεχνική, μπορέσαμε να ανιχνεύσουμε ποια προϊόντα συχνά αγοράζονται μαζί, αναδεικνύοντας πιθανές σχέσεις μεταξύ τους. Οι μετρήσεις όπως το support, η confidence και το lift βοήθησαν στην αξιολόγηση της σημαντικότητας αυτών των συσχετίσεων. Τέλος, με βάση αυτή την ανάλυση μπορούμε να προτείνουμε σημαντικές επιχειρηματικές εφαρμογές όπως την προώθηση στρατηγικών τιμολόγησης, τη βελτίωση της διάταξης των προϊόντων στα ράφια του καταστήματος, και την εξατομίκευση των προσφορών και της διαφήμισης για τους πελάτες.

Κεφάλαιο 4

4.1 Συμπεράσματα – Προτάσεις

Η μεταπτυχιακή διατριβή ανέλυσε εκτενώς τον τομέα της ανάλυσης δεδομένων και των εφαρμογών της στον επιχειρηματικό κόσμο, εστιάζοντας στην ανάλυση δεδομένων πελατών μίας διαδικτυακής πλατφόρμας παράδοσης φαγητού. Η εργασία αυτή αποσκοπούσε στην κατανόηση των προτιμήσεων και συμπεριφορών των πελατών, καθώς και στην ανάπτυξη στρατηγικών μάρκετινγκ βασισμένων σε δεδομένα.

Αρχικά, παρουσιάστηκε η έννοια των αποθετηρίων δεδομένων και των μεγάλων συλλογών δεδομένων, καθώς και ο ορισμός του analytics και της τεχνητής νοημοσύνης, με έμφαση στον επιχειρηματικό τομέα. Στη συνέχεια, παρουσιάστηκε η μελέτη περίπτωσης δεδομένων μίας ηλεκτρονικής πλατφόρμας διανομής φαγητού, με βασικό εργαλείο την RFM ανάλυση για την κατηγοριοποίησή τους σε διάφορα clusters.

Κατόπιν, επιχειρήθηκε η ανάδειξη των συνήθων συνδυασμών προϊόντων μέσω της μεθόδου Market Basket Analysis, προκειμένου να κατανοηθούν καλύτερα οι ανάγκες και οι προτιμήσεις των διαφόρων ομάδων πελατών. Τα συμπεράσματα που προέκυψαν από αυτήν την ανάλυση παρέχουν

σημαντικές εισηγήσεις για τη βελτίωση των επιχειρηματικών διαδικασιών και την ανάπτυξη αποτελεσματικών στρατηγικών μάρκετινγκ.

Βάσει των δεδομένων που προκύπτουν από την παρούσα έρευνα, προτείνονται τα παρακάτω:

- Η ευελιξία και η χρηστικότητα των διαδικτυακών πλατφορμών είναι κρίσιμα στοιχεία για τη βελτίωση της εμπειρίας του πελάτη. Η ευκολία χρήσης της πλατφόρμας, η ταχύτητα και η αξιοπιστία της επηρεάζουν άμεσα την ικανοποίηση και την εμπιστοσύνη των πελατών. Επιπλέον, η επέκταση των επιλογών παράδοσης και πληρωμής είναι καθοριστική για την προσέλκυση και τη διατήρηση πελατών. Για παράδειγμα, η προσθήκη επιλογών παράδοσης την ίδια μέρα ή σε συγκεκριμένες χρονικές ζώνες, καθώς και η προσφορά διαφόρων μεθόδων πληρωμής (όπως πιστωτικές κάρτες, PayPal, ψηφιακά πορτοφόλια, κ.λπ.), μπορεί να ενισχύσει την αίσθηση ευελιξίας και εξυπηρέτησης του πελάτη. Αυτές οι βελτιώσεις μπορούν να οδηγήσουν σε αυξημένες πωλήσεις και ενισχυμένη αφοσίωση των πελατών.
- Η προστασία των προσωπικών δεδομένων των πελατών αποτελεί θεμέλιο της εμπιστοσύνης τους και είναι ζωτικής σημασίας για την επιτυχία μιας επιχείρησης. Η ανάπτυξη ηθικών κατευθυντήριων γραμμών περιλαμβάνει την υιοθέτηση πολιτικών και πρακτικών που διασφαλίζουν τη συμμόρφωση με τους κανονισμούς προστασίας δεδομένων, όπως ο Γενικός Κανονισμός Προστασίας Δεδομένων (GDPR). Οι κατευθυντήριες γραμμές πρέπει να περιλαμβάνουν σαφείς διαδικασίες για τη συλλογή, την αποθήκευση και την επεξεργασία δεδομένων, καθώς και την ενημέρωση των πελατών για τα δικαιώματά τους. Επιπλέον, η εκπαίδευση του προσωπικού στις καλύτερες πρακτικές προστασίας δεδομένων και η χρήση τεχνολογιών ασφαλείας, όπως η κρυπτογράφηση και τα συστήματα ανίχνευσης απειλών, είναι κρίσιμες για την προστασία των προσωπικών πληροφοριών.
- Η δημιουργία ανεξάρτητων τμημάτων business analytics εντός μιας επιχείρησης μπορεί να προσφέρει σημαντική υποστήριξη στη λήψη αποφάσεων σε διάφορους τομείς, όπως το μάρκετινγκ, τα logistics και οι πωλήσεις. Τα τμήματα αυτά μπορούν να παρέχουν πολύτιμες πληροφορίες μέσω της ανάλυσης δεδομένων, βοηθώντας στην κατανόηση των τάσεων της αγοράς, των προτιμήσεων των πελατών και των επιχειρησιακών επιδόσεων. Με την ανάπτυξη μοντέλων πρόβλεψης και την αναγνώριση προτύπων, οι αναλυτές δεδομένων μπορούν να υποστηρίξουν τη στρατηγική σχεδίαση και τη βελτιστοποίηση των επιχειρησιακών διαδικασιών. Επιπλέον, η στενή συνεργασία με τα άλλα τμήματα της επιχείρησης διασφαλίζει ότι οι αναλύσεις ανταποκρίνονται στις συγκεκριμένες ανάγκες και προκλήσεις κάθε τομέα.
- Η εφαρμογή προηγμένων τεχνικών ανάλυσης δεδομένων και μηχανικής μάθησης στις επιχειρήσεις μπορεί να επιφέρει σημαντικά οφέλη σε διάφορους τομείς. Κατ'αρχάς, αυτές οι τεχνικές επιτρέπουν την ακριβέστερη πρόβλεψη της αγοραστικής συμπεριφοράς των πελατών. Χρησιμοποιώντας δεδομένα από παρελθοντικές αγορές, προτιμήσεις και συμπεριφορές, οι επιχειρήσεις μπορούν να αναπτύξουν μοντέλα που προβλέπουν με μεγαλύτερη ακρίβεια ποιες θα είναι οι μελλοντικές ανάγκες και επιθυμίες των πελατών τους. Αυτό όχι μόνο βελτιώνει τη διαχείριση των αποθεμάτων και τον σχεδιασμό των προϊόντων, αλλά επιτρέπει επίσης την ανάπτυξη στοχευμένων στρατηγικών μάρκετινγκ. Επιπλέον, οι τεχνικές μηχανικής μάθησης βοηθούν στην εφαρμογή στρατηγικών εξατομίκευσης, προσφέροντας μια πιο προσαρμοσμένη εμπειρία στον πελάτη. Αυτό μπορεί να επιτευχθεί μέσω της ανάλυσης δεδομένων για την κατανόηση των ατομικών προτιμήσεων και τη δημιουργία προσωποποιημένων προτάσεων και προσφορών. Η βελτιωμένη εμπειρία και ικανοποίηση των πελατών ενισχύει την αφοσίωση και μειώνει τα ποσοστά αποχώρησης. Επιπλέον, η ανάλυση δεδομένων μπορεί να παρέχει προτάσεις για cross sell και up sell,

αναγνωρίζοντας ποια προϊόντα ή υπηρεσίες μπορούν να ενδιαφέρουν περισσότερο έναν συγκεκριμένο πελάτη. Μέσω αυτής της στρατηγικής, οι επιχειρήσεις μπορούν να αυξήσουν τις πωλήσεις τους προτείνοντας σχετικά προϊόντα ή αναβαθμίσεις που ανταποκρίνονται στις ανάγκες και τις προτιμήσεις των πελατών τους. Συνολικά, η ενσωμάτωση προηγμένων τεχνικών ανάλυσης δεδομένων και μηχανικής μάθησης συμβάλλει στη δημιουργία πιο αποδοτικών, στοχευμένων και αποδοτικών στρατηγικών, οδηγώντας σε αυξημένα έσοδα και βελτιωμένη ικανοποίηση των πελατών.

- Η διεξαγωγή proof of concept (PoC) και pilot project είναι κρίσιμα βήματα για την αξιολόγηση της αποτελεσματικότητας των νέων τεχνολογιών και των μεθόδων ανάλυσης δεδομένων πριν από την πλήρη ενσωμάτωσή τους στην επιχείρηση. Ένα PoC επιτρέπει στην επιχείρηση να δοκιμάσει την πρακτική εφαρμογή μιας ιδέας ή μιας τεχνολογίας σε μικρή κλίμακα για να διαπιστώσει αν λειτουργεί όπως αναμένεται. Ένα πιλοτικό έργο, από την άλλη πλευρά, εφαρμόζει την ιδέα σε μια μεγαλύτερη, αλλά ακόμα ελεγχόμενη, κλίμακα για να αξιολογήσει την απόδοση και τα οφέλη σε πραγματικές συνθήκες. Αυτές οι προσεγγίσεις μειώνουν τον κίνδυνο αποτυχίας, καθώς επιτρέπουν την αναγνώριση και την επίλυση προβλημάτων νωρίς στη διαδικασία, παρέχοντας πολύτιμα δεδομένα και ανατροφοδότηση για την τελική υλοποίηση.

- Η ενσωμάτωση των διαδικασιών business analytics στις καθημερινές λειτουργίες μιας επιχείρησης μέσω model operations (ModelOps) διασφαλίζει ότι τα μοντέλα ανάλυσης δεδομένων είναι συνεχώς διαθέσιμα και μπορούν να ενημερώνονται εύκολα και γρήγορα. Το ModelOps είναι μια προσέγγιση που επικεντρώνεται στη διαχείριση, την ανάπτυξη και την παρακολούθηση των μοντέλων machine learning και analytics σε ένα παραγωγικό περιβάλλον. Περιλαμβάνει διαδικασίες για τη συνεχή ανατροφοδότηση και τη βελτιστοποίηση των μοντέλων βάσει νέων δεδομένων και αλλαγών στις επιχειρησιακές ανάγκες. Η ενσωμάτωση αυτή επιτρέπει στην επιχείρηση να διατηρεί την ακρίβεια και τη σχετικότητα των αναλύσεών της, εξασφαλίζοντας παράλληλα ότι οι πληροφορίες που προκύπτουν από τα analytics χρησιμοποιούνται ενεργά για τη λήψη αποφάσεων σε καθημερινή βάση.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Wixom, B. H., Watson, H. J., & Werner, T. (2011). Developing an enterprise business intelligence capability. *The Norfolk Southern Journey*, 10(2), 61–71.
- Baecker, J., Böttcher, T. P. and Weking, J., (2021). How Companies Create Value From Data – A Taxonomy on Data, Approaches, and Resulting Business Value. *ECIS 2021 Research Papers*. 124.
- Barbosa M. W., Vicente A. D., Ladeira M. B., and Oliveira M. P. (2018). Managing supply chain resources with big data analytics: a systematic review,” *International Journal of Logistics Research and Applications*, 21, 3, 177–200.
- Burton-Jones, A., and Grange, C. (2012). From Use to Effective Use: A Representation Theory Perspective. *Information systems research* (24:3), 632-658.
- Cano, J. (2014) The V's of Big Data: Velocity, Volume, Value, Variety, and Veracity. Available at: <https://www.xsnet.com/blog/bid/205405/the-v-s-of-big-data-velocity-volume-value-variety-and-veracity>
- Chen M., Mao S. , Liu Y. (2014). Big data: a survey. *Mobile Networks*, 19 , 171-209
- Chen S. and Sanweng I. (2021). Economic information analysis and its application in decision-making under the background of big data. *Mobile Information Systems*, 6.
- Darvazeh S. S., Vanani I. R., and Musolu F. M. (2020). Big Data Analytics and Its Applications in Supply Chain Management, *New Trends in the Use of Artificial Intelligence for the Industry 4.0*, , Eds., IntechOpen,.
- Desjardins, J. (2019). How much data is generated each day? *World Economic Forum*.
- Dihuni. (2020). Every day Big Data statistics – 2.5 quintillion bytes of data created daily. *Dihuni*.
- IDC, & Statista. (2020). Data created worldwide 2010-2024. *Statista*.
- Johnson, C. (2021). How much data is produced every day 2021? *The Next Tech*.
- Kshetri, N. (2014). Big Data's impact on privacy, security and consumer welfare. *Telecommun. Policy* , 38, 1134–1145.
- Kuo, Y.F.; Wu, C.M.; Deng, W.J. (2009). The relationships among service quality, perceived value, customer satisfaction, and post-purchase intention in mobile value-added services. *Comput. Hum. Behav.* 25, 887–896.
- Chongkolnee Rungruang, Pakwan Riyapan, Arthit Intarasit, Khanchit Chuarkham, Jirapond Muangprathub, RFM model customer segmentation based on hierarchical approach using FCA, *Expert Systems with Applications*, Volume 237, Part B, 2024
- Yadav, J., & Sharma, M. (2013). A Review of K-mean Algorithm. *Int. J. Eng. Trends Technol*, 4(7), 2972-2976.
- Demirović, D. (2019). An implementation of the mean shift algorithm. *Image Processing On Line*, 9, 251-268.
- Kaur, M., & Kang, S. (2016). Market Basket Analysis: Identify the changing trends of market data using association rule mining. *Procedia computer science*, 85, 78-85.
- Le, T., & Liaw, S.-Y. (2017). Effects of Pros and Cons of Applying Big Data Analytics to Consumers' Responses in an E-Commerce Context. *Sustainability*, 9(5), 798.
- Li, J.; Yu, H. (2013). An Innovative Marketing Model Based on AIDA: A Case from E-bank Campus-marketing by China Construction Bank. *iBusiness* , 5, 47–51.
- Llave, M. R. (2017). Business Intelligence and Analytics in Small and Medium-Sized Enterprises: A Systematic Literature Review. *Procedia Computer Science* (121), 194-205.
- Purcell B.M. (2013) Big Data using cloud computing. *Holy Family Univ. J. Technol. Res*.

Robinson, M. A. (2021). Thousands of pages of new data are created every second – and you can cash in on the company that's tracking it. Strategic Tech Investor.

Sakr S. (2016). Introduction Big Data 2.0 Processing Systems: A Survey, Springer International Publishing, Cham , pp. 1-13

Shafer, T. (2017) The 42 V's of Big Data and Data Science. Available at: <https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html>

Hossain, M., Sattar, A. S., & Paul, M. K. (2019, December). Market basket analysis using apriori and FP growth algorithm. In 2019 22nd international conference on computer and information technology (ICCI) (pp. 1-6). IEEE. Stoianov N. , Urueña M., Niemiec M. , Machnik P. , Maestro G. (2013) Integrated security infrastructures for law enforcement agencies. Multimedia Tools App. , 1-16

Vassakis K., Petrakis E., and Kopanakis I. (2018). Big data analytics: applications, prospects and challenges,” in Mobile Big Data. Eds., Springer, Cham

Wu, R. S., & Chou, P. H. (2011). Customer segmentation of multiple category data in e-commerce using a soft-clustering approach. Electronic Commerce Research and Applications, 10(3), 331-341

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

Khan, K., Rehman, S. U., Aziz, K., Fong, S., & Sarasvady, S. (2014, February). DBSCAN: Past, present and future. In The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014) (pp. 232-238). IEEE.

Tran, T. N., Drab, K., & Daszykowski, M. (2013). Revised DBSCAN algorithm to cluster data with dense adjacent clusters. Chemometrics and Intelligent Laboratory Systems, 120, 92-96.

Blömer, J., & Bujna, K. (2013). Simple methods for initializing the em algorithm for gaussian mixture models. CoRR.

Wang, K., Zhang, J., Li, D., Zhang, X., & Guo, T. (2008). Adaptive affinity propagation clustering. arXiv preprint arXiv:0805.1096.

Kanagala, H. K., & Krishnaiah, V. J. R. (2016, January). A comparative study of K-Means, DBSCAN and OPTICS. In 2016 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-6). IEEE.

Sasirekha, K., & Baby, P. (2013). Agglomerative hierarchical clustering algorithm-a. International Journal of Scientific and Research Publications, 83(3), 83.

Bouguettaya, A., Yu, Q., Liu, X., Zhou, X., & Song, A. (2015). Efficient agglomerative hierarchical clustering. Expert Systems with Applications, 42(5), 2785-2797.

Kurniawan, F., Umayah, B., Hammad, J., Nugroho, S. M. S., & Hariadi, M. (2018). Market Basket Analysis to identify customer behaviours by way of transaction data. Knowledge Engineering and Data Science, 1(1), 20.

Aldino, A. A., Pratiwi, E. D., Sintaro, S., & Putra, A. D. (2021, October). Comparison of market basket analysis to determine consumer purchasing patterns using fp-growth and apriori algorithm. In 2021 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE) (pp. 29-34). IEEE.

Sidhu, S., Meena, U. K., Nawani, A., Gupta, H., & Thakur, N. (2014). FP Growth algorithm implementation. International Journal of Computer Applications, 93(8).

Wahyuningsih, R., Suharsono, A., & Iriawan, N. (2023). COMPARISON OF MARKET BASKET ANALYSIS METHOD USING APRIORI ALGORITHM, FREQUENT PATTERN GROWTH (FP-GROWTH) AND EQUIVALENCE CLASS TRANSFORMATION (ECLAT)(CASE STUDY: SUPERMARKET “X” TRANSACTION DATA FOR 2021). Business and Finance Journal, 8(2), 192-201.

Birant, D. (2011). Data mining using RFM analysis. In Knowledge-oriented applications in data mining. IntechOpen.

Chen, D., Sain, S. L., & Guo, K. (2012). Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. *Journal of Database Marketing & Customer Strategy Management*, 19, 197-208.

Wei, J. T., Lin, S. Y., & Wu, H. H. (2010). A review of the application of RFM model. *African journal of business management*, 4(19), 4199.

Ha, S. H., & Park, S. C. (1998). Application of data mining tools to hotel data mart on the Intranet for database marketing. *Expert Systems with Applications*, 15(1), 1-31.

SAS White paper. The Machine Learning Landscape: A quick guide to the different types of learning.

Svetina, M., & Zupančič, J. (2005). How to increase sales in retail with market basket analysis. *Systems Integration*, 14(2), 418-428.